FIDELITY OF IMPLEMENTING AN ASSESSMENT TRANSLATION AND ADAPTATION FRAMEWORK IN A STUDY OF AN EMERGING INTERNATIONAL ASSESSMENT

by

MAGDA YANIRA CHIA

B.A., University of Maryland, 2002

M.A., New York University, 2003

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado at Boulder in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Education

2012

This thesis entitled:
Fidelity of implementing an assessment translation and adaptation framework in a study of an emerging international assessment

Written by Magda Chia
has been approved for the School of Education, University of Colorado, Boulder

---

Kathy Escamilla, Ph.D.

Bhuvana Narasimhan, Ph.D.

Richard Shavelson, Ph.D.

Guillermo Solano-Flores, Ph.D. (Chair)

Lucinda Soltero-González, Ph.D.

The final copy of this thesis has been examined by the signatories and we find that both the content and the form meet acceptable presentation standards
of scholarly work in the above mentioned discipline.

Chia, Magda Yanira (Ph.D., Education)

Assessment translation and adaptation: A staging and implementation study of an emerging international assessment

Thesis directed by Professor Guillermo Solano-Flores

This study addresses the complex process of translation and adaptation of two Collegiate Learning Assessment (CLA) performance tasks (PTs), originally developed in English for American students, into the languages and cultures of five participating countries.

Focusing on confirming evidence bits (CEBs), disconfirming evidence bits (DEBs), and no evidence bits (NEBs), I captured information regarding the factors that are critical to completing the translation and adaptation process as intended. Fidelity of Implementation (FOI) was examined by tasks to be completed and criteria to be met. Using a FOI framework created for this study, I examined information from 100 documents, emails, interviews, meetings minutes, and surveys and coded results according to evidence type. The framework includes 18 tasks that countries were to complete and 11 criteria with which countries were to comply. More specifically, I analyzed results from Spearman correlations for CEBs between countries and examined the relationship between CEBs and DEBs across tasks and criteria. In addition, I computed a fidelity of implementation (F) coefficient based on CEBs and DEBs by task and criterion as well as on aggregate for each category. Qualitative analysis provided contextual information explaining the F coefficients obtained, the Spearman correlation results, and the CEB to DEB ratios computed.

Results from this study point at specific aspects of countries' political, educational, linguistic, and cultural contexts that shape FOI during test translation and adaptation. Based on these findings, I share three recommendations for future test translation projects.

Dedication

To every parent and child who set-off to new lands with dreams of a brighter future.

And, to every person who ever helped these courageous dreamers along the way.

We are indebted to them all.

Acknowledgements

I am indebted to the team members from each of the five participating countries involved in the AHELO study, the Organisation for Economic Cooperation and Development (OECD), the Council for Aid to Education, and the team of experts in the U.S. for allowing me to write about the work completed throughout the Assessment of Higher Education Learning Outcomes (AHELO) feasibility study. Specifically, Dr. Roger Benjamin, Robert Keeley, and Amy Kurpius provided answers, documents, information, and kind words throughout my dissertation process. Without your support, this dissertation would not have been possible.

I am very thankful to the members of my committee. The guidance provided by Dr. Kathy Escamilla, Dr. Bhuvana Narasimhan, and Dr. Lucinda Soltero-Gonzalez on this dissertation—and beyond— has been invaluable to me personally and academically. In addition to serving on my committee, Dr. Richard Shavelson personally facilitated conversations with OECD and CAE that resulted in my ability to focus my studies on the AHELO project. His reminder that research is an important component of test development work is a lesson he continuously teaches to all around him. I am especially grateful to Dr. Guillermo Solano-Flores for his constant guidance, patience, and support. He has taught me so much more than research and academic writing. I am forever indebted.

These committee members are representative of a larger support system that I found within the EECD community—and its extended family. In particular, Dr. Leonard Baca, Dr. Susan Hopewell, Dr. Rachel Kachchaf, Juli Sarris, and Chao Wang have gently challenged me intellectually, provided invaluable guidance, and never failed to put a smile on my face. Dr.

Ruiz-Primo (Ayita) is a shining example for all females of color who dare to walk down the path of educational research. I am humbled and inspired by her dedication to the work.

I would be remiss if I did not acknowledge the support of two colleagues who have provided many warm memories and clarifications of advanced psychometric concepts. Thank you Dr. Weeks and Nathan Dadey.

I also recognize and am grateful to Bill, Connie, and Ofelia for their generosity and dedication to excellence in education. The Miramontes Fellowship allowed me to complete my doctoral studies. I look forward to the community of scholars that results from their belief in and support of research in educational equity.

I will also be forever grateful to my family. My parents taught me the importance of education and the social responsibility that comes with having acquired a certain amount of it. Finally, I want to thank my husband, Brian Skahan, who inspires me to do my best always.

Table of Contents

List of Tables x

List of Figures xi

Chapter 1: Introduction 1

Chapter Introduction 1

Assessment, Culture, and Language in an Era of Globalization 2

Increased Western Influence 2

Increased Use of Large-scale Assessments 3

Translation and the Problem of Construct Equivalence 5

Translation and Adaptation Procedures in Two Major International Assessment Programs: TIMSS and PISA 8

Historical Background 8

Target Populations, Constructs, and Formats 9

Translation and Adaptation Procedures 11

Limitations of Current Translation and Adaptation Procedures 15

Translation and Adaptation Procedures in an Emerging International Assessment Program: AHELO 18

Historical Background 18

Translation and Adaptation Procedures 22

Intended Innovative Translation and Adaptation Procedures 25

Purpose of the Study 28

Relevance of the Study 30

Chapter 2: Conceptual Framework 32

Chapter Introduction 32

Sociocultural and Sociolinguistics Perspectives 32

Measurement and Student Diversity 36

Cultural Validity 41

Theory of Test Translation Error 43

Fidelity of Implementation **45**
Summary **50**
Chapter 3: Literature Review **51**
Chapter Introduction **51**
Cultural responsiveness in test development **51**
Test Translation and Adaptation Procedures **57**
Measurement Issues **63**
Fidelity of Implementation **72**
Chapter 4: Methods **74**
Background Information **74**
Participants and Setting **76**
AHELO 76
Countries: Economic Context for Education 78
Countries: Higher education systems 80
Countries: AHELO areas of expertise 84
Researcher's Role **86**
Data Collection and Recording Procedures **87**
Data Analysis **92**
General Structure 92
Analysis - Phase 1 ..... **93**
Analysis - Phase 2 ..... **102**
Significance **110**
Chapter 5: Results **111**
Task FOI **112**
Discussion of Task FOI 118
Task completion: Symmetry graphs **119**
Discussion of task completion: Symmetry graphs 125
Task Completion: F coefficients **128**
Discussion of task completion: F coefficients 133

Criteria FOI **135**
Discussion of criteria FOI 141
Criteria compliance: Symmetry graphs **142**
Discussion of criteria compliance: symmetry graphs 147
Criteria compliance: F coefficients **149**
Discussion of criteria compliance: F coefficients 153
Chapter 6: Summary and Conclusions **155**
Motivation **155**
Context **156**
Methods **157**
Summary Results and Conclusion **159**
Limitations **164**
Future Research **165**
Recommendations for Funding and Organizing Agencies **167**
References **170**
Appendices **188**
Appendix A: Country lists 189
Appendix B: List of 100 documents included in the analysis. 191
Appendix C: Excerpts of conference call questions related to translation and adaptation 214
Appendix D: Excerpts of the task adaptation open-ended survey 215
Appendix E: Excerpts of the 'Task Adaptation Site Visit Interview Guide 217
Appendix F: Exhaustive list of tasks and sub-tasks most applicable to FOI criteria 219
Appendix G: Condensed list of tasks and subtasks included in translation and adaptation 262
Appendix H: Description matrices. 272
Appendix I: Detailed qualitative data about each cell. 319
Appendix J: Three matrices for each country, one for each evidence type: confirming evidence bit (CEB), disconfirming evidence bit (DEB), and no evidence bit (NEB). 477

List of Tables

Table 1: *Cross Country Comparison of Overall Education Expenditures, Languages Spoken, and Literacy Rates*. 79
Table 2: *Sample List of Tasks and Subtasks, Document, and Country*..... 95
Table 3: *Sample List of Tasks Organized by Task and Subtasks, Document, and Country* 96
Table 4: *List of Final 18 Tasks* 97
Table 5: List of 18 Tasks Organized by Three Categories 98
Table 6: *List of 11 Criteria* 99
Table 7: *List of 11 Criteria Organized by Three Categories*..... 100
Table 8: *Information on Task CEBs, DEBs, and NEBs According to Three Cut-Off Points Across All Countries* 115
Table 9: *Task CEB, DEB, and NEB Completion Percentages, Means, and Standard Deviations* 116
Table 10: *Individual Country and Average F Coefficient by Three Task Categories* 130
Table 11: *Criterion CEB, DEB, and NEB Cut-Off Points by Category* 137
Table 12: *Criteria Compliance Percentages, Means, and Standard Deviations of CEBs, DEBs, and NEBs for Each Country and on Average* 138
Table 13: *Percentages of CEBs for Each Criterion Across All Tasks (Rounded Percentages)* 140
Table 14: *Percentages of DEBs for Each Criterion Across All Tasks (Rounded Percentages)* 141
Table 15: *Percentages of NEBs for Each Criterion Across All Tasks (Rounded Percentages)* 141
Table 16: *F Coefficient Across Countries by Criterion.* 150

List of Figures

Figure 1. Staged factors and implementation factors that affect the FOI for test adaptation and translation (adapted from Ruis-Primo, 2006). 47
Figure 2. Example of events across 11 criteria for Task 1 associated with one country 102
Figure 3. Sample CEB Matrix for one country 104
Figure 4. CEBs for each criterion across tasks. 105
Figure 5. DEBs for each criterion across countries 107
Figure 6. NEBs for each criterion across countries. 108
Figure 7. Correlation Matrix of CEBs Across Countries for Tasks. 118
Figure 8. Country A's percentages of DEBs and CEBs across tasks. 120
Figure 9. Country B's percentages of DEBs and CEBs across tasks. 122
Figure 10. Country C's percentages of DEBs and CEBs across tasks. 123
Figure 11. Country D's percentages of DEBs and CEBs across tasks. 124
Figure 12. Country E's percentages of DEBs and CEBs across tasks. 125
Figure 13. Percentages of DEBs and CEBs across tasks for each country 125
Figure 14. Correlation matrix of CEBs across countries for criteria. 141
Figure 15. Country A's percentages of DEBs and CEBs across 11 criteria. 143
Figure 16. Country B's percentages of DEBs and CEBs across criteria. 143
Figure 17. Country C's percentages of DEBs and CEBs across criteria. 144
Figure 18. Country D's percentages of DEBs and CEBs across criteria. 145
Figure 19. Country E's percentages of DEBs and CEBs across criteria. 146
Figure 20. Percentages of DEBs and CEBs across criteria for all five participating countries. 147

Chapter 1

Introduction

Chapter Introduction

This chapter is divided into five major sections. The first section—Assessment, Culture, and Language in an Era of Globalization—discusses the relevance of the connection between student linguistic and cultural diversity in international assessment practices. The first section also discusses the increase in western influence on education policies and participation in large scale assessments and how these changes are creating challenges to test construct equivalence. The second section includes information regarding two major international assessments: the Trends in Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA). The section addresses the two assessments' measurement goals, participants, and translation and adaptation procedures. The third section provides information about the assessment that is the focus of this dissertation: the Assessment of Higher Education Learning Outcomes (AHELO). After discussing AHELO's background information, the section explains the assessment's innovative adaptation and translation procedures. The final two sections include an explanation of the purpose and relevance of the study.

Assessment, Culture, and Language in an Era of Globalization

Increased Western Influence

Globalization has increased the flows of knowledge and ideas through dense networks that exist as part of modern social life (Tomlinson, 2006). Increased access to new technology, such as the Internet, facilitates a greater exchange of information between people of different countries and cultures. The amount of influence that each country exerts on the lives of others varies as a result of complicated international relationships reflected in this exchange. For example, the increased use of a specific language is, at least partially, a result of globalization and, at the same time, the reason for linguistic hegemony around the world. Those whose mother tongue is a global language, such as English, have greater opportunities to share ideas and influence others (Crystal, 2003).

Increased globalization creates an environment in which few languages and international organizations influence education practices across countries. The use of English as a primary language in academic communication exchanges and literature has created an environment in which it is easier for English proficient scholars to publish in journals, present at conferences, and, as a result, influence research practices and policy. Similarly, international organizations can set goals and criteria to evaluate meeting those goals whose influence increases as the number of affiliated countries increases.

The impact of increased globalization is particularly evident in international educational practices. Entities such as the Organisation for Economic Co-operation and Development (OECD) emphasize compliance with standardization of accreditation and educational programs (Luke, 2008; Singh, 2004). International organizations often use large-scale assessments to

examine country progress in complying with these educational programs. The OECD is a case in point. As a result of these trends, many OECD member countries participate in OECD's international test comparison studies and educational research programs.

As the number of countries working with these international organizations increases—and as they strive to meet stated goals—so does the homogenization of some international education conventions. The assessments implemented by additional diverse countries need to go through adaptation and translation procedures to ensure that they reflect local cultural and linguistic norms.

Increased Use of Large-scale Assessments

There are several international assessments whose results have an important impact in countries because they are followed by popular media and quoted by policy-makers. I will discuss four of them: TIMSS, PISA, the Progress in International Reading Literacy Study (PIRLS), and the Latin American Laboratory for Assessment of the Quality of Education (LLECE). Given their relevance to the AHELO study, TIMSS and PISA will be discussed in great detail later in the chapter.

The numbers of countries and educational jurisdictions participating in TIMSS have increased considerably over the years. In 1995, there were 41 participating countries. In 1999, 38 countries participated and in 2003, 49 countries participated (Mullis, et al., 2007). In 2007, 58 countries and educational jurisdictions (Appendix A) participated in TIMSS's fourth and eighth-grade level studies (Mullis, Martin, Ruddock, O'Sullivan, Arora, A. & Erberber, 2007). The countries are from six continents and contain a great deal of diversity between and within them.

There are different languages spoken, cultural norms followed, and educational systems represented.

Likewise, the number of countries participating in PISA increased across test administrations. Since its inception in 2000, the number of countries participating in PISA has more than doubled (cf. OECD, 2006; OECD, 2009). In 2000, 32 countries and economies participated. In 2003, there were 41; in 2006, there were 57; and in 2009, there were 67 (Appendix A).

PIRLS and the LLECE have each only been administered twice. The number of participating countries, provinces, and sub-national entities increased between the two administrations. In 2001, 35 countries and two Canadian provinces administered PIRLS. In contrast, 40 countries and five Canadian provinces participated in 2006 (IEA, 2001; Martin, Mullis, & Kennedy, 2006) (Appendix A). The first LLECE assessment, titled the LLECE-1st Study, took place in 1997 with 13 countries. The second LLECE assessment, titled Segundo Estudio Regional Comparativo y Explicativo (SERCE), was administered in 2007 (Solano-Flores & Bonk, 2008). Nineteen countries and one sub-national entity participated in the study (Solano-Flores & Bonk, 2008; LLECE, 2001) (Appendix A).

As the number of countries participating in international assessments increases, so does the diversity of the population tested. This diversity stems from the fact that in some countries, such as Luxembourg, there is more than one official language (Horner & Weber, 2008; CIA, 2011). It also stems from the fact that in other countries, there is a considerable variation in the student population's proficiency of the dominant language. For example, from 1990 to 2000, there was a 152 percent increase in the number of students enrolled in English as a second language programs in United States K-12 grades (Tong, Lara-Alecio, Irby, Mathes, & Kwok, 2008). There

are more than 4.5 million English-language learners (ELLs) in U.S. public schools (Lenski, Ehlers-Zavala, & Sun-Irminger, 2006).

It is within this context that organizations affiliated with tertiary education are turning to large-scale summative assessments to measure outputs and outcomes of higher education. There has been a rapid increase in the number of students attending institutions of higher learning as well as growing internationalization in higher education (OECD, 2008). There has also been an increase in the amount of time students are spending in school and the number of HEIs that students can attend (OECD, 2008). This has resulted in an increase in the costs associated with tertiary education for governments, students, and their families (OECD, 2008). At the same time, little research has been conducted regarding the learning outcomes of higher education (OECD, 2008). In response, national and international organizations—such as OECD—have embarked on creating a system through which they can produce instruments to measure, evaluate, and report on student achievement and HEI outcomes (OECD, 2008). As the use of large-scale assessment increases globally, test developers need to address construct and difficulty level equivalence across all student populations.

Translation and the Problem of Construct Equivalence

An abuse of testing is using a fallible test to make inferences about student ability (cf. Shepard, 1992; Linn, 2003). The inferences would be based on faulty data and could result in harmful decisions made about groups of students. The challenges in ensuring construct equivalence across languages increase if careful attention is not given to local cultural and language use. A test not properly adapted and translated will be biased and unfair for some students. To diminish challenges to validity, linguistic, cultural, and local test formatting and

administration differences need to be addressed. It is important that test development and implementation processes take into account different ways in which students learn and express knowledge and skills.

One important challenge to validity is the multidimensionality of language. Language consists of various linguistic attributes: syntax, grammar, discourse, register. Furthermore, each language's attributes vary according to dialect norms. Therefore, tests with large amounts of required reading may use and expect knowledge of a discourse style with which students may not be familiar. These aspects of language multidimensionality make test translation particularly challenging (Solano-Flores, Backhoff, & Contreras-Niño, 2006). If a difference exists between the linguistic attributes of a test and the characteristics used by its respondents, the possibility of bias against a particular student group increases (Solano-Flores, 2006). Student groups with different linguistic and testing needs are disadvantaged when participating in a test that contains language, test format, or administration practices with which they are not familiar.

As with language, item format plays a critical role in test validity. Research shows that specific types of multiple-choice items may favor some students who are more familiar with a particular format and pose excessive cognitive demands to students who are not familiar with it (Abedi, Hofstetter, & Lord, 2004). The same can be said about multi-part questions, open-ended items, and any format that deviates from the conventional way of asking questions in students' school contexts. In addition, a student's knowledge of strategies best suited for specific item formats has a direct impact on student performance (Martinez, 1999). At times, tests can be measuring a student's familiarity and proficiency in answering a specific type of item and not the intended construct. Studies indicate that an unfamiliar item format may prevent students from responding in a way that shows true ability (Basterra, 2010).

Another challenge to validity stems from the way in which tests are administered. As new technologies become available, new demands on schools and students arise (Bachman, 2000). Students may not be familiar with computers, calculators, or other devices increasingly used in testing. Unfortunately, test developers may wrongly assume that all students have the same kind of access to and familiarity with those devices. A student's inability to properly apply a calculator to test taking may be a reflection of socioeconomic class, not the student group's mathematics ability (Lukyx et al., 2007). It could be argued that the same pattern exists across nations: Not all countries have the same level of access to calculators and not all curricula view the use of calculators in the same way (Mullis et al., 2007).

The fields of psychometrics and test development are yet to address sufficiently all concerns emerging with testing increasingly diverse students. For example, the International Testing Commission's (ITC's) guidelines for test translation and adaptation (Hambleton, 2005)—a document that took over three years to develop—are silent about selection and screening of translators, the challenges of using new technology, or the multidimensionality of language. Similarly, although it addresses the importance of considering language background during test development, the Standards for Educational and Psychological Testing do not provide details to avoid bias due to language demands (AERA, APA, & NCME, 1999). Using expertise from several disciplines can help address some of these challenges. Combining information from sociolinguistics, psychometrics, and second language acquisition can help to identify, change, and improve current assessment practices used with linguistically and culturally heterogeneous student populations (Solano-Flores, Backhoff, & Contreras-Niño, 2006).

Translation and Adaptation Procedures in Two Major International Assessment Programs: TIMSS and PISA

Historical Background

TIMSS and PISA are important cross-national studies with wide-range implications for education policy and practice. TIMSS is produced in large part with the support of the IEA. PISA is developed with support from OECD. Both assessments deserve further examination for several reasons. First, both assessments address domains that have received a great deal of attention and funding over the last half century: mathematics and science. PISA also examines reading literacy. Second, both assessments are intended to address cognitive skills. Third, they receive a great deal of participating countries' media attention, influencing people's views of their countries' educational systems. Finally, these assessments capture information over several testing occasions providing longitudinal information on achievement (Mullis et al., 1996; OECD, 2000). Each test disaggregates information by student subgroups such as socioeconomic status and gender (Martin, Hoyle, & Gregory, 1996; OECD, 2003).

One of the most interesting features of the PISA and TIMSS programs is that they collect information about learning context—albeit to various degrees. Students, teachers, and principals respond to surveys on attitudes towards learning, domain curriculum, teacher experience, resources available, school organization, and home life (Mullis et al., 2007; OECD, 2000). In 2006, PISA included a parent survey as well (Murphy, 2010). Survey information provides a context for interpreting student results as school and home backgrounds impact student test performance.

Target Populations, Constructs, and Formats

Although some countries participate in both programs, the populations targeted and the sets of participating countries are not identical. TIMSS targets one grade in each of its three assessments: 4th, 8th, and 12th, which is more commonly known as TIMSS advanced (Mullis et al., 2007). PISA has one test that is aimed towards the population completing compulsory education, which PISA defines as the age of fifteen (OECD, 2000).

Periodicity and test format are different for the two assessments. PISA has been administered every three years since 2000 (OECD, 2000). PISA uses two types of multiple-choice items and three types of constructed response items. Students answer traditional multiple-choice items and complex multiple-choice items (i.e. True/False, Yes/No). Students must also respond to closed constructed response items, which are straightforward short verbal or numerical answers, and open constructed (extended) response items. The extended response items require longer verbal explanations or show of work for mathematics items (OECD, 2000). PISA assessments are administered inside a student's regular school and have no strong time constraints or requirements (Grisay, 2003).

In contrast, TIMSS is geared towards different grade levels; its structure is more complex than that of PISA. First administered in 1995, the 4th and 8th grade tests are given every four years (Ruddock, O'Sullivan, Arora, & Erberer, 2007). The 12th grade assessment was given in 1995 and then again in 2008 (Arora, Foy, Martin & Mullis, 2009). The make-up of the 4th grade test is 50 percent multiple-choice and 50 percent constructed response (Ruddock et al., 2007). The percentage of multiple-choice items increases slightly in 8th grade (Ruddock et al., 2007). The 12th grade TIMSS assessment also contains a combination of multiple-choice and

constructed response items (Grisay, 2003). However, technical documents made widely available do not provide the number of each type of item included.

During each assessment cycle, PISA addresses three main domains but emphasizes one on each occasion: reading literacy in 2000; mathematics literacy in 2003; science literacy in 2006 (cf. OECD, 2000; OECD, 2003; OECD, 2006). Reading literacy uses a variety of tasks and uses several text types. The tasks in reading literacy range from identifying and retrieving information to demonstrating broad understanding and ability to interpret text (content and features) (OECD, 2000). The types of texts include standard prose, passages, and other documents like lists, forms, graphs, and diagrams (OECD, 2000). Although it is clear that PISA aims to measure a range of reading abilities, depending on the country, some of the document types—such as graphs and charts—may be more likely to appear in other domains such as mathematics or science.

PISA mathematics literacy also measures a range of skills and competencies. Students must be able to complete standard mathematical computations as well as use mathematical thinking and insight (OECD, 2003). PISA addresses chance, change and growth, space and shape, quantitative reasoning, uncertainty and dependency relationships (OECD, 2003). Three constructs are addressed: algebra, number sense, and geometry (OECD, 2003). The age at which compensatory education begins and the curricula implemented vary by country. Therefore, students may not have learned all mathematics constructs included in PISA prior to test implementation.

PISA science literacy focuses on key scientific concepts from hard sciences such as physics (OECD, 2003). However, it emphasizes the ability to apply scientific knowledge. PISA measures the way students use knowledge to make decisions about the natural world (OECD, 2003). Students must be able to recognize scientific questions, use scientific evidence, and draw

scientific conclusions (OECD, 2003). Students must also clearly communicate these conclusions. In PISA, scientific literacy addresses life and health, earth and the environment, as well as technology (OECD, 2003). In addition to concerns about contextualizing knowledge and skills, questions also arise regarding the specific technology tested. Students need to have access to the technology referenced in the test. Also, teachers must have expertise in teaching the technology that supports student test-taking.

Like mathematics literacy in PISA, the advance TIMSS assessment focuses on algebra and geometry; but, unlike PISA, it also includes calculus (Ruddock et al., 2008). The science test focuses on several components within physics: electricity, heat and temperature, mechanics, and atomic and nuclear physics (Ruddock et al., 2008). The advance TIMSS explicitly focuses on measuring cognition as well. It captures information regarding student's knowing, ability to apply knowledge, and reasoning processes (Arora & Mullis, 2008).

Translation and Adaptation Procedures

Various international assessments use common terms to discuss test adaptation and translation procedures. However, the terms are not operationalized in the same way. It is important to examine the way that cross-national studies most relevant to this study and operationalize their adaptation and translation processes.

PISA Procedures. I have organized PISA's current suggested adaptation and translation procedures into six phases. During the first phase of test translation, work is completed at the international center organized by specialists at OECD. The main goal of the first phase is to create two versions of the test; one version is in English and the other is in French (Grisay, 2003). The outcome is two source tests. During this phase, the international committee also

creates a list of potential challenges to the translation process, notes on the properties of each item, and the overall goal of the test (Grisay, 2003). Translators use this information to make decisions within an accurate test context and with an understanding of the test construct.

In the second phase, the national teams translate each of the two source language versions and create two individual translations for each language of instruction in each country (Arora & Mullis, 2008). For each target language, two translations are created, one from French and one from English. Translation procedures associated with each of the two different languages can provide helpful insight to the other translation team.

During the third phase, the two translated versions of the test (from English and French) are integrated into one version of the translated test (Grisay, 2003). The process can facilitate discussion regarding decisions made during initial translation from source languages. By comparing the two translations with the source language version it is easier to ensure that meaning is preserved.

During the fourth phase, the team at the international center and members of the national team negotiate how to handle different translation needs. At the same time, the international center compares national versions against the two source versions (Grisay, 2003). By the fifth phase, all changes must be completed. The international committee completely reviews the final version of the test. The international committee verifies that any translation or adaptation changes that were agreed upon in previous steps were completed, that the assemblage of materials is correct, and that there are no issues with test layout or graphics (OECD, 2000; 2003). Reviewing these aspects of the translation helps ensure that the translation is accurate.

The sixth phase of PISA's adaptation and translation process involves a field test of the measurement instruments created (OECD, 2006; OECD, 2000). Field tests can help gather

information on vocabulary and situational appropriateness for each student group (Hambleton, 2005).

TIMSS Procedures. The International Study Center (ISC) recommends that each country implement five main activities during the translation process (Chrostowski & Malak, 2003). First, each country should identify the target language(s) for translation by examining languages used in classroom instruction (Chrostowski & Malak, 2003). The ISC suggests that the translation process include input by professionals from each dialect identified (Chrostowski & Malak, 2003). By working with the variety of dialects that students use this activity can help create tests that are responsive to local students.

For the second activity, the country teams should identify two independent translators. Countries participating in both the mathematics and science assessment need to indentify two translators with experience in mathematics and two translators with experience in science. If a country is unable to identify a translator with experience in the subject matter, the translator is to work with a content expert (Chrostowski & Malak, 2003). If a country is unable to find all of the translators needed to complete this activity, the National Research Coordinator (NRC) assists (Chrostowski & Malak, 2003). Once the translators are identified, the translators are to translate the test independently (Chrostowski & Malak, 2003). Including content expertise helps ensure construct equivalence between the source and translated versions of the tests.

Translators are given six main aspects of translation to consider: identifying and minimizing cultural differences; finding equivalent words and phrases; ensuring the reading level remained the same; ensuring the essential meaning of the text does not change; ensuring the difficulty level does not change; changing the layout as required due to translation (Chrostowski & Malak, 2003). During this third activity, translators can make adaptations as they encounter unfamiliar

contextual terms (Chrostowski & Malak, 2003). Acceptable cultural adaptations include punctuation and notation, units of measurement, proper nouns, common nouns, spelling, verbs unrelated to content, and usage (Chrostowski & Malak, 2003). Ultimately, the translations cannot change the construct, content, or the difficulty level of the source item (Chrostowski & Malak, 2003). Using the cultural adaptation form during translation verification, translators document all adaptations and translations (Chrostowski & Malak, 2003).

During the fourth activity, a translation review team examines and compares the two translations and creates one final version of the test (Chrostowski & Malak, 2003). If disagreements arise during the translation review process, the translators contact a third translation expert to help determine the best course of action (Chrostowski & Malak, 2003). During this activity, the independent verifiers of the translation review team compare the translated items against the source test (Chrostowski & Malak, 2003). The verification process also considers layout issues: instructions, items, response options, graphics, font, word emphasis, shading, page and item identification, and pagination (Chrostowski & Malak, 2003). Finally, during this activity the ISC, the national centers, and the international quality control monitors also review and verify the translations (Chrostowski & Malak, 2003).

Verifiers supply the NRC's with verification reports that include the necessary corrections or improvements (Chrostowski & Malak, 2003). Once the instruments have been translated and internally reviewed, the text of the booklet, cover pages, directions, instructions, item blocks, and background questionnaires are submitted for international translation verification (Johansone & Malak, 2007). The ISC reviews the updated tests before providing additional suggestions and giving final approval (Chrostowski & Malak, 2003). The verifiers document all changes on the translation verification report for each instrument (Chrostowski & Malak, 2003). Translation

errors receive a code that provides guidance for future actions. 'Code 1' indicates a major change or error. 'Code 2' is used to classify a minor change or error. If the translation was adequate but could be worded differently, the verifier noted it with a 'Code 3.' Finally, 'Code 4' is used to when designating acceptable changes (Chrostowski & Malak, 2003). Verification helps identify any translation errors and make changes that best meet the needs of the local student population.

TIMSS guidelines include a list of qualifications for translators and verifiers. Translators are expected to have an excellent knowledge of both English and the target language, experience in the country's cultural context, and, if possible, experience in the subject matter, preferably at the level of the target grade (Johansone & Malak, 2007). More specifically, the translators should have first-language experience in the target language, formal credentials as translators working in English, and live and work in the target country (Chrostowski & Malak, 2003). Translation reviewers are expected to have experience in students in the target grade (Johansone & Malak, 2007). Including experts familiar with how students learn can help make adaptation and translation decisions most appropriate for the local student populations.

Limitations of Current Translation and Adaptation Procedures

Limitations of PISA Procedures. PISA translation procedures have three main limitations: having only English and French source tests, ignoring diversity in dialects, and not addressing linguistic multidimensionality.

PISA creates two source tests—one in English and one in French, which creates a hegemonic relationship (Graddol, 2008) favoring English and French over other languages. Similarly, by focusing on translating the test to the official standard versions of languages the guidelines ignore linguistic diversity within each language. What is deemed a standard language is simply

the dialect accepted and used by the dominant group within a community (Solano-Flores, 2006; Woolard, 1985). Ignoring dialect differences can be one of the most detrimental decisions made during test adaptation and translation (Hambleton, 2005; van de Vijver & Poortinga, 1997). Translating a test into one dialect can create a measurement tool that assesses student familiarity with and proficiency level in that specific dialect. The result can be a biased test that behaves differently across student groups of equal ability.

Other challenges to the translation process pertain to the negotiation process that takes place when competing ideas come about during the translation process. There are always tensions between each local country's needs and the intentions of the test development team. Curriculum, assessment practices, and communication styles may differ. There are also linguistic tensions that arise during test translation. Without clear information regarding PISA's handling of these competing needs, it is difficult to know how this challenge is met. For example, the source test may contain item formats with which local students are not familiar. Another issue that is likely to surface is a country's preference or reluctance to use new technology available for testing (Bachman, 2000).

Limitations of TIMSS Procedures. As with PISA, there are limitations in the TIMSS translation and adaptation process. TIMSS guidelines do not make clear how experts make decisions or the extent of each expert's involvement throughout the process. In addition, the review process may contribute to low FOI.

Several challenges to implementing the guidelines can be traced to the lack of clarity regarding several steps. For example, although TIMSS asks each country to identify local dialects, it is unclear who should complete this task or how the person(s) should identify the different dialects. Textbooks, common tests, and general curriculum may not reflect a

community's local dialect. Also, Browne Global Solutions-and Capstan (translation companies located respectively in England and Belgium) help the IEA Secretariat develop and manage the translation verification process (Chrostowski & Malak, 2003). However, it is unclear who makes final decisions when disagreements about the translation occur. Finally, TIMSS public documentation does not operationalize the four translation verification codes. The documents do not include information about how the verifiers involved in this process dealt with different opinions regarding the coding of an error. Available TIMSS documents should provide clarification regarding these aspects of the translation process.

Although translators' roles are fairly clear, the guidelines do not address two translation challenges that are certain to come about. First, it is unclear how or the degree to which the measurement expert takes part in the adaptation and translation. A local measurement expert would be best equipped to determine translation issues that may affect an item's validity. Second, the guidelines fail to address language multidimensionality. The tension between language dimensions—such as discourse, register, and style—challenges translators because they may require competing translations. Using technical register to follow content norms may cause difficulty with local discourse conventions.

The level of fidelity of implementation of the staged translation and adaptation procedures depends on the precision of the procedures, the material being translated, and the country context. Revisions for background questionnaires are usually identified and completed after the field tests are completed. As a result, it is necessary to conduct a second major translation effort with the questionnaires that does not take place during the translation of other materials (Chrostowski & Malak, 2003). This complicates the translation and translation verification process. Also, the complexity of implementing the translation guidelines grows as the number of

documents requiring translation and the number of countries with multiple dialects increases. As a result, the level of FOI has a tendency to decrease (Lynch & O'Donnell, 2005).

One of the greatest challenges with international test translation and adaptation is the ability to monitor its implementation. TIMSS and PISA provide countries with guidelines and strongly recommend that countries follow them. In fact, the procedures call for countries to report on their progress during the translation and adaptation process. However, there needs to be flexibility in the extent to which countries must follow each step of the guidelines and report on their implementation. Not all countries have access to the types of expertise demanded or the funds necessary to secure the expertise. Guidelines need mechanisms to monitor how the translation procedures are implemented in each country.

Translation and Adaptation Procedures in an Emerging International Assessment Program: AHELO

Historical Background

CLA. Richard Shavelson at Stanford University, Steve Klein at RAND, and Roger Benjamin at the Council for Aid to Education (a subsidiary of RAND at the time) created the Collegiate Learning Assessment (CLA). Since its inception over seven years ago, different types of institutions in the United States have implemented one of several versions of the CLA (Shavelson, 2010b). The number of participating high schools administering the CLA-based assessment increased from 10 in 2007 to 65 in the 2010-2011 school year (Shavelson, 2010b). In the 2011-2012 academic year, 300 institutions—including high schools, colleges and universities—administered the CLA (Keeley, personal communication). Given its success in the United States, it is not surprising that CAE received requests to implement the CLA

internationally. Of the 300 institutions that used the CLA during the 2011-2012 school year, eight were Canadian, one was in Australia, one in Abu Dhabi, one in Hong Kong (Keeley, personal communication).

The CLA was designed to measure critical thinking, analytic reasoning, problem solving, and written communication skills (Klein, Benjamin, Shavelson, & Bolus, 2007). The CLA's analytic writing tasks measures student ability to make and argument and critique an argument (CAE, 2008). The CLA's performance tasks place students in a real-world scenario seeking to measure analytic reasoning and evaluation, problem solving, writing effectiveness and writing mechanics (CAE, 2008). Students are expected to analyze and synthesize evidence presented in the PT and construct cogent arguments (CAE, 2008). Participating students from the initial five countries included in the AHELO feasibility study will be working with the CLA's PTs. Each PT contains general instructions, an introduction, and a unique library of documents. Using a propriety web interface, students are to gather information from the library of documents to complete the task (Benjamin, 2005). The document library includes letters, memos, summaries of research reports, newspaper articles, maps, photographs, diagrams, tables, charts, and interview notes or transcripts. Students need to evaluate the integrity of each document and identify credible and questionable assumptions made in each. In addition, students must justify their point of view with evidence from the documents provided. Each participating institution determines when during the school year they want to administer the assessment as well as the number of testing occasions in which students will participate. Students may take the assessment in the first year, last year, or at different times of their baccalaureate studies (Benjamin, 2005).

Emergence of AHELO. In late 2009, OECD contracted CAE to adapt and translate an international version of the CLA: the Assessment of Higher Education Learning Outcomes

(AHELO). In the context of international comparisons, no precedent exists of a performance-based assessment in higher education. Due to the unique characteristics of the assessment, the AHELO translation and adaptation procedures were intended to be more rigorous than those that have been in place for other international assessments. Therefore, CAE conducted a study to examine the feasibility of translating and adapting the CLA to languages other than English and cultures of countries other than the U.S. Although there was no intention of comparing countries based on AHELO results, one of the study's goals was to compare like institutions across countries. The long-term goal of this study was to put in place a performance-based, computer-administered assessment tool for international test comparisons at the higher education level.

AHELO is composed of three assessments: a generic strand (Module A), an economics strand (Module B), and an engineering strand (Module C) (Shavelson, 2010b). A contextual dimension strand (Module D) includes surveys distributed to students, faculty, program leadership and institutional leadership affiliated with participating institutions of higher learning (Lalancette, 2010). The three assessment strands consist of open-ended prompts making up PTs and analytic writing tasks (CAE, Architecture). Analytic writing tasks prompt students to make or critique an argument. PTs use complex real-world problems and ask students to make a recommendation or decision, reach a definitive conclusion, or solve a problem (Shavelson, 2010b).

Given the complexity of AHELO, various organizations shared the AHELO feasibility study responsibilities. CAE and the Australian Council for Educational Research (ACER) agreed to work on different tasks, yet, provide support to the country teams (Lalancette, 2010). In addition, OECD staff was responsible for helping countries acquire funds for the project (Lalancette, 2010).

As part of the instrument development process, for which CAE was responsible, each country selected two performance tasks, out of nine, that were best suited for their student population. Countries chose the two tasks they found most easily adaptable and translatable using guidelines CAE provided. During the initial "kick-off" meeting in New York, countries presented their choices, through consensus, chose the final two performance tasks that would be used in the study.

CAE was responsible for creating the adaptation and translation procedures as well as implementing the procedures with the two Module A PTs. CAE shared its adaptation and translation procedures with ACER, who was responsible for the other four modules. It is important to note that, although these organizations provided guidance and support, each country team was ultimately responsible for the adaptation, translation, and administration of each module.

Though team members had received literature concerning the PTs and translation recommendations, country teams began work on the AHELO feasibility study during a meeting in New York City that took place the second week of February 2010. During this meeting country teams shared thoughts about each of nine available PTs. One focus of the meeting was on choosing two PTs that were most culturally appropriate for all of the countries. Each country team shared any concerns they had about content or context of specific PTs. Team members also discussed reasons why they favored the use of specific PTs. The five country teams were able to decide on two PTs that could be adapted and used as common assessments.

All country team members reviewed and trained on adaptation and translation procedures. Team members worked together to discuss terms, context, and format. Country teams then proceeded to share country specific concerns and decisions made to deal with them.

Measurement experts helped country teams understand how changes could be made that could work for all countries. The assessment experts also explained why certain changes could not be done. At times changes would cause changes to the construct. Other changes were not relevant or fair to all countries. Country teams expressed that they understood the complexity of adaptation and translation and were ready to complete the task with translators.

Translation and Adaptation Procedures

Unlike other assessment procedures, AHELO's translation process is in large part based on the theory of test translation error (TTTE), which demands a more rigorous test translation process than PISA or TIMSS. The TTTE focuses on finding disconfirming evidence of adequate translation and addressing the fact that error-free test translations are not possible, though translation error can be minimized (Solano-Flores, Backhoff, & Contreras-Niño, 2006). As a result, the TTTE recommends that translation teams be aware of the number of errors and the severity of each. The errors should not cause biased items or tests that are not construct equivalent.

The procedures are also based on work used by other cross-national comparison studies. Six categories can help countries identify issues during adaptation: cultural differences, linguistic and cultural appropriateness, familiarity with computer-based assessment, context appropriateness, cognitive and linguistic equivalence, and appropriateness of procedures (Hambleton, 2005).

The PTs included in AHELO's general skills module were originally developed to measure learning among English speaking students attending college in the United States. At the foundation of each PT there is a real world problem to which U.S. college and university

students can relate. Therefore, each of the two PTs chosen by countries participating in AHELO had to go through individual adaptations and translations in five of the participating countries. The AHELO translation procedures used in Module A have crucial activities not included in other international assessment translation and adaptation procedures. AHELO divided adaptation and translation into two distinct yet interconnected phases.

AHELO also includes two translation verification procedures and talk-alouds. During talk-alouds researchers can gather real time-time information by asking students to share their thoughts while completing a task (Johnstone, Bottsford-Miller, & Thompson, 2006). Furthermore, researchers can ask students follow-up questions once they complete the task (Johnstone et. al., 2006). Talk alouds would provide valuable information about the ease or difficulty caused by translation. The information captured during think alouds could provide insight into potential required revisions. The second translation review would give country teams an opportunity to verify revisions made based on think aloud information.

Upon choosing appropriate PTs, each country was to adapt each task, response format, and scoring rubric to ensure a high quality national version. While following the guidelines, the adaptation team members looked for cultural differences between the original and target populations. They were also to provide evidence that the language, response formats, scoring rubrics, and directions were appropriate for the target population (CAE, GS.4, 2010). [1] In addition, they were to provide evidence that students would be familiar with computer based testing and PT context. Finally, the members in charge of adaptation were to provide evidence of linguistic and cultural equivalence of the source and target versions of the tasks.

[1] The general strand (GS) CAE documents developed for the AHELO project are cited following the format, GS.nn, in which nn denotes the ordering which a given document was created.

Based on information published by the International Test Commission (ITC), CAE also provided guidelines for the adaptation of computer-administered tasks. This was particularly important because the CLA is computer administered. The guidelines included twenty-three questions addressing available technology, control of computer administration, and security and privacy of the test and responses. Teams could assess their technical preparedness and needs by answering questions regarding staff knowledge, available technical infrastructure, and overall test security.

Once each country team completed the adaptation process, they were to submit their changes to CAE staff. The CAE team reviewed change requests to ensure that the task's construct was not changed and to ensure that a change that could help one country did not adversely impact another country. The CAE staff informed country teams of the changes that were possible or explained why a change could not be done. Country teams, incorporating the feedback that they received from CAE staff, produced and submitted the final adapted PTs.

The next major step consists of a number of activities (CAE, GS.14, 2010). First, each country team was to give the adapted PT to two separate and independent translators, Independent Translator 1 and Independent Translator 2. Each translator created a translated version of each PT according to guidelines that CAE provided. The translation process included two independent translations and a translation reconciliation step (CAE, 2010, GS.13). As a result of translation reconciliation each country team would complete one translated version of each performance task.

Upon completion of the initial translation, each of the independent translators submitted their translated tasks to the Translation/Adaptation Advisor. Along with the two translators, the advisor was to help reconcile the two translated versions of each of the two PTs. The guidelines

for the translation review asked the translators and advisor to examine errors related to: task layout (e.g., style, format, and local item conventions), characteristics of language (e.g., including grammar and syntax, semantics, and register), and content and culture (e.g., type of information included in the task, ensuring that the construct being measured remains the same while being sensitive to local curriculum and culture) (CAE, GS.4, 2010). Finally, the review and reconciliation of the translations should keep in mind any errors that were present in the source measurement tool.

The reconciled version of each task was then shared with two groups. First, the advisor shared the reconciled version with the rest of the country team members to prepare for the think alouds, which is the next activity in the AHELO translation process. After the initial translation, the translation advisor also submitted the reconciled version to an OECD-designated agency for translation verification (CAE, GS.13, 2010). The OECD-designated agency consists of an independent team of experts. This team looked for confirming evidence of translation appropriateness to assure quality control. The team verified each national version of each PT against the source version. Traditionally, this is the last step implemented in international assessment adaptation and translation procedures.

Intended Innovative Translation and Adaptation Procedures

Using think alouds, the PTs were then tested with a sample of students that came from each country's target population. The think alouds were added to gather information directly from students concerning the appropriateness of the translation. Through think alouds, the interviewers were able to learn about the translated tasks' readability, comprehensibility, and meaningfulness (CAE, GS.13, 2010). In addition, the think alouds could help verify that the

construct had not changed through the translation process (CAE, GS.13, 2010). Each country selected students to participate in the think alouds. Each student would respond to one PT, though both PTs went through the process. Initially, students were to share their thought process while responding to the PT. Upon completion of the PT, a country team member asked every student the same questions regarding the clarity, relevance, and interest level in the PT. The interviewer took note of all observations and responses on forms provided by CAE. The information captured would be reviewed and, if appropriate, used to revise the PTs.

Guidance CAE provided to conduct think alouds was adapted from the seven elements of universally-designed assessments proposed by Thompson, Johnstone, and Thurlow (2002). These guidelines addressed the inclusivity of the participant population, precision of the constructs, accessibility of the tasks, appropriate accommodations, clarity of test procedures, readability, and legibility (CAE, GS.4, 2010).[2] During think alouds, countries were to include a representative sample that took into account gender, socio-economic status, locale, type of institution, geographical area, ethnicity, and familiarity with computers (CAE, GS.4, 2010).

In addition, CAE created three worksheets to gather and organize information, a protocol script, and a training video specifically for the think aloud activity (CAE, GS.37, 2010). One worksheet allowed the interviewer to note the location in the PT during which the students behaved unexpectedly, the observations made by the interviewer, and possible follow-up questions. The second worksheet consisted of five questions that all students would answer. The questions addressed the task's instructions, the PTs, the student's decision making process, and

[2] The concept of test design appropriate for all students brings attention to important issues of validity. However, it is important to note that the goal of equitable testing across diverse student groups is not based on using the same test but instead symmetrical tests—each appropriate for each student group. Additional, more concrete, information should be included in the guidelines to improve universal design implementation.

the strategy employed when working through the task. In addition, the worksheet gathered information about how engaging the student found the task. The last worksheet helped the interviewer list proposed modifications as well as justification for them. Country teams were to share findings acquired through the think alouds with CAE staff. Focusing on fidelity to the construct being measured and the needs of each country, CAE and country teams could make changes to the translated PTs.

CAE's guide calls for a second translation review process after completion of the think alouds and subsequent revisions. The second translation review is a multidisciplinary approach that allows for the examination of disconfirming evidence of translation appropriateness. The focus is on finding any of several types of translation error: omission, insertion, alteration, inconsistency, inappropriate/imprecise, combination/conflation, substitution, or multiplicity (CAE, GS.13, 2010). The translation advisor, an assessment expert, and a third independent translator are to negotiate the requested changes and make final decisions keeping in mind the severity and number of errors found (CAE, GS.4, 2010). The country teams send the final versions of the PTs to CAE in preparation for the tasks' administration. A third party will review each of the country's translation (CAE, GS.14, 2010).

The PTs, the scoring rubric, and the computer platform and interface language go through the full process of adaptation and translation (CAE, GS.30, 2010). They should successfully pass through task adaptation, task translation, translation reconciliation, translation verification, talk-alouds, and a translation review. A mini PT created for tuning purposes, materials for the think alouds, the scoring handbook charts, the administrator manual, and the scorer training materials

underwent a dual translation (CAE, GS.30, 2010).[3] The dual translation had the two original translators reconcile their individual translated versions of each document.

CAE also provided country teams with recommended qualifications for the translators (CAE, GS.13, 2010). CAE categorized specific qualifications as indispensable given the scope of the project. CAE recommended the use translators who had obtained a national-language translation certificate by a professional translators organization. It was also important that each translator be a native speaker of the national language. Lastly, the procedure stressed that each translator should have ample experience performing translations. Other qualifications were deemed desirable. If possible, translators should have experience as translators of tests, educational material, and/or higher education documents.

Purpose of the Study

Despite the extensive staging that CAE created for the adaptation and translation process, the extent to which each country was able to implement the steps is unclear and requires further investigation. Similarly, although CAE provided a description of the responsibilities for each country team member, it was not clear that all country teams would be able to adhere to the suggestions. The extent to which countries implemented CAE's suggestions could impact the adaptation and translation process. As a result, the fidelity with which countries followed the guidelines is addressed in this study.

[3] Country teams expressed concerns about student unfamiliarity with the performance task format and requirements and the impact that it can have on test results. CAE created a mini-PT that all student participants could complete to become familiar with the PT format prior to participating in the AHELO study.

As a graduate assistant, starting in early 2010, I participated in a project whose goal was to provide technical assistance to CAE during AHELO's Module A adaptation and translation stage. I worked closely with the CAE assessment expert, the project manager for CAE, and an international assessment expert who specializes in issues of test adaptation and translation. I participated in meetings and communicated with the country teams. In this capacity, I was able to gather information relevant to how country teams implemented the CAE guidelines addressing aspects of the adaptation and translation procedures for the PTs included in the general skills strand.

This dissertation will document and examine the way that CAE staged the processes of task adaptation and translation for AHELO and the ways in which countries participating in an initial phase (a feasibility study) of the development of this assessment system interpreted and were able to implement the rigorous adaptation, translation, and translation review procedures designed for this endeavor. Specifically, I will examine the level of FOI attained or CAE's rigorous translation/adaptation procedures. Given that AHELO is a performance-based, higher education, computer-administered emerging international assessment, two research questions are posited:

- *How did participating countries vary as to the fidelity with which they implemented AHELO's translation and adaptation procedures?*
- *Based on the lessons learned, how can assessment translation and adaptation procedures be improved to ensure feasibility and validity across languages and cultures in international assessments?*

Because of the unique characteristics of the CLA, a study on the translation and adaptation of its tasks is an important contribution to the existing literature on test translation. First, unlike other international tests, such as TIMSS and PISA, the CLA consists of PTs and not multiple-

choice or essay tasks (Mullis et al, 2007; OECD, 2006; Shavelson, 2010). In CLA tasks, students are asked to read, interpret, and use diverse information in their written responses to several real-world situations. Second, the target population is also different. TIMSS measures elementary, middle, and high school students' proficiency levels in specific domains: mathematics and science. PISA measures proficiency in reading, math and science literacy among 15 year-old students (OECD, 2006). In contrast, CLA PTs assess students who are in their last year of studies at an institution of higher learning for their capacity to use, apply and act on their knowledge and reasoning. Finally, the way in which the PTs will be administered is also unique. Unlike other international tests, which use traditional pencil and paper, the CLA will be computer-administered.

Relevance of the Study

For the last fifty years there has been an increase in the number of studies and publications regarding adaptation and translation of measurement tools (Stansfield, 2003). Experts in evaluation, psychometrics, sociolinguistics, education, and cognitive psychology have written about the need for greater cultural responsiveness and linguistic sensitivity (Escamilla, 2000; Solano-Flores, 2010; Woolard, 1985; Hambleton, 2005). The majority of education publications dealing with this topic have focused on test reliability and validity. However, research often fails to discuss salient details on how to achieve culturally and linguistically sensitive tests. Current literature does not address adaptation and translation process with the level of rigor attempted in the AHELO study. In addition, no precedent exists on international assessments for higher education students, the use of PTs, and computer-based administration.

This study will contribute with a systematic approach to examining the complexities of linguistic and cultural diversity and their challenges to fidelity of implementation (FOI) in test translation and adaptation. The data analysis framework created and implemented in this study can serve as a model for other cross-national studies. Test developers can examine the FOI for their test translation and adaptation procedures. Furthermore, the lessons learned via this study can result in increased validity across languages and cultures in international assessments.

Chapter 2

Conceptual Framework

Chapter Introduction

Several disciplines and theoretical perspectives contribute to the conceptual framework that I am applying to this dissertation. I incorporate research from sociocultural theory and sociolinguistics relevant to testing, linguistic diversity and measurement, cultural validity, the theory of test translation error, and the importance of examining the FOI of staged test translation and adaptation. Though international assessments do not include emerging bilingual students in the United States, I include studies on English language learners because research from bilingual education focuses on addressing cultural and linguistic diversity. This multidisciplinary approach—which includes research findings from cognitive science, sociolinguistics, psychometrics, and program evaluation—creates a foundation for studying and possibly improving current translation and adaptation practices in international assessments.

Sociocultural and Sociolinguistics Perspectives

The conceptual framework is strongly grounded in sociocultural theory and the constructivist epistemology's premise that knowledge is constructed through life experiences and interactions (Crotty, 2003). The framework acknowledges the impact that culture and language have on student development and test performance.

The society in which children live shapes their mind. Students from different cultural backgrounds can develop different forms of reasoning (Cobb and Bowers, 1999). In formal schooling, students are exposed to symbolic tools—of which language is one—that help an individual's cognitive development and create the individual's reality (Egan & Gajdamaschko, 2003; Vygotsky, 1978). Performance within an academic setting is innately based on people's experiences with the community's social tools and norms. Social forces—such as teachers, parents, peers and community—have an impact on a child's development and learning (Kozulin, Gindis, Ageyev, & Miller, 2003; Chaiklin, 2003). The language people use will mediate the meaning made from an interaction as well as the interaction itself.

Language also reflects different epistemologies present across diverse cultural and linguistic groups. The words that students choose to use during communication are representative of the larger world-view to which a student prescribes (Smagorinsky, 1998; Vygotsky, 1978). Students' linguistic backgrounds will influence the way in which they respond to a question or stimulus (van de Vijver & Poortinga, 1997). Yet, during communication, students as well as test developers assume that the other party involved is following certain conversational maxims and creating the same meaning (cf. Wolfram, Adger, & Christian, 1999). Choices made during communication are based on and reflect the local dialect with which students are most familiar. Yet, little work has been done involving the different dialects that students use and the impact of assessment.

An assessment originating in one community is fundamentally based on that community's culture and language (Greenfield, 1997). Identical stimuli posed to diverse students will not necessarily obtain the same response despite ability (Cole & Bruner, 1971). To a large extent, the culture that generates a measurement instrument determines the goals, methods, discourse, and

format used in that instrument. When standardized exams do not reflect local community context, test results may not reflect a student's true ability (Trumbull & Solano-Flores, 2000). If an assessment includes cultural and linguistic attributes with which students are not familiar, the test may fail to measure the intended construct. A test measures, at least in part, student familiarity with the exam's culture and language (AERA, APA, & NCME, 1999).

Too often, test development fails to address the many dialects used within and across participating communities. Although there is often a dialect used for business and government interactions, it may not be the dialect of instruction that every student experiences. Each person involved in the testing process has his own language or dialect (Lukyx, Lee, Mahotiere, Lester, Hart, & Deaktor, 2007). Therefore, a student may use different grammatical structures, academic and non-academic terminology, discourse conventions, as well as idioms and colloquialisms from those found in an assessment (cf. Solano-Flores, 2006, 2009; Solano-Flores and Li, 2008).

Current translation practices stress the importance of achieving equivalence across the source language and the language of the targeted student population. Several measurement experts have published guidelines for linguistically responsive test translation practices. For example, the International Test Commission (ITC) recommends incorporating the expertise of translators who are knowledgeable of the source and target languages and cultures (Hambleton, 2005). The ITC guidelines also advise that the translators be trained on the test construct and goals. Indeed, the guidelines encourage that the translators participate in test development.

However, most literature does not include information regarding ways to address dialects during test translation. Rarely will a single translator possess detailed knowledge of various dialects. More often, translators are well versed in the dominant dialect. In fact, professional translation associations (e.g., American Translators Association, 2011) usually provide

certifications in one dialect of a language—the standard dialect. This creates additional challenges for test development organizations that attempt to find translators familiar with linguistic variation within and across communities. Current guidelines do not address the additional human and material resources necessary to accurately reflect the different dialects that students use. The inattention by researchers, test developers, and test users towards dialects is one of the major current problems with test validity and fairness (Hambleton, 2005).

People have their own views towards language even if they do not openly acknowledge them (Greenfield, 1997). This is true for test developers and the institutions for which they work. It is important to be aware of the strengths and limitations of their own perspectives (Greenfield, 1997).

Sociolinguistics provides two main orientations towards language (Ruiz, 1984) that are most appropriate to development of tests for cross-cultural studies. One view treats language strictly as a tool for basic communication (Tauli, 1978). This orientation focuses on efficiency and clarity in the use of language. Test developers that ascribe to this view of language would focus on developing one test that uses a 'superlanguage' that all students would need to know. This orientation ignores the cultural distinctness of language and demands that students be familiar with the superlanguage. The second orientation emphasizes language as a means for self-expression and self-identity (Kelman, 1972). This view acknowledges individual and group diversity in the creation and use of language. Test developers prescribing to this view would argue for simultaneous test development and no source test.

The treatment of language and culture reflects the power dynamic between languages and different types of cultural knowledge (cf. Foucault, 1989). The use of specific cultural norms and communication styles in testing is so prevalent that language proficiency exams include

questions on 'academic culture' (Aukerman, 2007). Students who come from a home environment that reflects the dominant culture are at an advantage (Gordon, 2008). These students will be well-versed in communication expectations, response styles, and cultural norms often valued in schooling experiences and reflected on assessments. Favoring one student group over another contributes to construct irrelevant variance and compromises the validity of the test.

In sum, in order to measure a student's proficiency within a domain, the test should not favor one student group's language and culture over another group. A test that only takes into account the dominant language and culture is merely perpetuating a system of inequality that favors few communities. Applying sociocultural theory and literature from sociolinguistics, test developers can better address the ways that linguistic inequity negatively impacts testing.

Measurement and Student Diversity

When making inferences about student ability based on test performance, it is essential that the test behave equivalently across student groups. The meaning of test scores should hold across different populations, settings, or contexts (Messick, 1995). Increased linguistic and cultural diversity in international assessments necessitates additional attention to test validity and fairness. Fairness can be described in terms of inferences drawn from test scores or test items in evaluating or selecting tests or test takers (Camilli, 2006). Bias relates to a test's or item's favoring of one student group over another, holding capacity constant. To discuss important test properties, three psychometric aspects of testing can serve as a framework: construct bias, method bias and item bias (van de Vijver and Poortinga, 2005).

Construct bias can result from variation in construct across student groups (van de Vijver and Poortinga, 2005). Construct bias can be related to the use of one construct definition despite a

difference in the interpretation and use of the construct across countries. Test developers may assume a specific definition is followed across all countries because of an ethnocentric view of the construct (Tucker, 2003). This can be connected to English and Western hegemony (Garcia, McKoon, & August, 2009). For some developers, it may not seem necessary to ensure that the way in which the construct is used on the test is the way in which it is addressed in classrooms across different student groups.

Construct bias can also be related to a test's limited coverage of the construct (Embretson & Reise, 2000; van de Vijver & Poortinga, 2005). Key aspects of the way in which students understand and express their knowledge of a construct may not be included. When student participants represent vast cultural diversity, construct underrepresentation can become a problem. Some students will not have an opportunity to demonstrate accurately their proficiency level. As a result, the test items may be biased and the test deemed unfair.

Method bias focuses on aspects of test implementation and other test properties that are not the object of measurement but can impact student performance (van de Vijver and Poortinga, 2005). This may result when the demands posed by a measurement tool are different from the practices with which students are familiar. For example, student response formats differ across and within countries. In some communities, paper and pencil may be the normal testing format while in other countries students may be accustomed to oral examinations. In addition, as there is an increase in the amount of technology available for testing, there are different demands made on schools and students (Bachman, 2000). Many students may not have access to the different test formats and technology increasingly used for test administration.

There may also be challenges in test administration (van de Vijver & Poortinga, 2005). Method bias due to test administration can result from several causes. Bias can occur when

instructions for test participation are not accurately conveyed due to challenges with the proctoring protocol itself. It can also arise because of linguistic or cultural differences between proctor and students. In international testing an additional source of bias from test administration can result from a strict test protocol that does not follow local norms. If test development uses testing norms only found in culturally and linguistically dominant groups, the possibility of method bias increases.

Item bias, the third type of bias discussed in this section, deals with how items behave across students. Items that behave differently across groups of equal ability levels are biased (van de Vijver and Poortinga, 2005). The item's format and context are variables that must be examined to ensure cultural responsiveness. Research has found large, statistically significant differences in performances across cultural groups on tests. A specific study showed that many items were biased in favor of white students (Mercer, 1984). The study found that although students of different cultural groups were considered to have equal ability levels, they performed differently on the items. White students outperformed their non-white peers despite other evidence that both groups were of equal proficiency levels.

Item format—such as the use of bold, italics, and hierarchical multipart questioning—can also have a different impact on student performance. Test scores are a function of items, the test respondent, and the assessment context (Messick, 1995; Cronbach, 1971). Item format and student background can impact the way in which students interact with a test item (Duran, 1983). As a result, the type of item can favor one student over another. Indeed, this is one reason why it is important to include different types of items on a test (e.g., multiple-choice items, ordered multiple-choice items, simple constructed response items, or extended constructed response items).

In an attempt to contextualize the test items, test development may create an item that favors some students over others. Grounding should be done in a way that is culturally appropriate for each student group (Lukyx et al., 2007). If not, the context may be a distraction or source of confusion for some student groups. Translation and adaptation are closely related to this type of item biased because a specific word can have different meanings across countries that use the same language. For example, the term *aggressive* can mean preparedness and determination in one country while being associated with an unprovoked attack in another. The adaptation and translation process must account for these contextual differences across regions and countries.

The psychometric tools most commonly used to examine bias in large-scale assessments come from item response theory (IRT). If mean test scores vary across groups of similar ability (e.g., males and females), this can be an indication of a problem with test validity. In this case, differential prediction in IRT can help determine whether test scores are invariant (Camilli, Briggs & Sloane, in press). For individuals from two groups with comparable skills on the construct measured invariance means that average item scores are about the same (Camilli et al., in press). When using IRT to scale student and item performance, researchers can use differential item functioning (DIF) to detect those items that behave differently across comparable individuals from different groups with similar ability thereby showing variance (Wilson, 2005).

Though widely used, there are limitations to DIF analysis. Although DIF analysis detects bias, it does not identify the source of the bias nor does it explain why items are behaving different than expected (Ercikan, 1998). Researchers can further study measurement bias by examining external relationships after detecting the possibility of DIF for an item (Zumbo, 1999). External evidence warrants examination of scores across various groups and cannot be done based on DIF. DIF analysis alone does not allow detection and examination of subtle

differences connected to language and culture. Lastly, because DIF requires a large number of items to produce accurate statistics, DIF cannot be used with many performance assessment tasks. Although multiple-choice tests often contain over twenty items, performance assessments contain much fewer tasks.

If the focus of research is to determine the possible causes of student score measurement error, generalizability (G) theory can prove useful. G-theory is a psychometric tool that focuses on measurement error and its sources (Cronbach, Gleser, Nanda, & Rajaratnam, 1971; see also Brennan, 2000; Shavelson & Webb, 1991). G-studies can help explain how well scores can be generalized across persons, groups, and settings. Typically, researchers conduct G-studies to determine how facets (factors) such as test item, testing occasion, rater, and the interactions among these facets and student—the object of measurement—contribute to measurement error. While these facets are always important when studying the performance of any group of students (even when linguistic diversity is not an issue), G-studies can incorporate the language of testing (or the dialect of that language) as a facet and examine the relationship between measurement error and language or dialect (Solano-Flores & Li, 2006, 2008).

Current psychometric tools used to detect bias do not provide fine-grained information about the challenges to validity. Validity studies should examine the cultural and linguistic reasons behind differences in student performance. In addition, it is arguably more important to examine how test items, format, and administration impact student test performance prior to test administration.

Cultural Validity

Since the advances in cultural studies that took place in the United States during the 1960s and 1970s, scholars studying cultural responsiveness in education have been calling for a comprehensive and consistent definition of multiculturalism (Ladson-Billings, 1995). Multicultural education researchers have provided a number of diverse definitions and goals (Banks & Banks, 2003). Some researchers have focused on the relationship between the level of student engagement, the types of institutional opportunities available to all students, and student academic performance (Jolly, 2001). Other researchers recognize the importance of community-specific approaches to educational research because of the individuality of each group's meanings and activities (Greenfield, 1995). Still others focus on the fluid relationship that exists between a student's home and school cultures (Ladson-Billings, 1995).

Additional research applying multicultural frameworks to test development, implementation, and use is needed. Tests are cultural artifacts reflecting the language and beliefs of those who develop them (Solano-Flores, 2006). Test content is a reflection of the skills, forms of knowledge, and communication styles that a society—or the influential group of a society—values. There is often an assumption that all test takers are familiar with the contexts used to frame problems, the ways in which questions are worded, and the most appropriate way of responding to the problems (Solano-Flores, 2011a). Without implementing an appropriate adaptation and translation process, there is a decreased likelihood that tests will be culturally responsive.

The way that student dialect and cultural background impact education is evident when testing linguistically diverse students. Students from different backgrounds will view test

prompts differently and will vary in cognitive processes and problem solving (Hawkins, 2004). For example, there is evidence that Spanish-speaking students compartmentalize language syllabically, where English-speaking students compartmentalize language by words (Escamilla, 2000). An exam that requires extensive reading and uses a language different from that of students can cause challenges and distractions for students. Also, there is evidence that Spanish-speaking students use a nonlinear discourse to organize their communication, where English users emphasize a linear discourse (Escamilla, 2000). This can also be a problem with testing. Thus, if a test demands writing and uses a grading system that values linear over nonlinear discourse organization, the test item may be biased against certain students.

Cultural validity should be a foundational component of test development and not a matter addressed once the test is used. Applying the concept of cultural validity during each stage of test development or translation can improve cultural responsiveness (Solano-Flores & Nelson-Barber, 2001). An assessment's cultural validity is determined by how well an assessment addresses the socio-cultural influences that shape student thinking and the ways in which students make sense of test items or tasks and respond to them (Solano-Flores & Nelson-Barber, 2001). The concept of cultural validity takes into account that these socio-cultural influences include sets of values, communication patterns, teaching and learning styles, and views of the student's community. Ensuring validity across all subgroups participating in the test requires greater scrutiny of items, test format, and grading system. Test items vary in linguistic demands, item context is shaped by language, and raters need skills to properly interpret bilingual students' responses (Solano-Flores & Li, 2008).

Taking into account the sociocultural factors that exist within each student population community can lessen issues of bias and construct validity. Too often these issues are discussed

once the test has been created or translated and ready to be implemented. Ideally, cultural and linguistic factors should be addressed during every stage of test development.

Theory of Test Translation Error

The theory of test translation error (TTTE) posits that language in tests is multidimensional and its properties can be analyzed in dimensions of different kinds, such as style, format, conventions, grammar and syntax, semantics, register, information, construct, curriculum, and origin (Solano-Flores, Backhoff, & Contreras-Niño, 2006). The TTTE dimensions can serve as guides for translation review team members. The TTTE breaks down language into three dimensions: grammar and syntax, semantics, and register (Solano-Flores et. al., 2006). The incorrect use of tense, prepositions, idiomatic expressions, and terms can result in invalid test items. In addition, the use of accents, punctuation, font and visual size and layout, and writing practices should follow local norms to minimize the possibility of bias.

TTTE's attention to content is particularly rigorous. There are four dimensions associated with content: information, construct, curriculum, and origin (Solano-Flores et. al., 2006). Unlike other translation review approaches, the TTTE takes into account an item's discursive style. In addition, the TTTE takes into account the possibility that an error may originate in the source item. The TTTE suggests that the reviewers examine each item in both the source and translated languages.

Applying the TTTE in the process of test translation and test translation review entails paying special attention to the qualifications of the translation review team members. The TTTE indicates the need for using a multidisciplinary translation review team that includes independent translators and local curriculum specialists in the translation process. It also promotes the

inclusion of content specialists (e.g., mathematics professors), a sample of local teachers who represent the diversity of teaching across the community, a psychometrician, and a sociolinguist. All members of the review team should be familiar with the test properties and goals as well as the local culture and language.

The TTTE also differs from other review systems in the way that it views translation goals and translation errors. According to the theory, there can be no perfect test translation because of the multidimensionality of language. For example, in an attempt to avoid a syntactical error, a discursive error (e.g., a discursive form grammatically correct but not common in the target language) may take place.

The TTTE allows coding and counting the frequency and severity of errors observed in the translation of a test (Solano-Flores, et al., 2006). The number or the severity of errors should not necessarily cause any student group confusion as to what the test is requiring. Nor should the errors necessarily guide students towards an incorrect response. According to the theory, tests with a high frequency of errors or items with high error severity may result in biased item and invalid scores. Unlike other procedures, the TTTE actively seeks disconfirming evidence of correct translations. Other translation review procedures look for confirming evidence that the translated test is accurate.

A student's familiarity with semantics, word frequency, idiomatic expressions, notation, conventions, syntactical structures, and ways of building arguments will influence test performance (Solano-Flores, 2006). These aspects of language provide contextual information or provide clues or can be distractors for bilingual students. By guiding translation verifiers to look for disconfirming evidence, the TTTE can help test translators and verifiers identify errors.

Fidelity of Implementation

During the last several decades, funding organizations and researchers have become increasingly interested in the evaluation of social service programs. Stakeholders are asking for information that indicates how well a program functions. In general, evaluations should include systematic observation and logical rules to draw inferences from those observations (Rossi, Lipsey, & Freeman, 2004).

Although guidelines may appear straightforward, challenges often arise during their implementation. The number of program participants may change, which could alter the amounts or types of resources necessary or available. Qualifications originally required for those implementing the program may have to be relaxed. In addition, the time frame for deliverables may need to be adjusted.

Given the potential changes that take place during implementation, the evaluation process should include some flexibility. The evaluation design should have a balance of scientific and pragmatic considerations (Rossi et. al., 2004). It is important to include rigor that is similar to that found in scientific research. It is also important that the evaluation design take into account the needs of program stakeholders. A challenge to evaluation is the limited resources discussing details of evaluation for different types of programs despite the existence of many approaches (cf. Rossi et. al., 2004; Nevo, 1983).

In the field of education there is limited guidance on ways to evaluate test development. Though numerous studies examining test and item performance are available, literature addressing the evaluation of the test development, adaptation, and translation is almost non-existent. There is a dearth of research published regarding the evaluation of assessment of

linguistic minorities. Often the literature pertains to English language learners in the United States and only addresses very broad criteria (e.g., Wolf, Herman, & Dietel, 2010). The focus remains on aligning assessments with standards, including essential academic language, and avoiding linguistic complexity. However, often researchers do not offer criteria to evaluate how well test development procedures incorporate these suggestions or how well they are implemented.

Even within fields that have rigorous evaluation demands—such as medicine and educational intervention programs—there is no single definition of fidelity of implementation (FOI) (Century et. al., 2008). However, there are some common characteristics of FOI that can be applied to international test development. First, there is a relationship between the FOI and process effectiveness (Keith, Hopp, Submaranian, Wita & Lower, 2010). Second, a relationship may exist between a procedure's FOI and the usefulness of the procedure to the entities using them (Lynch & O'Donnell, 2005). Third, research also indicates that users will adapt specified processes to suit their local needs (Lynch & O'Donnell, 2005). This is particularly important when the procedures involved are highly complex (Keith et. al., 2010; Lynch & O'Donnell, 2005).

Taking into account these common assumptions, FOI can be viewed as a mediating variable between context and the effectiveness of process (Keith et al., 2010). The context within which a process is implemented impacts these factors as well as the measurement of FOI (Lynch & O'Donnell, 2005). Without descriptive information and measurement of implementation, it is difficult to know if ill-achieved outcomes are due to inadequate programs or improper implementation (Century, Freeman, & Rudnick, 2008).

Several factors are connected to FOI of translation and adaptation procedures. Factors deal with the staged implementation process and a country's willingness and ability to follow the process (Figure 1). The graphical representation for FOI of the translation and adaptation is based on one for science education programs (Ruiz-Primo, 2006). The first set of factors deal with the staged translation and adaptation process provided to countries: complexity of the process, materials provided, review stages included, and training involved. The second set of factors address the country implementation process: country team members' theoretical mindset, available expertise, and external pressures found within the country.

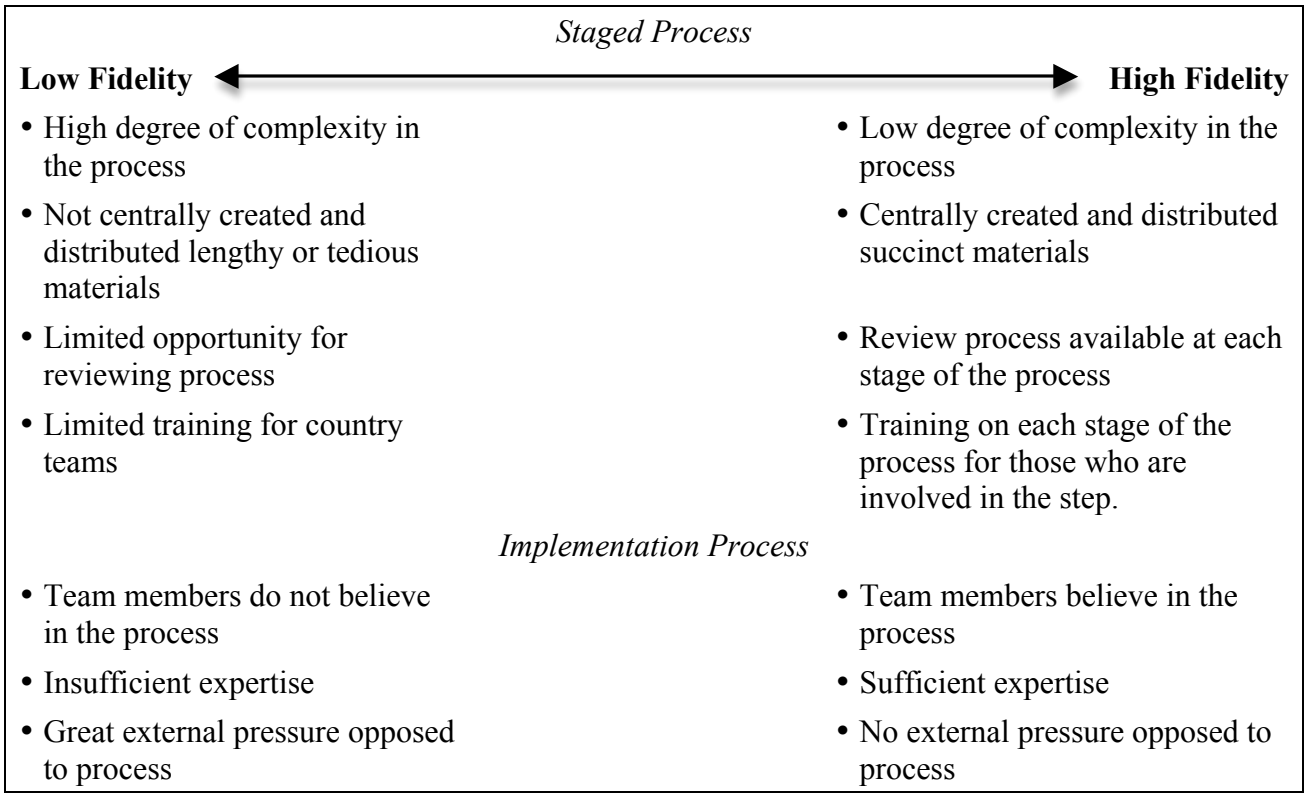

Figure 1. Staged factors and implementation factors that affect the FOI for test adaptation and translation (adapted from Ruis-Primo, 2006).

The first factor in staging, the complexity of the staged process, addresses the demands of the steps involved in the process. It includes a number of sub-factors: the difficulty involved in each

step and overall process; the degree of precision required; the amount and type of coordination needed (e.g., how many entities involved); the amount of time necessary. The more intricate and demanding the steps involved in the process, the greater the challenge to achieve high FOI. Likewise, if the amount of time given per task or overall project is not enough, this can also result in low FOI.

The provision of useful materials is also important to FOI. Organizations should ask if the materials are provided by the same entity so as to increase continuity and decrease errors throughout the documents. They should also ensure that the implementation manuals or guides address each step of the process. The materials should include a system to document each step. Finally, there should be a level of specificity in the materials that provides detailed information without overwhelming the users.

A third sub-factor of staging involves incorporating the opportunity for reviewing the implementation throughout the process. Developers of procedures need to determine the frequency of reviews, who should be involved in the review, and reviewer qualifications. Each review should have criteria for the tasks involved that include the goal of the review and how to check that the goal of each stage is met.

FOI can be impacted by the extent to which all of the materials are covered, the number of people performing the training, and the number of people attending with specific expertise. It is also important that those attending the training will be responsible for completing the steps they are getting trained on. Finally, those attending the training should be able to connect the staged processes to success of the process.

The context in each of the countries will also impact FOI. The mindset of the individuals involved in the adaptation and translation process is important. If those responsible for

implementation perceive effectiveness in the process they tend to have a higher commitment to, and better implementation of, the process. It is also important for a country to have necessary expertise available. Ideally, in the adaptation and translation of international assessments, countries should have access to expertise in measurement and psychometrics, certified translators familiar with source and local languages and dialects, content specialists, and team members with experience in large-scale assessments (Solano-Flores, 2008).

There may also be external pressures over which the country does not have control that can impact FOI. For example, without sufficient financial support, country teams would face obstacles in gathering experts needed. Limited funding can prevent countries from acquiring and using technology needed to complete the translation and adaptation process. This becomes a greater issue as computer administered assessments gain popularity. Countries also face political involvement and subsequent pressure associated with it (Solano-Flores, 2008b). International assessments can have political implications due to the cross-national comparison results published in the media and among policy-makers. Finally, general education legislation can also impact a country's ability to implement the staged procedures.

Incorporating a FOI framework in international assessment development can help improve adaptation and translation procedures. However, it is challenging—and perhaps not possible—to analyze fully a country's FOI. Future research should list the challenges faced when evaluating the fidelity of adaptation and translation as well as how the challenges were met or why it was not possible to meet the challenge.

Summary

The framework presented in this study incorporates aspects of all fields of study discussed. Recognizing the impact that a student's linguistic and cultural background can have on test performance, the study focuses on examining the translation and adaptation of an assessment used internationally. Local educational and assessment policies and practices impact student testing strategies and performance. Therefore, the framework includes the examination of the inclusion of local educational, translation, and measurement experts. In addition, since current popular psychometric tools do not provide fine-grained information about the challenges to validity, this framework examines how each country implemented talk-alouds. Furthermore, based on the TTTE, the framework includes ways to examine confirming and disconfirming information for each country's implementation of the translation and adaptation procedures. The amount of confirming and disconfirming evidence related to each step of the translation and adaptation process explain country and project-wide level of fidelity of implementation.

Chapter 3

Literature Review

Chapter Introduction

Experts in evaluation, psychometrics, sociolinguistics, education, and cognitive psychology have written about the need for greater cultural responsiveness and linguistic sensitivity in assessment. However, most research dealing with this topic focuses on test reliability and validity. Future research should address salient details on achieving culturally and linguistically sensitive tests. The research I discuss in this chapter addresses different foundational concepts behind the processes planned in the AHELO study: the importance of cultural validity, translation procedures, language and cognition, the use of Performance Tasks, test development processes, and fidelity of implementation. Since researchers and advocates for emerging bilingual students have focused on cultural and linguistic responsiveness, I include work from literature addressing English language learners in the United States. Still, available literature does not address the issues that are relevant to the translation and adaptation of the kinds of tasks (performance-based), the form of test administration (computer-based), and the translation process that make AHELO such a unique assessment endeavor.

Cultural responsiveness in test development

Since researchers and advocates for emerging bilingual students have brought attention to cultural and linguistic responsiveness, literature addressing English language learners in the

United States can also guide international test translation and adaptation. Grosjean (1998) discusses the way in which different variables may mediate student performance on an assessment. Using the 'complementary principle,' Grosjean (1998) describes how different domains, purposes, and people influence how bilinguals acquire each language. According to the complementary principle bilinguals acquire and use their languages for different purposes, in different domains of life, with different people (Grosjean, 1998).

There are person specific variables that also mediate an individual's interaction with language and help describe the heterogeneity that exists among bilinguals. Fluency and proficiency of each language differs in each bilingual; language repertoire—changes over time depending on the bilingual's needs of each language (Grosjean, 1998). In addition, the communication context also has an impact on a bilingual's communication. For example, bilinguals interact differently with bilinguals than the way they do with monolinguals (Grosjean, 1998). Finally, the author posits that the type of task engaging a bilingual will also activate language differently. Although the author addresses the existence of these variables, he does not provide guidance on how to use this information during test development, during interpretation of student performance on an exam, or how this information can be used formatively in the classroom. Although the author addresses different aspects of research and student-language interaction in this article, he does not provide enough information on the way that bilinguals navigate each language or both languages.

During the 1990s, some research was guided by language and culture, language and identity, language socialization, and the distinction between learning and using a language while examining second language acquisition (Kramsch & Whiteside, 2007). As a result, questions arose about non-native speakers being viewed as language learners versus language users.

Researchers used sociocultural theory, language emergence theory, conversation analysis, language socialization, and language ecology as fundamental guides during bilingual study design and implementation (Kramsch & Whiteside, 2007). The focus was on viewing real-life encounters and social interactions as processes for learning and for socialization. Therefore, researchers had to question the appropriateness of using native speakers as a standard to which they could compare non-native speakers regarding standard grammar and idiomatic lexicon—given that they would have such distinctly different life experiences (Kramsch & Whiteside, 2007). In addition, there is discussion around the heterogeneity of the non-native speaking population. There is no general, universal set of rules that can by applied to all English language learners given the richness in native language (L1) diversity (Kramsch & Whiteside, 2007). Though the authors focus on theories arising from a constructivist epistemology in bilingual education, they only address research regarding teaching and learning.

Other researchers examine the power dynamics that exist between languages in the United States. As evidence, Escamilla (2000) argues that Spanish and English have unequal status in the U.S. Often, educators and policy makers view a student's L1 as a barrier that interferes with learning English. In addition, despite evidence of the student's L1 helping to learn to read in the second language (L2), some educators view teaching of the L1 as a waste of time (see Escamilla, 2000). Although assessments should include the full range of knowledge and skills that a student possesses in both languages, raters often disregard knowledge from the first language. In a study of emerging bilinguals and native speakers, Escamilla (2000) found that teachers considered monolingual native speakers who knew five colors more advanced than bilingual students who knew three colors in their first language and three colors in their second language. This is clear evidence that raters value certain languages more than others. Additional research is needed in

order for test developers and graders to understand how to use and assess student proficiency in both languages. Specifically, we need to learn how to judge student proficiency accurately when student language varies and needs to be reflected in grading across translated exams.

It is important to understand the greater context of cultural validity in psychometrics. Although Kirkhart (1995) concentrates on culturally appropriate evaluations, her reasonings can also be applied to large-scale summative assessments. In fact, like evaluations, assessment systems and their measurement tools are not culturally neutral or culture-free. Kirkhart does not discuss specific research, but, instead defines multicultural validity and discusses the complexities of trying to define culture. She argues that multicultural validity should be a central dimension of validity as it is a way to organize concerns about pluralism and diversity in evaluation as well as the cultural boundness of evaluation work. Furthermore, Kirkhart connects culture to three major dimensions of validity that can provide a framework for assessment creation and implementation. Methodological validity examines the soundness of findings from methods of inquiry through two areas. Measurement validity focuses on the tools and procedures used to gather data. Design logic validity asks the researchers to question the research design. Both types of methodological validity focus on relevance and equivalence. Interpersonal validity, Kirkhart's second type of validity, addresses data from personal interactions. Finally, consequential validity addresses the change exerted on systems by the evaluation itself—both positive and negative unintended consequences of evaluation.

Kirkhart argues that any evaluation must include staff that is knowledgeable of the culture of the people who are being evaluated. Given the use of large-scale assessments with international student populations, researchers should constantly question the validity of the assessment system. The assessment design, the way in which the assessment is adapted, the items and test

format, and the way in which assessments are used should always be questioned. Unfortunately, Kirkhart does not provide any guidance on how to incorporate questions of validity during assessment creation or implementation. Additional information about specific steps involved in test development is needed.

There is research that focuses on more specific issues of cultural responsiveness in assessment. In her article, Peña (2007) addresses cultural bias and validity threats that can result from poor translation processes during cross-cultural research. The author argues that all aspects of the assessment can be potential areas for bias; instructions as well as the content of instruments. The author proposes four areas of equivalence that researchers should examine for threats to validity. The article examines linguistic equivalence in instructions or elicitation procedures. Functional equivalence deals with ensuring that the translated tool measures the same construct as the original instrument; the elicitation frames for each language in the final version should elicit linguistically similar responses. Cultural equivalence examines the ways that varying cultural and linguistic groups interpret the meaning of an item. Also, metric equivalence refers to item or question difficulty. This can be examined by creating psychometrically parallel tests. Finally, Peña recommends debriefing with respondents during the pilot stage to understand how respondents might interpret the assessment as well as the response patterns. This last piece of the article is the only attempt at offering any suggestion as to how to avoid bias and threats to validity. Though the attention to linguistic cultural aspects of measurement are important, additional research is needed to address best practices for ensuring all students have equivalent assessments and how to do so when working with heterogeneous cultural and linguistic minority groups.

The impact of linguistic and cultural diversity across student populations on assessment practices are important to consider when using performance assessments to evaluate student progress. Performance assessments have several strengths. They can help teachers evaluate higher order cognitive skills and complex learning outcomes that usually cannot be measured with traditional closed-ended questions (Gronlund & Waugh, 2009). These types of assessments encourage the application of knowledge and skills to real-world situations (Gronlund & Waugh, 2009). However, scoring student performance can be subjective and produce low reliability (Gronlund & Waugh, 2009). The authors do not address specific rater-student differences—either cultural or linguistic—that can impact scoring and test results.

Some research addresses the complex nature of performance assessments. Solano-Flores and Shavelson (1997) discuss three dimensions that impact performance assessments and the tension that can occur between them: content, equipment, and use. They provide examples of the cost of materials and difficulty with scorer inter-rater reliability. When developing performance assessments the authors recommend using an iterative process that includes gathering feedback from students with different schools, classrooms, and backgrounds (Solano-Flores & Shavelson, 1997). However, the authors do not provide more details about sampling students from different linguistic and cultural populations. In addition, the authors do not discuss the way in which the task's content can impact traditionally under-represented students. Finally, the article addresses assessing a specific domain, science, within a k-12 grade environment. Additional research is needed to examine more carefully issues of language and their interaction in performance assessments when used with linguistically diverse students. Furthermore, other studies should examine measuring complex critical thinking and problem solving skills. Also, more work is

needed with students who are completing tertiary education. Finally, more work is required addressing the use of performance assessments as large-scale tests.

Another important work addressing performance assessments specifically address linguistic challenges involved with performance assessment (Abedi, 2010). The author presents information indicating that performance tasks may contain unnecessary linguistic complexity in test directions and contextual information (Abedi, 2010). However, the author also sites research indicating that performance assessments help English language learners by providing relevant contextual information and various ways to express their knowledge (Abedi, 2010). Although this information is relevant to second language learners in the United States, the results of various studies may not apply to culturally diverse students living throughout other countries. Additional research with students in other countries is necessary before measurement experts can determine the impact of using performance assessments internationally. Furthermore, Abedi (2010) sites work conducted with k-12 students. Research with performance assessments also should take place with students attending higher education students.

Test Translation and Adaptation Procedures

Since the mid-1990s, research has shown that there is also a connection between culture and performance on exams. Almost 20 years ago, Geisinger (1994) examined the reasons for adapting and translating measurement tools when using them with different cultural and linguistic groups. In fact, in this seminal piece, the author listed several specific situations when test adaptation would be required and provided general guidelines for the adaptation and translation process. For example, Geisinger suggests that if the assessment tool will be used with a group that has a different cultural background, country, and language or if the language

remains the same but the culture or life experiences differ the assessment should undergo an adaptation stage. It is important to consider cultural and linguistic differences between the original and new populations. Adaptation should also take place when working with subpopulations within a given nation or within a single language. Geisinger recommends that when looking at subgroups test users examine educational background, parents' socioeconomic status, student familiarity with tests, and specific test-taking skills and abilities. More specifically, the test translators need to look at the new target population's familiarity with item and test formatting as well as vocabulary difficulty and connotation.

Geisinger goes further listing ten suggested steps for adaptation and translation, with a caution that some assessment tools may require more or less stages and often require an iterative process. The initial stage Geisinger suggests is translating and adapting the measurement tool with a team that is knowledgeable of culture and fluent in both languages as well as an expert in content and characteristics of tool. Finally, the experts involved should also have knowledge of how the assessment tool will be used. Though the author briefly shares criticism of backward translation he does not provide any specific steps that translators should follow during the initial adaptation and translation process. Instead he moves on quickly to the test translation review phase. He suggests that the translation review occur with a group that does not include the translator. The group members share thoughts about the items as a group and then reconciling. The group can then discuss their findings with the translator so that he/she can explain reasons for original translation results. Afterwards, the group adapts the draft.

At this stage the focus moves to the population that will be using the assessment. First, the author recommends piloting the new instruments with a small but comparable student sample. A few trial administrations can help developers learn about potential problems. Once changes from

the pilot have been created, the test creators can field test the measurement tool with a large sample to examine differential item analysis (DIF). Geisinger also addresses psychometric principles that can help with measurement error that may be a result of the adaptation and translation processes. He advises that psychometricians examine the standardized test scores for comparability because there can be systematic differences between groups. Furthermore, test makers should perform validation research to ensure that the test is measuring same construct and that the scores still mean the same thing. Finally, test makers should create a manual (and any other documents) for the users and train users against improper implementation or misuse. Finally, they should collect reactions from users to check for misuse or misinterpretation on their part and to gather information on the implementation process.

Despite the impact that Geisinger has had on test translation, there are some aspects of translation that demand further clarification. Geisinger does not describe the skills and expertise that the translators, the translation review team, and those in charge of pilot testing should have—or how test developers should go about securing these resources. In addition, Geisinger does not address the demands on time, finances, and intellectual resources that his process requires. Finally, it is important to note that this article was written specifically about psychological measurement tools—not educational assessment. Since the article is not based on a specific study, but rather on the author's thoughts regarding translation, there needs to be work on the challenges due to implementing the recommended process.

Auchter and Stansfield (1997) examined the feasibility and process of translating the General Education Development (GED) assessment from English into Spanish and the appropriateness of using it with different populations. First, the study stressed the importance of using an interdisciplinary approach—including a linguist who specializes in the original and second

languages as well as experts in second language testing, cross-lingual assessment, and psychometrics. A panel working on the study argued that though the constructs measured in the English and Spanish writing skills are similar, they were tapping different abilities. To ensure that English and Spanish speaking students are being measured for the same skills and content, the assessment tools required adaptation.

Auchter and Stansfield provide some guidance for a test translation process. The researchers proposed that the professional translators should have orientation about the tool on which they will be working. The authors also argue against literal translations and stress the importance of appropriate syntax, register, and dialect. These are three important areas of language systems and shows that the authors are addressing linguistic diversity among the test participants. According to the article, the translation process itself should begin with a forward translation and continue with several stages of review and revisions. The review process should include a reviewer and contractor each identifying and sharing any concerns about the translated document with the project manager. It is important that all suggestions and changes be documented along with any reasons for rejecting any suggestion. At the end of the review process, the translated document should be completed.

Although the authors address linguistic diversity, they do not address cultural differences that exist between students. The authors do not acknowledge or provide any guidance on how to address these differences during the adaptation stage. There is an unstated assumption that students will read, interpret and respond to the test without any need to account for variations across communities in different parts of the United States or the level of acculturation that the student may have undergone. Also, the authors do not address the next stage of the test development process. They fail to acknowledge the value in pilot testing, field testing, or

cognitive labs with a representative student sample. Lastly, although the test addressed content areas and general writing skills, it did not focus on cognitive skills.

Similarly, Cook, Schmitt-Cavallari, and Brown (2005) discuss the ITC guidelines on translation and administration procedures. The authors specifically address the importance of highly qualified translators when accounting for linguistic and cultural differences between students during the adaptation process. They also stress the importance of gathering evidence that item content and the language used in all parts of the test are appropriate for each local student population. In addition, the authors stress the importance of using statistical tools available to those examining item and test performance across diverse groups. Although the authors focus on very important aspects of cross-national test development additional research is needed to provide specific actions and guidelines. Finding qualified translators and gathering evidence for cultural responsiveness are complex and challenging, particularly internationally.

In one of the most cited works regarding international test translation, Hambleton (2005) shares a fairly comprehensive list of challenges that must be addressed in cross-national comparisons. Focusing his discussion on the International Test Commission's (ITC) guidelines for test adaptation, Hambleton argues that test development must carefully examine testing context, local needs, and test score validity. Hambleton agrees with the ITC's argument that testing context, development and adaptation, implementation, as well as score interpretation and documentation can contribute to score measurement error. For example, the author stresses that ignoring dialect differences can be one of the most detrimental decisions made during the adaptation and translation process. Hambleton provides some potential ways to alleviate measurement error. He argues that the test adaptation and translation process should address these issues. Including stringent requirements for translators and pilot testing the instrument prior

to implementation can decrease measurement error associated with imprecise translation. The author stresses that field tests can help gather information on vocabulary and situational appropriateness for each student group. However, although Hambleton lists these broad categories that can help with cross-national comparison studies, he does not provide specific information about translator expertise. In addition, the importance of student sampling and the use of the exact measurement tool and conditions are not stressed. As in other literature, the reader is left without clear implementation guidelines of the important suggestions that the author provides.

A better awareness and comprehension of linguistic and cultural richness can help researchers understand important aspects of assessment adaptation and translation process. Having accurate expectations of the adaptation and translation procedures is important. In their article, Solano-Flores, Backhoff and Contreras-Niño (2006) present the theory of test translation error (TTTE). The authors organize ten translation error dimensions—style, format, conventions, grammar and syntax, semantics, register, information, construct, curriculum, and origin—into three larger categories: item design, language, and content. In the TTTE Solano-Flores, Backhoff, and Contreras-Niño posit that given the multidimensionality of language, it is impossible to have error-free translated assessments. For example, in avoiding an error related to syntax or style may require that translators make an error related to discourse or register.

The authors posit that instead of focusing on finding confirming evidence of a perfect translation, translation review should intentionally seek disconfirming evidence of perfect translation. The authors remind researchers to discuss the number and severity of errors that should be allowed when translating their measurement tool. The number and severity of errors should not cause any student group confusion or guide them towards an incorrect response.

Developers and translators must be careful with the way that errors impact score measurement error. Conceptually the authors advanced translation work for assessments. However, the authors acknowledge the need for empirical examination of translation error as a way to explain test invalidity. Also, although—as the authors suggest—the work is critical for studying sources of score measurement error after test administration, the TTTE can be beneficial during the test development process. Translators and translation reviewers can focus on the dimensions that the authors list during test adaptation and translation.

Measurement Issues

Performance assessments have been receiving greater attention as interest in their use has increased. The CLA involves PTs that may be difficult to translate and challenging to have high inter-rater reliability. Messick (1994) argued that though validity criteria—both evidentiary and consequential—for performance assessments may differ from those of other assessments, the criteria itself must remain stringent. Simply stating that a performance assessment is authentic and/or direct does not justify a lack of "evidential grounding." Messick equally emphasizes the entanglement of test validity and social issues and that this relationship must be examined when addressing unintended consequences of a test. Messick connects a construct validity framework to the assessment's criterion and purpose while minimizing construct underrepresentation (authenticity) and construct irrelevant variance (directness). Messick argues for "more realistic item context" to improve the item's meaningfulness and transparency.

Scoring must be met with the same rigor as creating the PT due to the subjectivity that can occur during the scoring process. Messick gives an example of handwriting impacting the way in which a rater grades persuasive writing. Since the CLA is computer administered, handwriting

will not be an issue but dialect and discourse can be a problem if these vary between the student and rater. Given the diversity within and between the participating countries, the rubric used to score student responses must measure the same writing constructs while respecting cultural differences. Messick's reasonings fall short about the implications of differences between raters and student responses they may score. Assessing persuasive writing may be biased by the raters' perceptions of, familiarity with, or attitudes towards non-standard dialects.

Messick addresses authentic assessments, which are supposed to replicate challenges and standards of performance that people typically face in real world situations, and student use of skills in different contexts. According to Messick, "What is important is that the skill…changes nonrandomly with conditions and hence correlates with construct-relevant various." This is also applicable to the CLA because it uses 'real-life' scenarios. However, Messick's work does not express how to ensure validity of tasks that use these types of scenarios when 'real-life' differs greatly between students. Additional work is needed addressing the best ways to adapt and translate assessments across languages and cultures.

Some authors have examined physical features of tests in conjunction with psychometric tools. After briefly reviewing quantitative and qualitative methods, Bachman (2000) discusses validity issues related to an increase in use of emerging technology in assessment. There is concern regarding the generalizability of newer types of testing formats. Bachman argues that an important aspect of any test is its format, including the way it looks as well as the types of items it contains. Advances in multimedia and computer-based testing have allowed new task formats and modes of presentation (Bachman, 2000). In addition, with increased use of performance assessments we must examine more closely inter-rater reliability. Other validity issues arise with these new formats. For example, aspects of reading measured may change across paper-and-

pencil and computer screen formats. Also, computer-based formats may be biased in favor of students who have regular or greater access to this type of technology. Changing formats may also require different test preparation and administration demanding that schools have personnel with different computer abilities and pedagogical skills. Although Bachmann argues for increased attention to these issues the author gives limited guidance on how to address them. Bachman suggests greater training of professionals and the use of Item Response Theory (IRT). However, the author does not discuss the shortcomings of IRT. Bachman also fails to discuss specific information that should be included in the training or the challenges of offering additional training. More research addressing the complexities of these topics is required.

Sévigny, Savard, & Beaudoin (2009) examine the validity of holistic writing scores across over 1,500 English responses and over 1,400 French ones. The over 3,100 essays came from the 1994 School Achievement Indicators Program (SAIP) given to students between the ages of thirteen and sixteen. In addition to holistic rating, students were also evaluated for voice, content, organization, vocabulary, structure, and rules. The authors found that holistic criteria do not appear to bias the results. The authors address inter-rater reliability, rater training, and the potential for bias due to linguistic differences. However, they do not explain if either the English and French version of the test was a translation or what adaptation procedure was used. As such, as the authors acknowledge, it is difficult to say with certainty if the raters in both languages were truly examining the same construct.

Hambleton, Yu, and Slater (1999) conducted a field test of the 1994 International Test Commission's (ITC) guidelines for adapting educational and psychological tests. They examined the adaptation of 69 items from the 1992 8th grade Mathematics National Assessment of Educational Progress (NAEP). The multiple choice and short constructed-response items had

been translated and adapted from English into Chinese. The twenty-two ITC guidelines were organized into four categories: context, instrument development and adaptation, administration, and documentation/interpretation. The authors found that the guidelines could benefit from clarity. They also posited that the guidelines could avoid overlap by eliminating some of them—arguing that fewer guidelines would prove more manageable. However, there are a couple of challenges to creating one set of guidelines to be used by all countries. First, countries may not have the human and financial resources or time necessary to apply the guidelines. Second, given the difference in expertise and resources, countries may benefit from more concrete and detailed steps on how to implement the suggestions.

Some authors focus on using psychometric tools to detect issues caused by poor test translation. Performing DIF analysis, Allalouf, Hambleton, & Sireci (1999) identified items that behaved differently across different student groups and identify the causes of DIF. They analyzed results from approximately eight thousand students participating in a large-scale high-stakes test: over 6,400 examinees using the Hebrew (source) test and close to 1,700 using the Russian (translated) test. The researchers found DIF for over forty items on each form. They used translators familiar with Hebrew and Russian to examine items showing DIF to identify the causes for the different item behavior across groups. As a result of the study, the authors created a chart to be used during translation so that problematic items could be identified prior to pilot testing or test administration.

Allalouf, Hambleton, & Sireci's chart consists of a sequence of four questions posed of each item that to identify problematic items. The first question is problematic: Is the translation correct? First, the question is geared towards confirming evidence that the translation is accurate. It does not help translators focus on disconfirming evidence of correct translation. Second, the

publication does not define correct translation. Given the language multidimensionality, there is no error-free translation. The second question—Did the format stay exactly the same?—is also problematic. Translation and adaptation should not strive for exactness but rather equivalence across groups so acknowledge and follow local testing format norms. The third and fourth questions should be asked for any translation, Do the words have the same level of difficulty? Are there differences in cultural relevance? However, these are challenging questions to answer and the authors do not provide guidance on how to deal with translations if the answer to either is yes.

Ercikan (1998) examined item performance in a cross-national comparison study. To examine language issues, the author conducted DIF analysis on 70 science items that were created in English and translated into French. The author argues that using DIF analysis will detect translation errors. It is important to note, however, that DIF only indicates that an item is behaving differently across student groups of equal ability. DIF analysis does not explain the reasons behind item performance. Ercikan acknowledges that after DIF analysis was completed further review of the items was necessary. However, the author does not detail how the researchers determined if the translation was an issue affecting student performance nor does he provide criteria for determining the aspect of translation causing measurement error. Likewise, although Ercikan argues for careful translation procedures and pilot testing the author does not provide guidance on how to achieve successful translations or field tests. Also, Ercikan focuses on the fact that linguistic differences across countries can cause problems. Yet, the author does not address the diversity that exists across dialects used within each country. Studying translation issues must include the examination of linguistic and cultural differences within and across countries.

Ercikan (2002) also conducted a cross-validation study using test outcomes from different countries. Ercikan compared English-French results from Canada, England, France, and United States. The researcher performed differential item functioning (DIF) analysis on items used in the different countries. She also asked translators to identify adaptation-related problems associated with the items. Finally, the author examined the relationship between curricular differences and DIF in the Canadian items. In total, Ercikan completed DIF analysis, judgmental reviews, and curricular reviews on approximately 150 mathematics and 140 science TIMSS items.

The DIF analysis identified different numbers of items for each content area across the three comparison groups. In addition, DIF analysis indicated that the language favored also varied. The judgmental reviews found that in mathematics 27% of the DIF items had adaptation-related differences. The translators also found adaptation-related differences in 37% of the DIF science items. When the author disaggregated information during the Canadian curriculum review, she found that 26% of geometry items showed DIF. The vast majority of those items, 83%, favored the French-speaking group. Similarly, 75% of chemistry items and 91% of earth science items showing DIF favored the French-speaking group. Since the results from the judgmental reviews varied across countries, Ercikan posits that there are other possible explanations for DIF: differences in instruction methods and cultural backgrounds.

One of the limitations of this article is that it does not provide information on the process that the translators followed. Examining cultural and contextual differences are a part of the TIMSS test adaptation process—and should, therefore, be included in judgmental reviews examining adaptation issues related to DIF. In addition, the article does not explain if and how translators dealt with the different dialects represented by the different locations. Greater detail is needed

concerning the criteria used during the judgmental review. As with the examination of curricular differences, it is important to disaggregate the information found in the adaptation-related issues process. This could help identify, in greater detail, issues associated with test adaptation.

In a different study Ercikan and Roth (2006) examine the relationship between language and differential item functioning (DIF) by using, in part, think alouds. Used as one of three methods to identify sources of DIF, researchers conducted think alouds with fifty students. Students were asked to voice their thoughts as they completed each item. Upon completion of the test the interviewer asked a series of questions posed by an interviewer. The questions addressed student understanding of an item's intent, the steps students took to answer the item, and aspects of the item that helped or hindered the problem-solving process. Conducting think alouds allowed the researchers to examine how students interpreted questions, used information from the test, and detected differences in the methods used by students from different language groups.

The authors share that conducting think alouds is time consuming and requires a high degree of researcher interpretation and inference. The authors do not address how to conduct interviews with students from different cultural groups. Verbalizing thoughts about the complexity or usefulness of a test item may not be customary in some countries. Also, the authors do not discuss where in the adaptation and translation process think alouds should be integrated or how results from the think alouds should be incorporated into the translation revision process.

In a seminal piece, Solano-Flores (2006) questions the way in which current assessment policies deal with English language learners and includes a discussion of sociolinguistics that is relevant to large-scale international assessments. The concepts of dialect and register are particularly important when adapting and translating an assessment into several languages. Dialect refers to the way in which socioeconomic status, gender, and origin impacts a person's

communication (Halliday, 2003). As part of the language system dialects are also dependent upon rule-governed arrangements but are varieties of a same given language (Solano-Flores, 2006). However, dialects can be distinguished from one another due to the way users pronounce words, vary in grammar usage, vocabulary use, discourse conventions, and the use of certain sets of idiomatic expressions and colloquialisms (Solano-Flores, 2006).

Examining local community linguistic diversity is important in education measurement because, as a part of a system, each dialect is connected to a social group or social class. The 'standard' language is usually the dialect used in the mainstream and is the most socially acceptable dialect of a given society (Solano-Flores, 2006). Like a dialect, register is an aspect of language that impacts the way in which students communicate with each other and comprehend text. Register refers to a variation of a language that is determined by use. Register reflects social processes that are part of a larger language system (e.g., division of labor, specialty, contexts, and specific activities) and are associated with the characteristics of the language (especially academic language) used in tests (Solano-Flores, 2006). Assessments are written in a 'standard' language that is often associated with mainstream society—or the most socially acceptable dialect with a register used by academics. This has implications when creating procedures for test adaptation and translation. For example, idiomatic expressions, notation, conventions, syntactical structures, phrases, ways of building arguments, may provide scant contextual information or provide unintended clues that become distractors. Students not familiar with the mainstream dialect in which the test was written may be at a disadvantage.

Solano-Flores merges advanced psychometric principles with concepts from sociolinguistics to explain the connection between language and test development. This is an important contribution to the literature. However, there are no clear guidelines on how to feasibly go

through an adaptation and translation process for an international assessment that contains many technical and opinion-based documents. Also, though the article provides detail information about aspects of language, it does not address other aspects of culture.

In an earlier article, Solano-Floes and Nelson-Barber (2001) create a clear connection between the psychometric principle of validity and a student's cultural background. The authors explain that culture and society shape the way in which students learn and think. As a result, the authors argue that a paradigm shift must occur around the concepts of independence and validity. Some of the sociocultural factors that impact cultural validity include communication patterns, teaching and learning styles, values, beliefs, and experiences. Current practice often creates a test in an original language and then translates the test into other languages. However, the authors question the homogeneity assumed among different cultural and linguistic groups by test makers. Although the authors address cultural validity within the area of science, their five areas of reasoning—student epistemology, student language proficiency, cultural world views, cultural communication and socialization styles—and student life context and values are applicable to other content areas as well. Using these areas during test development helps avoid erroneous preconceptions and assumptions about different cultures. That being said, the authors acknowledge the difficulty in balancing cultural responsiveness with test standardization. Additional research needs to address how test development can respect cultural and pedagogical differences that exist among schools and countries while still measuring a common construct.

Despite the importance of assessments in the international context, very little research has been completed regarding cultural and linguistic adaptation process or translation that goes beyond examining DIF. Research on assessing university students using PTs that measure skills required for real life after graduation has not been conducted. In addition, to date, no publications

examine the use of think alouds during the translation process of an international assessment as is done in the AHELO feasibility study. Although research discusses the importance of cultural responsiveness in assessment creation and implementation, it fails to provide detailed guidelines for adaptation and translation that reflect best practices. As a result, current literature does not address the feasibility of implementing more rigorous translation procedures across countries.

Research acknowledges the difficulty in balancing cultural responsiveness with test standardization. However, additional research needs to address how test development can respect cultural and pedagogical differences that exist among schools and countries while still measuring a common construct. In essence this is what the AHELO study is attempting to accomplish.

Fidelity of Implementation

There is limited research regarding fidelity of implementation in the field of education. Studies addressing fidelity of implementation in other disciplines can provide insight into the criteria that is important to consider when studying implementation of any process. One article addresses the fidelity of implementation of medical interventions in reducing patient morbidity, mortality, and resource utilization (Keith, Hopp, Submaranian, Wita & Lower, 2010). The authors examined organizational members' level of commitment to using specific components of an intervention (Keith et. al., 2010). To do so they interviewed 18 organizational members in four different medical centers using three dimensions of commitment to use: satisfaction, consistency, and quality. The authors also examined the statistical significance of FOI as a predictor of intervention effectiveness. Similar studies are needed within the area of educational assessment. For example, research can consider the relationship between FOI and a system of assessment development's effectiveness across different locations.

Within the field of education, some research has been published regarding fidelity of implementation of specific curriculum. Lynch & O'Donnell (2005) measured FOI of a four-year scale-up study of a middle school science curriculum applied with diverse students in K-12 classrooms. Lynch & O'Donnell (2005) observed teachers implementing the curriculum in the classroom. The authors had provided teachers with FOI guidelines that were based on work conducted in public health and mental health fields: adherence, exposure, quality of delivery, participant responsiveness, and program differentiation (Lynch & O'Donnell, 2005). By examining work completed on FOI from different fields of study, education researchers can conduct research regarding FOI in educational measurement.

In an article addressing educational program implementation, Ruiz-Primo (2006) reviews five aspects of FOI used in other work: adherence, exposure, quality of program delivery, participant responsiveness, and program differentiation. More specifically, the author provides program and context characteristics that impact the degree of FOI. The factors include level of complexity, required time, material required, number of people needed to implement a program, training, participant satisfaction, perceived effectiveness, and amount of supervision. Furthermore, Ruiz-Primo (2006) argues that it is important to differentiate between critical and related components. Ruiz-Primo (2006) states that there is a relationship between differential patterns of implementation and differential program effectiveness. The author posits that tracking FOI can provide information about improving a program. Ruiz-Primo (2006) applies these concepts to examining the implementation of a specific science curriculum. However, future research can apply the same concepts to studying FOI of test development and implementation across culturally and linguistically diverse students.

Chapter 4

Methods

Background Information

I collected qualitative data for this study. However, I used qualitative and quantitative methods during data analysis. Qualitative research methods are well suited for working on a case study in which the researcher explores the nature of a new form of program, activity, or process that involves one or more individuals (Creswell, 2009; LaCompte & Schensul, 1999).

Quantitative analysis allows for replication of findings as well as the ability to generalize findings (Creswell, 2009). My analyses examined qualitative data to detect patterns common and unique situations among countries' implementation of the AHELO translation and adaptation procedures. I used dichotomous coding of information to identify national, cross-country, and assessment system level confirming and disconfirming evidence of the fidelity of implementation (FOI) of the translation and adaptation procedures. I identified task-level and criterion-level confirming, disconfirming, and lack of evidence of fidelity of implementation. Based on this information, I computed a fidelity of implementation (F) coefficient for tasks and criteria, defined in terms of the percentages of confirming and disconfirming evidence found at the country and cross-country level.

To examine the consistency and accuracy of the initial coding, a second coder independently coded ten percent of the task-criterion intersections. The second person, familiar with the framework and translation study, coded a sample of 50 task-criterion intersections. Ten task-

criterion intersections were selected randomly and coded across the five countries. The second coder was given the name of the task-intersection and documents relevant for coding. The second coder independently coded without consulting with the first coder. Ten task-criterion intersections were randomly selected and coded across the five participating countries:

1. Adapt the test (T5) – Timely Communication (C1)
2. Changes from verification (T12) – Translation expertise (C3)
3. Attend meetings (T16) – Project management expertise (C4)
4. Adapt the test (T5) – Review opportunities (C5)
5. Review translation (T10) – Training opportunities (C6)
6. Review translation (T10) – Apt deadlines (C8)
7. Changes validation (T12) – Apt deadlines (C8)
8. Configure team (T1) – In-country support (C10)
9. Technical infrastructure (T4) – In-country support (C10)
10. Select items (T3) – Outside country support (C11)

The second coder proceeded to examine documents and use the framework to code CEBs, DEBs, and NEBs, for each country. Results show an 80 percent match in the coding, 40 of the 50 cells were coded the same.

The majority of data used in this investigation was collected in 2010 and 2011; additional information needed during data analysis was collected in 2012. Data came from multiple formal and informal sources of information generated during the implementation of test translation and adaptation of CAE PTs in five countries. Data analysis began in September of 2011. During analysis I examined information from emails, meeting notes, working and normative documents, interviews with team members, and a survey offered to all country teams.

Participants and Setting

AHELO

Several entities collaborated to carry out the AHELO study: CAE staff, a translation technical assistance team (TAT) hired by CAE, an OECD representative, the Australian Council for Educational Research (ACER) staff, and several country teams (Country A, Country B, Country C, Country D, Country E, and the United States).

CAE staff had expertise in psychometrics and statistics, project management, and computer technology. CAE psychometricians involved with AHELO had extensive knowledge of the original test development, test implementation, and subsequent test results and analysis. Several CAE staff members had experience working with international organizations and were familiar with the implementation process of international assessments. CAE staff also possessed technical expertise needed to administer the test using a proprietary platform over the Internet.

The TAT had expertise in measurement and project management. One of the TAT members was both a psychometrician and one of developers of the original performance assessment, CLA. Another psychometrician from the TAT had written extensively on issues of test translation, and had experience training multidisciplinary teams on test translation and adaptation. Both psychometricians had conducted analysis of international assessment results. The initial OECD liaison to the AHELO feasibility study had twenty years of experience working as a specialist in measurement and evaluation, having worked with national and international education assessment experts. This individual had extensive experience managing national and international assessment programs and held a Master's degree in education with an emphasis in measurement and evaluation.

There were also country teams who partially or fully participated in the adaptation and translation process. Three team members representing universities from four U.S. states took part in a few tasks. The three team members participated in the selection of the two PTs to be used in the AHELO feasibility study and in an initial workshop on test translation and adaptation facilitated by the TAT at the beginning of the project. Since the PTs were created for U.S. students attending U.S. institutions of higher education, translation and adaptation was not an issue for the U.S. Therefore, given the scope of the study, they will not be discussed in this dissertation.

Team members from the other five countries—Country A, Country B, Country C, Country D, and Country E—were responsible for the entire translation and adaptation process in their own countries. After selecting two PTs, country teams were to adapt them, find and hire qualified translators to translate them, and conduct think alouds. In addition, country teams were to use the theory of test translation error as part of their translation review procedures. During each task, country team members communicated their progress to CAE staff and the TAT. When it was deemed helpful, CAE staff and the TAT shared successful strategies used by one country team with other country teams.

The AHELO study was divided into two phases that were the focus of my analysis: 1) test translation and adaptation and 2) translated test implementation. CAE and the TAT were responsible for staging as well as training and supporting country teams in the translation and adaptation procedures. Together, CAE staff and the TAT monitored and noted each country team's progress throughout the study. At least one psychometrician from the TAT traveled to each country to provide training on the translation and adaptation procedures. The project

manager met with country teams via in-person and telephone meetings. In addition, CAE staff and the TAT were available to country team members via email and telephone.

Countries: Economic Context for Education

The overall political, economic and educational systems of the participating countries are diverse, as reflected by the overall funds spent on education, the languages spoken, and overall literacy (Table 1).

Table 1
Cross Country Comparison of Overall Education Expenditures, Languages Spoken, and Literacy Rates.

Country	Education expenditures (as % of GDP)	Languages	Literacy rate (%)
Country A	5.9	Finnish, Swedish	100
Country B	Undisclosed	Country B	99
Country C	3.8	Arabic, English	Male: 94 Female: 91
Country D	4.8	Spanish, Indigenous languages	Male: 86.9 Female: 85.3
Country E	6.8	Bokmal Norwegian, Nynorsk Norwegian, Sami, Finnish	100

Country A. Country A has a highly industrialized, mainly free-market economy and is composed of nineteen regions (CIA, 2011). Almost 97% of the residents speak official languages of Finnish or Swedish and has a 100 percent literacy rate (CIA, 2011).[4] In addition, primary through tertiary education—which includes a doctorate or licensure—on average, spans eighteen

[4] The U.S. Central Intelligence Agency (CIA) determines a country's literacy rate by the number of people who are fifteen years old and are able to read and write.

years with minimal difference by gender (CIA, 2011; Ministry of Education and Culture, 2011). Education expenditures are 5.9% of the country's gross domestic product (CIA, 2011).

Country B. Thanks to close collaboration between government and business, Country B's economy has improved from financial difficulties it faced during the first half of the twentieth century (Bae & Lawler, 2000). Country B's population is almost fifty million and is fairly homogenous ethnically and linguistically. Records indicate that 100% of their population speaks Country B (CIA, 2011). Country B has a 99% literacy rate among males and a 96% rate among females (CIA, 2011).

Country C. Country C has a relatively open economy with approximately 9% of the world's oil reserves. Despite Country C's relatively small size, its economy is strong and its population is quite diverse (Fearon, 2003). Of the 3.4 million total people living in the country, only about a third are native to Country C. 35% of the people living in Country C belong to another Arab group, 9% are South Asian, 4% are Iranian, and 7% belong to other groups (CIA, 2011). Although Arabic is the official language, English is widely spoken. The literacy rates among males and females is respectively 94% and 91% (CIA, 2011). The school life expectancy is twelve years and the national education expenditure is 3.8% of the gross domestic product (CIA, 2011).

Country D. Country D has a free market economy that is a mixture of industry and agriculture. 60% of the population is Amerindian-Spanish, 30% are Amerindian or predominantly Amerindian, 9% are white, and 1% are of another ethnicity (CIA, 2011). 92.7% of the population speaks only Spanish, 5.7% speak Spanish and indigenous languages, .8% speaks only indigenous languages, and another .8% speaks other, unspecified languages (CIA,

2011; Bernard, 1999). 86.9% of males and 85.3% of females are literate (CIA, 2011). The school life expectancy is fourteen years and the education expenditure is 4.8% (CIA, 2011).

Country E. Country E's economy is a combination of free market activity and government involvement resulting in welfare capitalism (CIA, 2011). 94.4% of its inhabitants are from the country, 3.6% belong to another European ethnic group, and 2% are from another, non-European ethnic group (CIA, 2011). Country E has two official languages: Bokmal Norwegian and Nynorsk Norwegian; a small number of communities speak Sami or Finnish (CIA, 2011). Country E has a 100% literacy rate and has an education expenditure that is 6.8% of the gross domestic product (cf. CIA, 2011; Aamodt, 2008). On average, females spend 18 years through tertiary education and males spend 17 years (CIA, 2011; Aamodt, 2008).

Countries: Higher education systems

Systems of higher education are diverse within and across countries. The number and types of institutions vary greatly as does the focus of individual schools. The system-wide and institution-level organization impacts student opportunities. For example, the cost to individual students to attend an institution of higher education varies by country and at times by type of institution within a country. These variables can impact the overall type of student who attends and completes higher education—and can determine the emphasis on cultural and linguistic adaptation work required during the translation and adaptation process.

Country A. Country A's higher education system consists of sixteen universities and twenty-six polytechnics/universities of applied sciences (Ministry of Education and Culture, 2011). Universities focus on scientific research and experience a high degree of autonomy. Polytechnics train professionals in response to labor market needs. To attend an institution of higher learning,

students participate in entrance tests and must possess a general secondary or vocational diploma (Ministry of Education and Culture, 2011). In addition to free basic health care, reductions in public transportation, and meals in campus restaurants, Country A's students receive other direct financial support (Ministry of Education and Culture, 2011). In addition to receiving a study grant of almost 300 Euro a month, the state guarantees any student loans (Ministry of Education and Culture, 2011).

Since the 1990s, Country A has seen an increase in the number of people attending higher education institutions. In 2005, 2.8 million people had completed post-comprehensive level; 25 percent of persons 15 years old and older had completed a tertiary level of education (Statistics, http://www.stat.fi/tup/suomi90/marraskuu_en.html). This increase is due in large part to the rise of permanent polytechnic institutions, which replaced vocational post-secondary and vocational tertiary level education. The education system in Country A is not as stratified as in other countries. The difference in performance between the top and bottom 20 percent of students is 62 point; the average difference among other students in other OCED countries is 99 points (OECD, http://www.oecdbetterlifeindex.org/countries/Country A/). In 2005, 35 percent of 25-64 year olds in Country A completed tertiary education. In the same year, 41 percent of 35-44 year olds, 38 percent of 25-34 years olds, 34 percent of 45-54 year olds, and 27 percent of 55-64 year olds completed tertiary education (OECD, 2008).

Country B. Country B has hundreds of public and private universities as well as vocational schools throughout the country. The academic experience of students attending these institutions of higher learning will vary greatly depending on the academic vigor of the individual institution. As a result, Country B is implementing a quality assurance program to which all public and private universities will adhere (Country B Council for University Education, 2011). Students are

largely responsible for expenses associated with attending an institution of higher learning. Country B's government is oversees the distribution of available funding opportunities to ensure institutions have appropriate research financial support.

Over 50 percent of Country B's 25-34 year olds attain a tertiary education (Lee, 2009). However, 10 percent of 55-64 year olds attained tertiary education (Lee, 2009). Almost 41 percent of the students in tertiary education are females (Lee, 2009). As of 2008, approximately 70 percent of vocational high school graduates moved on to higher education; over 85 percent of high school graduates continued on to higher education (Lee, 2009).

Country C. Country C's higher education system consists of public universities, vocational schools, and private institutions. However, the majority of the students attend the country's vocational schools (Country C Cultural Office, 2011). In addition, students also look for scholarships to study abroad. There is diversity in the type of education that private post secondary institutions offer: liberal art model, professional education, vocational, open education, and a branch of MSM-Holland (private business college) (Ministry of Higher Education, 2011). The country is currently working on establishing standards, increasing the number of institutions of higher education, and creating an accreditation process for private universities (Ministry of Higher Education, 2011).

During the 2005/2006 academic year there were over 27,000 students enrolled in the four state-supported higher education institutions in Country C. Total enrollment in all tertiary educational programs reached 11 percent during the same academic year (World Bank, http://databank.worldbank.org/ddp/home.do?Step=3&id=4).

Country D. Of Country D's 2,539 institutions of higher education, over half are private. They consist of federal public, state public, public technological institutes, public technological

universities, public polytechnic universities, public intercultural universities, public teacher colleges, private teacher colleges, public research centers, and other public institutions (Subsecretaría de Educación Superior, 2011). Although over half of the institutions are private, sixty-seven percent of university students attend public higher education institutions at a very low cost (Subsecretaría de Educación Superior, 2011). Country D uses several standardized tests for accountability purposes (Subsecretaría de Educación Superior, 2011).

In 2005, on average, 15 percent of 25-64 year olds completed tertiary education in Country D (OECD, 2008). During the same year, 25-34 year olds made up the highest percentage of those who graduated from a higher education institution: 18 percent (OECD, 2008). All other age groups had a lower percentage of graduates; 16 percent of 35 to 44 year olds completed tertiary education, 14 percent of 45 to 54 year olds, and eight percent of 55 to 64 year olds (OECD, 2008).

Country E. Country E has seven universities, seven specialized institutions at the university level (one is private), 24 state university colleges, two national institutes of the arts, two private university colleges, and 31 private institutions with accredited study programs (Statistics Country E, 2009). Of 31 private institutions, 22 receive some public funding (Ministry of Education and Research, 2011). Although each university has a board that is responsible for operations, they are directly subordinate to the Ministry of Education and Research (Ministry of Education and Research, 2011), which coordinates admissions and courses at state colleges and universities. Country E plans to be a part of the Bologna Process with a free-flow of lecturers and students across Europe (Ministry of Education and Research, 2011).

In 2005, 33 percent of 25-64 year olds graduated from a tertiary education in Country E (OECD, 2008). In the same year, 25-34 year olds had the highest percentage of graduates: 41

percent. 35-44 year olds had a graduation rate of 35 and 45-54 year olds 30 percent. In 2005, 55-64 year olds in Country E had a graduation rate from higher education institutions of 24 percent (OECD, 2008).

Countries: AHELO areas of expertise

Five areas of expertise were deemed helpful for country teams to possess for participation in AHELO: statistics and measurement, test adaptation and translation, local curriculum and instruction, research methodology, and project management.[5] Although country teams tended to have expertise in most of the five areas, only two country teams had expertise in test translation and adaptation.

Country A. The three Country A team members had diverse backgrounds and experiences. One team member had a doctorate in statistics, taught at the university level, and was a senior researcher at the Country A Institute for Educational Research. The second member's research focused on quality assurance in higher education, mergers of universities, and assessment of higher education learning outcomes. This team member had a doctorate in education. The third person had experience as a project researcher for the country's Center for Research and Development of Higher Education and studied university pedagogy. This team member had experience in other international test translation and implementation. In addition, this team member completed a masters degree and began doctoral studies while working in the study.

[5] *United States.* Although the U.S. team did not participate in all tasks involved in the AHELO feasibility study, team members helped select the PTs used. The four-member team had extensive background in project management and local curriculum and instruction. Some members also possessed experience in research methodology. The team did not have a background in statistics and measurement or in test adaptation and translation.

Collectively, the three Country A team members had knowledge of local curriculum and instruction. They also possessed extensive knowledge of statistics, research methods, and project management. In addition, the team had experience in test translation and adaptation.

Country B. The Country B country team members had expertise in research, higher education curriculum and instruction, project management, and educational measurement. The first team member was an expert in national assessment and statistics tools. This person had experience in development of achievement tests that included a collegiate higher order thinking ability scale, a national entrance exam, and the national system of school evaluation. The second member was head of research for the Country B Educational Development Institute with a focus on quality assurance and public funding of higher education. Altogether, the two Country B team members have expertise in measurement and statistics, project management, and local curriculum. The team did not have expertise in test adaptation and translation procedures.

Country C. Both team members from Country C had experience in administration for university systems and individual universities. One member worked for the Ministry of Higher Education and the Private Universities Council. Another member was an Assistant Dean for Accreditation for a private university based on U.S. higher education. Collectively, given the team members' work background, the Country C team had knowledge of local curriculum and instruction. They also had expertise in research and managing projects. The team did not possess expertise in measurement or test adaptation and translation.

Country D. One of Country D's team members had extensive experience in statistics. The second Country D team member specialized on techniques for online teaching. The final person completed graduate studies in technologies for learning, teaching, as well as online course design, evaluation and facilitation. Jointly, the three team members had expertise in statistics,

local curriculum and instruction, research methodology, and project management. Altogether, the team had knowledge of the cultural and linguistic diversity present in Country D, and understood the importance of accounting for that in test development. However, the team did not have expertise in test translation.

Country E. One Country E team member was the head of research at the Country E Institute for Studies in Innovation, Research and Evaluation (NIFU STEP). This team member focused on education policy, equity, and student finance. The second team member taught teacher education and learning strategies at the university level. Furthermore, since 1998, this person had been involved in the implementation of PISA in Country E. Jointly, the team possessed expertise in research methodology, project management, educational measurement, and local curriculum and instruction.

Researcher's Role

As a former high school teacher who worked with socially, culturally, and economically diverse students in urban areas of the U.S. east coast, I was interested in student performance on high stakes exams. I became increasingly interested in test development as well as test result analysis and use.

As a doctoral student, I took particular interest in studying educational measurement. I focused on examining assessment practices used with diverse student groups—particularly ELLs and students participating in cross-national studies. I was able to take doctoral courses and participate in research that focused on test development, test translation and translation review procedures, test accommodation practices, and test implementation procedures. I was also able to analyze test results using statistical tools, item response theory, and generalizability theory.

My experience in the AHELO feasibility study has been as both a graduate assistant and a researcher. Initially, I participated in the study as a graduate research assistant of the TAT. In that capacity, I helped create documentation regarding test adaptation and translation for meetings and trainings with country teams. I also co-authored the documentation and video instructions used to train on and conduct think alouds. In addition, I documented the questions posed and information shared by country teams and organizations during in-person meetings and conference calls. My work as a member of the TAT gave me access to country team members and all documentation pertaining to the adaptation and translation staging and implementation processes.

As a researcher, my focus is on capturing and analyzing information about the staged and implemented adaptation and translation processes. Being from another country can help me anticipate and understand the variety of available resources and cultural differences that exist across countries. My background can also help me identify and interpret the different communication styles used by country team members. However, I must be aware of my personal experiences and worldviews in order to avoid misinterpreting communication. It is important that I respect the specific set of communication styles and cultural norms that make each country unique.

Data Collection and Recording Procedures

My dissertation examined the fidelity with which each country implemented the AHELO adaptation and translation procedures. I documented the challenges to fidelity of implementation encountered while implementing AHELO-established translation and adaptation process at the

country and system, task, and criteria levels. The sources of information used include (Appendix B):

- In-person meetings
- Conference call meetings
- 100 documents created by organizing entities to assist country teams with the translation and adaptation process and to document the progress
- An open-ended survey administered via Internet
- Group interviews
- Email communication

In-person meetings. One of the first opportunities for data collection took place during an initial, "kickoff" meeting for the AHELO feasibility study attended by the country teams, CAE staff, the TAT, and the OECD. This formal gathering of teams from the U.S. and the five countries that are the object of this study took place from February 15-18, 2010 in New York City. Prior to the meeting, CAE staff provided country teams with documents addressing the study's theoretical framework and timeline. During this meeting, countries discussed, chose, and began adapting the PTs. In addition, each country team presented important information about their systems of higher education. The detailed notes that I took during this meeting are organized by day and date as well as the presentations. Notes from the meeting include details about questions and concerns expressed by country team members, CAE representatives, the TAT, and the OECD representative.

Conference call meetings. A series of conference calls took place in the fall of 2010. The calls were usually held with a person from the country team. However, country team members from Country E and Country A were able to share information during the same call. Along with CAE staff and the OECD representative, country representatives spoke about several important aspects of the project. The meeting agenda contained several items directly related to PT adaptation and

translation (Appendix C). Participants discussed each country's progress or any other translation tasks that they were still in the process of implementing. Country teams were also asked to share their reflections on the AHELO adaptation and translation process. Specifically, teams were asked about what they would change about communication, documentation, suggested tasks, and meetings.

Country team members were also given an opportunity to discuss the think alouds. They were asked about how they chose students who participated in the think alouds, trained the think aloud interviewer, and used the forms created for conducting the think alouds. In addition, the team members were asked about how the information collected could reflect on the adaptation and translation processes. Finally, country teams were given an opportunity to ask or comment about any aspect of the process not already addressed in the initial agenda. This information will give further insight into the differences and similarities in how the countries completed the think alouds.

100 Documents. To guide country teams through, and report on, PT adaptation and translation and document their progress throughout the study, CAE created 100 documents. These documents include different iterations of the PTs, the scoring rubric, published research on the theories behind the translation and adaptation process, work planning documents, and training materials. Of critical importance is to examine the ways in which the TAT provided support by giving a theoretical foundation for the translation and adaptation process, suggesting requirements for country teams, and detailing information on the procedures' implementation. The 100 documents also include reports created by the organizing agencies that were submitted to the international coordinating agency. The reports included information about the progress made, challenges encountered, and unplanned steps completed by countries and agencies.

Also, since the documents are archived chronologically, these documents will show how the process unfolded (e.g., the challenges that each country faced, how they dealt with them, and the results).

Open-ended survey. During the spring of 2010, I created an open-ended survey and administered it via Qualtrics, a web-based survey tool (Appendix D). I sent an email asking that each team answer questions pertaining to the study. I was able to include two links in the email that connected each respondent to questions about the adaptation phase for each PT. Given that each of the two tasks is based on a different context, contains different documents, and asks for different types of responses, it was important to disaggregate data for each. The survey included questions regarding visual literacy, technological literacy, tenor appropriateness, subject matter appropriateness, context familiarity, and task familiarity. The survey contained six open-ended questions addressing different aspects of test adaptation. All responses are recorded on the Qualtrics server.

Group interviews. Visits from project staff to each country for training purposes during the summer and fall of 2010 provided an opportunity to collect data via an in-person group interview with the country teams and translators (Appendix E). The interview protocol consisted of two main sections. The first section included questions about the way in which country teams made decisions about who would participate in the adaptation process, what was culturally appropriate, and lessons learned about translation and adaptation. Although the on-line survey asked about the adaptation process, it focused on specific aspects of cultural relevance, not details about the process. The second section addressed task translation. Participants were asked to explain the student sampling strategy, which impacts adaptation and translation procedures. All interviews were recorded and transcribed.

In addition, country team members discussed the difficulties translators faced as well as the differences and similarities between the adaptation and translation processes. This will provide information on how well the countries were able to follow the adaptation and translation guidelines that the CAE provided. This also provides insight into how country resources impact their ability to follow the recommended CAE procedures.

Email communication. Country team members regularly used email as a way to communicate with the CAE staff and each other. I have access to emails between countries, between countries and CAE staff, and between TAT staff. The emails address a variety of topics in four general categories. One category is directly related to the tasks involved in adaptation and translation implementation: adaptation, translation, and translation review. A second category has to do with results from those tasks. Another has to do with questions about talk aloud procedures. Finally, other emails are related to funding, time constraints, and gathering necessary expertise to complete the translation process.

Additional data collection. I had access to CAE staff, the study's principal investigator, and several country teams throughout the course of the dissertation study. I was able to gather additional information about rubric appropriateness and challenges with its translation. I was also able to gain insight into the timeliness of communication between country teams and CAE. In addition, I was involved in work following the translation verification process that CAE organized and an external translation company completed. Finally, I was able to ask three of the four translation team members about their application of the TTTE throughout the translation review process.

Data completeness. Given the uniqueness of the AHELO feasibility study, the availability of information varies by country. Although every effort was made to collect data from country

teams, it is important to note that country teams had a certain degree of autonomy. The countries were responsible for the costs associated with adapting, translating, and implementing the PTs. Therefore, the degree of oversight and reporting was different than usually found on projects with a central funding agency. Second, given that this was a feasibility study, CAE staff and the TAT understood that it could be difficult and even inappropriate to demand that each country strictly implement every task listed in the staged adaptation and translation guidelines. Missing data is another way to analyze fidelity of implementation by country, task, and criteria.

Data Analysis

General Structure

The organization and analysis of data occurred in two distinct yet interrelated phases. During Phase 1, I captured and organized information about country teams' implementation of the translation and adaptation procedures. I created a list of 18 general tasks that any system of test development that includes diverse students should incorporate. Furthermore, I identified 11 criteria associated with fidelity of implementation with which those involved in test translation should comply. Furthermore, I organized, sorted, reduced, and patterned data into a "story" or interpretation of all of the information collected (LeCompte & Schensul, 1999). For example, I gathered data about the extent to which local measurement and translation experts were involved in the process or the amount of local support country teams received. I then examined the successes and difficulties experienced with completion of the tasks and compliance with criteria.

During Phase 2, for each of the two dimensions—tasks and criteria—I examined confirming evidence bits (CEBs) and disconfirming evidence bits (DEBs). I also accounted for occasions across the two dimensions when there was insufficient information, or no evidence bits (NEBs).

When coding each cell I first examined all of the documents, emails, meeting notes, and results from the survey and interviews. I wrote a narrative capturing all evidence associated with each task-criterion intersection (Appendix I). Once all of the evidence was collected into one narrative, I was able decide if the narrative showed evidence that a particular task was completed while complying with the specific criterion. If the evidence showed that a country met the task-criterion intersection, I coded the cell as a CEB. If no CEB was coded for a particular cell, I re-examined the narrative—and material if necessary—for disconfirming evidence. If I found disconfirming evidence, I coded the cell with a DEB. If no disconfirming evidence was found, I coded the cell with a NEB—an indication that no evidence, either confirming or disconfirming, existed for a particular task-criterion intersection. Perhaps due to the fine-grained coding at the task-criterion intersection that took place for each country, I did not encounter conflicting evidence within cells. I did not find myself needing to deliberate between CEB-DEB-NEB coding possibilities at the task-criterion level. The challenge came in the need to review all information.

The number of CEBs, DEBs, and NEBs provided a way to quantify the level of fidelity of implementation. When taken together across the two dimensions, the three types of evidence helped quantify the fidelity of implementation coefficient across the study. For each of the two dimensions, I examined results at the country level and cross-country level across all countries, using symmetry graphs, and FOI (F) coefficient.

Analysis - Phase 1

Phase 1 focused on organizing data and examining the information for patterns. The first phase consisted of creating a list of all tasks and as well as identifying events for each task and

sub-task associated with each country. I initially generated one table including the tasks and subtasks captured from all documents, the document where I found the information, and the country to which the task applied (Table 2). I color-coded the tasks and subtasks to indicate which belonged to the same general topic. For example, Table 2 includes two tasks highlighted in yellow indicating that both are related to country teams choosing PTs: *Summary descriptions of nine CLA PTs provided by CAE read by country teams*; *Recommendations and rationales for selecting from 4-5 subset PTs read by country teams at the NYC meeting*. Another two tasks are highlighted in grey showing they both refer to countries needing measurement expertise to understand psychometric properties associated with the PTs.

Table 2
Sample List of Tasks and Subtasks, Document, and Country.

Task/Subtask	Doc	Country	Color
Initial details CAE provided for Feb 2010 meeting CAE read by country teams.	Intr ltr	All	Green
Conceptual framework for PT translation provided by CAE read by country teams.	Intr ltr	All	Pink
General instructions for CLA administration on Internet provided by CAE read by country teams.	Intr ltr	All	Blue
Summary descriptions of nine CLA PTs provided by CAE read by country teams.	Intr ltr	All	Yellow
Recommendations and rationales for selecting from 4-5 subset PTs read by country teams at the NYC meeting.	Intr ltr	All	Yellow
Explanation of complexities of generic strand (intended constructs) read by country teams.	Intr ltr	All	Grey
Mentions issues of validity in addition to those associated with cross-cultural appropriateness, and linguistic transferability—to be read by country teams.	Intr ltr	All	Grey
Country teams to agree to sign and abide by confidentiality agreement.	Annex D	All	none

The initial complete list contained 992 total tasks and subtasks (see Appendix F). However, not all tasks were directly related to test translation and adaptation. For example, Table 3 includes a task that is not highlighted, *Country teams to agree to sign and abide by confidentiality agreement*. Although this task needed to be completed at the beginning of the study, it was not integral to the translation and adaptation process.

Using the initial table, I organized tasks by task and sub-tasks after eliminating tasks and subtasks not directly related to the translation and adaptation process. The table included the general task with subtasks below, the document where the information was located, and the country to which the task applied (Appendix F). As an example, Table 3 includes the task, *Familiarize team representatives with CAE provided conceptual framework for adaptation*, and a sample of four subtasks associated with the task.

Table 3
Sample List of Tasks Organized by Task and Subtasks, Document, and Country

Task/Subtask	Doc	Country
Familiarize team representatives with CAE provided conceptual framework for adaptation.	Annex D	All
Country teams had a list of four key documents that informed the translation and adaptation procedures' conceptual framework.	GS.4	All
Country teams were to read and understand the ideas shown in four different pieces of literature contributing to adaption and translation conceptual framework.	GS.7-10	All
Country teams were to become familiar with issues, designs, and technical guidelines for test translation and adaptation (Hambleton, 2005) • carefully choose test administrators • use appropriate item formats • control for speed effect • translators should be familiar with target group, their culture, test content, have some training in test development, and are most capable in test adaptation. • Choose judgmental designs appropriately	GS.7	All

Task/Subtask	Doc	Country
• Choose appropriate data collection designs • Choose statistical analysis appropriately (differing curricula, cultural backgrounds, levels of motivation, socio-political factors) • Use appropriate ITC Guidelines for Test Adaptation		
Country teams were to become familiar and use applicable features of universal design (Thomson, Johnston, & Thurlow, 2002): • Design instruments so allow participation of widest range of students (and flexible enough to allow for changing student populations) • Precisely defined constructs • Accessible non-biased items • Amendable accommodations • Simple, clear, and intuitive instructions—and procedures • Maximum readability • Maximum legibility • Careful use of results	GS.10	All

To make the list of tasks more manageable, and applicable to future studies, I further narrowed the list to the 18 most essential tasks directly associated with the translation process (Table 4) (Appendix G).

Table 4
List of Final 18 Tasks

Task
Task 1. Configure a team according to coordinating group's specifications
Task 2. Acquire funding for all steps of the process.
Task 3. Select test items based on specific criteria established by the coordinating group.
Task 4. Acquire necessary technical infrastructure.
Task 5. Adapt test based on agreed upon cultural adaptation suggestions.
Task 6. Hire translators possessing qualifications set by coordinating group.
Task 7. Translate the assessment.
Task 8. Review translation and notes from translation process.
Task 9. Translate ancillary materials as described by coordinating group.
Task 10. Review translation of material for assessment implementation.
Task 11. Implement changes based on verification procedures prescribed by coordinating group.

Task
Task 12. Make agreed upon changes resulting from validation procedure established by coordinating group.
Task 13. Test assessment implementation process for target population usability.
Task 14. Provide students with an opportunity to become familiar with test format and expectations.
Task 15. Hire scorers according to coordinating group's specifications.
Task 16. Attend kick-off meeting as well as in-person and phone meetings to discuss progress.
Task 17. Submit feedback on process and technical reports regarding progress.
Task 18. Recruit institutions and students to participate in the assessment.

After narrowing the tasks to those most applicable to the study, I organized the final list of 18 tasks into three categories that emerged (Table 5): by logistical tasks, tasks associated with actual translation and adaptation steps, and tasks associated with translation review procedures. The Logistical category included eight tasks: *configure team* (T 1), *acquire funding* (T 2), *acquire technical infrastructure* (T 4), *hire translators* (T 6), *hire scorers* (T 15), *attend meetings* (T 16), *submit feedback* (T 17), and *recruit higher education institutions* (T 18). The Translation and Adaptation category four tasks: *select items* (T 3), *adapt test* (T 5), *translate assessment* (T 7), and *translate ancillary materials* (T 9).

Table 5
List of 18 Tasks Organized by Three Categories

Category	Task
Logistical	1-Configure team
	2-Acquire funding
	4-Technical infrastructure
	6-Hire translators
	15-Hire scorers
	16-Attend meetings
	17-Submit feedback
	18-Recruit higher education institutions
Translation and Adaptation	3-Select items
	5-Adapt test
	7-Translate assessment
	9-Translate ancillary
Review Procedures	8-Review translation
	10-Review implementation materials
	11-Verification changes
	12-Validation changes
	13-Test implementation
	14-Student familiarity

The Review Procedures category included six tasks: *review test translation* (T 8), *review implementation materials* (T 10), *make changes based on verification procedure* (T 11), *make changes based on validation procedure* (T 12), *test computer implementation* (T 13), *give students opportunity to become familiar with PT* (T 14). Creating the three categories allows for the examination of challenges and successes by each of the three general areas associated with the study's translation and adaptation tasks.

To determine the level of fidelity of implementation achieved for each task I created a list of 11 criteria (Table 6). The greater the number of criteria met for a particular task, the higher the level of FOI achieved for that particular task. Ultimately, the greater the number of criteria met

across all 18 tasks, the higher the fidelity of implementation achieved for the entire translation and adaptation process.

Table 6
List of 11 Criteria

Criterion
Criterion 1. Timely communication.
Criterion 2. Expertise in measurement.
Criterion 3. Expertise in translation.
Criterion 4. Expertise in project management.
Criterion 5. Review opportunities.
Criterion 6. Training opportunities.
Criterion 7. Opportunities to document progress.
Criterion 8. Apt deadlines.
Criterion 9. User-friendly materials
Criterion 10. In-country support external to the team.
Criterion 11. Support external to country and country team.

As with tasks, I organized the 11 criteria by three categories: Support Received while completing the task, the Ease of the translation and adaptation steps, and Logistics (Table 7). There were five criteria associated with Support: measurement expertise (C2), translation expertise (C3), project management expertise (C4), support from entities within the country (C10), and support from entities outside of the country (C11). There were five criteria included in the Ease of Implementation category: review opportunities (C5), training opportunities (C6), and user-friendly materials (C9). The Logistics category included 3 criteria: timely communication (C1), document progress (C7), and apt deadlines (C8). Creating the three categories allows for the examination of challenges and successes by each of the three general areas associated with the study's criteria.

Table 7
List of 11 Criteria Organized by Three Categories

Category	
Support	Criteria
	Criterion 2-Measurement expertise
	Criterion 3-Translation expertise
	Criterion 4-Project management expertise
	Criterion 10-Support in country
	Criterion 11-Support out of country
Ease/Use	
	Criterion 5-Review opportunities
	Criterion 6-Training opportunities
	Criterion 9-User-friendly materials
Logistics	
	Criterion 1-Timely communication
	Criterion 7-Document progress
	Criterion 8-Apt deadlines

To facilitate coding, I created a matrix containing 18 tasks and 11 criteria in which each cell included evidence for the intersection of a criterion and task. By arranging all tasks and criteria respectively in rows and columns I constructed a *description matrix* that can be examined at the task and criteria level as well as holistically. A row provides a detailed description of a particular task according to different criteria; a column provides a detailed description of performance of the different 18 tasks according to one criterion. For each country I identified events relevant to each task according to each criterion and captured it in the country's description matrix (Appendix H). Figure 2 includes an example of the events that I identified for one of the five countries as the team completed Task 1, *Configure a team according to coordinating group's specifications*.

Sample Country	1	2	3	4	5	6
Tasks	Timely communication	Expertise-measurement	Expertise-translation	Project management	Review opportunities	Training opportunities
Configure a team according to coordinating group's specifications.	Country provided the US organizing agency limited communication about team members.	Neither of the two primary country team members had the required or desired measurement qualifications.	The country team had translation experience.	The country NPM met all suggested qualifications.	Country reviewed and chose persons for both positions--CAE did not review CVs afterwards	All country team members attended taining; no evidence of reading theoretical literature--Solano-Flores, Site, 2010

Sample Country	7	8	9	10	11
Tasks	Opporutnities to document	Apt deadlines	User friendly materials	In-country support external	Support external to team outside
Configure a team according to coordinating group's specifications.	Country contact list and confidentiality agreements were emailed for each team member	Country team was not able to fill team positions in time for initial meetings and document review	Country did not express difficulty using documents provided for team configuration	Country team had support from government and academia to configure team; however, assessment expert did not participate as originally planned	Country received support from the international coordinating agency and US organizing agencywhen configuring the team.

Figure 2. Example of events across 11 criteria for Task 1 associated with one country

Columns 2-11 contain events identified for individual criterion affiliated with a particular row, or tasks. For example, Figure 2 includes an event found for this country team as it tried to complete Task 1, *Configuring a team*, and comply with Criterion 10, *in-country support external to team*. This country received support from its government and higher education institutions to configure the country team.

In addition to the information in the matrix, I created a document with detailed qualitative data about each cell (Appendix I). The document contains all of the evidence for the final entry in the description matrix. For example, the following is information gathered for the sample

country when trying to *configure the team* (T 1) while *gathering support from the country that was external to the team* (C 10):

> Sample country was able to find team members from various national agencies and universities. During the year prior to beginning the AHELO study, the Sample country Ministry of Education (MinEdu) authorized that personnel be made available from the Sample country's Institute for Educational Research (FIER) of the University of the sample country. FIER then coordinated with the sample country's University Centre for Research and Development of Higher Education and the Sample country's Higher Education Evaluation Council (FINHEEC) (Author, personal communication, January 12, 2010). The AHELO national team also received support from seven members from higher education institutions and student unions) (Author, personal communication, January 12, 2010). As a result the country had all of the support necessary to create the AHELO team in sample country.

The additional information provides context and explanation for the quantitative analyses completed in phase 2.

Analysis - Phase 2

Once events were noted for each intersection of task and criterion, I created three matrices for each country, one for each evidence type: confirming evidence bit (CEB), disconfirming evidence bit (DEB), and no evidence bit (NEB) (Appendix J). The CEBs, DEBs, and NEBs are irreducible pieces of information associated with each task. For the purposes of this study, a bit is a binary unit of an event that is well defined and can be described by its presence or absence. These discrete units can provide great detail about how each country implemented each step of the translation and adaptation process. It can also help detect subtle differences that may exist throughout the process implementation.

I coded each event dichotomously as having or not having confirming, disconfirming, or no evidence bits. If an event for the intersection of a particular task and criterion was a CEB or NEB, the event was coded as a '1.' If there was the presence of a DEB, the event was coded as a '-1'. A 0 indicates that there was no evidence bit for that matrix's evidence type. Figure 3 is an example of a CEB matrix for one country. The matrix is organized by capturing information regarding criteria down columns and tasks across rows. Categorizing events into binary bits provided quantitative information regarding the level of fidelity of implementation achieved by each country.

		c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	
	t_1	1	1	1	1	1	1	1	1	1	1	1	11
	t_2	1	1	1	1	1	0	1	1	1	1	1	10
	t_3	1	1	0	1	1	1	1	1	1	0	1	9
	t_4	1	1	1	1	1	1	1	1	1	1	0	10
	t_5	1	1	0	0	1	1	1	1	0	1	1	8
	t_6	1	1	1	1	0	1	1	0	1	1	1	9
	t_7	1	0	1	1	0	1	1	0	1	1	1	8
	t_8	1	1	1	1	1	1	1	0	1	1	1	10
	t_9	1	0	1	1	0	1	1	0	1	1	1	8
Country X	t_{10}	1	1	0	0	1	1	1	1	1	0	1	8
CEBs=144	t_{11}	1	1	0	1	0	0	0	0	1	0	1	5
	t_{12}	1	1	0	1	1	1	1	1	1	1	1	10
	t_{13}	0	0	0	0	0	0	0	0	0	0	0	0
	t_{14}	1	1	0	1	1	1	1	1	1	1	1	10
	t_{15}	1	1	0	1	1	1	0	1	1	0	1	8
	t_{16}	1	1	1	1	0	0	1	1	1	0	1	8
	t_{17}	0	1	1	1	0	0	1	1	0	1	0	6
	t_{18}	1	0	0	0	0	1	1	0	1	1	1	6
		16	14	9	14	10	13	15	11	15	12	15	

Figure 3. Sample CEB Matrix for one country.

Given the variety in the types of challenges found across the five countries, there were different sets of CEBs and DEBs for each task. For example, confirming evidence would include having found certified translators or psychometricians with all of the suggested credentials.

Disconfirming evidence would include the case in which a country does not have access to translators with desired credentials or a psychometrician with suggested expertise. Using binary bits provided quantitative information regarding the level of fidelity of implementation achieved across countries. To gather FOI information across all five countries, I first added each set of five matrices by evidence type (Figures 4-6).

Within the CEB matrix, the greatest value possible for each cell is a 5 (Figure 4). This would indicate that documents showed all five countries had confirming evidence for a particular intersection of task and criterion; all countries completed a particular task while complying with a particular criterion.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	
t_1	3	4	4	4	5	4	5	3	5	5	5	47
t_2	5	4	3	5	5	0	5	5	5	5	5	47
t_3	5	4	1	5	5	5	5	5	5	0	5	45
t_4	5	4	5	5	5	5	5	5	5	5	0	49
t_5	4	4	1	0	5	5	5	5	0	2	5	36
t_6	5	3	3	4	0	5	5	2	5	5	5	42
t_7	2	1	3	4	1	4	5	3	5	5	5	38
t_8	5	4	3	1	5	5	5	0	5	4	5	42
t_9	3	1	3	3	1	5	5	0	5	5	5	36
t_{10}	5	3	1	0	5	5	5	5	5	0	5	39
t_{11}	3	4	1	3	0	0	2	1	2	0	5	21
t_{12}	3	3	1	4	1	5	5	1	3	5	5	36
t_{13}	0	0	0	0	0	0	0	0	0	0	0	0
t_{14}	5	4	1	5	5	5	5	4	5	4	5	48
t_{15}	5	5	0	4	4	4	0	4	5	0	5	36
t_{16}	5	4	1	5	0	0	5	5	5	0	5	35
t_{17}	0	4	3	5	0	0	5	5	0	5	0	27
t_{18}	5	0	0	0	0	5	5	0	5	5	5	30
	68	56	34	57	47	62	77	53	70	55	75	

Figure 4. CEBs for each criterion across tasks.

A value of 0 in a particular cell indicates that there was no confirming evidence found in any of the data for any of the five countries while they tried to complete a task and meet a criterion. A

value of between 1 and 4 showed that, for a particular task-criterion intersection, documents included confirming evidence for some but not all countries. Values included in the final column indicate the fidelity of implementation achieved across all five countries for each task. The maximum value possible for a cell in the final column is 55. Values included in the final row indicate how well all countries complied with each criterion. The maximum possible value for a cell in the final row is 90. Values included in the final row indicate how well all countries completed each task.

Within the DEB matrix, the greatest value possible for each cell is a -5 (Figure 5), indicating that all five countries had disconfirming evidence for a particular intersection of task and criterion. A -5 means that none of the countries completed a particular task while complying with a particular criterion.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	
t_1	-2	-1	-1	-1	0	-1	0	-2	0	0	0	-8
t_2	0	-1	-2	0	0	-5	0	0	0	0	0	-8
t_3	0	-1	-4	0	0	0	0	0	0	-5	0	-10
t_4	0	-1	0	0	0	0	0	0	0	0	-5	-6
t_5	-1	-1	-4	-5	0	0	0	0	-5	-3	0	-19
t_6	0	-2	-2	-1	-5	0	0	-3	0	0	0	-13
t_7	-3	-4	-2	-1	-4	-1	0	-2	0	0	0	-17
t_8	0	-1	-1	0	0	0	0	-5	0	-1	0	-8
t_9	-2	-4	-1	-2	-4	0	0	-5	0	0	0	-18
t_{10}	0	0	0	0	0	0	0	0	0	0	0	0
t_{11}	-1	0	0	-1	0	-5	-3	-4	-3	0	0	-17
t_{12}	-2	0	-3	0	-4	0	0	-4	-2	0	0	-15
t_{13}	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-55
t_{14}	0	0	-1	0	0	0	0	-1	0	0	0	-2
t_{15}	0	0	-5	-1	-1	-1	-5	-1	0	0	0	-14
t_{16}	0	-1	-4	0	-5	-5	0	0	0	-5	0	-20
t_{17}	5	-1	-2	0	-5	-5	0	0	-5	0	-5	-18
t_{18}	0	-1	-5	0	0	0	0	0	0	0	0	-6
	-21	-24	-42	-17	-33	-28	-13	-32	-20	-19	-15	

Figure 5. DEBs for each criterion across countries

A 0 indicates that there was no disconfirming evidence found in any of the data for any of the five countries while they tried to complete a task and meet a criterion. A value of -1 to -4 showed that, for a particular task-criterion intersection, documents included disconfirming evidence for some but not all countries. Values included in the final column, maximum -55, indicate the total DEBs across all five countries for each task. Values included in the final row, maximum -90, the total DEBs across all five countries for each criterion.

The greatest value possible for each cell within the NEB matrix is 5 (Figure 6). This would indicate that there was no evidence found across all five countries for a particular intersection of

task and criterion. A 5 means that no conclusive information about how well the countries were able to comply with a criterion while completing a task was found.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}
t_1	0	0	0	0	0	0	0	0	0	0	0
t_2	0	0	0	0	0	0	0	0	0	0	0
t_3	0	0	0	0	0	0	0	0	0	0	0
t_4	0	0	0	0	0	0	0	0	0	0	0
t_5	0	0	0	0	0	0	0	0	0	0	0
t_6	0	0	0	0	0	0	0	0	0	0	0
t_7	0	0	0	0	0	0	0	0	0	0	0
t_8	0	0	1	4	0	0	0	0	0	0	0
t_9	0	0	1	0	0	0	0	0	0	0	0
t_{10}	0	2	4	5	0	0	0	0	0	5	0
t_{11}	1	1	4	1	5	0	0	0	0	5	0
t_{12}	0	2	1	1	0	0	0	0	0	0	0
t_{13}	0	0	0	0	0	0	0	0	0	0	0
t_{14}	0	1	3	0	0	0	0	0	0	1	0
t_{15}	0	0	0	0	0	0	0	0	0	5	0
t_{16}	0	0	0	0	0	0	0	0	0	0	0
t_{17}	0	0	0	0	0	0	0	0	0	0	0
t_{18}	0	4	0	5	5	0	0	5	0	0	0
T	1	10	14	16	10	0	0	5	0	16	0

Figure 6. NEBs for each criterion across countries.

A 0 in a particular cell indicates that there was either confirming or disconfirming evidence for all of the five countries while they tried to complete a task and meet a criterion. A value between 1 and 4 showed that, for a particular task-criterion intersection, there was a lack of information for some but not all countries. Values included in the final column indicate the amount of information missing across all five countries for each task. Values included in the final row indicate the amount of missing information across all countries for a particular criterion.

After examining the coded data included in the descriptive and quantitative matrices, I was able to judge FOI by task. Based on data from the matrices I set cut-off points for tasks that would indicate high or low FOI. If a country, or countries, contained CEBs across 75% of a task I judged this as indicative of high FOI. If a country, or countries, had DEBs for at least one-third of criteria when completing a particular task, that was indicative of low FOI. If a country, or countries, contained NEBs for at least one-third of criteria when completing a task it showed that there was difficulty in capturing information.

The data also helped set cut-off points to determine FOI through criteria. If a country, or countries, had CEBs for more than 75% of criteria across all tasks it was evidence of high FOI. If a country, or countries, had DEBs for at least one-third of the criteria across all tasks it was indicative of low FOI. Finally, if a country, or countries, contained NEBs for at least one-third of criteria across all tasks it showed that there was difficulty in capturing information.

Having quantitative information about country performance allowed me to calculate a Spearman rank correlation between the five countries according to CEBs across the 18 tasks. I also calculated a Spearman rank correlation between the five countries according to CEBs across the 11 criteria. The coefficients indicate the similarity or difference in the rank ordering of tasks or criteria. If the correlation coefficient between two countries is high it is because the criteria or tasks with which the countries experienced success or challenges were similar. If the coefficient is low, it indicates that the tasks and criteria with which the two countries experienced success or challenges were not similar.

I included symmetry graphs comparing CEBs with DEBs for each task by each of the five countries as well as for all countries. I also created symmetry graphs comparing CEBs with DEBs for each criterion by each of the five countries and for all countries. The symmetry graphs

allow the reader to identify quickly the ratio of CEBs to DEBs for a particular task or particular criterion. Symmetry graphs show the pattern of differences and similarities across tasks, criteria, and countries. Similarities across tasks, criteria, or countries would be represented by similar CEB to DEB ratios and would have symmetry graphs with identical lines across CEBs and DEBs. Differences across tasks, criteria, or countries would have symmetry graphs with CEB and DEB lines that vary.

Finally, I calculated the fidelity of implementation (F) coefficient for each task and each criterion by individual country—as well as the average across all countries. This F coefficient was computed as the percent of CEBs observed divided by the sum of the percent of CEBs and the percent DEBs observed. The F coefficient ranges from 0 to 1:

$$F = \frac{C}{C + D} \tag{1}$$

where *C* is the percentage of observed CEBs, and D is the percentage of observed DEBs,

I also calculated a F coefficient for each task, criterion, and country:

$$F_{t_i} = \frac{C_{t_i}}{C_{t_i} + D_{t_i}}, \tag{2}$$

$$F_{r_i} = \frac{C_{r_i}}{C_{r_i} + D_{r_i}}, \tag{3}$$

$$F_{n_i} = \frac{C_{n_i}}{C_{n_i} + D_{n_i}} \tag{4}$$

where t_i = individual task, r_i = individual criterion, n_i = individual country,

Since there were few cases of absent evidence, NEBs are not included in these formulas. NEBs are discussed only for the few cases in which NEB frequency was high—at least one-third,

as the extent of missing information can provide context for discussing a high F coefficient. A high-NEB% may render a spuriously high F coefficient as this allows for a high CEB to DEB ratio.

Significance

This study contributed with a systematic approach to examining the complexities of test translation and their challenges to FOI in test translation and adaptation. The data analysis framework created and implemented in this study can serve as a model for other international studies. Test developers can examine the FOI for their test translation and adaptation procedures. Furthermore, the lessons learned via this study can result in increased validity across languages and cultures in international assessments.

I was able to provide information about the relationship between the complexity of the translation and adaptation process and FOI. Specifically, I was able to identify aspects of the translation and adaptation process most challenging to FOI. Furthermore, I was able to illustrate how to measure FOI within current translation and adaptation procedures by creating specific ways to monitor FOI throughout implementation of the procedures.

Chapter 5

Results

In this chapter, I examine fidelity of implementation (FOI) based on the number of pieces of evidence that confirm and disconfirm the assumption that the test was translated and adapted according to the intended process. As part of the analysis, I also take into consideration the instances in which information to determine confirming or disconfirming evidence was not available.

This chapter is divided into two main sections, one for each of the two main types of analysis: task fidelity of implementation and criterion fidelity of implementation. Each of the two sections discusses completion by country, completion by symmetry graphs, and F coefficients by country. Finally, each section also discusses those task-criteria intersections that I regard as being of special interest because valuable lessons can be learned from them about what worked well and what proved challenging to implement.

To examine the consistency and accuracy of the initial coding, a second coder independently coded ten percent of the task-criterion intersections. Results show an 80 percent match in the coding, 40 of the 50 cells were coded the same. Given the vast amount of documents as well as the diverse types of material, an 80 percent match in coding is reasonable.

The greatest discrepancy occurred while coding the Technical infrastructure (T4) – In-country support (C10) intersection. Coder one assigned DEBs across all five countries while coder two assigned CEBs across all five countries. The first coder used production of materials

on in-country computers as evidence of CEBs. However, the second coder was seeking evidence referring explicitly to the acquisition of the required technical infrastructure. There was discrepancy in coding the Adapt the test (T5) – Review opportunities (C5) intersection across Country C and Country E. The first coder coded these as NEBs while the second coder coded them as CEBs. In this second case, the first coder was seeking explicit information about the review of adapted materials in the documents. The second coder, knowing the study well used personal knowledge that the task had been completed meeting the criteria as evidence. Finally, the intersection Changes validation (T12) – Apt deadlines (C8) had differences between the two coders for Country C, Country D, and Country E. Again, the first coder coded these as NEBs. However, the second coder coded them as DEBs. Similar to the previous task-criterion intersection, one coder looked explicitly to documents and finding none coded NEBs. Again, the second coder used personal knowledge about the project to code DEBs.

Each case during which discrepancies in the coding were found can be explained by one of the coders seeking explicit mention of the task-criterion intersection versus the other coder using personal knowledge of the study's progress. In each case of coding differences, using personal knowledge of work completed during the study would produce a very slight increase in the level of FOI attributed to a country. The use of personal knowledge versus documented progress will need to be addressed in future work prior to beginning coding.

Task FOI

When examining data from the AHELO study, I formalized the tasks in a way that was specific to the AHELO project but would still provide guidance to future translation research (Table 8). To organize my discussion, I classified the 18 tasks into three categories, based on

their function in the study: Logistical aspects, the Translation and Adaptation process, and Review Procedures (Table 8). In this section, I discuss task results by country, using symmetry graphs, and by F coefficients.

Task completion: Countries

While examining task information by country, I created three cut-off points, one cut-off point for each of the evidence types. The cut-off points are directly related to the level of FOI. For tasks, I determined that CEBs across 75% of criteria was indicative of a high FOI. I also determined that tasks performed with DEBs or NEBs for at least one-third of criteria suggested a lower FOI for completing tasks.

Overall, countries were able to comply with over 75% of the criteria when completing four of the eight tasks included in the Logistical category. Countries were able to *configure teams according to criteria* (T 1) that the Translation technical assistance team (TAT) hired by CAE provided, *acquire necessary funding* (T 2), *provide the necessary technical infrastructure* (T 4), and *hire translators* (T 6) with high FOI. Country teams met over 75% of criteria when completing one of four tasks included in the Translation and Adaptation category: *selecting items* (T3). Countries were also able to comply with over 75% of criteria when performing two of six tasks in the Review Procedures category: *review translation* (T8) and *student familiarization with test administration* (T14).

Table 8
Information on Task CEBs, DEBs, and NEBs According to Three Cut-Off Points Across All Countries

		Cut-off points		
Category	Task	>75% CEB	≥ 1/3DEB	≥ 1/3 NEB
Logistical	1-Configure team	1		
	2-Acquire funding	1		
	4-Technical infrastructure	1		
	6-Hire translators	1		
	15-Hire scorers			
	16-Attend meetings		1	
	17-Submit feedback		1	
	18-Recruit higher education institutions			1
Translation and Adaptation	3-Select items	1		
	5-Adapt test		1	
	7-Translate assessment			
	9-Translate ancillary		1	
Review Procedures	8-Review translation	1		
	10-Review implementation materials			
	11-Verification changes			
	12-Validation changes			
	13-Test implementation		1	
	14-Student familiarity	1		

There were also tasks that were challenging across countries (Table 8). Countries had disconfirming evidence for at least one-third of the criteria when completing two of eight tasks in the Logistical category: *participating in in-person and telephone meetings* (T 16) and *submitting feedback* on the study's processes and individual country team progress throughout the project (T 17). Countries also had difficulty completing two of four tasks in the Translation and Adaptation category: *adapting the test* (T5) and *translating ancillary materials* (T9). Finally, countries experienced challenges when performing one of six tasks included in the Review Procedures category: *providing opportunities for students to become familiar with test implementation* (T 13).

In addition, there was insufficient information for at least one-third of the criteria while countries *recruited higher education institutions and students* to participate in the study (T 18), which is part of the Logistical category.

Results by individual country across tasks provide another perspective through which fidelity of implementation can be studied. The total number of CEBs, DEBs, and NEBs that each country attained for each task helps to explain the level of fidelity of implementation with which the criteria from this framework were met. More specifically, counting the number of tasks that met particular criteria created for this study provides insight into the FOI for each country (Table 9). Although the percentages I chose as guidelines are based on the specific results of this study, they provide a framework through which to discuss findings concerning FOI in other contexts. To achieve a high FOI, a country needed to meet at least 8 of the 11 criteria while completing at least 75% of the 18 tasks included in the framework.

Countries varied in the percentage of tasks for which they were able to meet at least 8 of the 11 criteria (Table 9). On average, all five countries met at least eight criteria for 54% of tasks. However, the standard deviation was 21, indicating a high FOI variability across country task completion. Country A was able to meet at least 8 of the 11 criteria for 78% of the 18 tasks. Country E met at least eight criteria for 72% of the 18 tasks. Information pertaining to Country D revealed that the country team met at least eight criteria for 50% of the tasks. Country C met at least 8 of 11 criteria for 44% of the tasks. Finally, Country B met at least eight criteria for 28% of the tasks. Country A was the only country to meet the cut-off for high FOI: meeting 8 of 11 criteria for at least 75% of tasks.

Table 9
Task CEB, DEB, and NEB Completion Percentages, Means, and Standard Deviations

Evidence bit type	Country A	Country B	Country C	Country D	Country E	Mean (SD)
CEBs ≥ 8	78	28	44	50	72	54 (21)
DEBs ≥ 3	39	61	61	56	22	48 (17)
NEBs ≥ 3	28	17	17	17	6	17 (8)

Examining the number of times that countries had DEBs for at least 3 of 11 criteria when completing a task helps discover challenges to fidelity of implementation. The greater the number of tasks with three or more DEBs across criteria, the lower the fidelity of implementation (Table 9). The mean across the five countries was 48% with a standard deviation of 17. Country E had DEBs for at least three criteria when completing 22% of the 18 tasks. Data showed that, for 39% of tasks, Country A did not meet at least three criteria. Documents with information about Country D's progress indicate that the team had disconfirming evidence for at least three criteria when completing 56% of the 18 tasks. Lastly, Country B and Country C each did not meet at least three criteria for 61% of tasks.

Documenting the number of tasks for which there was missing information about meeting criteria by country contributed to the study (Table 9). As with disconfirming evidence, the number of times that countries had NEBs for at least three criteria when completing a task helps examine possible challenges to fidelity of implementation. The higher the number of tasks with three or more NEBs is indicative of greater difficulty in documenting possible progress. The mean across the five countries was 17 with a standard deviation of 8. For 39% of tasks, Country A had missing information for at least three criteria. For 17% of tasks, Country B, Country C, and Country D had missing information for at least three criteria. Finally, for 6% of tasks, Country E had missing information for at least three criteria.

Another way to compare countries consisted of examining the relationship between countries according to the rank ordering of tasks and criteria by CEBs and DEBs. A high correlation between countries indicates that the tasks that countries found easy or difficult to complete were similar. A weak correlation demonstrates that tasks that countries found challenging varied; tasks that they were successful in completing also varied. Figure 7 shows a Spearman rank correlation matrix between the five countries according to CEBs across the 18 tasks included in this framework.

	Country A	Country B	Country C	Country D	Country E
Country A	1	0.56	0.67	0.83	0.69
Country B		1	0.65	0.66	0.42
Country C			1	0.73	0.58
Country D				1	0.72
Country E					1

Figure 7. Correlation Matrix of CEBs Across Countries for 18 Tasks.

Country A and Country D had a strong correlation, 0.83, indicating a strong similarity in the tasks that the countries were able to complete and tasks they found challenging. Country D and Country C also had a strong relationship with a correlation of 0.73. This coefficient indicates that Country D and Country C were able to complete similar tasks. Similarly, Country D and Country E also had a strong association with a correlation of 0.72. Country D and Country E were able to complete similar tasks. The weakest relationship existed between Country E and Country B who had a correlation of 0.42. The second weakest relationship occurred between Country A and Country B with a correlation of 0.56. These coefficients indicate that these countries varied in

the tasks that they were able to complete. Some countries are similar and some countries are different as to the tasks that they were able to complete.

Discussion of Task FOI

On average, the five participating countries were able to meet 75% of criteria when completing 7 of 18 tasks. Of the seven tasks for which countries met 75% of criteria, four were components of the Logistical category: *configure team* (T1), *acquire funding* (T2), *acquire the necessary technical infrastructure* (T4), and *hire translators* (T6). The Review Procedures category included two tasks for which countries met 75% of criteria: *review translation* (T8) and *allowing student to become familiar with the performance task structure* (T14). The Translation and Adaptation category included one task for which countries met 75% of criteria: *select items* (T3). All of these tasks were completed throughout the first half of the AHELO project. During this time, there was a team in the United States with extensive experience in project management, measurement, and test translation supporting countries working with the Translation technical assistance team (TAT) hired by CAE. Furthermore, all country teams had members participate in training addressing *translation review* (T8). Country team members also worked collaboratively, and with the U.S. TAT's support, when *selecting items* (T3). The targeted support from the Translation technical assistance team (TAT) throughout the completion of these early tasks may have contributed to country teams' understanding and ability to meet criteria.

Individually, each country varied in their ability to meet criteria while completing tasks. Country A had confirming evidence for at least 8 of the 11 criteria for 78% of the 18 tasks. Country B had confirming evidence for at least 8 of 11 criteria for 28% of tasks. Country A had

the most frequent communication throughout the study with the Translation technical assistance team (TAT) hired by CAE. Country A asked questions about team configuration, hiring translators, and the technical infrastructure needed. Conversely, Country B experienced the greatest difficulty communicating with the TAT. Similarly, Country C was not able to communicate with the TAT in a timely manner and had CEBs for 8 of 11 criteria for 44% percent of tasks.

Task completion: Symmetry graphs

The symmetry graphs in Figures 8, 9, 10, 11, and 12 compare CEBs with DEBs for tasks by each of the five countries. The positive percentages at the bottom right of the symmetry graphs represent CEBs; the negative percentages at the bottom left of the symmetry graphs represent DEBs.

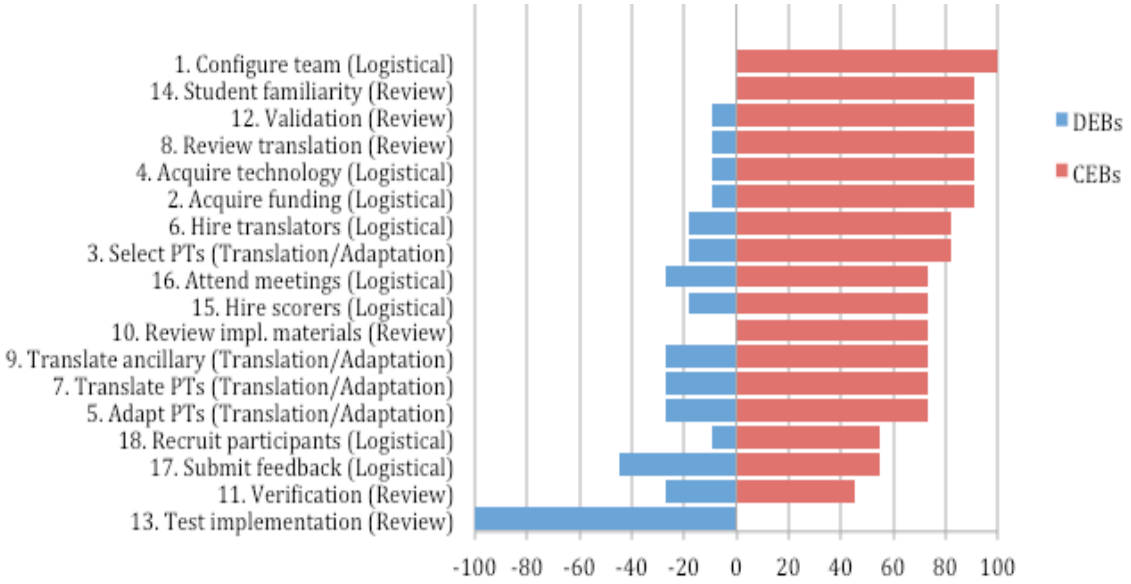

Figure 8. Country A's percentages of DEBs and CEBs across tasks.

Country A was successful in completing several tasks as per the guidelines in this framework. Country A had confirming evidence across 100% of observations when *configuring their country team* (T 1). Country A also had confirming evidence for 91% of the observations regarding *giving students an opportunity to become familiar with the PTs* (T 14), *completing validation procedures* (T 12), *reviewing translations* (T 8), *acquiring technological infrastructure* (T 4), and *acquiring funding throughout the project* (T 2).

Data also indicated that Country A had difficulty with two tasks (Figure 8). Country A was not able to complete Task 13, *Testing the performance task implementation with students*. Country A had disconfirming evidence for 100% of observations for Task 13. Country A also had disconfirming evidence across 55% of observations when *submitting feedback* (T 17).

There was one task during which Country B was successful in meeting criteria (Figure 9). Country B had confirming evidence for 91% of observations regarding *acquiring technology necessary for the project* (T 4). The positive percentages at the bottom right of the symmetry graphs represent CEBs. Country B also had confirming evidence across 82% of observations when *selecting PTs* (T 3) and *acquiring funding* (T 2).

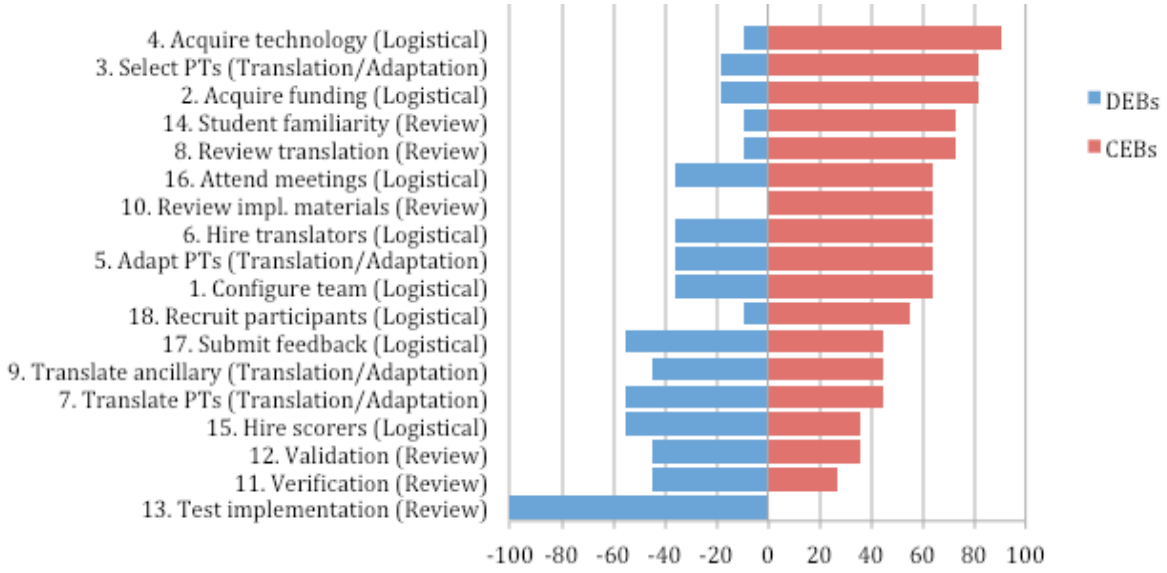

Figure 9. Country B's percentages of DEBs and CEBs across tasks.

There were four tasks during which Country B had disconfirming evidence for over 50% of the observations (Figure 9). The negative percentages at the bottom left of the symmetry graphs represent DEBs. Country B was unable to *test PT implementation with students* (T 13). As a result, there was disconfirming evidence for 100% of observations for Task 13. In addition, there was disconfirming evidence across 55% of observations when Country B, *submitted feedback* (T 17), *hired scorers* (T 15), and *translated the PTs*, (T 7).

Country C was equally successfully during the completion of two tasks (Figure 10). For 82% of the observations, there was confirming evidence when Country C *provided students with an opportunity to become familiar with the PT format* (T 14) and *acquired technology* (T 4).

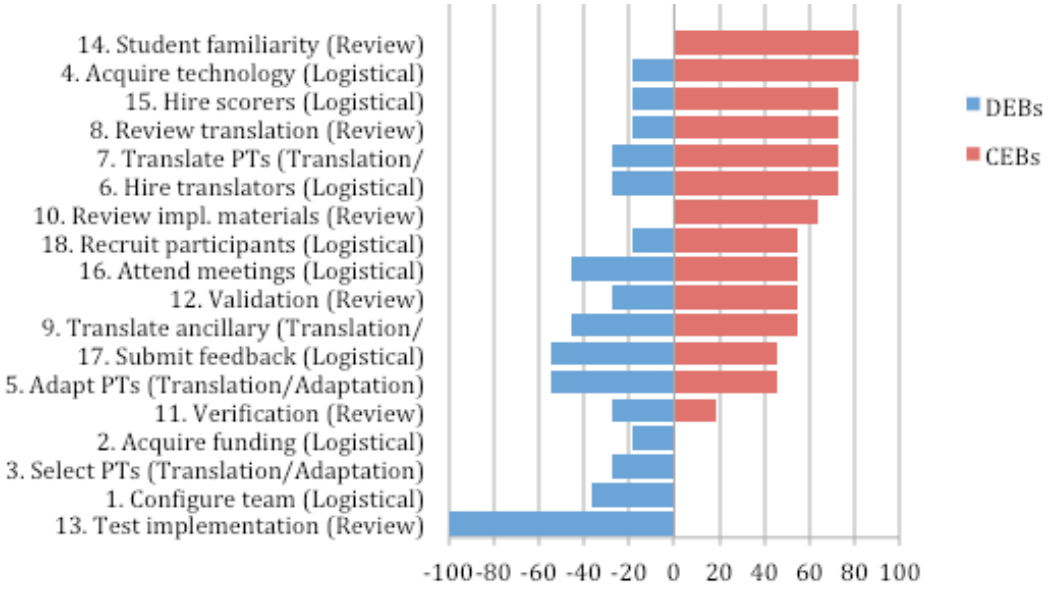

Figure 10. Country C's percentages of DEBs and CEBs across tasks.

However, Country C found complying with criteria while performing Task 13, Task 17, and Task 5 challenging (Figure 10). There was disconfirming evidence across 100% of observations regarding the *testing of PT implementation* (T 13). Also, there was disconfirming evidence across 55% of the observations regarding the *submission of feedback* (T 17) and *adaptation of the PTs* (T 5).

Documents showed that there were three tasks Country D completed while meeting criteria in this framework (Figure 11). There was confirming evidence across 100% of observations addressing Country D's *configuring of the country team* (T 1). While completing Task 14, *Providing students with an opportunity to become familiar with the PTs*, and Task 4, *Acquiring technology*, Country D had confirming evidence across 91% of observations.

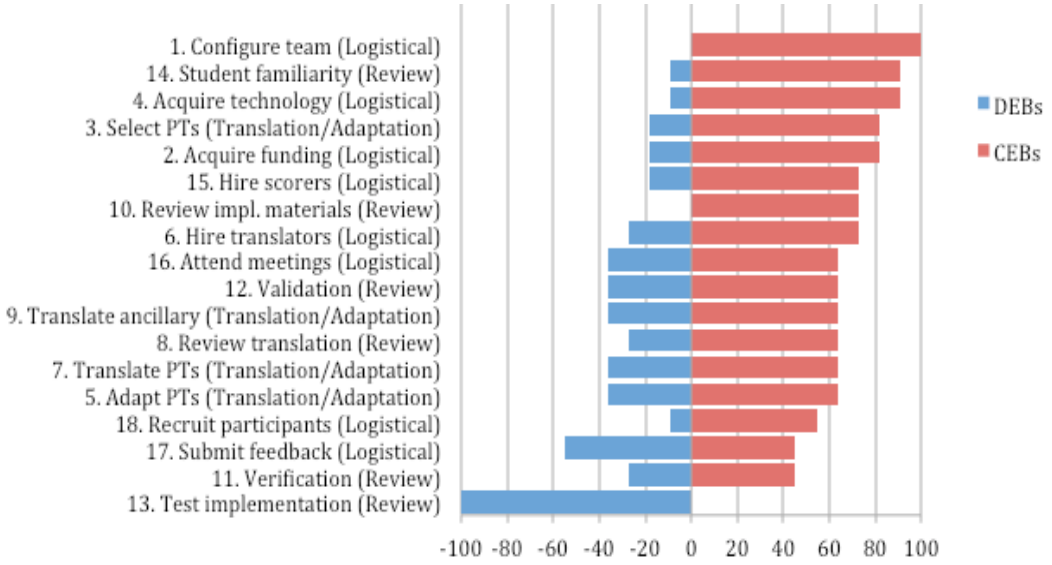

Figure 11. Country D's percentages of DEBs and CEBs across tasks.

Evidence also indicated that meeting criteria while completing two tasks was challenging for the Country D team (Figure 11). There was disconfirming evidence across 100% of the observations addressing Task 13, *Testing the PT implementation*, showing that the team was not able to complete any part of the task. Documents included disconfirming evidence for Task 17, *Submit feedback*, across 55% of the observations.

Country E had a high amount of confirming evidence for across 8 of the 18 tasks included in this framework (Figure 12). Data showed confirming evidence across 100% of observations when *configuring the team* (T 1) and when *providing students with an opportunity to become familiar with the PT structure and response requirements* (T 14). Information also included confirming evidence for 91% observations when *acquiring funding* (T 2), *selecting the PTs* (T 3), *acquiring technology* (T 4), *hiring translators* (T 6), *translating the PTs* (T 7), and *translating ancillary material* (T 9).

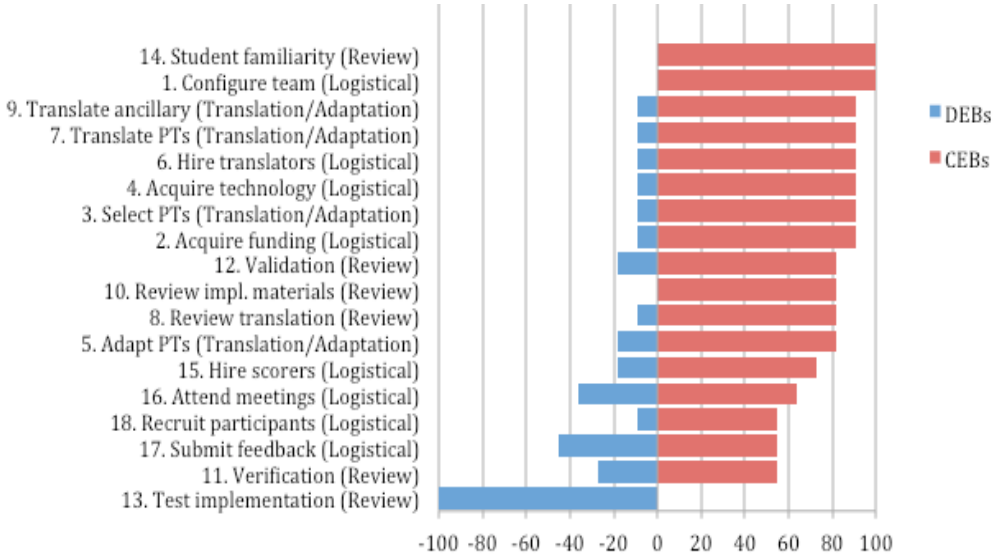

Figure 12. Country E's percentages of DEBs and CEBs across tasks.

There was only one task included in this framework with which Country E experienced challenges (Figure 12). Documents included disconfirming evidence across 100% of the observations regarding *testing PT implementation* (T 13). Country E was not able to complete any component of this task and, as a result, the country team was not able to test the computer interface with its students.

Using symmetry graphs comparing total CEBs and DEBs for each country provides an overall summary view of the CEB to DEB ratio for each country across all tasks (Figure 13). Countries with a high percentage of CEBs and low percentage of DEBs had the best CEB to DEB ratio. As a result these countries had a high fidelity of implementation.

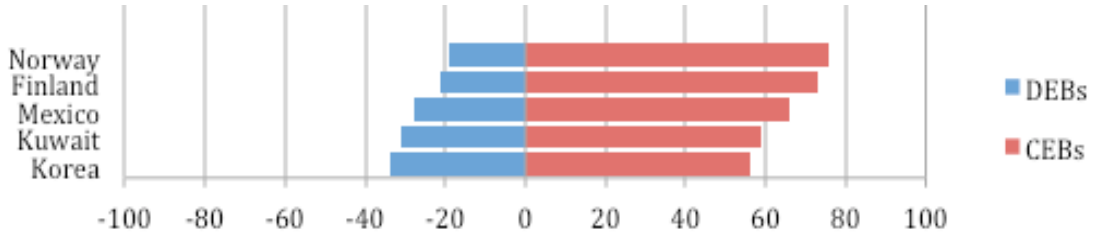

Figure 13. Percentages of DEBs and CEBs across tasks for each country.

Across the 18 tasks, Country E had the highest CEB to DEB ratio. Country E was the only country to meet criteria while completing over 75% of the 18 tasks. Country E had CEBs across 76% of the intersections of a criterion and task and DEBs for 19% of the intersections of a criterion and task. Country E was able to meet criteria when completing over three-quarters of the tasks successfully, approximately 14 tasks, successfully; Country E had difficulty meeting criteria when performing 3 of the 18 tasks. Country A, with the second highest CEB to DEB rate, had CEBs for 73% of the intersections of a criterion and task and DEBs for 21%. Country A was able to comply with criteria when completing 13 of the tasks included in this framework and had difficulty with meeting criteria while completing four tasks.

According to data, Country B had the lowest CEB to DEB rate across criteria and tasks. Country B had CEBs for 56% of the intersections of a criterion and task and DEBs for 34%. Country C also experienced some challenges when completing tasks while meeting criteria. Country C had CEBs for 59% of the intersections of a criterion and task and DEBs for 31% of the intersections of a criterion and task.

Discussion of task completion: Symmetry graphs

Symmetry graphs provide a visual representation of the relationship between CEBs and DEBs for each task for each country and across countries. Country E and Country A were highly

successful in completing a large number of tasks as exhibited by their CEBs and DEBs. However, Country B had the greatest difficulty across the greatest number of tasks.

Country E had the highest number of tasks with a high percentage of CEBs while also obtaining the lowest number of tasks with a high percentage of DEBs. Documents showed that Country E had a high percentage of CEBs across eight tasks: *configure a team* (T1), *allowing students to become familiar with PT format* (T14), *acquire funding* (T2), *select PTs* (T3), *acquire necessary technology* (T4), *hire translators* (T6), *translate PTs* (T7), and *translate ancillary material* (T9).

Country E's success in completing these tasks may be explained, in part, by the prior experience that the Country E team members had in international test comparisons. Country E was able to configure a national team that included expertise in various key areas. Several team members had previously worked on translation of PISA and TIMSS assessments. The team's national project manager previously worked with OECD to author background reports and help Country E participate in international comparison studies. As a result, the team was well prepared to select, translate, and review PT translations. In addition, the team had already put in place the type of technical infrastructure required to work on an international test translation project. Also, the team also understood the importance of familiarizing students with item and test requirements and how to accomplish this task. Finally, as with other assessment programs, the county's Ministry of Education and Research assisted the team in securing funds necessary to participate in the international comparison study.

Similarly, Country A had a high percentage of CEBs while performing six tasks: *configure a team* (T1), *allowing students to become familiar with PT format* (T14), *validation of changes* (T12), *review translation* (T8), *acquire technical infrastructure* (T4), and *acquiring funding*

(T2). The Country A team's experience with international test comparisons also explains the team's ability to complete these tasks. The Country A national project manager had experience with higher education assessments that took place across diverse European countries. Specifically, the national project manager had examined the impact of the Bologna Process. This experience helped Country A with test translation, translation review, and validation procedures. In addition, the Country A Ministry of Education selected the national coordination group and helped secure funds needed to participate in the AHELO study. Finally, given Country A's experience with participation in other international test comparisons they had the technical infrastructure in place required to participate in the study.

Both Country A and Country E had only one task with a high percentage of DEBs: *test PT implementation* (T13). In fact, none of the five countries were able to complete Task 13. A challenge to completing this task was the need for access to the computer version of each translated test. An external contractor located in the U.S. uploaded all of the translated documents onto a proprietary computer platform. Loading documents that required different language characters and text direction into a computer system initially created solely for English proved challenging. Due to the amount of time required to upload each set of translated documents, countries were not able to test the computer system with their students. As a result, none of the countries were able to complete any part of Task 13.

Country B had a high percentage of CEBs across three tasks: *acquire technology* (T4), *select PTs* (T3), and *acquire funding* (T2). The Country B team's success in completing Task 4, Task 3, and Task 2 stems from the external assistance that the team received. The selection of PTs was completed through a consensus-based activity that included all country teams and was guided by the translation technical assistance team (TAT) hired by CAE. Additionally, the

Country B Educational Development Institute (KEDI) and the Seoul National University in Country B help finance the project by providing direct funds, staff members, and their technical infrastructures.

Country B had greater difficulty completing tasks when relying solely on the country team expertise. The Country B team had a high percentage of DEBs across four tasks: *test implementation* (T13), *submit feedback* (T17), *hire scorers* (T15), and *translate PTs* (T7). There are two contributing factors to Country B's difficulty in completing these tasks. First, Country B had difficulty when transferring national project management activities from an interim to a permanent staff member. As a result, it was difficult for the team to submit feedback and transfer the knowledge necessary to hire scorers in a timely manner. Second, the staff that the country team hired to complete translation work did not possess the qualifications provided by the Translation technical assistance team (TAT) hired by CAE. Instead, the team was guided strictly by local cultural norms when hiring translation staff. The team hired professors with doctoral degrees in literature instead of experienced translators with a translation certification. Therefore, the team was not able to complete the translations as expected.

Task Completion: F coefficients

There were several tasks for which countries, individually and collectively, were able to meet a high percentage of the criteria (Table 10).

Table 10

Individual Country and Average F Coefficient by Three Task Categories

Category	Task	Country A	Country B	Country C	Country D	Country E	Average
Logistical	1-Configure team	1	.64	.64	1	1	.86
	2-Acquire funding	.91	.82	.82	.82	.91	.86
	4-Technical infrastructure	.91	.91	.82	.91	.91	.89
	6-Hire translators	.82	.64	.73	.73	.91	.77
	15-Hire scorers	.80	.40	.80	.80	.80	.72
	16-Attend meetings	.73	.64	.55	.64	.64	.64
	17-Submit feedback	.55	.45	.45	.45	.55	.49
	18-Recruit HEIs	.86	.86	.75	.86	.86	.84
Translation and Adaptation	3-Select items	.82	.82	.73	.82	.91	.82
	5-Adapt test	.73	.64	.45	.64	.82	.66
	7-Translate assessment	.73	.45	.73	.64	.91	.69
	9-Translate ancillary	.73	.50	.55	.64	.91	.67
Review Procedures	8-Review translation	.91	.89	.80	.70	.90	.84
	10-Review impl. materials	1	1	1	1	1	1
	11-Verification changes	.63	.38	.40	.63	.67	.54
	12-Validation changes	.91	.44	.67	.64	.82	.70
	13-Test implementation	0	0	0	0	0	0
	14-Student familiarity	1	.89	1	.91	1	.96
	Average	.78	.63	.66	.71	.81	.72

I calculated a F coefficient for each task by dividing the CEBs by the total number of CEBs and DEBs for each task by each country and across all countries. For special cases, it was necessary to examine the number of NEBs associated with specific tasks to help provide context and meaning for the F coefficient.

Initially, countries appeared to have the highest F coefficient for Task 10, *Review of implementation materials*. Each country achieved a F coefficient of 1 for Task 10. According to the information available for CEBs and DEBs, a F coefficient of 1 for this task indicates fully successful implementation. However, a closer examination shows that none of the countries met all 11 criteria when completing Task 10. The number of NEBs associated with the completion of

Task 10 across countries explains the misleadingly high F coefficient. On average, there were NEBs for 29% of criteria for this task.

Countries had an F coefficient of .85 or higher for four tasks. Three of the four tasks were in the Logistical category (T1, T2, and T4) and one was in the Review Procedures category (T14). The five participating countries were able to meet closely the criteria presented in this framework while completing these four tasks. For Task 14, *Providing students with an opportunity to become familiar with the assessment*, countries had an F coefficient of .96. Individually, Country A, Country C, and Country E had an F coefficient of 1 for Task 14. With an F coefficient of .89 and .91 respectively, Country B and Country D were also able to comply with most criteria when completing Task 14. When performing Task 4, *Acquiring the necessary technical infrastructure*, countries collectively obtained the next highest F coefficient. For Task 4 countries had, on average, an .89 F coefficient. Country A, Country B, Country D, and Country E each had an F coefficient of .91 for Task 4; Country C had an F coefficient of .82. Two tasks, *configuring a team* (T 1) and *acquiring funds* (T 2) had the same results each had an F coefficient of .86. When *configuring a team* (T1), Country A, Country D, and Country E each had an F coefficient of 1; Country B and Country C each had a coefficient of .64.

There were four tasks for which countries had an F coefficient between .75 and .84. Two of the tasks were in the Logistical category (T6 and T18), one was in the Translation and Adaptation category (T3), and one was in the Review Procedures category (T8). Countries had an F coefficient of .84 for Task 8, *Translation review*. Country A had an F coefficient of .91, Country B a .89, and Country E .90, indicating a high FOI during translation review. With an F coefficient of .80 and .70 respectively, Country C and Country D appear to have experienced difficulty in meeting criteria when completing translation review.

Countries also had a high F coefficient, .84, for Task 18, *Recruiting higher education institutions*. Country A, Country B, Country D and Country E each had a F coefficient of .86 when completing Task 18. Country C had an F coefficient of .75 indicating that Task 18 was more challenging for Country C than the other countries. However, this result needs to be interpreted with caution, as there was a NEB rate of 35%.

Data showed that countries obtained a .82 F coefficient for Task 3, *Select items*. Country E met criteria successfully with an F coefficient of .91. Country A, Country B, and Country D had an F coefficient of .82. Country C had an F coefficient of .73 indicating that the country team had less success than the other countries meeting criteria when selecting items.

When completing Task 6, *Hiring translators*, countries had a .77 F coefficient. However, there was variability in individual country ability to meet criteria when performing Task 6. Country E was highly successful in meeting criteria, which is reflected in a .91 F coefficient. Country A was also able to meet most criteria and had a F coefficient of .82. Country C and Country D each had a F coefficient of .73 indicating that they experienced more difficulty in complying with criteria. Finally, Country B met the least number of criteria when *hiring translators*, which is reflected with a .64 F coefficient.

There were seven tasks for which countries had FOI indices between .50 and .74. The greatest number of tasks fell in this F coefficient range. Two tasks are in the Logistical category, three in the Translation and Adaptation category, and two in the Review Procedures category. For Task 15, *Hire scorers*, countries had a .72 F coefficient on average but there was variability across countries. Country A, Country C, Country D, and Country E each had a F coefficient of .80. However, Country B was not able to meet most criteria, which is reflected in the .40 F coefficient. When completing Task 12, M*ake changes based on validation procedures*, countries

had a .70 F coefficient. Country A met a high number of criteria and had a FOI of .91and Country E was successful with a .82 F coefficient. However, Country C had an F coefficient of .67, Country D a .64, and Country B a .44 indicating difficulty in meeting criteria.

For Task 7, *Translate assessment*, country teams obtained a .69 F coefficient. Country E was very successful in complying with criteria and had an F coefficient of .91. Country A and Country C each had a F coefficient of .73, Country D a .64, and Country B a .45 F coefficient. When completing Task 9, *Translate ancillary materials*, country teams had a .67 F coefficient. Country E was very successful at meeting criteria and had a F coefficient of .91. Country A had a F coefficient of .73, Country D .64, Country C .55, and Country B .50, showing that they had difficulty meeting criteria.

For Task 5, A*dapt a test*, countries had a .66 F coefficient. Countries greatly varied in their ability to meet criteria when performing Task 5. Country E had an F coefficient of .82 and Country A a .73. Country B and Country D each had a .64 F coefficient. Country C had a .45 F coefficient.

For Task 16, *Attend meetings*, countries had a .64 F coefficient. Country A had an F coefficient of .73. Country B, Country D, and Country E each had a .64 F coefficient. Country C had a .55 F coefficient. Finally, when completing Task 11, *Translation verification*, country teams achieved a .55 F coefficient. However, it is important to note that for Task 11 countries had a 31% NEB rate, which results in an F coefficient that is artificially high. Country E met six criteria, Country A and Country D five criteria, Country B three criteria, and Country C two criteria.

There were two tasks for which countries had FOI indices below .50. For Task 13, *Testing the performance task implementation*, country teams had the lowest F coefficient. When trying to

complete Task 13, country teams achieved a CEB to DEB ratio of 0 to 11. Since there were 0 CEBs and 0 NEBs, there is no CEB to NEB ratio. None of the five countries were able to complete the task and as a result the F coefficient for Task 13 was 0. When completing Task 17, *Submitting feedback on the process and progress*, country teams had the second lowest F coefficient: .49. Country A and Country E had an F coefficient of .55. Country B, Country C, and Country D had a .45 F coefficient. Collectively and individually, countries were able to meet less than half of the criteria when completing Task 17.

Discussion of task completion: F coefficients

Calculating the F coefficient allowed me to quantify each country's ability to perform tasks and examine patterns between country teams' performance. Task 14, *allowing students to become familiar with the PTs*, had the highest F coefficient, .96. Task 4, *Acquiring technical infrastructure*, had a .89 F coefficient. Task 2, *Acquiring funds*, each had a F coefficient of .86. Consistently high success in completing these tasks can be attributed to several factors. First, the Translation technical assistance team (TAT) hired by CAE supported countries while completing these tasks. The translation technical assistance team (TAT)supplied timely detailed information on the type of technology that the countries would need throughout the project. Members of the TAT created a mini-PT that students could use to practice prior to taking the actual test. Second, each country's ministry of education and one local university provided their existing technical infrastructure to country teams for use throughout the study. Also, country teams received financial assistance, in the form of staff and direct funds, from their ministries of education.

Task 1, *Configuring a team*, had a .86 F coefficient. However, there was inconsistency across countries in completing this task. Country A, Country D, and Country E each had an F coefficient of 1. These three countries were able to configure a team with project management

experience, measurement experience, and knowledge of language and culture. The members of the team with the expertise in each field were able to contribute throughout the duration of the study. The three teams had some experience with large-scale assessment. Country B and Country C each had a .64 F coefficient indicating difficulty in completing Task 1. Neither team had members with prior experience in international testing. Country B did not include members with expertise and experience in translation or project management. Country C was not able to include members with expertise in measurement.

There were two tasks that had consistently low F coefficients across countries. Task 13, *Test assessment implementation*, had a F coefficient of 0 for each country. Task 17, *Submit feedback*, had an average F coefficient of .49 and individual country F coefficients for .55 or .45. The difficulty with Task 13 arose because of the technological demands placed on the contracted technical team. As a result, the country teams did not have access to the computer interface with translated PTs to share with their student participants. The low F coefficients for Task 17 were due to several reasons. First, only Country A and Country D were able to complete a survey asking about the study's adaptation process or participate in an interview regarding rubrics. These two countries were the only ones with a member of the team dedicated to communication responsibilities. Second, there was no formal support for countries to provide ongoing feedback. All country teams gave feedback during in-person and telephone meetings; however, the countries were not aware of diaries they could use to provide constant feedback until well into the study. These diaries were not put in place by the Translation technical assistance team (TAT) hired by CAE, but rather by an international agency working with the countries on other assessments.

Task 11, *implement changes based on verification procedures*, had a low F coefficient on average, .54. However, there was evidence of some inconsistency between countries. Country E, Country A, and Country D had F coefficients of .67, .63, and .63 respectively. Country C achieved a .40 F coefficient and Country B a .38. Upon closer examination, the differences in the coefficients were a result of a 31% NEB rate. There was limited information regarding how the external verification was to be completed, how countries were to review the suggested changes, or how countries were to use the resulting suggested changes.

Criteria FOI

Examining results by criterion provides another way to quantify the general level of fidelity of implementation. There were 11 criteria included in this framework, which I grouped into three categories dealing with general aspects of any project: Support for Task Completion, Ease of Use and Implementation, and Logistics (Table 11). After examining criteria according to the three categories, I analyzed results at the individual criterion level for individual evidence types as well as the F coefficient.

At both the category-level and individual criterion-level analysis, I recorded the number of confirming (CEBs) and disconfirming (DEBs) evidence bits as well as the occasions for which there was insufficient information (NEBs). The higher the percentage of CEBs associated with a criterion and the lower the DEBs and NEBs associated with a criterion, the higher the degree of FOI with which countries were able to follow the framework.

Criteria compliance: Countries

While examining criteria compliance by country, I created three cut-off points, one cut-off point for each of the evidence types (Table 11). The cut-off points are directly related to the level

of FOI. For criteria, I determined that CEBs across 75% of tasks was indicative of a high FOI. I also determined that criteria with DEBs or NEBs for at least one-third of tasks suggested a lower FOI.

Across the five countries, there were four criteria with more than 75% CEBs indicating a high FOI. One criterion with high FOI was in the Support category: *acquiring support from outside of the country* (C 11). One criterion with high FOI was in the Ease of Use category: *having user-friendly materials* (C9). Two criteria with high FOI were in the Logistics category: *participating in timely communication* (C 1) and *documenting the study's progress* (C 7). The Logistics category had the most number of criteria with high FOI; however, the difference in the number of criteria meeting the cut-off did not vary greatly across categories.

Table 11
Criterion CEB, DEB, and NEB Cut-Off Points by Category

Category		Cut-off points		
Support	Criteria	>75% CEB	≥ 1/3 DEB	≥ 1/3 NEB
	C2-Measurement expertise	1		
	C3-Translation expertise		1	
	C4-Project management expertise			
	C10-Support in country			
	C11-Support out of country	1		
Ease/Use				
	C5-Review opportunities		1	
	C6-Training opportunities			
	C9-User-friendly materials	1		
Logistics				
	C1-Timely communication			
	C7-Document progress	1		
	C8-Apt deadlines		1	

On average, three of the eleven criteria had DEBs for at least one-third of the tasks completed. There was one criterion in each of the three categories. Countries experienced difficulty when trying to *include translation expertise* (C3), *participating in review procedures* (C5), and *meeting apt deadlines* (C8). Criterion 3 is part of the Support category, Criterion 5 part of the Ease of Use category, and Criterion 8 is part of the Logistic category. None of the 11 criteria had NEBs across at least one-third of the tasks.

Results by individual country across all criteria provide another perspective through which fidelity of implementation can be examined. The total number of CEBs, DEBs, and NEBs that countries attained for each criterion helps to explain the level of fidelity of implementation with which the criteria from this framework were met. Although the percentages I chose as guidelines are based on the specific results of this study, they provide a framework through which to discuss findings. For a high FOI, a country needed to meet at least eight criteria when completing at least 75% of the tasks included in the framework.

Countries varied in their ability to meet criteria (Table 12). On average, all five countries had CEBs for 51% of criteria. However, the standard deviation was 24 indicating variability across criteria compliance success. Documents show that Country E met 82% of criteria for at least 13 of the 18 tasks. Country A complied with 64% of criteria when performing at least 13 of 18 tasks. Information pertaining to Country D revealed that the country team met 55% of criteria for at least 13 tasks. Country C and Country B each met 27% of criteria for at least 13 tasks.

Table 12
Criteria Compliance Percentages, Means, and Standard Deviations of CEBs, DEBs, and NEBs for Each Country and on Average

Evidence bit type	Country A	Country B	Country C	Country D	Country E	Mean (SD)
CEBs ≥ 13	64	27	27	55	82	51 (24)
DEBs ≥ 5	36	64	55	55	18	46 (18)
NEBs ≥ 5	0	9	0	0	0	2 (4)

Examining the number of times that countries had DEBs across criteria when completing at least five of 18 tasks helps to identify challenges to fidelity of implementation. On average, countries had DEBs for at least five tasks across 46% of criteria with a standard deviation of 18. Country B experienced the greatest difficulty complying with criteria. Country B had DEBs across 64% of criteria for at least 5 of the 18 tasks. Country C and Country D also experienced difficulty complying with criteria. Each country had DEBs for 55% of criteria across at least five tasks. Country A had DEBs for 36% of criteria for at least five tasks. Finally, Country E experienced the least difficulty meeting criteria. Country E had DEBs for only 18% of criteria across at least five tasks.

On average, all countries had missing information for at least five tasks across 2% of criteria. However, a standard deviation of 4 indicates there was variability between the countries. Country B had NEBs for at least five tasks for 9% of criteria, indicating difficulty documenting their progress. However, none of the other four countries had missing information for at least five tasks.

Another way to compare countries was by examining the relationship between countries according to the rank ordering of tasks and criteria by CEBs and DEBs. A high correlation between countries indicates that the criteria that countries were able to meet were similar.

Likewise, a high correlation would also indicate that the criteria that countries were not able to comply with were similar. A low correlation coefficient would indicate that the criteria countries were able to complete varied.

Figure 14 is a correlation matrix using Spearman rank correlation between the five countries according to CEBs across the 11 criteria included in this framework. The Spearman correlation value helps explain the similarity and differences in the criteria with which countries were able to comply.

	Country A	Country B	Country C	Country D	Country E
Country A	1	0.74	0.45	0.95	0.65
Country B		1	0.69	0.82	0.62
Country C			1	0.43	0.11
Country D				1	0.75
Country E					1

Figure 14. Correlation matrix of CEBs across countries for 11 Criteria.

The Spearman correlations of CEBs showed that Country D and Country A were most similar in the criteria that they were able to meet or with which they had difficulty: 0.95. The correlation between Country D and Country B was 0.82, also indicating a strong similarity in complying with criteria. The correlation between Country A and Country B, 0.74 also confirms a similarity in the ability to meet criteria. These strong correlations indicate that countries were similar in their ability to meet criteria.

Several countries had weak correlations, indicating dissimilarities in the way that countries were able to meet criteria. With a correlation of 0.11, Country C and Country E had the weakest relationship between countries. The next weakest correlation, 0.43, indicated that Country D and

Country C had the second weakest relationship. There was also a weak correlation of 0.45 between Country C and Country A.

Examining specific counts of CEBs, DEBs, and NEBs for each criterion across tasks for all countries provides specific information about the ability to apply criteria with specific work associated with translation and adaptation. This analysis also provides context for the correlation results. Country teams had high percentages of confirming evidence when meeting four criteria (Table 13). Country teams were able to *document progress* (C 7) for 86% of the tasks. For another criteria, *acquiring support from outside of the country* (C 11) countries had CEBs for 83% of tasks across countries. Country teams were able to *use materials* that the Translation technical assistance team (TAT) hired by CAE provided (C 9) for 78% of the tasks. In addition, for 76% of the tasks, country teams were able to have *timely communication* (C 1).

Table 13
Percentages of CEBs for Each Criterion Across All Tasks (Rounded Percentages)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
All Tasks	76	62	38	63	52	69	86	59	78	61	83

Countries had high percentages of disconfirming evidence for three criteria (Table 14). The criterion inclusion of *translation expertise* (C 3) throughout the study was challenging for countries to meet. There was disconfirming evidence for countries when completing Criterion 3 across 85% of tasks. Countries found *reviewing work* throughout the project (C 5) challenging. There were DEBs for 67% of tasks across all countries. Finally, when *meeting deadlines* (C 8) countries had DEBs for 65% of tasks.

Table 14
Percentages of DEBs for Each Criterion Across All Tasks (Rounded Percentages)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
All Tasks	23	27	47	19	37	31	14	36	22	21	17

There were two criteria for which there was insufficient information (Table 15). There was a small amount of information missing for the criterion *project management expertise* (C 4). Information was missing across 18% of tasks for all countries. Information was also missing for 18% of the tasks *support found in the country external to the team* (C 10).

Table 15
Percentages of NEBs for Each Criterion Across All Tasks (Rounded Percentages)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
All Tasks	1	11	16	18	11	0	0	6	0	18	0

Discussion of criteria FOI

Across the five countries, there were four criteria that had CEBs for more than 75% of observations. Throughout the study, there was successful *inclusion of measurement expertise* (C 2), *support from out of country* (C 11), *access to user-friendly materials* (C 9), and *documentation of progress* (C7) across countries. Success in meeting the four criteria is due in large part to the work accomplished within country teams with support from the translation technical assistance team (TAT) hired by CAE. The TAT created 76 documents that country teams could use while completing tasks. Also, in addition to two psychometricians working on the TAT, Country A, Country B, Country D, and Country E included measurement experts as part of their core team throughout most of the study. Finally, the Translation technical assistance

team (TAT) hired by CAE solicited information from country teams and wrote seven reports documenting country teams' progress throughout the study.

On average, countries met criteria for at least 13 tasks across 51% of observations when completing tasks. However, a standard deviation of 24 indicates that the degree to which individual countries met criteria while completing tasks varied. Country E complied with criteria for at least 13 tasks across 82% of observations. Country B and Country C experienced some difficulty and met criteria across at least 13 tasks across 27% of observations. Country E's vast experience with international comparison studies, research project management, and test translation facilitated this country's compliance with criteria. Country B and Country C did not have deep knowledge or experience concerning international test comparisons or managing large research projects.

Criteria compliance: Symmetry graphs

A symmetry graph shows the differences between the criteria by percentages and rank individual criterion according to their number of CEBs from the highest at the top to lowest percentage at the bottom. The symmetry graphs in Figures 17, 18, 19, 20, and 21 compare CEBs with DEBs for criteria by each of the five countries. The visuals show the differences between the criteria in percentages.

Documents showed that Country A had confirming evidence for *timely communication* (C 1) across 89% of observations (Figure 15). In addition, there was confirming evidence for 83% of observations for three criteria: acquiring *out of country support* (C 11), *applying user-friendly materials* (C 9), and *documenting progress* (C 7) (Figure 15).

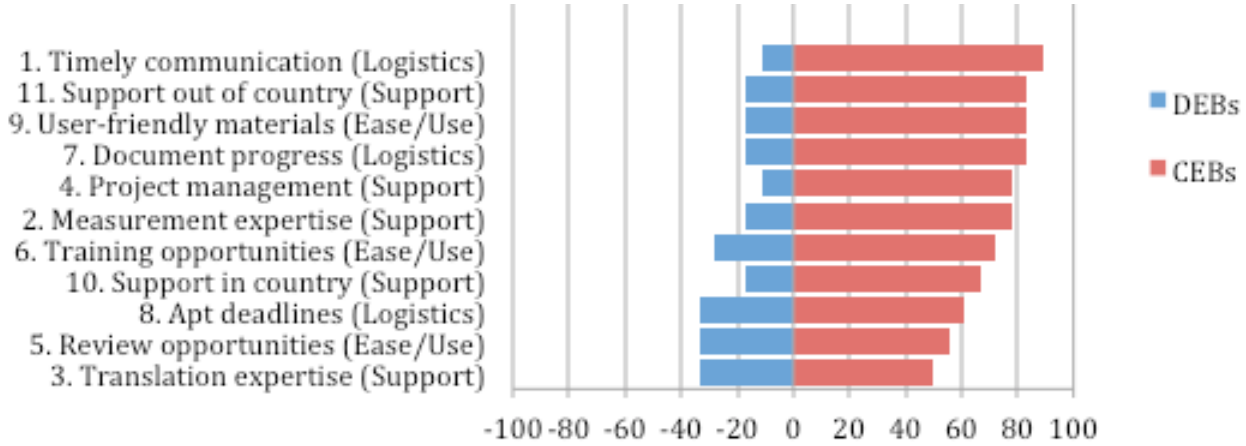

Figure 15. Country A's percentages of DEBs and CEBs across 11 criteria.

Data included information about criteria with which Country A experienced challenges as well (Figure 17). Country A had disconfirming evidence for 33% of data when *including translation expertise* (C 3), incorporating *review opportunities* (C 5), and *working within deadlines* (C 8).

Data showed Country B had a high amount of confirming evidence for two criteria (Figure 16). Country B had confirming evidence for 89% of observations when *documenting progress* (C 7). Documents also included confirming evidence across 83% of data when Country B was gaining *support from outside of the country* (C 11).

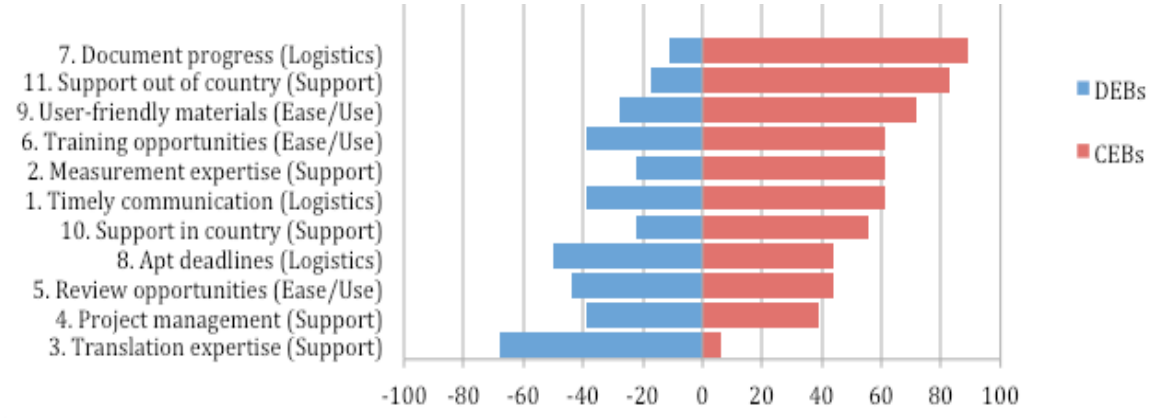

Figure 16. Country B's percentages of DEBs and CEBs across criteria.

Documents indicated that Country B experienced difficulty with six criteria (Figure 16). There was disconfirming evidence for 67% of data when Country B was supposed to be *including translation expertise* (C 3). There was also disconfirming evidence across 50% of observations when Country B tried to *meet deadlines* throughout the project (C 8) and across 45% of observations when complying with *review opportunities* (C 5). Additionally, there was disconfirming evidence for 37% of data when including *project management* (C 4), participating in *training opportunities* (C 6), and participating in *timely communication* (C 1).

There were three criteria for which Country C had confirming evidence across 83% of observations (Figure 17). Documents included confirming evidence for Criterion 11, *acquiring out of country support throughout the project*, Criterion 9, *Accessing user-friendly materials*, and Criterion 7, *Documenting progress throughout the project*.

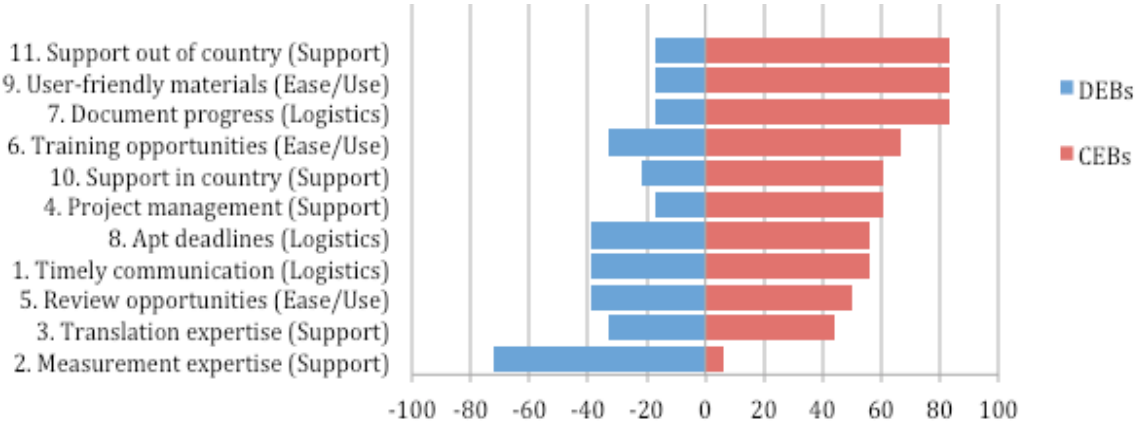

Figure 17. Country C's percentages of DEBs and CEBs across criteria.

There were four criteria that proved challenging for Country C (Figure 17). Data showed that Country C had 72% disconfirming evidence when *including measurement expertise* (Criteria

2). Country C also had disconfirming evidence for 39% of observations when *communicating in a timely manner* (T 1), *reviewing translation* (T 5), and *meeting deadlines* (T 8).

Country D successfully met three of the 11 criteria included in this framework (Figure 18).

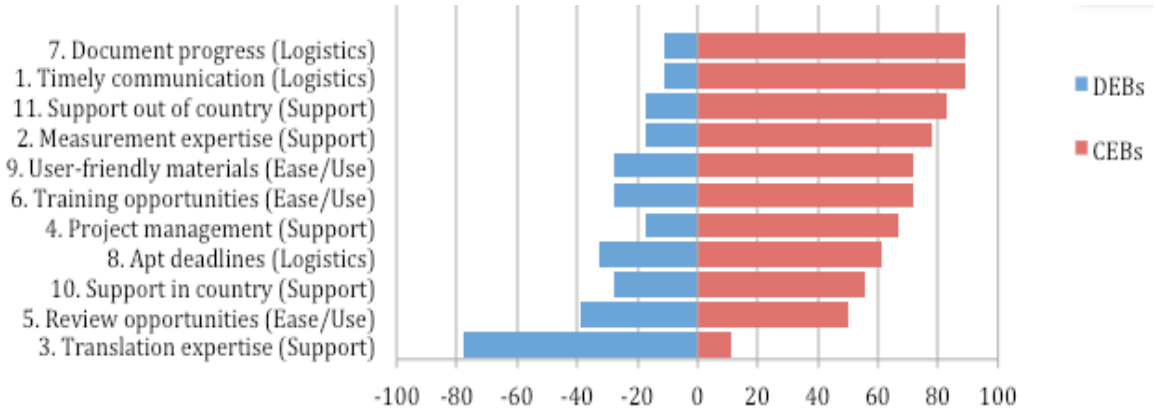

Figure 18. Country D's percentages of DEBs and CEBs across criteria.

Documents include confirming evidence across 89% of observations while Country D was *documenting progress* (C 7) and *communicating in a timely fashion* (C 1). In addition, for Criterion 11, *gaining out of country support*, there was confirming evidence across 83% of data observed.

Three criteria were particularly challenging for Country D (Figure 18). Data showed disconfirming evidence for Criterion 3, *Including translation expertise*, across 78% of data. There was disconfirming evidence across 39% of observations for *reviewing translations* (C 5) and 33% for *meeting deadlines* (C 8).

Country E was successful with four criteria included in this framework (Figure 19). Data included confirming evidence throughout 89% of observations for Criterion 2, *Including measurement expertise throughout the project*. Documents also included confirming evidence for

83% of observations when *communicating in a timely fashion* (C 1), *documenting progress* (C 7), and when *acquiring support from outside of the country* (C 11).

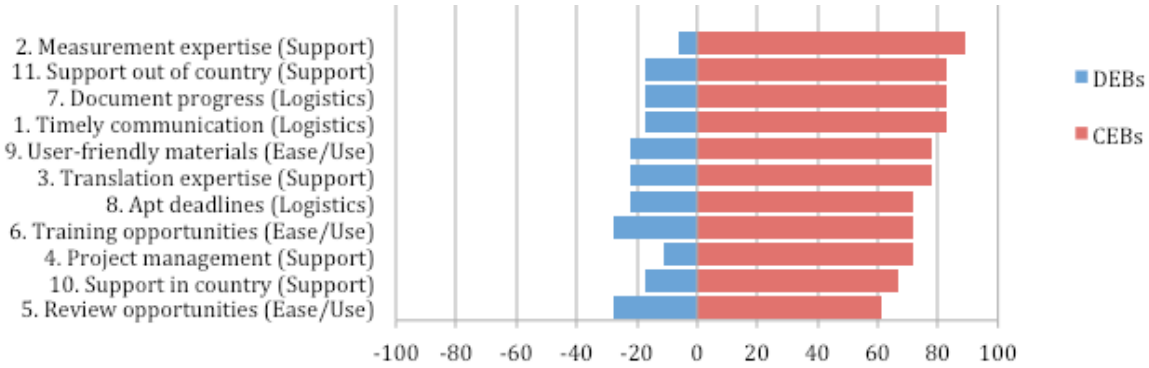

Figure 19. Country E's percentages of DEBs and CEBs across criteria.

For Country E, there was limited disconfirming evidence across all criteria included in this framework (Figure 19). There were two tasks for which there was disconfirming evidence across 28% of data: Criterion 5, *Participate in opportunities for review*, and Criterion 6, *Participation in training opportunities*.

Using symmetry graphs comparing CEBs to DEBs across countries can also provide a CEB to DEB ratio criteria for each country (Figure 20). Data showed that Country E had the highest CEB to DEB ratio across criteria included in this framework. Country E had CEBs for 76% of the 11 criteria and DEBs for 19% of the criteria. Country A had the second highest CEB to DEB ratio across criteria. Documents included CEBs for 73% of criteria and DEBs for 21% of criteria included in this framework.

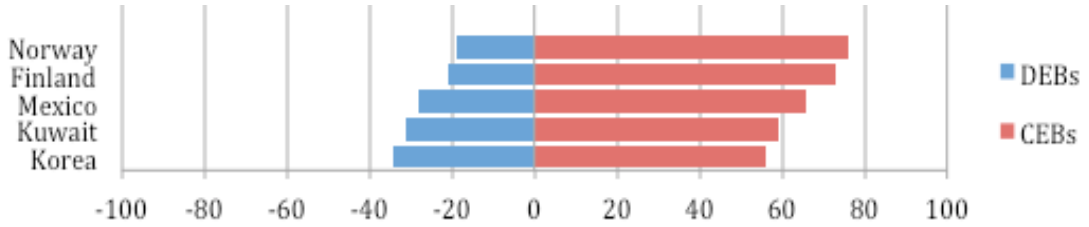

Figure 20. Percentages of DEBs and CEBs across criteria for all five participating countries.

Country B had the lowest CEB to DEB ratio across criteria. Data showed that Country B had CEBs for 56% of criteria and DEBs for 34% of the criteria. Country C had the second to lowest CEB to DEB ratio across criteria. Documents showed that Country C had CEBs for 59% of criteria and DEBs for 31% of criteria. Country D had CEBs for 66% of criteria and DEBs for 28% of criteria.

Discussion of criteria compliance: symmetry graphs

Symmetry graphs provide a visual representation of the relationship between CEBs and DEBs for each criterion for each country and across countries. Country A and Country E were highly successful in meeting a large number of criteria as exhibited by their CEBs and DEBs. However, Country B had the greatest difficulty across the greatest number of criteria.

Country A and Country E were each successful in meeting four criteria across a high number of tasks. Country A contributed to *timely communication* (C 1) across 89% of observations. Country A also had CEBs for 83% of observations when *acquiring out of country support* (C 11), *applying user-friendly materials* (C 9), and helping *document progress* (C 7). Country E had confirming evidence throughout 89% of observations for Criterion 2, *Including measurement expertise throughout the project*. They also had confirming evidence for 83% of

observations when *communicating in a timely fashion* (C 1), *documenting progress* (C 7), and when *acquiring support from outside of the country* (C 11). Having worked on large-scale assessment projects, Country A and Country E were able to work and communicate with the TAT regarding questions and progress throughout the study. Country A and Country E requested information about technical demands, translation requirements, and scoring and was able to include responses as they proceeded. Team members from both countries showed deep understanding of the documents made available by the TAT. Country teams were prepared for meetings and training sessions. Finally, having assessment experts on each team allowed the countries to examine changes to the construct and creating the sampling framework.

The Country B team had difficulty complying with 6 of 11 criteria. The team was not able to *include translation expertise* (C 3) or *project management expertise* (C4). The team did not always *meet deadlines* (C 8), communicate in *timely* communication (C 1), complete *review opportunities* (C 5), or fully participate in *training opportunities* (C 6). The team did not always follow guidelines provided by the TAT and instead followed local cultural practices. For example, the team focused on including faculty from universities instead of professionals with certifications in translation because degrees have a greater value than certificates—even if the degrees did not address translation. Also, much of the work required open critique and discussion regarding translation errors. For example, during an in-person country visit by two members of the TAT, the translators rarely spoke and then only when directly elicited. Instead, discussion was conducted among the team in Country B and a summary was filtered through the national project manager and then reported to the visitors. Also, the team shared that they had not followed the translation guidelines created for the study. The team had not conducted the review

process that required identifying and openly discussing translation error regardless of whose work had resulted in the error.

Criteria compliance: F coefficients

I calculated a F coefficient for each criterion by dividing the CEBs by the total number of CEBs and DEBs for each criterion. I completed this calculation by individual country and across all countries. For special cases, it was necessary to examine the number of NEBs associated with specific tasks to help provide context and meaning for the F coefficient.

Examining each criterion individually permits the analysis of an index of evidence bits providing an overall fidelity of implementation index across criteria (Table 16). To obtain a F coefficient for each criterion, I divided each criterion's CEBs by the sum of the number CEBs and the number of DEBs. Although NEBs were not included in the F coefficient calculation, it was important to take into account the extent of missing information to provide context for discussing the F coefficient. When examining documents during criterion analysis, I decided that meeting 85% of the tasks represented the high end of FOI for meeting criteria. I also decided that having DEBs and NEBs for more than 50% of the criteria indicated a low F coefficient.

Table 16

F Coefficient Across Countries by Criterion.

Category	Criterion	Country A	Country B	Country C	Country D	Country E	Average
Support	2- Measurement expertise	.82	.73	.07	.82	.94	.68
	3- Translation expertise	.60	.08	.57	.13	.78	.43
	4- Project management	.88	.50	.79	.80	.87	.77
	10-Support in country	.80	.71	.73	.67	.80	.74
	11- Support out of country	.83	.83	.83	.83	.83	.83
Ease and Use	5- Review opportunities	.63	.50	.56	.56	.69	.59
	6- Training opportunities	.72	.61	.67	.72	.72	.69
	9- User-friendly materials	.83	.72	.83	.72	.78	.78
Logistics	1- Timely communication	.89	.61	.59	.89	.83	.76
	7- Document progress	.83	.89	.83	.89	.83	.86
	8- Apt deadlines	.65	.47	.59	.65	.76	.62
	Average	.77	.60	.64	.70	.80	.70

There was only one criterion for which countries had an F coefficient of above .85. Therefore, the five participating countries were able to meet closely only one criterion presented in this framework. Countries had a .86 F coefficient for Criterion 7, *Document progress*. Examination of Criterion 7 by country reveals that Country B and Country D met *documentation of progress* most closely with a .89 F coefficient. Country A, Country C, and Country E each had a .83 F coefficient. Overall, there was little variability meeting Criterion 7 across countries, as there was only a .03 difference in the F coefficients.

Countries obtained an F coefficient of .75 to .84 for four criteria. For Criterion 11, *acquiring out-of-country support*, on average countries had an F coefficient of .83. In fact, there was no variability in between individual country F coefficients for this criterion. Each country had an F coefficient of .83 when *acquiring support from out of the country*. For Criterion 9, *Accessing user-friendly materials*, all countries on average had a .78 F coefficient. However, there was some difference in how well individual countries met Criterion 9. Country A and Country C had

a .83 F coefficient; Country E had a .78 F coefficient; Country B and Country D each had a .72 F coefficient. For Criterion 4, *Acquiring the necessary technical infrastructure*, countries had a .77 F coefficient. When examining results by country, it is clear that there were great differences in individual country ability to meet Criterion 4. Country A and Country E had a high F coefficient of .88 and .87 respectively. Country C and Country D were similar in their ability to include project management throughout the study with a F coefficient of .79 and .80. Country B had the greatest difficulty in meeting Criterion 4, as evidenced by a F coefficient of .50. For Criteria 1, *timely communication*, country teams had an F coefficient of .76. Country A and Country D were successful in contributing to *timely communication*. Each had a F coefficient of .89. Country E was fairly successful and had a .83 F coefficient. Country B and Country C had difficulty meeting the *timely communication* criterion. Country B had a .61 F coefficient and Country C a .59.

Countries obtained an F coefficient between .50 and .74 for five criteria. The greatest number of criteria fell in this F coefficient range. For Criterion 10, *Support within the country external to the team*, country teams had an F coefficient of .74. Country A and Country E each had some success in meeting Criterion 10; each achieved a .80 F coefficient. Country C had a .73 F coefficient, Country B a .71, and Country D a .67. For Criterion 6, *training opportunities,* countries had a .69 F coefficient. Country A, Country D, and Country E each had a .72 F coefficient, indicating some difficulty in meeting Criterion 6. Country B, with a .61 F coefficient, and Country C, with a .67, found *participation in training opportunities* more challenging than the other countries.

For Criterion 2, *Including measurement expertise*, country teams had a .68 F coefficient. This criterion had the greatest variability across countries. Country E was very successful in

meeting Criterion 2 and achieved a .94 F coefficient. Country A and Country D were also successful as indicated by their .82 F coefficient. With a .73 F coefficient, Country B had some challenges when trying to include measurement expertise across all activities in the study. However, Country C experienced the greatest challenge as indicated by the .07 F coefficient.

For Criterion 8, *Having and meeting apt deadlines*, country teams had a .62 F coefficient, which is evidence of having experienced difficulty meeting the criterion. Though Country E achieved the highest F coefficient, .76, the country found the criterion challenging. Country A and Country D were similar in their ability to *meet apt deadlines*. Each achieved a .65 F coefficient. Country C had greater difficulty still with a .59 F coefficient. Country B experienced the greatest difficulty *meeting apt deadlines* and achieved a .47 F coefficient. Finally, for Criterion 5, *Review opportunities*, countries had a .59 F coefficient. This low F coefficient shows that countries experienced difficulty in meeting Criterion 5. Country E had a .69 F coefficient, Country A a .63, Country C and Country D a .56, and Country B had the greatest difficulty meeting *review opportunities* with .50 F coefficient.

Countries had an F coefficient of less than .50 for only one criterion. For Criterion 3, *Including translation expertise* throughout the project, country teams had the lowest F coefficient: .43. Upon closer examination there is evidence that countries did not find Criterion 3 equally challenging. Country E was successful in including *translation expertise* and achieved a .78 F coefficient. Country A had some difficulty as their .60 F coefficient shows. Country C achieved a .57 F coefficient also indicating that the country found *including translation expertise* throughout the study challenging. However, Country D and Country B had the greatest difficulty *including translation expertise* as stated in the study's instructions. Country D achieved a .13 F coefficient and Country B a .08 F coefficient.

Discussion of criteria compliance: F coefficients

Calculating the F coefficient allowed me to quantify each country's ability to comply with criteria and examine patterns between country teams' performance. Criterion 7, *Document progress*, and Criterion 11, *Support from outside country*, had high F coefficients: .86 and .83 respectively. Countries were provided a way to systematically document their progress and findings during PT adaptation, cognitive lab interviews, working on the mini-PT, translation reconciliation, and translation review. In addition, the Translation technical assistance team (TAT) hired by CAE documented gathered information regarding country teams' selection of PTs, acquisition of funds, and meeting discussions. Also, the international organizing agency documented each country team's progress in student sampling. The international and U.S. organizing agencies provided country teams with a great deal of support. They wrote reports, assisted with financial support, external translation review, and provided expertise in measurement, research, and translation.

Criterion 3, *Including translation expertise*, had the greatest variability in F coefficients and the lowest average F coefficient. The average F coefficient for Criterion 3 was .43. However, Country E was fairly successful in meeting this criterion and achieved a F coefficient of .78. Country A and Country C experienced some challenges in trying to *include translation expertise* throughout the study and had a F coefficient of .60 and .57 respectively. Country B, with a F coefficient of .08, and Country D, with a F coefficient of .13, experienced great challenges in *including translation expertise* throughout the project. Country E was able to include qualified translators with experience in test translation during most of the study's tasks because members of the Country E team had a background in large-scale test translation. Neither Country A nor Country C included people with experience or expertise in translation as part of their team. As a

result, Country A and Country C did not include translation expertise during item selection, adaptation, or incorporating changes resulting from external validation procedures. These tasks came before or after the actual translation process. The reason Country D and Country B had such low F coefficients is that neither country hired translators with the desired qualifications. Country D focused on including people who were bicultural but who did not have test translation experience. Country B focused on including people with advanced degrees who did not have a background in test translation or cross-cultural large-scale assessment projects.

Chapter 6

Summary and Conclusions

Motivation

Given the increase in the number of countries participating in international test comparison studies, there is an increasing concern regarding the validity of translated tests. Collecting, organizing, and analyzing data from the AHELO study resulted in gathering information on how test translation and adaptation procedures can be improved and, ultimately, create more culturally and linguistically responsive assessments. By examining variation in fidelity of implementation (FOI) of translation and adaptation procedures and the contextual factors affecting FOI, I am able to provide funding and organizing agencies with recommendations to improve current translation practices.

Goals and Research Questions

This dissertation addresses the need for improved test translation and adaptation procedures as critical to ensuring test validity in international comparisons. I examined how countries participating in an initial stage (a feasibility study) of the development of this assessment system interpreted and were able to implement the rigorous adaptation, translation, and translation review procedures designed for this endeavor. To systematically move work addressing test translation forward, I answered two questions:

1. *How did participating countries vary as to the fidelity with which they implemented AHELO's translation and adaptation procedures?*

2. *Based on the lessons learned, how can assessment translation and adaptation procedures be improved to ensure feasibility and validity across languages and cultures in international assessments?*

Context

In late 2009, the OECD solicited CAE to adapt and translate a version of the CLA that could be implemented internationally. As a result, CAE organized the Assessment of Higher Education Learning Outcomes (AHELO)—a study whose goal was to examine the feasibility of translating and adapting the CLA respectively to the languages and the cultures of countries other than the U.S. Five countries other than the U.S. (Country A, Country B, Country C, Country D, and Country E)[6] agreed to participate in the translation of the generic strand component of AHELO. CLA performance tasks assess students who are in their last year of studies at an institution of higher learning for their capacity to use, apply and act on their knowledge and reasoning. In these tasks, which are computer-administered, students are asked to read, interpret, and use diverse information in their written responses to several real-world situations.

Two tasks from the generic strand of the CLA were chosen, adapted, and translated to meet the linguistic and cultural needs of the five participating countries. The ways in which CAE's performance tasks were adapted and translated are distinctive to the AHELO project. Country teams first adapted each performance task to ensure cultural adequacy. Country teams were also to have qualified translators independently participate in an iterative translation process that included a translation reconciliation process. In addition, the translated PTs went through two

[6] A U.S. team also participated in the overall study but not in the translation and adaptation process.

independent review procedures. The first occurred within each country using the theory of test translation error (TTTE), which demands a more rigorous test translation review process focused on seeking evidence of disconfirming evidence of adequate translation. The second review served as a verification process and was completed by an external translation agency. In addition, AHELO incorporated talk-alouds conducted with students from the participating institutions once the CAE performance tasks were translated and reviewed.

Methods

CAE staff, a translation technical assistance team (TAT) hired by CAE, an OECD representative, and the Australian Council for Educational Research (ACER) staff worked with all of the five countries. Together, these entities provided measurement, translation, and project management support.

My dissertation addresses the complexity of test translation and adaptation by examining the factors that shaped the fidelity of implementation (FOI) of translation and adaptation procedures were implemented during the AHELO feasibility study. From January, 2010 through December, 2011, I collected data that allowed me to examine how participating countries interpreted and were able to implement the rigorous adaptation, translation, and translation review procedures created for the AHELO study. I documented the challenges and successes encountered by each country during translation and adaption. The sources of information I used included in-person meetings, conference call meetings, 100 documents created by organizing entities to assist country teams with the translation and adaptation process and to document the progress, an open-ended survey administered via Interne, group interviews, and email communication.

Both quantitative and qualitative information allowed me to perform several types of analyses by tasks and criteria that would provide insight into FOI by country and across countries. I examined FOI by setting cut-off percentages. I established the following set of rules for interpreting evidence bits (pieces of evidence):

1 CEBs for more than 75% of tasks indicates high FOI.

2. CEBs for more than 75% of criteria, indicates high FOI.

3. DEBs for at least one-third of tasks indicates low FOI.

4. DEBs for at least one-third of criteria indicates low FOI.

5. NEBs for at least one-third of tasks or one-third of criteria indicates a challenge in documenting information.

I also examined country performance by calculating two Spearman pairwise correlations between the five countries: one according to CEBs across the 18 tasks; one according to CEBs across the 11 criteria. The correlation coefficient between countries across the 18 tasks provides information about the similarity in the rank ordering of tasks; the coefficient for the 11 criteria indicates the similarity in the rank ordering of criteria. If the correlation coefficient between two countries is high it is because the criteria or tasks that were easiest or most problematic in one country were also easiest or most problematic in the other country. Conversely, two countries sharing a low correlation coefficient experienced dissimilar success when complying with criteria and completing tasks.

I also compared the relationship between CEBs and DEBs for each country. One set of symmetry graphs provided a graphic representation of the CEB-to-DEB ratio by task. A second set of symmetry graphs provided a graphic representation of the CEB-to-DEB ratio by criterion.

The higher the number of CEBs and the lower the number of DEBs, the greater the level of FOI for a task or criterion.

Finally, I calculated the fidelity of implementation (F) coefficient for each task and each criterion by country. I then computed the average F coefficient across countries.

Summary Results and Conclusion

Participating countries differed in the fidelity with which they implemented AHELO's full translation and adaptation procedures. Not all countries were able to recruit and hire translators with the credentials required. In addition, countries were not always able to take advantage of the suggested translation validation procedures. Finally, countries were not consistently able to communicate their progress and answer questions in a timely manner. The countries that experienced the greatest success during test translation (i.e., completed the greatest number of tasks while complying with the greatest number of criteria) shared certain characteristics. Two countries with evidence of high FOI were Country A and Country E. In contrast, Country B had low FOI across tasks and criteria.

The results, by research question, can be summarized as follows:

1. *How did participating countries vary as to the fidelity with which they implemented AHELO's translation and adaptation procedures?*

The results indicate that countries varied in their ability to hire translators with specified qualifications, to apply the theory of test translation error (TTTE) throughout the translation review, and use of cognitive labs. Successful countries were able to hire and use the suggested translation expertise throughout the study. These countries had a background working on international test translation and their translators had experience in translating educational material. Successful countries also organized their translation teams so that they could

implement the TTTE during translation review. These countries experienced open discussion about translation error and worked collaboratively to make decisions about translation changes. In addition, successful countries were able to use cognitive interviews as a validation procedure. These countries either had extensive experience conducting cognitive interviews or with qualitative research.

Countries with lower FOI did not hire translators with recommended qualifications. Instead, countries with lower FOI focused on hiring bilinguals or academics with experience in literature or language. In addition, countries with low FOI did not use the TTTE model of open discussion regarding translation error. Finally, countries with low FOI expressed that conducting cognitive labs was challenging because of their limited experience with them.

Strong similarities appear when comparing results from the verification process and F coefficients. Countries that received positive results in the verification process also achieved a high fidelity of implementation (F) coefficient; countries that received concerning feedback from the verification process achieved a low F coefficient. The external translation verification process examined the overall quality of the translation, the professionalism and consistency of the terminology used as well as mistranslations and omissions.

Country A had an average task completion F coefficient of .78 and Country E an F coefficient of .81. During the translation verification process, the external review company provided positive feedback regarding the translations completed by each country. External reviewers found Country A's translations to be extremely good with use of appropriate words and construction of phrases there were very natural (Quality, 2011). The external reviewers found that the translations read as if they were original text, not translations. Results from verification of Country E's work were similar. External reviewers found that the translator

captured the tone of the English versions and found appropriate versions of names and places (Quality, 2011). Reviewers expressed that the translators kept stylistic aspects of the different original texts (Quality, 2011). Both Country A and Country E were able to hire qualified translators. Their translators had extensive academic training in translation and experience translating large-scale educational assessments.

Of the five countries, Country B achieved the lowest average task completion F coefficient: .63. The external translation verification team found serious challenges in Country B's translation. The translation team found that the translation followed an English structure and too literary resulting in low readability (Quality, 2011). Verification also showed lack of consistency in style across the translations and, at times, provided additional unnecessary verbiage (Quality, 2011). Country B had to revise their translations of both performance tasks. This reflects the tasks and criteria with which Country B had difficulty completing and following. Country B was not able to hire qualified translators to assist with adaptation or translation.

Local social structures impacted FOI across the different countries. For example, countries varied in the way that their team members communicated with each other and the TAT. Given the powerful and strict hierarchical social structure of some countries, there was difficulty in openly discussing certain aspects of the translation process. More specifically, in one country, it was not customary to have an open discussion about errors perceived to have been committed by a professional of a high social standing.

Social structure also became a noticeable factor during translation because of the deference and politeness used with different members of society. Given that the PTs were geared towards college students but at times were asking them to take on the role of a person requiring a higher level of deference, some country teams struggled with the appropriate social register to use in the

translation. Also, a country with a dyadic language needed to decide the language to use for PT translation.

Culture also impacted the translation and adaptation process. For example, countries' ability to complete tasks and comply with criteria were dependent, at least in part, on the time of year when work was to be done. Specifically, local norms regarding the observance and celebration of religious or national holidays impacted countries' ability to meet deadlines. Therefore, for each country it is important to examine when work—especially training, meetings, team translation work, and cognitive labs—is scheduled to be completed so that local schedules are taken into account.

Local linguistic attributes also had an impact on the way in which country teams were able to perform tasks and meet criteria. First, additional work on the computer interface was required to comply with the way in which text is presented across countries (e.g., left-to-right or right-to-left). Second, linguistic attributes also affected the scoring rubric. Some components of the rubric (e.g., writing mechanics) were not applicable in some countries despite being emphasized in other countries. As a result changes needed to be made to the rubric to ensure equity across countries with regard to the intended constructs being measured.

Educational contexts can also impact fidelity of implementation. First, countries with greater diversity in their student population found completing certain tasks and meeting criteria more challenging. Second, because each country's education systems has specific ways to assess learning, one country's team members expressed concern that students may not be familiar with performance tasks. Finally, differences in the beginning and end dates of the school year and in the number of days of the school year caused some challenges to implementation.

Economic climate can also impact translation and adaptation procedures. One of the greatest concerns that country team members discussed constantly was funding. Several countries expressed concern over the appropriate compensation for student participation in the study and cognitive labs. Another country experiencing challenges in political and economic stability was concerned about the fluctuation in the currency exchange rate for the duration of the study.

2. Based on the lessons learned, how can assessment translation and adaptation procedures be improved to ensure feasibility and validity across languages and cultures in international assessments?

Results indicate that the hiring of qualified translators was key to completing the translation and adaptation process successfully. First, it is imperative that translators hired to complete the translation process have formal training in translation procedures. If possible, the persons hired should have credentials in the form of certifications or degrees in translation. Second, translators should have experience in test translation within a large-scale assessment context. Third, teams should include translation expertise throughout the entire study, from item selection to adaptation to translation, translation review, and translation validation. Fourth, all members of the translation and translation review team must be comfortable implementing procedures in which error is the focus of translation review.

Another way to improve translation is to include a project manager with experience in managing large-scale research projects within the education field—with a focus on the grade levels for which translation will occur. This expertise would be useful in conducting cognitive labs or similar validation procedures. Project management experience with a background in educational research would also help countries complete work in a timely manner. The person

would understand the importance of timely communication and the potential challenges related to student diversity, linguistic differences, and social norms.

There were two additional main issues that affected fidelity of implementation: duration of study and continuity of staff at the country as well as the funding or organizing agency. These two challenges are interconnected; often, the longer the duration of the study, the higher the turnover among country team staff. Countries with low turnover had greater success in completing tasks on time. They accumulated intellectual capital through training and were able to retain knowledge of the reasons that specific tasks were chosen and certain decisions about final translations were made.

As with continuity in country teams, continuity of staff working with, or supporting, a funding or organizing agency for the duration of the study is also important. Low turnover would ensure that a core group of people who are knowledgeable about country teams' strengths and weaknesses, the relationship between different work groups, and decisions made during translation procedures can inform the translation process.

Limitations

Some limitations need to be recognized concerning the methods used in this study and the types of generalizations that can be made based on the results obtained. Concerning the methods used, the full coding was performed by only one person. Complete double, independent coding was not feasible because of the vast number of documents and the need to be thoroughly knowledgeable of the study's planned and implemented procedures, country teams, agencies, and tasks. There was consultation with a member of the technical advisory team regarding the coding scheme and initial coding decisions. In addition, the technical advisor also coded a sample of 100

criterion-task intersections. However, there was no verification of the way in which the complete coding was implemented. However, in practice, it is unlikely to expect that evaluations with this level of detail of analyses use two or more coders.

Also, some data collected were provided by participants. During in-person meetings, conference calls, surveys, and interviews, participants reported the information that they viewed as most relevant and felt comfortable sharing. It is possible that countries may have not provided information on sensitive issues. However, given the vast amount of data and site visit reports by technical advisory team members, it is unlikely that self-reporting may have biased substantially the kind of data collected and the kinds of conclusions made.

Finally, concerning the types of generalizations that can be made from this study, it is important to recognize a possible self-selection bias. All of the five participating countries had the financial means needed to participate in this economically demanding study. The results obtained and the conclusions obtained might not apply to countries that may not have the same kinds of resources. Thus, there may be other challenges to the fidelity of implementation of test translation and adaptation procedures that did not arise in this study.

Future Research

Several possibilities for future research arose as I completed this study. First, it would be important for agencies to use the FOI framework as an evaluative tool during the translation and adaptation process. During the AHELO study, the framework was used to examine FOI after the translation and adaptation process was completed by all of the five participating countries. Using the framework during translation and adaptation could help detect areas where countries are

experiencing the greatest challenges. In principle, an agency could use the information to provide appropriate support and, ultimately, improve the quality of translation.

Second, as mentioned before, the AHELO study involved countries that had the financial resources to participate. As a result, the five participating countries are not reflective of all countries that may take part in cross-national comparison studies. In the future, researchers should apply the FOI framework across a greater number of countries that vary in their financial and human resources.

Third, AHELO examined performance tasks intended to measure general problem solving and analytic skills. Therefore, the study's assessment is not reflective of all the current international assessments, such as PISA and TIMSS, which do not use hands-on, performance tasks. Future studies are needed in which the framework is applied in international testing projects in which conventional, multiple-choice items are used.

Fourth, there are alternate methodological approaches that can be applied to the coding during this type of study. An alternative approach that should be examined requires a more fine-grained approach to the cells and coding technique. Some cells (task-criterion intersections) may be more applicable to the translation and adaptation process. For purposes of this study, all cells were treated the same. Another change that can be done in future research is to examine the possibility of assigning different weights when the coder finds different entries within a specific cell. During this study, coding was clear for CEBs or DEBs within each cell. However, with the addition of countries, items, or team members the coding can include more variation in the evidence found for each cell. Although, this different approach will still include some subjectivity due to the different weights that a researcher can assign, it may be relevant for different reasons.

Recommendations for Funding and Organizing Agencies

Based on results from this study, I provide recommendations for funding agencies and organizing agencies assisting countries in translating assessments to be used in international comparison studies.

- **Recommendation 1: Funding and organizing agencies should have a fidelity of implementation (FOI) framework to be used throughout the process to ensure compliance with required work and completion of quality deliverables.**

Given the complex nature of test translation, it is important for agencies to create and use a FOI framework throughout the translation process. The framework should clearly establish the basic set of tasks to perform and criteria to meet. The list of tasks should address test adaptation, test translation, test translation review, process documentation, external test translation verification, cognitive labs, and testing of final translated test and administration system.

Criteria should address the expertise needed in translation, measurement, and project management. In addition, the set of criteria should reflect the knowledge of, and experience with large-scale assessment translation and implementation needed. Translators should have earned a certification from a professional agency to translate from the source language to the desired language. If no such professional agency exists within a country, translators should have formal academic training in translation. Agencies should not substitute translation training with a background in literature, language, or linguistics. Country teams should also have measurement expertise. Measurement experts should have deep knowledge of the test and focus on maintaining the intended construct and difficulty levels across the two versions of the test. Country teams should also have a person dedicated to managing the translation and adaptation

process—documenting progress, meeting deadlines, and meeting specified tasks. Ideally, measurement experts, translation experts, and project management experts should have experience with large-scale assessment translation and implementation.

- **Recommendation 2: Funding and organizing agencies should require that, prior to engaging in any translation activities, country team members complete training focusing on test translation as a multidisciplinary, iterative endeavor, the focus on disconfirming evidence, and the use of cognitive labs.**

At times, translation and adaptation require skills and knowledge that country teams may not possess. Also, translation teams should participate in training addressing the need to focus on disconfirming evidence of adequate translation.

Finally, agencies should provide country teams with training on how to conduct cognitive labs. Specifically, cognitive labs used for translation review should capture information on the impact of item wording and format on student's interpretation of the tasks and performance. Based on their examination of the data captured, the team members should identify ways in which the translation of the tasks needs to be improved.

- **Recommendation 3: Funding and organizing agencies should require that country teams provide documentation of the actions they take to comply with the translation and adaptation procedures.**

As countries attempt to complete necessary tasks and comply with criteria, it is important that they document and report their progress. Funding and organizing agencies should require from countries documentation of test translation as a multidisciplinary, iterative endeavor, the focus of disconfirming evidence, and the use of cognitive labs.

Funding and organizing agencies should also require documentation regarding the field-testing of the test administration process employed. This is particularly important for computer administered assessments. Among the many important issues to document is information on the way in which countries' local language appear on computer screens or tablets.

References

Aamodt, P. O. (2008). Access to higher education within a welfare state system: Developments and dilemmas. In P. N. Teixeira, D. B. Johnstone, M. J. Rosa, & J. J. Vossentsteijn (Eds.), *Cost-sharing and accessibility in higher education: A fairer deal?* Dordrecht, The Netherlands: Springer.

Abedi, J. (2010). *Performance assessments for English language learners*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

Abedi, J. (2007). In J. Abedi (Ed.), *English language proficiency in the nation: Current status and future practice*. Davis, CA: University of California, Davis.

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), 4-14.

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, 74 (1), 1-28.

Allalouf, A., Hambleton, R. K., & Sireci, S. G. (1999). Identifying the causes of DIF in translated verbal items. *Journal of Educational Measurement*. 36 (3), 185-198.

American Education Research Association, American Psychological Association, & National Council on Measurement in Education. 1999. Standards for educational and psychological testing. Washington, D.C.: American Education Research Association.

American Translators Association. http://www.atanet.org/. Last accessed August 15, 2011.

Arora, A., Foy, P., Martin, M. O., & Mullis, I.V.S. (2009). *TIMSS Advanced 2008 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Arora, A. & Mullis, I.V.S. (2008). *Developing the TIMSS advanced 2008 instruments*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Auchter, J. E., & Stansfield, C. W. (1997). Developing parallel tests across languages: Focus on the translation and adaptation process. Paper presented at the 27th Annual Large Scale Assessment Conference, Colorado Springs, Colorado, June 15-18, 1997.

August, D., & Shanahan, T. (2006). Executive summary: Developing literacy in second-language learners (pp. 1-9). *Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Assoc.

Aukerman, M. (2007). A Culpable CALP: Rethinking the controversial/academic language proficiency distinction in early literacy instruction. The Reading Teacher, 60(7), 626-635. http://web.ebscohost.com/ehost/pdf?vid=2&hid=116&sid=f1565737-84f8-478b-95c2-1c4a16d98d06%40sessionmgr108

Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*. 17 (1), 1-42.

Bae, J, & Lawler, J. J. (2000). Organization and HRM strategies in Country B: Impact on firm performance in an emerging economy. *Academy of Management Journal*. 43 (3), 502-517.

Banks, J. A., & McGee Banks, C. A. (2003). *Handbook of Research on Multicultural Education*. San Francisco, CA: Jossey-Bass.

Bernard, H. R. (1999). Languages and scripts in contact: Historical perspectives. In D. A. Wagner, R. L. Venezky, & B. V. Street (Eds.). *Literacy: An international handbook*. New York, NY: Westview Press.

Briggs, D. & Alonso, A. Building a learning progression as a cognitive model. Paper presented at

the annual meeting of the National Council for Measurement in Education, San Diego, CA, April 16, 2009.

Camilli, G. (2006). "Test Fairness." In R. L. Brennan, *Educational Measurement*. American Council on Education, Washington, DC, 2006.

Camilli, G., Briggs, D., Sloane, F., & Chiu, T. W. (in press) Psychometric Perspectives on Test Fairness: Shrinkage Estimation.

Central Intelligence Agency Factbook. https://www.cia.gov/library/publications/the-world-factbook/geos/nl.html. Last accessed October 21, 2010.

Central Intelligence Agency Factbook. https://www.cia.gov/library/publications/the-world-factbook/geos/us.html. Last accessed October 21, 2010.

Century, J., Freeman, C., & Rudnick, M. (2008). A framework for measuring and accumulating knowledge about fidelity of implementation (FOI) of science instructional materials. Paper presented at the National Association for Research in Science Teaching Annual Meeting, March 31, 2008.

Chaiklin, S. "The zone of proximal development in Vygotsky's analysis of learning and instruction." In A. Kozulin, B. Gindis, V. Ageyev, & S. Miller, *Vygotsky's Educational Theory in Cultural Context*. New York, NY: Cambridge University Press, 2003.

Chrostowski, S. J., & Malak, B. (2003). "Translation and Cultural Adaptation of the TIMSS 2003 Instrument" in M.O. Martin and I.V.S. Mullis *Third International Mathematics and Science Study: Quality Assurance in Data Collection.* Chestnut Hill, MA: Boston College.

Cobb, Paul and Bowers, Janet. (1999). Cognitive and Situated Learning Perspectives in Theory and Practice. *Educational Researcher*. 28(2), 4-14.

Cole, M. & Bruner, J. S. (1971). Cultural differences and inferences about psychological processes. *American Psychologist*. 26 (10), 867-876.

Collins, Kathleen. "Introduction: A Sociocultural Perspective on Disability." In K.M. Collins, *Ability Profiling and School Failure*. Mahwah, NJ: Lawrence Erlbaum Press, 2003.

Council for Aid to Education, Roger Benjamin. (2008). *CLA: Collegiate Learning Assessment, UNC Asheville*. Retrieved from http://www.collegiatelearningassessment.org/

Council for Aid to Education, Roger Benjamin. (2005). *CLA in context*. Retrieved from http://www.collegiatelearningassessment.org/

Council for Aid to Education. Architecture of the CLA tasks. Retrieved from http://www.collegiatelearningassessment.org/files/Architecture_of_the_CLA_Tasks.pdf

Council for Aid to Education. (2010). *GS.4 Conceptual framework: AHELO Module A adaptation and translation of performance tasks.*

Council for Aid to Education. (2010). *GS.13 Adaptation and translation of performance tasks, CAE AHELO generic strand meeting.*

Council for Aid to Education. (2010). *GS.14 Adapted performance task in English (CAE/AHELO Team).*

Council for Aid to Education. (2010). *GS.37 Cognitive labs guidelines, AHELO Module A.*

Council for Aid to Education. (2010). *GS.30 Group of national experts on the AHELO feasibility study: Progress report on generic skills.*

Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA: SAGE Publications, Inc.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York:

Wiley.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.

Cross, C. T. (2004). *Political education: National policy comes of age*. New York, NY: Teachers College Press.

Crotty, M. (2003). *The Foundations of Social Research: Meaning and Perspective in the Research Process*. Thousand Oaks, CA: SAGE Publications, Inc.

Crystal, D. (2003). *English as a Global Language*. New York, NY: Cambridge University Press.

Del Rosario Basterra, M. (2010). "Cognition, culture, language, and assessment: How to select culturally valid assessments in the classrom." In M. del Rosario Basterra, E. Trumbull, & G. Solano-Flores, *Cultural Validity in Assessment*. New York, NY: Routledge.

Duran, R. L. (1983). Communicative ability: A measure of social communicative competence. *Communication Quarterly*. 31, 320-326.

Egan, K. & Gajdamaschko, N. (2003). "Some cognitive tools of literacy." In A. Kozulin, B. Gindis, V. Ageyev, & S. Miller, *Vygotsky's Educational Theory in Cultural Context*. New York, NY: Cambridge University Press.

Embretson, S. E., and Reise, S. P. (2000) *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ercikan, K. & Roth, W. M. (2006). Constructing data. In C. Conrad & R. Serlin (Eds.), *SAGE Handbook for research in education: Engaging ideas and enriching inquiry* (pp. 451-475). Thousand Oaks, Ca: Sage.

Ercikan, K. (2002). Disentangling sources of differential item functioning in multilanguage assessments. *International Journal of Testing*. 2 (3&4), 199-215.

Ercikan, K. (1998). Translation effects in international assessment. *International Journal of Educational Research*. 29, 543-553.

Escamilla, K. (2000). Bilingual means two: Assessment issues, early literacy and Spanish-speaking children. *Reading Research Symposium for Second Language Learners (pp. 1-16).* Washington, D.C.: National Clearinghouse for Bilingual Education.

Fearon, J. D. (2003). Ethnic culture and cultural diversity around the world: A cross-national data set on ethnic groups. Paper presented at the 2002 Annual Meeting of the American Political Science Association, August 29-Sept. 1, Boston.

Country A Ministry of Education and Culture, http://www.minedu.fi/OPM/Koulutus/yliopistokoulutus/yliopistot/?lang=en. Last accessed September 5, 2011.

Foucault, M. (1989). *Power/Knowledge: Selected Interviews and Other Writings 1972-1977*. (Ed.) Colin Gordon. New York, NY: Pantheon Books.

Freeman, D. E. & Freeman, Y. S. (2004). *Essential linguistics: What you need to know to teach reading, ESL, spelling, phonics, and grammar*. Portsmouth, NH: Heinemann.

Garcia, G. E., McKoon, G., & August, D. (2006). Language and literacy assessment of language minority students. In August, D. & Shanahan, T. (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*. 6 (4), 304-312.

Gordon, J. V. (2008). Performance on large-scale science tests: Item attributes that may impact achievement scores. (Doctoral dissertation).

Graddol, D. (2008). *The Future of English? A Guide to Forecasting the Popularity of English in the 21st century.* The British Council.

Grambs, J. D. *Multicultural Education: Issues Without Answers*, The Illinois Association of Colleges for Teacher Education Monograph, No. 1 (January, 1979).

Greenfield, P. M. (1997). Culture as process: Empirical methods for cultural psychology. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology, Second Edition. Vol. 1:Theory and method.* (pp. 301-346). Needham Heights, Massachusetts: Allyn & Bacon.

Grek, S., Lawn, M., & Ozga, J. (2009). *Study on the use and circulation of PISA in Scotland*. Know & Pol, Knowledge and Policy in Education and Health Sectors, European Commission.

Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing*. 20, 225-240.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1,* 131-149.

Halliday, M. A. K. (2003). Language as social semiotic. In J. Maybin (Ed.), *Language and literacy in social practice* (pp. 23-43). Great Britain: Short Run Press, Ltd.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning.* London: Edward Arnold.

Hambleton, R.K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R.K. Hambleton, P.F. Merenda, & C.D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Hambleton, R. K., Yi, J., & Slater, S. C. (1999). Fieldtest of the ITC guidelines for adapting educational and psychological tests. *European Journal of Psychological Assessment*. 15 (3), 270-276.

Hambleton, R K. (1994). The rise and fall of criterion referenced measurement? *Educational Measurement: Issues and Practice*. 13 (4), 21-26.

Hambleton, R. K. & Kanjee, A. (1993). Enhancing the validity of cross-cultural studies: Improvements in instrument translation methods. Paper presented at the 1993 Annual Meeting of the American Educational Research Association, April 13-15, Atlanta, GA.

Hawkins, M. R. (2004). Researching English language and literacy development in schools. *Educational Researcher*. 33 (3), 14-25.

Horner, K. & Weber, J. J. (2008). The language situation in Luxembourg. *Current Issues in Language Planning*. 9 (1), 69-128.

Hwang, J. R. International Journal of the Sociology of Language. Volume 1990, Issue 82, Pages 41–56, ISSN (Online) 1613-3668, ISSN (Print) 0165-2516, DOI: 10.1515/ijsl.1990.82.41, October 2009. Last accessed July 22, 2012.

International Test Commission (2005). *International guidelines on computer-based and internet delivered testing*. Granada, Spain.

International Association for the Evaluation of Educational Achievement (2001). *PIRLS International Report Executive Summary*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Johansone, I. & Malak, B. (2007). "Translation and national adaptations of the TIMSS 2007 assessment and questionnaires." *Overview of TIMSS 2007*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Johnstone, C. J., Bottsford-Miller, N. A., & Thompson, S. J. (2006). *Using the think aloud method (cognitive labs) to evaluate test design for students with disabilities and English language learners* (Technical Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved [May 16, 2012], from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/Tech44/

Jolly, E. (2001). *On the quest for cultural context in evaluation: Non ceteris paribus*. National Science Foundation. (2001). *The Cultural Context of Educational Evaluations: The Role of Minority Evaluation Professionals.* Workshop Proceedings. June 1-2.

Keeley, R. (personal communication, May 16, 2012).

Keith, R. E., Hopp, F. P., Submaranian, U., Wita, W., & Lower, J. C. (2010). Fidelity of implementation: Development and testing of a measure. *Implementation Science*. http://www.implementationscience.com/content/5/1/99.

Kelman, H. C. (1972). Language as an aid and barrier to involvement in the national system. In: *Advances in the Sociology of Language*, Vol. II, edited by Joshua A. Fishman. The Hague: Mouton. 185-212.

Kincheloe, J. L., & Steinberg, S. R. (1997). *Changing Multiculturalism*. Bristol, PA: Open University Press.

Kirkhart, K. E. (1995). Seeking multicultural validity: A postcard from the road. *Evaluation Practice*, 16 (1), 1-12.

Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). The collegiate learning assessment: Facts and fantasies. White Paper. *Evaluation Review*.

Kopriva, R. J., Emick, J.E., Hipolito-Delgado, C. P., & Cameron, C.A. (2007). Do proper accommodation assignments make a difference? Examining the impact of improved decision making on scores for English language learners. *Educational Measurement: Issues and Practice*, Fall 2007, 11-20.

Country B Council for University Education, http://english.kcue.or.kr/resources/resources_01_01.php/. Last accessed September 1, 2011.

Kozulin, A., Gindis, B., Ageyev, V. & Miller, S. (2003). "Introduction: Sociocultural Theory and Education: Students, Teachers, and Knowledge." In A. Kozulin, B. Gindis, V. Ageyev, & S. Miller, *Vygotsky's Educational Theory in Cultural Context*. New York, NY: Cambridge University Press, 2003.

Kurpius, A. & Shavelson, R. (2010). Criteria for choosing CLA performance tasks and recommended tasks for the AHELO feasibility study.

Country C Cultural Office, http://www.Country Cculture.com/About%20Us/higher.htm. Last accessed September 5 2011.

Country C Ministry of Higher Education, Private Universitites Council, http://www.puc.edu.kw/en/index.php?TP=ecommittees. Last accessed September 5,

2011.

Laboratorio Latinoamericano de Evaluacion de la Calidad de la Educacion (LLECE). (2001). *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados, para alumnus del tercer y cuarto grado de la educació básica.*

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal, 32*(3), 465-491.

Lalancette, D. (2010, February). OECD AHELO overview for CAE generic skills meeting. Presented at the CAE AHELO February meeting, New York, NY.

LeCompte, M. D., & Schensul, J. J. (1999). *Designing and Conducting Ethnographic Research, Ethnographer's Toolkit*. New York, NY: Altamira Press.

Lee, O. (2003). Equity for linguistically and culturally diverse students in science education: A research agenda. *Teachers College Record*, 100 (3), 465-489.

Lemke, J. L. (2001). Articulating communities: Sociocultural perspectives on science education. *Journal of Research in Science Teaching*, 38 (3), 296-316.

Lenski, S. D., Ehlers-Zavala, F., Daniel, M. C., & Sun-Irminger, X. (2006). Assessing English-language learners in mainstream classrooms. *The Reading Teacher, 60* (1), 24-34.

Lidz, C. S. & Gindis, B. (2003). "Dynamic assessment of the evolving cognitive functions in children." In A. Kozulin, B. Gindis, V. Ageyev, & S. Miller, *Vygotsky's Educational Theory in Cultural Context*. New York, NY: Cambridge University Press, 2003.

Linn, R. (2003). Accountability: responsibility and reasonable expectations. *Educational Researcher*, 32 (7), 3-13.

Luke, C. (2008). *Globalization and Women in Academia: North/West-South/East*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lukyx, A. Lee, O., Mahotiere, M., Lester, B., Hart, J., & Deaktor, R. (2007). Cultural and home language influences on children's responses to science assessments. *Teachers College Record, 109*(4), 897–926.

Lynn, P. (2003). Developing quality standards for cross-national survey research: Five approaches. *Social Research Methodology*. 6 (4), 323-336.

Lynch, S. & O'Donnell, C. (2005). "Fidelity of implementation" in implementation and scale-up research designs: Applications from four studies of innovative science curriculum materials and diverse populations. Paper presented at AERA Annual Meeting, Montreal, Canada, April 14, 2005.

Martin, M.O. & Mullis, I.V.S. (2006). *Overview of TIMSS 2007*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Martin, M.O., Mullis, I.V.S., Kennedy, A. M., (Eds.). (2006). *PIRLS 2006 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Martin, M.O., Hoyle, C.D., and Gregory, K.D. (1996). "Monitoring the TIMSS Data Collection" in M.O. Martin and I.V.S. Mullis *Third International Mathematics and Science Study: Quality Assurance in Data Collection.* Chestnut Hill, MA: Boston College.

Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*. 34 (4), 207-218.

Mercer, J. R. (1984). What is a racially and culturally nondiscriminatory test? A sociological and pluralistic perspective. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing*. New York: Plenum Press.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*. 50 (9), 741-749.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). (pp.13–103). Washington, DC: American Council on Education & National Council on Measurement in Education.

Country D Subsecretaría de Educación Superior, 2011, http://www.ses.sep.gob.mx/wb/ses/educacion_superior_publica. Last accessed September 5, 2011.

Mullis, I.V.S., Martin, M.O., Ruddock, G.H., O'Sullivan, C.Y., Arora, A., & Erberber, E. (Eds.). (2007). *TIMSS 2007 Assessment Frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Kelly, D.L., and Haley, K. (1996). Translation verification procedures. In M.O. Martin and I.V.S. Mullis (Eds.), *Third International Mathematics and Science Study: Quality Assurance in Data Collection.* Chestnut Hill, MA: Boston College.

Murphy, S. (2010). The pull of PISA: Uncertainty, influence, and ignorance. *Revista Interamericana de Educación para la Democracia--Interamerican Journal of Education for Democrazy,* 3(1)*,* 27-44.

Country E Ministry of Education and Research, http://www.regjeringen.no/en/dep/kd/Selected-topics/higher-education.html?id=1415. Last accessed September 5, 2011.

Organisation for Economic Co-operation and Development (OECD). (2008). *Roadmap for the OECD Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study (As of 3 July 2008)*.

Organisation for Economic Co-operation and Development (OECD). (2006). *PISA 2006 Technical Report*.

Organisation for Economic Co-operation and Development (OECD). (2003). *Learning for tomorrow's world: First results from PISA 2003*.

Organisation for Economic Co-operation and Development (OECD). (2000). *OECD Programme for International Student Assessment National Project Manager's Manual*.

Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*. 78 (4), 1255-1264.

Raabe, M. (2009). Facts about education in Country E 2010-key figures 2008. Country E: Ministry of Education and Research.

Rossi, P. H., Lipsey, M. W., and Freeman, H. E. (2004). *Evaluation: A Systematic Approach*. Thousand Oaks, CA: Sage Publications, Inc.

Ruddock, G. J., O'Sullivan, C.Y., Arora, A., Erberer, E. (2007). *Developing the TIMSS 2007 Mathematics and Science Assessments and Scoring Guides*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Ruiz, R. (1984). Orientations in language planning. *NABE Journal*, 8 (2), 15-34.

Ruiz-Primo, M. A. (2006). A multi-method and multi-source approach for studying fidelity of implementation. CSE Report 677, SEAL, Stanford University/CRESST, Palo Alto, CA.

Sévigny, S., Savard, D., & Beaudoin, I. (2009). Comparability of writing assessment scores across languages: Searching for evidence of valid interpretations. *International Journal of Testing*. 9, 134-150.

Singh, P. (2004). Globalization and Education. *Educational Theory*, 54(1), 103-115.

Shavelson, R. J. (2010, December 27). Personal communication email.

Shavelson, R. J. (2010, March). The collegiate learning assessment. Invited presentation, Helsinki, Country A.

Shavelson, R. J., & Towne, L. (Eds.) (2002). *Scientific research in education*. Washington, D.C.: National Academy Press.

Shepard, L.A. (1992). Uses and abuses of testing. In Marvin C. Alkin (Ed.), *Encyclopedia of Educational Research*, Sixth Edition, pp. 1477-1485. New York: MacMillan.

Smagorinsky, P. (1998). Thinking and speech and protocol analysis. *Mind, Culture, and Activity*. 5 (3), 157-177.

Solano-Flores, G. (In Press). Successive test development. In C.R. Reynolds, R.W. Kamphaus, C. DiStefano (Eds.), *Encyclopedia of psychological and educational testing: Clinical and psychoeducational applications.* New York: Oxford University Press.

Solano-Flores, G. (2011). *Assessing the cultural validity of assessment practices: A critical guide for educators.* In M. R. Basterra, E. Trumbull, & G. Solano-Flores. (Eds.), *Cultural validity in assessment: Addressing linguistic and cultural diversity.* New York: NY: Routledge.

Solano-Flores, G. (2010). Function and form in research on language and mathematics education. In Moschkovich, J. (Ed.), *Language and mathematics in education: Multiple perspectives and directions for research* (pp. 113-149). Charlotte, NC: Information Age Publishing, Inc.

Solano-Flores, G. (2009). Function and form in research on language and mathematics education in Language and Mathematics Education: Multiple Perspectives and Directions for Research. Ed. J. N. Moschkovich. Information Age Publishing Inc.

Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language learners. *Educational Researcher*, 37 (4), 189-199.

Solano-Flores, G. (2008b). A conceptual framework for examining the assessment capacity of countries in an era of globalization, accountability, and international test comparisons. In 6th Conference of the International Test Commission, Liverpool, UK.

Solano-Flores, G. (2006). Language, dialect, and register: Sociolinguistics and the estimation of measurement error in the testing of English-language learners. *Teachers College Record.* 108 (11), 2354-2379.

Solano-Flores, G. & Bonk, W. (2008). *Evaluation of the Latin American Laboratory for the Evaluation of Educational Equity (LLECE)*. United Nations Educational, Scientific, and Cultural Organization, Santiago, Chile.

Solano-Flores, G. & Li, M. (2008). *Examining the dependability of academic achievement measures for English language learners.* Assessment for Effective Intervention33 (3), 135-144.

Solano-Flores, G., Backhoff, E., & Contreras-Niño, L.A. (2006). Theory of test translation error. *International Journal of Testing*, 9, 78-91.

Solano-Flores, G. & Li, M. (2006). The use of generalizability (G) theory in the testing of linguistic minorities. *Educational Measurement: Issues and Practice*. 25 (1), 13-22.

Solano-Flores, G. & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher*. 32 (2), 3-13.

Solano-Flores, G. & Nelson-Barber, S. (2001). Cultural validity of science assessments. *Journal*

of Research in Science Teaching. 38 (5), 553-573.

Solano-Flores, G. & Shavelson, R. (1997). Development of performance assessments in science: conceptual, practical, and logistical issues. *Educational Measurement: Issues and Practice.* Fall 1997, 16-25.

Stansfield, C. W. (2003). Test translation and adaptation in public education in the USA. *Language Testing*. 20 (2), 189-207.

Subsecretaría de Educación Superior. (2011). Fondo para elevar la calidad de la educación superior de las universidades públicas estatales (UPE).

Tauli, V. (1978). The theory of language planning. In Fishman, J. A. (Ed.), *Advances in language planning* (pp. 49-67). Mouton: The Hague.

Tomlinson, J. (2006). *Globalization and culture.* Presented at University of Nottingham Ningbo China (UNNC) Research Seminar Series 2006–2007, Ningbo, China.

Tong, F., Lara-Alecio, R., Irby, B., Mathes, P., & Kwok, O. (2008). Accelerating early academic oral English development in transitional bilingual and structured English immersion programs. *American Educational Research Journal,* 45(4), 1011-1044.

Trumbull, E. & Solano-Flores, G. "Defining Good Assessment," *Making Assessment Work for Everyone: How to Build on Student Strengths*. Assessment Laboratory Network Project of the Regional Educational Laboratories. 2000. http://www.sedl.org/pubs/tl05/welcome.html.

Tucker, G. R. (2003). A global perspective on bilingualism and bilingual education. In Paulston, C. B. & Tucker, G. R. (Eds.), *Sociolinguistics: The essential readings*, (pp. 464-471). Charlotte, NC: Information Age Publishing, Inc.

van de Vijver, J. R. & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. In R.K. Hambleton, P.F. Merenda, & C.D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 39-64). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

van de Vijver, J. R. & Poortinga, Y. H. (1997) Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*. 13 (1), 29-37.

Vygotsky, L. S. (1978). Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard College Press.

Wilson, M. (2005). Constructing Measures: An Item Response Modeling Approach. Mahwah, NJ: Erlbaum.

Wolfram, W., Adger, C. T., & Christian, D. (1999). Dialects in Schools and Communities. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Woolard, K. A. (1985). Language variation and cultural hegemony: Toward an integration of sociolinguistic and social theory. *American Ethnologist*. 12 (4), 738-748.

Appendices

Appendix A: Country lists

TIMSS 2007 Participating Countries*

- Algeria
- Armenia
- Australia
- Austria
- Bahrain
- Bosnia and Herzegovina
- Botswana
- Bulgaria
- Chinese Taipei
- Colombia
- Cyprus
- Czech Republic
- Denmark
- Egypt
- El Salvador
- England
- Georgia
- Germany
- Ghana
- Hong Kong SAR
- Hungary
- Indonesia
- Iran, Islamic Rep. of
- Israel
- Italy
- Japan
- Jordan
- Kazakhstan
- Country B, Republic of
- Country C
- Latvia
- Lebanon
- Lithuania
- Malaysia
- Malta
- Mongolia
- Morocco
- Netherlands
- New Zealand
- Country E
- Oman Palestinian National Authority
- Qatar
- Romania
- Russian Fedtion
- Saudi Arabia
- Scotland
- Serbia
- Singapore
- Slovak Rep.
- Slovenia
- Sweden
- Syrian Arab Republic
- Thailand
- Tunisia
- Turkey
- Ukraine
- United States
- Yemen

* Information from http://timss.bc.edu/timss2007/countries.html.

PISA 2009 Participating Countries*

- Albania
- Azerbaijan
- Canada
- Czech Republic
- Country A
- Hong Kong-China
- Ireland
- Jordan
- Liechtenstein
- Macao-China
- New Zealand
- Poland
- Russian Federation
- Slovak Republic
- Switzerland
- Tunisia
- United Kingdom
- Argentina
- Belgium
- Chile
- Denmark
- France
- Hungary
- Israel
- Kazakhstan
- Lithuania
- Country D
- Country E
- Portugal
- Republic of Serbia
- Slovenia
- Chinese Taipei
- Turkey
- Australia
- Brazil
- Colombia
- Dubai (UAE)
- Germany
- Iceland
- Italy
- Country B
- Luxemburg
- Republic of Montenegro
- Panama
- Qatar
- Shanghai-China
- Spain
- Thailand
- United States
- Austria
- Bulgaria
- Croatia
- Estonia
- Greece
- Indonesia
- Japan
- Kyrgyz Republic
- Latvia
- The Netherlands
- Peru
- Romania
- Singapore
- Sweden
- Trinidad and Tobago
- Uruguay

* Information from http://www.oecd.org/document/4/0,3343,en_32252351_32236225_39758660_1_1_1_1,00.html

PIRLS 2011 Participating Countries*

- Australia
- Austria
- Azerbaijan
- Belgium, French
- Botswana
- Bulgaria
- Canada
- Chinese Taipei
- Colombia
- Croatia
- Czech Republic
- Denmark
- England
- Country A
- France
- Georgia
- Germany
- Honduras
- Hong Kong SAR
- Hungary
- Indonesia
- Iran, Islamic Rep. of
- Ireland
- Israel
- Italy
- Country C
- Libya
- Lithuania
- Malta
- Morocco
- Netherlands
- New Zealand
- Northern Ireland
- Country E
- Oman
- Poland
- Portugal
- Qatar
- Romania
- Russian Federation
- Saudi Arabia
- Singapore
- Slovak Republic
- Slovenia
- South Africa
- Spain
- Sweden
- Trinidad and Tobago
- United Arab Emirates
- United States

PIRLS 2011 Benchmarking participants

- Alberta, Canada
- Ontario, Canada
- Quebec, Canada
- Andalusia, Spain
- Abu Dhabi, United Arab Emirates
- Dubai, United Arab Emirates
- Florida, United States

SERCE 2007 Participating countries

- Argentina
- Bolivia
- Brasil
- Chile
- Colombia
- Costa Rica
- Cuba
- Ecuador
- El Salvador
- Honduras
- Country D
- Nicaragua
- Paraguay
- Peru
- Republica Dominicana
- Uruguay
- Venezuela

Appendix B: List of 100 documents included in the analysis.

Author unknown. (2010). *Criteria for Selecting Country Assessment Experts*. Assessment of Higher Education Learning Outcomes Generic Strand. January.

Author unknown. (2010). *AHELO National Project Managers Paris, List of Participants*. Assessment of Higher Education Learning Outcomes Generic Strand. October.

Author unknown. (2010). *AHELO NC-NPM role description*. Assessment of Higher Education Learning Outcomes Generic Strand. August.

Author unknown. (2010). *AHELO Detailed Workplan*. Assessment of Higher Education Learning Outcomes Generic Strand. February.

ACER. (2011). *OECD AHELO Feasibility Study: November NPM and Lead Scorer Training*. Assessment of Higher Education Learning Outcomes Generic Strand. October.

ACER. (2011). *OECD AHELO Feasibility Study: Strands and Languages*. Assessment of Higher Education Learning Outcomes Generic Strand. January.

ACER. (2010). *OECD AHELO Feasibility Study: Approach and Resources*. Assessment of Higher Education Learning Outcomes Generic Strand. December.

ACER. (2010). *OECD AHELO Feasibility Study: Selecting and Engaging Institutions*. Assessment of Higher Education Learning Outcomes Generic Strand. December.

ACER. (2010). *Translation, adaptation and verification*. Assessment of Higher Education Learning Outcomes Generic Strand. November.

ACER. (2010). *OECD AHELO Feasibility Study: Sampling Manual*. Assessment of Higher Education Learning Outcomes Generic Strand. October.

Al-Atiqi, I. (personal communication, January 7, 2011). [Re: "Country C change of AHELO representative"]. Imad leaving.

Al-Atiqi, I. (personal communication, January 1, 2011). [Re: "CAE AHELO Final Adaptation/Translation Document"]. One PT attached. 01:06 PM.

Al-Atiqi, I. (personal communication, January 1, 2011). [Re: "CAE AHELO Final Adaptation/Translation Document"]. One PT attached.

Al-Atiqi, I. (personal communication, January 13, 2010). [Re: "CAE AHELO FOLLOW UP"]. Call set and Asad's CV.

Al-Rashed, A. (personal communication, February 13, 2011). [Re: "AHELO NPM Diaries"]. Problem access due to password.

Al-Rashed, A. (personal communication, August 3, 2010). [Re: "CAE AHELO Translation Update 6"]. Delay rough translation two PTs.

ACER. (2010). *OECD Higher Education Update*. Assessment of Higher Education Learning Outcomes Generic Strand. September.

Benjamin, R. (personal communication, February 14, 2011). [Re: "OECD comments on Milestone 3 report"].

Bily, C. (personal communication, September 13, 2010). [Re: "Three more nominations"]. New countries added to AHELO.

CAE. (2011). *AHELO Update 22 December 2011*. Assessment of Higher Education Learning Outcomes Generic Strand. January.

CAE. (????). *AHELO Feasibility Study Analysis Plan*. Assessment of Higher Education Learning Outcomes Generic Strand. ??.

CAE. (2010). *OECD-AHELO Module A Progress Report, Milestone 3*. Assessment of Higher Education Learning Outcomes Generic Strand. December.

CAE. (2010). *On-Site Translation Review Training: Items for Translation*. Assessment of Higher Education Learning Outcomes Generic Strand. Summer.

CAE. (2010). *Summary of Progress: AHELO Generic Skills Strand*. Assessment of Higher Education Learning Outcomes Generic Strand. November.

CAE. (2010). *Appendix D: CAE AHELO Work Plan*. Assessment of Higher Education Learning Outcomes Generic Strand. January.

CAE. (2010). *Appendix E: Criteria for selecting country assessment representatives*. Assessment of Higher Education Learning Outcomes Generic Strand. January.

CAE. (2010). *Appendix F:NYC meeting logistics information*. Assessment of Higher Education Learning Outcomes Generic Strand. January.

CAE. (2010). *CAE Biographical Sketches*. Assessment of Higher Education Learning Outcomes Generic Strand. January.

CAE. (2010). *Initial Project Letter Final*. Assessment of Higher Education Learning Outcomes Generic Strand. January.

CAE. (personal communication, July 28, 2010). [Re: "CAE AHELO Translation Update 6"].

CAE. (personal communication, May 21, 2010). [Re: "CAE AHELO Translation Update 2"]. Mini PT, PT Response Features, Cog labs guide.

CAE. (personal communication, April 26, 2011). [Re: "CAE AHELO Online Platform Review (Pre-Implementation)"].

CAE. (personal communication, February 11, 2011). [Re: "AHELO NPM Diaries"]. ACER request flexible.

CAE. (personal communication, January 24, 2011). [Re: "CAE AHELO Internet Platform Documents (Pre-Implementation)"].

CAE. (personal communication, January 24, 2011). [Re: "CAE AHELO Internet Platform Documents (Pre-Implementation)"] password.

CAE. (personal communication, January 20, 2011). [Re: "CAE AHELO Final Adaptation/Translation Document"] Country A reminder for cog lab synthesis feedback.

CAE. (personal communication, January 20, 2011). [Re: "CAE AHELO Final Adaptation/Translation Document"] Country B reminder for cog lab synthesis feedback.

CAE. (personal communication, January 20, 2011). [Re: "CAE AHELO Final Adaptation/Translation Document"] Country D reminder for cog lab synthesis feedback.

CAE. (personal communication, January 20, 2011). [Re: "CAE AHELO Final Adaptation/Translation Document"] Country E reminder for cog lab synthesis feedback.

CAE. (personal communication, January 12, 2011). [Re: "CAE AHELO Pre-Implementation"].

CAE. (personal communication, December 28, 2010). [Re: "CAE AHELO Final Adaptation/Translation Document"]. Cog labs synthesis.

CAE. (personal communication, November 16, 2010). [Re: "CAE AHELO: Example Student Responses—Country A Cognitive Labs"].

CAE. (personal communication, November 1, 2010). [Re: "CAE AHELO Translation Update 8"]. Computer needs.

CAE. (personal communication, October 22, 2010). [Re: "CAE AHELO Translation Update 8"]. Correction attachments: GS.43, GS.44.

CAE. (personal communication, October 22, 2010). [Re: "CAE AHELO Translation Update 8"]. GS.43.

CAE. (personal communication, October 14, 2010). [Re: "CAE AHELO Translation Update 7"]. Country A cog lab obs, summary of teleconferences, module A activities 10/2010-5/2011.

CAE. (personal communication, September 30, 2010). [Re: "Pre-meeting Gathering on 27 October for Generic Skills Strand NPMs"].

CAE. (personal communication, September 22, 2010). [Re: "Teleconference call on 27 September"]. Logistics Country A/Country E, France, USA.

CAE. (personal communication, September 22, 2010). [Re: "Teleconference call on 29 September"]. Logistics Country D, France, USA.

CAE. (personal communication, September 22, 2010). [Re: "Teleconference call on 28 September"]. Logistics Country B, France, USA.

CAE (2010). *AHELO GS Progress Report-Milestone 2*. Assessment of Higher Education Learning Outcomes Generic Strand. August.

CAE. (personal communication, August 18, 2010). [Re: "CAE AHELO Preparation for Implementation"].

CAE. (personal communication, July 9, 2010). [Re: "CAE AHELO Translation Update 5"].

CAE. (personal communication, July 9, 2010). [Re: "CAE AHELO Translation Update"]. GS.33a, 34a, 33b, 34b, 35, 36.

CAE. (personal communication, June 4, 2010). [Re: "CAE AHELO Translation Update 3"]. GS.40, 41.

CAE. (personal communication, May 11, 2010). [Re: "CAE AHELO GS Reminder: Translation Team Qualifications"].

CAE. (personal communication, April 28, 2010). [Re: "CAE AHELO Generic Strand Update 6"]. Scoring.

CAE. (personal communication, April 27, 2010). [Re: "CAE AHELO Generic Strand Update 6"]. GS.31.

CAE. (personal communication, April 21, 2010). [Re: "CAE AHELO GS Reminder: Translation Team Qualifications"].

CAE. (personal communication, April 19, 2010). [Re: "CAE AHELO PT ADAPTATION UPDATE"]. To Country D, Attachments: compiled both PT adaptations.

CAE. (personal communication, April 18, 2010). [Re: "CAE AHELO PT ADAPTATION UPDATE"]. Next steps after review of suggested modifications.

CAE. (personal communication, January 29, 2010). [Re: "CAE AHELO Generic Strand Update 2"]. GS.4, 5, 6.

CAE. (2010). *GS.1 AHELO Generic Strand timeline UPDATED.docx.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.1 Project Work Plan with Dates_UpdateNeeded.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.2 CAE Confidentiality Agreement AHELO_ForColombiaOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.3 Campus Responsibilities_UpdateNeeded.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.4 Conceptual Framework.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.5 Criteria and Recommendations for PT Selection.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.6 CAE AHELO Meeting Agenda_ForReferenceOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.7 Hambleton_2005_Issues_design_and_technical_guidelines_1.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.8 ITC-2005_Guidelines_on_Computer_Delivered_Testing.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.9 Solano-Flores_et_al._2009-Theory_of_test_translation_error-final.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.10 Thompson_et_al._2002_Universal_designed_applied_to_large-scale_assessments.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.11 Biographical Sketches of Meeting Participants_For Reference Only.pdf (35).* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.12 AHELO Module A Meeting Overview_ForReferenceOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010) GS.13 Translation and Adaptation Guidelines.pdf. Assessment of Higher Education Learning Outcomes Generic Strand. February.

CAE. (2010). *GS.14 Translation Flowchart.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.15 Communication Log for AHELO Study_ForReferenceOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.16 Country A HE Overview_ForReferenceOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.17 Country B HE Overview_ForReferenceOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.18 Country C HE Overview_ForReferenceOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.19 Country D HE Overview_ForReferenceOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.20 Country E HE Overview Description_ForReferenceOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.21 US HE Overview_ForReferenceOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.22 OECD AHELO Overview for CAE GEN Skills Mtg_For Reference Only.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.23 AHELO GS Lead Scorer.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.24 Generic Strand Scorer Estimates_ForReferenceOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.25 AHELO Module A 17 March Meeting Agenda_For Reference Only.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.26 AHELO Generic Strand NY Meeting Notes_ForReferenceOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.27 CLA Invited Presentation_Helsinki_ForReferenceOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.28 CLA Invited Presentation_Oslo_ForReferenceOnly.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.29 US CLA Admin Manual_REFERENCE ONLY.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.30 GNE(2010) - Progress Report on Generic Skills_For Reference Only.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.31 Materials Translation List_UpdateNeeded.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.32 CAE Representative Site Visit_UpdateNeeded.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.33a Catfish Performance Task_English_CONFIDENTIAL.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.33b Catfish Performance Task_Template_CONFIDENTIAL.docx.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.34a Lake to River Performance Task_English_CONFIDENTIAL.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.34b Lake to River Performance Task_Template_ CONFIDENTIAL. docx.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.35 AHELO Scoring Rubric_CONFIDENTIAL.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.36 AHELO Translation Guide_UpdateNeeded.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand. February.

CAE. (2010). *GS.36 and GS.37 Addendum.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.37 AHELO Cognitive Labs Guide.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.38 Mini Performance Task.Final English.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.38 Mini Performance Task_Template.doc.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.39 Performance Task Response Features_CONFIDENTIAL.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.40 Motivation Items.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.40 Motivation Items_Template.docx.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.41 U.S. CLA Internet Platform Tutorial (Confidential).pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.42 AHELO Internet Platform Instructions_CONFIDENTIAL.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.43 Pre-Implementation Process - Internet Platform _Update Needed.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.44 Pre-Implementation Process - Supplemental Materials Update Needed.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.45 Module A Activities October 2010 to May 2011_ For Reference Only.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.46 Summary of AHELO Generic Strand Teleconferences_ For Reference Only.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

CAE. (2010). *GS.47 Summary of Final PT Changes_CONFIDENTIAL.pdf.* Assessment of Higher Education Learning Outcomes Generic Strand.

Chia, M. (2011). *AHELO – Country D Rubric Interview*. Assessment of Higher Education Learning Outcomes Generic Strand. October.

Chia, M. (2011). *AHELO – Country A Rubric Interview*. Assessment of Higher Education Learning Outcomes Generic Strand. October.

Choi, J. (personal communication, February 1, 2010). [Re: "Signed confidentiality agreement forms of Country B participants"]. Team members.

Choi, J. (personal communication, January 21, 2010). [Re: "From Country B AHELO team_Summary of the phone meeting on January 18"].

Choi, J. (personal communication, January 14, 2010). [Re: "GREETINGS FROM CAE AND THE GENERIC STRAND OF AHELO"]. Initial logistics.

Coates, H. (personal communication, October 3, 2011). [Re: "AHELO update and resources"]. Country E logistics documents.

Coates, H. (personal communication, July 12, 2011). [Re: "AHELO Contextual Dimension Instruments"].

Coates, H. (personal communication, February 11, 2011). [Re: "draft NPM agenda fro review/feedback"].

Coates, H. (personal communication, January 17, 2011). [Re: "AHELO 2011, teleconference"].

Coates, H. (personal communication, December 20, 2010). [Re: "AHELO Update"] 8:20pm.

Coates, H. (personal communication, November 17, 2010). [Re: "AHELO NPM information"] 8:18pm.

Coates, H. (personal communication, September 28, 2010). [Re: "AHELO NPM meeting, 27-28 October, Paris"].

Coates, H. (personal communication, September 27, 2010). [Re: "October NPM meeting agenda"].

Coates, H. (personal communication, September 17, 2010). [Re: "AHELO documents and agenda"] times.

Coates, H. (personal communication, September 15, 2010). [Re: "AHELO documents and agenda"].

Gallimore, T. (personal communication, February 18, 2010). [Re: "AHELO requirements"]. Some items from follow-up NY meeting (e.g., HEIs).

Hyytinen, H. (personal communication, September 23, 2011). [CAE AHELO Rubric Interview on Translation].

Jan-Schreiner, L. (personal communication, January 28, 2010). [Re: "Visit to Country E March 4/5"]. Country E Ministry seeking meeting.

Keeley, R. (personal communication, September 13, 2011).

Keeley, R. (personal communication, March 2, 2011). [Re: "CAE AHELO: Phase 1 User Acceptance Testing"] example of same email sent to all countries.

Keeley, R. (personal communication, September 28, 2011). [Re: "Inquiry regarding a modification of assessment tools"] Country B second review—help by Buros.

Keeley, R. (personal communication, June 6, 2011). [Re: "CAE AHELO: Phase 2 User Acceptance Testing (Pre-Implementation)"] All-directions for Phase 2.

Keeley, R. (personal communication, June 2, 2011). [Re: "Updated maps of the PT's"] Country D GIF and edits.

Keeley, R. (personal communication, June 2, 2011). [Re: "Evidence on feasibility of Generic Skills test"] country evaluations and conclusion on process.

Keeley, R. (personal communication, May 26, 2011). [Re: "Verification of the docs"] Country A xls to doc ITS documents.

Keeley, R. (personal communication, May 20, 2011). [Re: "Test Language Pt's 1 and 2"] Country D feedback on internet PTs.

Keeley, R. (personal communication, May 19, 2011). [Re: "VL: dokut"] Country A feedback on internet PTs.

Keeley, R. (personal communication, April 28, 2011). [Re: "CAE AHELO Online Platform Review (Pre-Implementation)"] Country D computer security.

Keeley, R. (personal communication, April 26, 2011). [Re: "CAE AHELO Online Platform Review (Pre-Implementation)"] internet docs and info on 2 phases.

Keeley, R. (personal communication, April 18, 2011). [Re: "CAE AHELO Phase 1 User Acceptance Testing"]. Country E Word-document for comments.

Keeley, R. (personal communication, March 4, 2011). [Re: "CAE AHELO Online Platform Review (Pre-Implementation)"] Country E ready-external trans. service.

Keeley, R. (personal communication, February 15, 2011). [Re: "Word Version of the final Country D PTs"] km-miles conversion.

Kim, K.S. (2010). *Kyung-Sung Kim CV Measurement Expert Country B*. Assessment of Higher Education Learning Outcomes Generic Strand. January. (78)

Kurpius, A. (personal communication, March 29, 2011). [Re: "Communicaid Inc Project 61636-GS.33 – catfish and GS.34-lake to river Portfolios"].

Kurpius, A. (personal communication, February 8, 2011). [Re: "Status of AHELO Translation"] Country B cog labs feedback clarifying information.

Kurpius, A. (personal communication, December 7, 2010). [Re: "Translator 1 documents"] Country D Rubrics and Instructions translation process.

Kurpius, A. (personal communication, December 3, 2010). [Re: "Status of AHELO Translation"] Country B reminder for feedback on cog labs and Country A example.

Kurpius, A. (personal communication, October 7, 2010). [Re: "Additional observations of PTs"]. Country A's observations. Also, other countries missing deadline.

Kurpius, A. (personal communication, September 22, 2010). [Re: "Translator 1 documents"] Country D Rubrics and Instructions translation process.

Kurpius, A. (personal communication, August 8, 2010). [Re: "Milestone 2 Progress Report"] Exchange between Amy and Jim.

Kurpius, A. (personal communication, June 10, 2010). [Re: "CAE AHELO Translation Update 3"] Country A prep for site visit.

Kurpius, A. (personal communication, June 9, 2010). [Re: "CAE AHELO Translation Update 3"] mini PT, cog labs, conf agreements, Willy visit.

Kurpius, A & Shavelson, R. (personal communication, March 22, 2010). [Re: "CAE AHELO Generic Strand Update 5"].

Kurpius, A. (personal communication, March 17, 2010). [Re: "AHELO Generic Strand March 17 (Paris, OECD) Meeting Information"].

Kurpius, A. (personal communication, February 8, 2011). [Re: "Translated PTs"] copy of all translated PTs minus Country B.

Kurpius, A. (personal communication, February 2, 2011). [Re: "CAE AHELO Final Adaptation/Translation Document"] Country B.

Kurpius, A. & Shavelson, R. (personal communication, January 19, 2010). [Re: "CAE AHELO Generic Strand Update 1"]. GS.1, 2, 3.

Kurpius, A. (personal communication, January 5, 2011). [Re: "CAE AHELO Final Adaptation/Translation Document"] Country C.

Kurpius, A. (personal communication, October 8, 2010). [Re: "Additional observations of PTs"]. Revised timeline-Country A only country on time.

Kurpius, A. & Shavelson, R. (personal communication, September 6, 2010). [Re: "A small question"]. Response to Country A question about mini-PT error.

Kurpius, A. (personal communication, August 10, 2010). [Re: "Translations of PTs"].

Kurpius, A. (personal communication, February 12, 2010). [Re: Country teams laptops NY].

Kurpius, A. (personal communication, February 26, 2010). [Re: update 4 nyc meeting].

Kurpius, A. (personal communication, February 1, 2010). [Re: to Imad NYC attendance].

Lenth, C. (personal communication, February 21, 2010). [Re: Review of materials and update]. Cog labs.

Country D. (2010) *Caracteristicas de las respuestas de los PTs*. Assessment of Higher Education Learning Outcomes Generic Strand. November.

OECD. (2010). *OECD AHELO Brochure 2010-2011*. Assessment of Higher Education Learning Outcomes Generic Strand. January. (80)

Opheim, V. (personal communication, April 18, 2011). [Re: "Changes in the Country E AHELO team"]. Vibeke leaving.

Opheim, V. (personal communication, March 4, 2011). [Re: "CAE AHELO Internet Platform Documents (Pre-Implementation)"]. 11 documents (including libraries, scoring, interface).

Opheim, V. (personal communication, January 26, 2011). [Re: "CAE AHELO GS.2 Confidentiality Agreement Document"]. signed.

Opheim, V. (personal communication, January 21, 2011). [Re: "CAE AHELO Final Adaptation/Translation Document"]. PTs and mini-PT.

Opheim, V. (personal communication, December 22, 2010). [Re: "Status of AHELO Translation"]. Cog lab results and "other translated documents".

Opheim, V. (personal communication, August 26, 2010). [Re: "CAE AHELO Translation Update 6"]. PTs and Draft Cog Lab reconciled translations.

Opheim, V. (personal communication, January 25, 2010). [Re: "CAE AHELO Final Adaptation/Translation Document"]. IT instructions GS.42.

Opheim, V. (personal communication, January 13, 2010). [Re: "CAE-AHELO FOLLOW UP"]. CV Roe and Turmo—assessment experts.

Richardson, Sarah. (personal communication, November 30, 2010). [Re: "AHELO Teleconferences with NPMs - Update"].

Richardson, Sarah. (personal communication, November 29, 2010). [Re: "AHELO Teleconferences with NPMs"].

Roe, A. (personal communication, January 21, 2011). [Re: "CAE AHELO Final Adaptation/Translation Document"]. Country E explanation for missing deadline due to PISA.

Roe, A. (personal communication, April 28, 2010). [Re: "CAE AHELO Generic Strand Update 6"]. Country E-new team member.

Roe, A. (2010). *Astrid Roe CV*. Assessment of Higher Education Learning Outcomes Generic Strand. January. (81)

Rosas Chavez, P. (personal communication, April 26, 2010). [Re: "CAE AHELO GS: Willy Solano-Flores Visit"]. PTs, mini-PT, Museum PT, Willy visit.

Rosas Chavez, P. (personal communication, April 21, 2010). [Re: "CAE AHELO GS Reminder: Translation Team Qualifications"].

Rosas Chavez, P. (personal communication, January 25, 2010). [Re: "CAE AHELO GS.2 Confidentiality Agreement Document"]. PTs, mini-PT, Museum PT, Willy visit.

Rosas Chavez, P. (personal communication, January 14, 2010). [Re: "Getting in touch"]. Initial phone conference call.

Sanchez-Gomez, R. (personal communication, December 20, 2011). [Re: "Your Test Preview Authorization"]. Acknowledgement.

Sanchez-Gomez, R. (personal communication, April 21, 2010). [Re: "CAE AHELO PT ADAPTATION UPDATE"]. Country D two PT adaptations compiled.

Sanchez-Gomez, R. (personal communication, January 21, 2011). [Re: "CAE AHELO Final Adaptation/Translation Document"]. PTs, mini-PT, responses, Internet platform.

Shavelson, R. (personal communication, October 28, 2010). [Re: "CAE AHELO Translation Update 8"]. Country A follow-up to Paris NPM meeting in October and their update.

Shavelson, R. (personal communication, September 23, 2010). [Re: "Translator 1 documents"]. Agreement to meeting time.

Shavelson, R. (personal communication, July 27, 2010). [Re: "WORKSHOP IN COUNTRY D"]. To Willy—Country D's translation work status.

Shavelson, R. (personal communication, July 26, 2010). [Re: "CAE AHELO Translation Update 5"]. Country D call planning.

Shavelson, R. (personal communication, June 18, 2010). [Re: "Questions for survey"]. Country E-challenges finding items for on-site training w/ Willy.

Shavelson, R. (personal communication, April 19, 2010). [Re: "CAE AHELO PT ADAPTATION UPDATE"]. Country D adaptation arrangements.

Shavelson, R. & Kurpius, A. (personal communication, February 9, 2010). [Re: "CAE AHELO Generic Strand Update 3"]. Prep for NY office.

Shavelson, R. (personal communication, January 26, 2010). [Re: "From Country B AHELO team_Summary of the phone meeting on January 18"]. Country B moving GS testing into spring 2011.

Shavelson, R. (personal communication, January 18, 2010). [Re: "Jeung's email"]. Choosing one member of Country B's team.

Shavelson, R. (personal communication, January 14, 2010). [Re: "CAE AHELO FOLLOW UP"]. Country C initial contacts.

Shavelson, R. (personal communication, January 11, 2010). [Re: "GREETINGS FROM CAE AND THE GENERIC STRAND OF AHELO"]. Country A—scheduling phone meeting.

Shavelson, R. (personal communication, January 9, 2010). [Re: "GREETINGS FROM CAE AND THE GENERIC STRAND OF AHELO"]. Initial email January 9-11.

Shavelson, R. (personal communication, March 10, 2011). [Re: "AHELO GENERIC SKILLS TRANSLATION VERIFICATION"]. Wanting external verification.

Shavelson, R. (2010). *Country C Conference Call Notes*. Assessment of Higher Education Learning Outcomes Generic Strand. October. (82)

Shavelson, R. (2010). *Country A/Country E Conference Call Notes*. Assessment of Higher Education Learning Outcomes Generic Strand. September.

Shavelson, R. (2010). *Country B Conference Call Notes*. Assessment of Higher Education Learning Outcomes Generic Strand. September.

Shavelson, R. (2010). *Country D Conference Call Notes*. Assessment of Higher Education Learning Outcomes Generic Strand. September.

Solano-Flores, G. (personal communication, June 10, 2010). [Re: "CAE AHELO Translation Update 2"]. Willy and Rich analytic units during translation.

Solano-Flores, G. (personal communication, May 26, 2010). [Re: "Questions about translation"]. Country A-translation process and site visit questions.

Solano-Flores, G. & , M. (2010). *Country C Interview Transcription regarding translation process*. Assessment of Higher Education Learning Outcomes Generic Strand. September. (86)

Solano-Flores, G. & Chia, M. (2010). *Country A Interview Transcription regarding translation process*. Assessment of Higher Education Learning Outcomes Generic Strand. XXX.

Solano-Flores, G. & Chia, M. (2010). *Country B Interview Transcription regarding translation process*. Assessment of Higher Education Learning Outcomes Generic Strand. XXX.

Solano-Flores, G. & Chia, M. (2010). *Country D Interview Transcription regarding translation process*. Assessment of Higher Education Learning Outcomes Generic Strand. XXX.

Solano-Flores, G. & Chia, M. (2010). *Country E Interview Transcription regarding translation process*. Assessment of Higher Education Learning Outcomes Generic Strand. September.

Solano-Flores, G., Shavelson, R., & Chia, M. (2010). *Country E and Country A Site Visit Report*. Assessment of Higher Education Learning Outcomes Generic Strand. July.

Solano-Flores, G. (2010). *Country B Site Visit Report*. Assessment of Higher Education Learning Outcomes Generic Strand. XXX.

Solano-Flores, G. (2010). *Country C Site Visit Report*. Assessment of Higher Education Learning Outcomes Generic Strand. XXX.

Solano-Flores, G. (2010). *Country D Site Visit Report*. Assessment of Higher Education Learning Outcomes Generic Strand. XXX.

Throndsen, I. (personal communication, April 28, 2010). [Re: "adaptations-Country E"]. PTs with dropdown menus.

Tremblay, K. (personal communication, February 18, 2011). [Re: "Agendas for the March round of meetings"]. March meeting schedule.

Tremblay, K. (personal communication, January 7, 2011). [Re: "AHELO-Welcome, introduction to the project and next steps"]. OECD new country Slovak Republic.

Turmo, A. (personal communication, June 3, 2010). [Re: "CAE AHELO Translation Update 2]. Mini-PT, no adaptations needed.

Turmo, A. (2010). *Dr. Are Turmo CV*. Assessment of Higher Education Learning Outcomes Generic Strand. January. (95)

Urrea, B. (personal communication, December 15, 2010). [Re: "Final PTs and Mini PTs from Country D"].

Urrea, B. (personal communication, November 4, 2010). [Re: "GS.37 cognitive lab reconciled"].

Urrea, B. (personal communication, November 3, 2010). [Re: "CAE AHELO Translation Update"]. 2 PTs and mini-PT reconciled versions.

Urrea, B. (2010). *Excel translation materials, responsibilities, and timeline*. Assessment of Higher Education Learning Outcomes Generic Strand. September.

Urrea, B. (personal communication, September 28, 2010). [Re: "Documents needed by UdeG as soon as possible concerning PT3"]. Payment and devaluation of peso.

Urrea, B. (personal communication, September 27, 2010). [Re: "internet platform instructions conciliated translation"].

Urrea, B. (personal communication, September 27, 2010). [Re: "Documents we send before tomorrow's meeting"]. GS.31 updated.

Urrea, B. (personal communication, September 22, 2010). [Re: "Reconciled translation PT1, PT2 and Mini PT from Country D Team"].

Urrea, B. (personal communication, September 22, 2010). [Re: "Translator 1 Spanish versions of GS.35 and GS.42"]. Rubric, IT instructions.

Urrea, B. (personal communication, September 20, 2010). [Re: "Translator 1 documents"]. Rubric, IT Platform, CV.

Urrea, B. (personal communication, September 9, 2010). [Re: "CAE AHELO – Minor Edit in GS.38 Mini PT Document"]. Forward to translation team.

Urrea, B. (personal communication, July 27, 2010). [Re: "CAE AHELO Translation Update 5"]. Payment CAE, 3rd PT, new timeline.

Urrea, B. (personal communication, July 16, 2010). [Re: "CAE AHELO Translation Update 5"]. Conference call timing.

Urrea, B. (personal communication, June 1, 2010). [Re: "CAE AHELO Translation Update 2"]. GS.38 mini PT adapted.

Urrea, B. (personal communication, April 28, 2010). [Re: "CAE AHELO Generic Strand Update 6"]. Questions about GS.31-unclear.

Urrea, B. (personal communication, January 25, 2010). [Re: "Confidentiality Agreements signed"].

Urrea, B. (personal communication, April 19, 2010). [Re: "CAE AHELO PT ADAPTATION UPDATE"]. Questions about Country D's adaptations

Ursin, J. (personal communication, January 21, 2011). [Re: "Country A PTs"]. Final versions of PTs and mini-PTs.

Ursin, J. (personal communication, January 20, 2011). [Re: "CAE AHELO Final Adaptation/Translation Document"]. Quick Jani response to reminder.

Ursin, J. (personal communication, November 8, 2010). [Re: "Scanned examples of answers to Pts in Finalnd"].

Ursin, J. (personal communication, October 7, 2010). [Re: "Additional observations of PTs"]. Country A-feedback from expert group meeting, including translation review team.

Ursin, J. (2010). *Country A Excel work with demographic information for cognitive labs*. Assessment of Higher Education Learning Outcomes Generic Strand. September.

Ursin, J. (2010). *Country A results from cognitive labs*. Assessment of Higher Education Learning Outcomes Generic Strand. September.

Ursin, J. (personal communication, September 24, 2010). [Re: "VS: Teleconference call on 27 September"]. Country A-Observations from cog labs.

Ursin, J. (personal communication, September 7, 2010). [Re: "A small question"]. Country A-found minor error on mini-PT; also completed cog labs.

Ursin, J. (personal communication, August 19, 2010). [Re: "CAE AHELO Preparation for Implementation"]. Country A-third translator hired.

Ursin, J. (personal communication, August 11, 2010). [Re: "Translations of PTs"]. Country A-changes on national team.

Ursin, J. (personal communication, August 10, 2010). [Re: "Translations of PTs"]. Country A-reconciled PTs.

Ursin, J. (personal communication, June 28, 2010). [Re: "VS: CAE AHELO Translation Update 4"].

Ursin, J. (personal communication, June 8, 2010). [Re: "VS: CAE AHELO Translation Update 3"]. Incl Amy response to Jani questions in thread.

Ursin, J. (personal communication, June 8, 2010). [Re: "VS: CAE AHELO Translation Update 3"]. Logistics for translation team and cog labs.

Ursin, J. (personal communication, April 22, 2010). [Re: "CAE AHELO GS Reminder: Translation Team Qualifications"]. Seeking clarification on translation duties.

Ursin, J. (personal communication, April 28, 2010). [Re: "CAE AHELO: Willy Solano-Flores Visit"]. Country A-process of getting translation team together.

Ursin, J. (2010). *Jani Ursi CV*. Assessment of Higher Education Learning Outcomes Generic Strand. January.

Ursin, J. (2010). *Jani Ursi publications and presentations*. Assessment of Higher Education Learning Outcomes Generic Strand. January.

Ursin, J. (personal communication, March 25, 2010). [Re: "CAE AHELO Generic Strand Update 5"]. Country A-recommended modifications for both PTs in response to synthesized notes.

Ursin, J. (personal communication, January 18, 2010). [Re: "GREETINGS FROM CAE AND THE GENERIC STRAND OF AHELO"]. NY meeting logistics.

Ursin, J. (personal communication, January 11, 2010). [Re: "GREETINGS FROM CAE AND THE GENERIC STRAND OF AHELO"]. To Rich.

Ursin, J. (personal communication, January 12, 2010). [Re: "Responses from Country A"]. Country A team contact list and Ursin CV and publications and meeting times.

Ursin, J. (personal communication, January 27, 2010). [Re: "Conf Agrmts signed"]. Country A team.

Young, E. (personal communication, June 4, 2010). [Re: "CAE AHELO Translation Updae 2"]. Revised mini-PT.

Young, E. (personal communication, April 28, 2010). [Re: "Modified/adapted PT PDFs"]. PTs attached.

Young, E. (personal communication, March 26, 2010). [Re: "Modified Performance Tasks"]. Suggested modifications.

Young, E. (personal communication, February 8, 2011). [Re: "CAE AHELO Final Adaptation/Translation Document"]. Instructions and PTs attached.

Young, E. (personal communication, January 25, 2011). [Re: "CAE AHELO Internet Platform Documents (Pre-Implementation"]. Teleconference, lunar new year, password.

Appendix C: Excerpts of conference call questions related to translation and adaptation

1. Update on progress from each of the country teams.
    a. Dates of importance
2. Reflections on and review of AHELO Module A adaptation and translation process.
    a. What would you change?
        i. Communications/documentation
        ii. Procedures/process
        iii. Meetings
        iv. other
    b. any clarifications/questions about think alouds or any other translation procedures?
    c. What else can CAE do to help facilitate completion of the translation phase?

Appendix D: Excerpts of the task adaptation open-ended survey

1. Visual Literacy

Describe cases in which the style with which textual and visual information was presented in the original version of the task might prevent students in your country from properly interpreting information presented in the tasks. For each case, describe how your team addressed the challenges encountered.

Click for the Visual Literacy Example

2. Technological Literacy

Describe cases in which the technological skills required to navigate through the task's information might impact the students' ability to meet the task's requirements. For each case, describe how your team addressed the challenges encountered.

Click for the Technological Literacy Example

3. Tenor Appropriateness

Describe cases in which the style in which information is presented does not reflect your country's culture and may prevent students from properly interpreting information. For each case, describe how your team addressed the challenges encountered.

Click for the Tenor Appropriateness Example

4. Subject Matter Appropriateness

Describe cases in which your students may consider the subject matter offensive or inappropriate and may cause them difficulty when trying to respond to a task. For each case, describe how your team addressed the challenges encountered.

Click for the Subject Matter Appropriateness Example

Appendix E: Excerpts of the 'Task Adaptation Site Visit Interview Guide

Task Adaptation Site Visit Interviewer Guide

1. Explain to participants the purpose of the interview

   As you know, there are two main components in the feasibility study using Performance Tasks —adaptation and translation. This interview will address each of the two processes.

   During the first half of the interview, we will ask you to share all pertinent information regarding the adaptation process that your team implemented to decide what was and was not culturally appropriate in the two tasks. During the second half of the interview, we will ask you to share what your team has done thus far for the translation process and what you plan on doing as we move forward with it.

2. Ask the general questions (shown in bold letters) as shown. Direct the interview to ensure that the participants address the issues shown in italics. Check the boxes to indicate that the corresponding issues have been addressed.

   Adaptation:

   - **Regarding the participants in your team, how did you decide on who would participate in the adaptation process?**
     - ❑ *How many people participated in the adaptation process?*
     - ❑ *How did you choose the participants for the adaptation process?*
     - ❑ *What were their backgrounds?*

   - **Regarding the adaptation process, how did you decide on what was culturally appropriate or inappropriate?**
     - ❑ *What specifically were you looking for during the adaptation process?*
     - ❑ *How did you ensure that these issues were addressed during the adaptation process?*
     - ❑ *Please explain the process itself (some of this information may come about during the previous question).*
     - ❑ *Please explain the role of each person during the process.*
     - ❑ *How did you decide what needed to be changed?*

   - **Regarding your evaluation of the adaptation process, what are some of the lessons learned?**
     - ❑ *What worked well during the process?*
     - ❑ *What did not work well during the process?*

Translation

- **Regarding students, how are you selecting the students who will participate in the study so that you obtain a representative sample (e.g., gender, SES, geographic areas, type of education institution)?**
  - ❑ *What dialect or cultural differences exist among your university students that may impact their understanding of the performance tasks?*
  - ❑ *How are you planning to address the factors suggested for selecting students to participate in the study?*
    - ❑ *Based on gender?*
    - ❑ *Based on SES?*
    - ❑ *Based on geographic areas?*
    - ❑ *Based on type of institution?*
    - ❑ *What will your sample look like (e.g., size by each of the above factors)?*

- **Regarding the translation process, how will you ensure that during the translation process you address all of the suggestions that we have provided?**
  - ❑ How will you determine or how have you determined who will be on the translation team?
  - ❑ What difficulties has your team had finding people to become members of the translation team?
  - ❑ How do you see the translation process as being similar and different to the adaptation process?

Appendix F: Exhaustive list of tasks and sub-tasks most applicable to FOI criteria

Task/Subtask	Doc	Cty
Initial details provided for Feb 2010 meeting by CAE read by country teams.	Intr ltr	All
Conceptual framework for PT translation provided by CAE read by country teams.	Intr ltr	All
General instructions for CLA administration on Internet provided by CAE read by country teams.	**Intr ltr**	**All**
Summary descriptions of 9 CLA tasks provided by CAE read by country teams.	Intr ltr	All
Recommendations and rationales for selecting from 4-5 subset PTs read by country teams at the NYC meeting.	Intr ltr	All
Explanation of complexities of generic strand (intended constructs) read by country teams.	Intr ltr	All
Mentions issues of validity in addition to those associated with cross-cultural appropriateness, and linguistic transferability—to be read by country teams.	Intr ltr	All
Points out the role of contextual information in using test results for policy and practice-related decisions—to be read by country teams.	Intr ltr	All
Indication of information found in other documents—country teams to cross reference and read.	**Intr ltr**	**All**
Ask for each country team to provide availability for conference call to discuss future Feb 2010 meeting.	Intr ltr	All
Country teams to nominate professionals to be assessment representatives: • person must be available ½ time during 2010 calendar year • country teams must provide each nominee's CV to CAE	Intr ltr	All
Familiarize team with 16 contacts (2 CA, 2 CO, 12 NY)	**Annex B**	**All**
Country teams to read about CLA and its role in the larger context of assessment in higher education accountability.	**Annex C**	**All**
Country teams to understand the constructs being assessed in the CLA.	Annex C	All
Country teams to understand the criterion sampling approach.	Annex C	All
Country teams to understand student operant responses and the context of performance tasks and multiple skills.	Annex C	All
Country teams to be able to read tables, plot charts, and research abstracts.	Annex C	All
Country teams to understand complexity of performance tasks: time required, cost,	Annex	All

Task/Subtask	Doc	Cty
and scoring time requirements.	C	
Country teams to understand natural language processing software and its relationship to reliability or validity.	Annex C	All
Country teams to understand how matrix sampling can reduce testing time.	Annex C	All
Country teams to understand the impact of paperless administration.	Annex C	All
Country teams to understand the impact of online rater scoring and calibration of the PT.	Annex C	All
Country teams to understand the accountability aspects of the CLA: signaling, benchmarking, value added focus	Annex C	All
Country teams to understand the issues associated with summative function of accountability: stakes, common set of indicators, incentive or punishment.	Annex C	All
Country teams to understand formative function of accountability at the student and school levels: diagnosing and providing feedback, monitoring change.	Annex C	All
Country teams to be aware of the importance of benchmarking or examining value added.	Annex C	All
Country teams to become familiar with the controversial aspects of high-stakes testing (e.g., cheating, sanctions) to understand CLA's position on the issue.	Annex C	All
Country teams to understand the limitations of the CLA (needs measures for specific majors; measures of social, moral, and civic outcomes).	Annex C	All
Country teams to agree with the timely fashion in which activities must be completed.	**Annex D**	**All**
Country teams to agree to sign and abide by confidentiality agreement.	Annex D	All
Country teams to prepare for the meeting in New York City (February 2010).	Annex D	All
Countries must have access to software and computer/telecom equipment.	Annex D	All
Familiarize team representatives with CAE provided conceptual framework for adaptation.	**Annex D**	**All**
Familiarize team representatives with CAE provided procedures for training country members to adapt the PTs.	Annex D	All
Team representatives to participate in New York City meeting (February 2010).	Annex D	All
During February 2010 meeting, country teams to become more familiar with CLA.	Annex D	All
Country team representatives to review a subset of at least nine CLA PTs.	Annex	All

Task/Subtask	Doc	Cty
	D	
Country teams to select four suitable PTs—considered valid in an international context.	Annex D	All
Country teams to select two suitable PTs that will be used in AHELO.	Annex D	All
Country teams to modify the two suitable PTs for use in AHELO.	**Annex D**	**All**
Country teams to become familiar with the three criteria to select the subset of PTs. (universality of PT theme; ease of translation [based on complexity of language in PT]; ease of scoring based on US experience w/ CLA).	Annex D	All
Country assessment experts choose Assessment Adaptation Group (AAG).	**Annex D**	**All**
AAG to train (by CAE) in task adaptation.	Annex D	All
AAG to train (by CAE) in translation process.	Annex D	All
AAG to train translation review process.	Annex D	All
Country team to learn about recruiting of test translation team.	**Annex D**	**All**
Country teams to review contents of two selected/agreed upon PTs.	Annex D	All
After NYC meeting, country teams to create list of modifications for each PT to have each PT fit country context.	Annex D	All
Via telecom meetings AAG member to present recommended modifications of CLA PTs.	Annex D	All
Via telecom meetings, based on CAE and Country Team evaluations and recommendations, gain consensus on the two PTs.	Annex D	All
CAE to fully modify two tasks following agreed upon modifications.	Annex D	All
CAE to fully modify scoring rubrics following agreed upon modifications.	Annex D	All
CAE to fully modify IT administration procedures following agreed upon modifications.	Annex D	All
CAE to send modified tasks, rubrics, and IT administration procedures for review, comment, and revision as they are completed.	Annex D	All
Country teams to review, comment, and revise the two tasks when received modified versions from CAE.	Annex D	All

Task/Subtask	Doc	Cty
Country teams to review, comment, and revise the rubrics when received modified versions from CAE.	Annex D	All
Country teams to review, comment, and revise the IT administration procedures when received modified versions from CAE.	Annex D	All
Translation team translates modified PTs using guidelines.	Annex D	All
a. Country teams, w/ CAE help, to recruit translation team of at least 2 people.	Annex D	All
a. Country teams, w/ CAE help, to recruit translation review team (translator, university prof knowledgeable of content area, at least one assessment expert).	Annex D	All
a. Country teams responsible for compensating (if applicable) translation team and translation review team.	Annex D	All
b. CAE representatives and translation expert to visit, assist, train, and guide translation team and translation review team.	Annex D	All
c. Via telecom meetings w/ translation teams, CAE will provide support of translation process.	Annex D	All
c. Via telecom meetings w/ translation teams, CAE will provide share translations for review and finalization.	Annex D	All
d. Country teams, with guidance from CAE, to conduct small pilots (~10 students) in home country. This will include cognitive workshops.	Annex D	All
d. Country teams, with guidance from CAE, to conduct cognitive workshops with a "small sample of students".	Annex D	All
d. Country teams to become familiar with benefits of think alouds—insure that the thinking elicited by the performance task is the thinking sought; see if the PT measures the same thinking across countries.[7]	Annex D	All
e. Each country team to select only one language for field test.[8]	Annex D	All
f. Translation teams will translate the on-line testing instructions early in the process. These take longer to adapt onto the test delivery website.	Annex D	All
g. CAE produces final versions of translated PTs for incorporation into the internet testing application.	Annex D	All
CAE to participate in two AHELO GNE meetings in Paris.[9]	Annex D	All

[7] Willy, I thought the think alouds were to find out if translation caused any problems—and not to seek general information about thinking procedures.
[8] What is this field test referring to?
[9] Document only states that, "funding provided under this contract" for the CAE travel and participation in meetings held in Paris.

Task/Subtask	Doc	Cty
Country teams aware of separate optional faculty academies.	Annex D	All
Country teams to familiarize themselves with CAE staff and TAT backgrounds and responsibilities (Roger Benjamin, Ph. D., James Hundley, Rich Shavelson, Ph. D., Guillermo Solano-Flores, Ph. D., Amy Kurpius, Stephen Kelin, Ph. D., Marc Chun, Ph. D., Jeffrey Steedle, Ph. D.	Annex D	All
Country teams to be familiar with other technical advisors (Ron Hambleton, Ph. D.[10], Scott Elliot, Ph. D., other program managers/associates.	Annex D	All
Country teams to be aware that the test translation team minimally required two certified professional translators (English-target language), one of the two serving as a leader, and desired additional team members (content specialist, reconciler, linguist)	Annex D	All
Country teams to be aware that the test translation review team minimally required one translation review leader (assessment specialist), one specialist in the content area (university professor)[11], one certified professional translator (English-target language), and desired additional team members (linguist, second translator).	Annex D	All
Country teams each to nominate several persons with expertise in assessment who will represent country team in AHELO.	**Annex E**	**All**
One assessment representatives will be selected based on formal training, professional experience, and relevant set of technical qualifications.[12]	Annex E	All
Country teams to nominate people with required qualifications: • education and formal training (Ph.D. in psychometrics, statistics, social-science measurement or related field) • English proficiency (fluent in conversational and reading/writing English) • Professional experience (higher education, int'l or multicultural education; academic achievement assessment; assessment development—preferable constructed response tasks, computer simulations or computer-based training, test translation and test translation review	Annex E	All
Nominees for assessment representative will hopefully have desired qualifications: • experience coordinating the collection, management, and analysis of data for educational projects. • Ability to successfully work with multidisciplinary and/or multicultural teams • Record of technical publications in the areas of expertise	Annex E	All

[10] Document indicates, pg 8, that this function was now taken on by 'Consortium's technical advisory group. Is this ACER? Did Hambleton ever serve? Why the change?
[11] How did they handle content area specialization for generic strand?
[12] Unclear who does the selection of the assessment representative.

Task/Subtask	Doc	Cty
Assessment representative requires 50% of full time for calendar year 2010 (continuous communication with both AHELO and national project staff and ensure highest quality standards and ensure proper implementation of AHELO procedures).	Annex E	All
CAE (or TAT) to reserve conference rooms, sleeping rooms for NYC meeting.	Annex F-NYC mtg	All
CAE (or TAT) to reserve airline tickets for select CAE staff and TAT members.	Annex F-NYC mtg	All
CAE and TAT to create agenda and logistics (travel) information for NYC meeting.	Annex F-NYC mtg	All
Country team members traveling to the United States (for the NYC February 2010 meeting) were required to obtain the appropriate nonimmigrant visa for temporary stay or a visa waiver.	Annex F-Travel Visas	All
Mentions that in April 2009 in Paris AHELO GNE (Group of National Experts) agreed on proposed division of work by several entities: • Secretariat • Contractors • National experts and higher education institutions (HEIs) • GNEs The GNEs at that time also asked for clarification of: • the NPMs role. • Associated costs for countries • Associated costs for HEIs	GNE 19	All
The OECD Secretariat asked bidders for clarifying information regarding costs to countries, costs to HEIs, and the role of NPMs.	GNE 19	All
Country teams were to use information regarding costs (to country and to HEIs) and the role of NPMs to plan for the AHELO project.	GNE 19	All
AHELO GNEs were to comment on the NPM role as tentatively defined in the OECD document (GNE 19).	**GNE 19**	**All**
AHELO GNEs were to take note of estimated implementation resource needs for their country.	GNE 19	All
Country teams were to establish a National Centre to provide appropriate infrastructure for managing key facets of the AHELO Feasibility Study.	**GNE 19**	**All**
When choosing NC site, each country team was to consider financial demands as	GNE	All

Task/Subtask	Doc	Cty
well as the purpose and positioning of the AHELO feasibility study.	19	
Country teams were to select an NC site that would allow each team to engage institutions in a scholarly and quality improvement perspective.	GNE 19	All
Country teams were to choose an NC site that would allow for efficient communication with OECD, government agencies, contractors, and HEIs.	GNE 19	All
Country teams were to staff the NC with people who had nuanced knowledge of the system, effective leadership capacity, --optimally—established relationships with opinion leaders, and sound technical footings.	GNE 19	All
Country teams were to staff the NC with a core staff to include: • NPM • Research assistant • Administrative assistant • Translation/adaptation advisor • Technical advisor • Editorial support	GNE 19	All
Country teams will nominate the NPM.	**GNE 19**	**All**
The NPM will: • be responsible for implementation of AHELO at national level. • Ensure tasks are carried out as per the schedule • Implementation follows technical standards • Implementation follows survey operations guidelines • Documents processes implemented at national level (to be used towards AHELO Feasibility Study's final reports) [13]	GNE 19	All
NPM role in implementation of AHELO at national level includes: • primary contact for OECD Secretariat for day-to-day decisions • primary contact for Contractors for day-to-day decisions	GNE 19	All
NPMs should have several qualifications: • experience in planning, organizing, and conducting large-scale assessment • identify, select, and manage team of project staff • experience with successfully handling multiple tasks simultaneously • have excellent oral and written communication in local language and	GNE 19	All

[13] Will be important to gather this 'documentation' or find out what the documentation was.

Task/Subtask	Doc	Cty
English—be able to represent country at international meetings • helpful if have previous work in the fields of higher education, educational assessment, and contextual surveys. • Helpful if familiar with data processing, survey quality control procedures, and data file structures.		
If appropriate, the same person can serve as NPM and GNE.	GNE 19	All

Task/Subtask	Doc	Cty
Contractors must provide detailed, timely, and accurate information to NPMs, which will guide the NPM with implementation (1 of 2 imp roles): • regular and direct email • telephone and video-based meetings • maintaining an AHELO Feasibility Study website (e.g., with meeting papers and meeting records) • conducting briefing and training meetings (NPMS must give info to contractors so that contractors know how to help)	GNE 19	All
Second major NPM role is providing a channel through which national interests are represented in the implementation of the study through: • surveys and review documents • online discussion forum • meeting sessions	GNE 19	All
Additional information about NPM responsibilities and timelines would be provided after contracts were signed with other entities.	GNE 19	All
Contractors will examine and if necessary revise technical standards for the AHELO study.	GNE 19	All
NPMs would also assist with context instruments (contextual survey): • provide background materials to inform conceptual and practical designs • consult w/ institutions for context and valid indicators • supply and review items and instruments • assist w/ qual and quant validation activities • review final instrument prior to implementation • help w/ collecting and verification of existing system data • contribute to data entry, verification, and coding • review results and reports	GNE 19	All
National committee for each country may include expertise in survey research, education management, relevant gov't agencies, teachers' associations, relevant university departments.	GNE 19	All
National committee should offer advice to the project and ensure national views are represented.	GNE 19	All
National committee should be able to provide input and feedback through regularly scheduled meetings (or some other alternative).	**GNE 19**	**All**
National committee will review progress, procedures, and results throughout the project.	GNE 19	All

Task/Subtask	Doc	Cty
National committee could help raise profile of AHELO and gain cooperation of HEI.	GNE 19	All
Country teams (National Centre or GNE) will coordinate the collection of national examples of country assessments. • id appropriate national experts to contribute to the review process. • Collate review of assessment materials and contextual questionnaires • Id national experts who can contribute to the review process • Collate review response and communication w/ contractors	GNE 19	All
Contractors should schedule and coordinate meetings with NPMs.	GNE 19	All
NPMs will participate in five meetings throughout the study: 1. to become familiar and discuss project, assessment framework, sample items, and give national presentations. 2. discuss fieldwork procedures, sampling, national reports, data management and analysis systems. 3. Update on implementation updates, coder training, report on fieldwork. 4. Review fieldwork implementation and outcomes, review data and initial analysis, consider practical and scientific feasibility 5. Review and debrief on findings, results, and outcomes.	GNE 19	All
NPM must fill at least 1/5 to 2 full time positions by: • working full-time or • hiring support staff such as administrative assistant or data manager who the NPM can oversee (most cost effective).	GNE 19	All
NPM is responsible for communication and reporting: • country's official position on the project to contractors and at NPM meetings • with other international committees as needed • by preparing reports on study's preparation and implementation • using the AHELO website • with local media to promote the study • review technical reports and draft of the final report.	GNE 19	All
NPM is responsible for test development • review all material for accuracy and relevance and cultural appropriateness • recruit assessment/subject specialist to review instruments and surveys	GNE 19	All
NPM is responsible for carrying out translation and adaptation:	GNE 19	All

Task/Subtask	Doc	Cty
• monitor and coordinate translation and adaptation of instruments, training materials, and administration materials as per the guidelines provided • document proposed changes to instruments for verification • communicate with contractors about any translation and adaptation issues		
NPM is responsible for sampling activities based on provided guidelines: • monitor sample design and selection process • recruit HEIs based on convenience sample • review student sampling process, selection of faculty, selection of programme, and selection of institutional leadership staff.	GNE 19	All
NPM must follow survey operations: • oversee production, dispatch, and receipts of materials to HEIs • ensure Institutional Coordinators (ICs) are properly trained • recruit scorers based on guidelines	GNE 19	All
NPM is responsible for following data file preparation activities • communicate data entry procedures • supervise data entry • validate data • organize the dispatch data files to contractors • respond to questions in timely fashion	GNE 19	All
NPM is responsible for ensuring quality control throughout survey implementation.	GNE 19	All

Task/Subtask	Doc	Cty

NPM must perform tasks for computer-assisted delivery:
• may choose to appoint an information technology coordinator to handle several activities
  • ensure access to enough computers that meet certain technical requirements
  • configure computers w/ software
  • train ICs on how to use computers and software
  • train ICs on assessment instruments
  • operate national helpdesk to help ICs w/ any technical issues
  • extract data from computers
  • manage formatting and cleaning of data for delivery to contractors
	GNE 19	All

NPMs had to be available from Jan 2010 through Dec 2011.	GNE 19	All
NPMS are to fill-out Table 2-National Project Manager Major Tasks as they proceed in the project.	GNE 19	All
Country teams will need staff: • NPM: 100-400 days • Research assistant: 100-400 days • Administrative assistant: 80-150 days • Translation/adaptation advisor: 0-30 days • Technical advisor: 10-30 days • Editorial support: 10-20 days	GNE 19	All
Country teams will need to cover costs (EUR): • Low: 2010=200K; 2011=200K; total: 400K • High: 2010=450K; 2011=450K; total: 900K	GNE 19	All

Country teams will need to determine:
• local staff rates
  • government staff
  • institution staff as test administrators
  • experts that take part
• translation; printing costs
• availability of premises or equipment needed
	GNE 19	All

Task/Subtask	Doc	Cty
• if they will participate in the contextual dimension • if HEIs need incentive		
2/19-3/12/2010: Country NPM/GNE and assessment experts begin recruiting" • translation team • translation review team	GS.1 Workplan	All
Country teams were to receive updated timeline (initial implementation expectations) from CAE by 7/28/10.	GS.1 Timeline	All
Country teams were to receive platform instructions for translation and addendums (GS.36/.37) from CAE by 7/28/10.	GS.1 Timeline	All
Countries were to send rough—or reconciled—translations of two PTs to CAE 8/15/10	GS.1 Timeline	All
Countries were to have translations of cognitive labs materials to CAE 8/15/10	GS.1 Timeline	All
Countries were to conduct cognitive labs and revise two PTs based on results between 8/15/10-10/1/10	GS.1 Timeline	All
Countries were to participate in tele or video conference call with CAE 9/27-9/29/11 to: • reflect on trans & adap translation process • discuss implementation phase expectations • ask any final questions • transitioning to CAE implementation team (from CAE adaptation team)	GS.1 Timeline	All
Countries were to make final revisions to two PTs and instructions to CAE 10/1/-10/15/10 (based on teleconference and cog labs)	GS.1 Timeline	All
Countries were to complete final translations of performance tasks, scoring rubric, and instructions sent to CAE for review by 10/18/10	GS.1 Timeline	All
Final translations of mini PT (GS.38) and scoring handbook charts/response features (GS.39) sent to CAE for review by 12/1/10.	GS.1 Timeline	All
NPMs work with campus teams for each HEI.	GS.3 Campus	All
Campus teams: • should choose primary and secondary campus contact persons	GS.3 Campus	All

Task/Subtask	Doc	Cty
• maintain regular contact with NPMs • submit required materials and forms by deadlines • provide dates for when students will begin testing • provide dates for when students will end testing • secure physical location(s) for computers for testing • organize proctors and have them view the training video • set-up all testing sessions • select and recruit student sample (n=200 seniors/HEI, but oversampling) • report testing irregularities to NPM		
Country teams are to become familiar with three classifications of performance tasks by nature of the task.	GS.5 Crit Selection	All
Country teams are to use five criteria for reducing the number of PTs from nine to four to two.	GS.5 Crit Selection	All
Country teams can add PT reducing criteria in NYC Feb 2010 meeting	GS.5 Crit Selection	All
At NYC meeting country teams will: • become more familiar with PTs • select final two PTs • get an overview of the CLA proctor interface • get an overview of CLA PTs online interface • learn about security issues • learn about logistical issues	GS.5 Draft Mtg	All
Each country team was to prepare and present information about the structure of the higher education system.	GS.16	A
Each country team was to prepare and present information about the structure of the higher education system.	GS.17	B
Each country team was to prepare and present information about the structure of the higher education system.	GS.18	C
Each country team was to prepare and present information about the structure of the higher education system.	GS.19	D

Task/Subtask	Doc	Cty
Each country team was to prepare and present information about the structure of the higher education system.	GS.20	E
Each country team was to prepare and present information about the structure of the higher education system.	GS.21	US
Country teams were to become familiar with issues, designs, and technical guidelines for test translation and adaptation (Hambleton, 2005) • carefully choose test administrators • use appropriate item formats • control for speed effect • translators should be familiar with target group, their culture, test content, have some training in test development, and are most capable in test adaptation. • Choose judgmental designs appropriately • Choose appropriate data collection designs • Choose statistical analysis appropriately (differing curricula, cultural backgrounds, levels of motivation, socio-political factors) • Use appropriate ITC Guidelines for Test Adaptation	GS.7	All
Country teams were to become familiar and use applicable features of universal design (Thomson, Johnston, & Thurlow, 2002): • Design instruments so allow participation of widest range of students (and flexible enough to allow for changing student populations) • Precisely defined constructs • Accessible non-biased items • Amendable accommodations • Simple, clear, and intuitive instructions—and procedures • Maximum readability • Maximum legibility • Careful use of results	GS.10	All
Country teams were to become familiar with ITC guidelines (2005): • Consider the technological issues (CBT) and internet • Consider quality issues in CBT and internet testing • Provide appropriate level of control over CBT and Internet testing • Provide appropriate for security and safeguarding privacy	GS.8	All
Country teams were to become familiar with TTTE (Solano-Flores, 2008)	GS.9	All

Task/Subtask	Doc	Cty
• disconfirming evidence		
• multidimensionality of language		
• multidisciplinary-team approach		
• tension among error dimensions		
• attention to language usage, culture, and local curriculum		

Task/Subtask	Doc	Cty
Country teams were to understand: • performance task have three parts: task, response format, scoring system • CLA PT: holistic, complex, real-world task, students write a recommendation or decision, reach a conclusion or solve a problem supported with facts and evidence • CLA measures analytic reasoning and evaluation, problem solving, writing persuasiveness, and writing mechanics	GS.12	All
Country teams were to become familiar with and understand the translation process: • two independent translations • reconciliation of first two translations • conduct cognitive labs • make changes to PTs based on cognitive labs • conduct translation review • create final version of PTs based on AHELO • translation verification by OECD – designated agency • create final version of PTs	GS.14	All
Country teams to turn in translations of each PT.	GS.33	E
Country teams to turn in translations of each PT. (CF)	GS.33b	D
Country teams to turn in translations of each PT with comments from reconciliation process. (CF)	GS.33b leng	D
Country teams to turn in translations of each PT. (LTR)	GS.34b	D
Country teams to turn in translations of each PT with comments from reconciliation process. (LTR)	GS.34b leng	D
Country teams to turn in translations of each PT. (CF)	GS.68 Ar	C
Country teams to turn in translations of each PT with comments from reconciliation process. (CF)	GS.68 Fin	A
Country teams to turn in translations of each PT with comments from reconciliation process. (CF)	GS.68 Nor	E
Country teams to turn in translations of each PT with comments from reconciliation process. (LTR)	GS.69 Spa	D
Country teams to turn in translations of each PT with comments from reconciliation process. (LTR)	GS.69 Ara	C
Country teams to turn in translations of each PT with comments from reconciliation process. (LTR)	GS.69 Fin	A
Country teams to turn in translations of each PT with comments from reconciliation process. (LTR)	GS.69 Nor	E
Country teams were to examine and correct discursive structure.	GS.13	All
Country teams had to compensate for ideas and idiomatic expressions that cannot be	GS.13	All

Task/Subtask	Doc	Cty
translated.		
Country teams were to keep in mind the dialect or dialects used by students	GS.13	All
Country teams were to keep in mind using appropriate register.	GS.13	All
Country teams need to know if students are used to proposing, challenging, or critiquing ideas, persons or institutions.	GS.13	All
Country teams needed to examine appropriateness of context and procedures.	GS.13	All
Country teams needed to ensure cognitive and linguistic equivalence.	GS.13	All
In step 1 of the translation process two translators independently translate material according to OECD translation guides.	GS.13	All
Country teams were to look for translators with the following "indispensible" qualifications: • English to natl language translation certificate from prof trans org. • Native speakers of the nat'l language • Ample experience as translators	GS.13	All
Country teams were to look for translators with the following "desirable" qualifications: • Experience with test translation • Experience with educational material • Experience with higher education documents	GS.13	All
Translators were to translate the three PT components (task, response format, and scoring rubric) simultaneously.	**GS.13**	**All**
Translation reconciliation was to merge the two independent translations (done by translation/adaptation advisor and translators).	GS.13	All
Translators were to reconcile the three PT components (task, response format, and scoring rubric) simultaneously.	GS.13	All
An OECD designated agency was to verify the reconciled translation of all material against source versions to assure quality control.	GS.13	All
Country teams were to conduct cognitive labs—have student talk as they work through the tasks and then be interviewed—to examine if the task elicits targeted reasoning.	**GS.13**	**All**
Country teams, through cognitive labs, are to gather information about: • cognitive processes used (analytic reasoning, problem solving, communication) • self-reported strategies for approaching task • readability	GS.13	All

Task/Subtask	Doc	Cty
• comprehensibility • meaningfulness		
Country teams are to conduct translation reviews to: • **examine disconfirming** • **focus on error, not appropriateness—assume that translation error is inevitable** • **examine tension among error dimensions** • **pay specific attention to language usage, culture, and local curriculum**	**GS.13**	**All**
Translation review team is to look for different types of errors: • omission • insertion • alteration • inconsistency • inappropriateness/imprecision • combination/conflation • substitution • multiplicity	GS.13	All
Translation review team is to look for layout dimension errors: • style • format • conventions	GS.13	All
Translation review team is to look for language dimension errors: • grammar and syntax • semantics • register	GS.13	All
Translation review team is to look for content dimension errors: • information • construct • curriculum and culture • origin	GS.13	All
Country team must examine translation quality by applying probabilistic space taking	GS.13	All

Task/Subtask	Doc	Cty
into account • translation error frequency • translation error severity		
PTs should go through: • task adaptation • task translation • translation reconciliation • translation verification • try-aloud and talk-alouds • translation review	GS.13	All
Country teams with ACER were working on: • Module A: Generic Skills strand (AHELO) • Module B: Economics strand • Module C: Engineering strand • Module D: Contextual dimension surveys	GS.22	All
Country teams must write and submit: • **Technical reports** for each strand they are implementing (GS; Econ; Eng; Cont surv; Proj mgmt, survey ops, and analyses of results) • final report (based on all technical reports)	GS.22	All
Country team representatives were to attend the Generic Strands meeting in Paris (OECD) on March 17, 2010	GS.23	All
Country teams must have hired and trained scorers close to the date when they will be doing actual scoring.	GS.23	All
It was recommended that country teams have scorers, or at least the lead scorer, take part in translation and piloting. This is particularly helpful to get them familiar with the rubric.	GS.23	All
Lead scorer can help: • evaluate recommendations for rubric translation • observe pilot studies • review some CAE previously scored responses. • Provide insight during scorer training	GS.23	All
Country teams were to get a maximum of 10 HEIs with 200 students at each.	GS.24	All
Based on 10 HEIs and 200:	GS.24	All

Task/Subtask	Doc	Cty
• 100 students would take each PT=1000 PTs • each PT gets double scoring. • Each scoring takes 10 minutes • PT1 will need 333 hours of scoring (X2 or PT2=666 total scoring hours)		
Each country team needs to hire 10 scorers (5/each PT) for training and recalibration—country teams pay	GS.24	All
Each scorer will have two days of training (16 hrsX10=160 hours)—country teams pay	GS.24	All
Each scorer will have 6 hours for recalibration (6X10=60)—country teams pay	GS.24	All
Country teams will pay for approximately 886 hours for scorers (11 days—8 hrs—per scorer)	GS.24	All
Country team member(s) participated in OECD, Paris AHELO Module A meeting from 9:00am-12:30pm.	GS.25	All
Country teams need to determine the best way to recruit students for participation.	GS.29	All
Some countries may have the added expense of financially compensating students to participate (money or gift).	GS.29	All
Country teams may have to acquire the assistance of HEI personnel to send emails, put up flyers, or make phone class.	GS.29	All
Country teams also need to recruit HEI faculty and staff.	GS.29	All
All country teams must agree on the questions to be included in the survey.	GS.29	All
Country teams must train proctors: • review the proctor guide • watch the online training video • be familiar with technical assistance • be familiar with PTs	GS.29	All
Proctors should be trained on how to use: • the proctor guide • proctor checklist • CLA configuration requirements • Irregularity report forms	GS.29	All
Country teams must set-up computers in concentrated area with high speed internet access.	GS.29	All
Country teams must ensure that each student has enough space to work on computer w/out distractions	GS.29	All

Task/Subtask	Doc	Cty
Country teams should provide a telephone in the testing area.	GS.29	All
Country teams must work with HEI IT dept to make sure technology is taken care of.	GS.29	All
Country teams must ensure each proctor has register as a user.	GS.29	All
Country teams must create own consent forms that conform to HEI's Institutional Research office.	GS.29	All
Each country team must have a CLA rep for HEI proctors.	GS.29	All
Country teams will review the translation (done by CAE contracted company) of the following between 11/1-3/2011: • IT/Computer Platform • Interface language • Survey questions • Administrator manual	GS.31	All
5/10-10/18/2010 country teams will conduct full translation of both PTs	GS.31	All
5/15-12/1/2010 country teams will conduct full translation of scoring rubric	GS.31	All
7/10-12/1/2010 country teams will conduct dual translation of cognitive lab	GS.31	All
5/20-12/1/2010 country teams will conduct dual translation of mini PT for tuning	GS.31	All
7/28-10/18/2010 country teams will conduct dual translation of internet interface instructions	GS.31	All
Country team members will participate in a two-day CAE representative site visit in June, July, or August 2010: • NPM • Assessment expert • Translators • Translation review team • Anyone else NPM thinks should be there (e.g., linguist, lead scorer)	GS.32	All
Country teams needed to collect 10 items in the country's language from PISA or TIMSS	GS.32	All
Country teams, with the help of the CAE team, will also collect the English versions of the 10 PISA or TIMSS items	GS.32	All
Country teams must have a minimum of 3 pages of each CAE PT translated for the CAE rep visit meeting	GS.32	All
During the CAE rep visit meeting country team members will: • get trained on the translation procedure • get trained on the translation review procedure	GS.32	All

Task/Subtask	Doc	Cty
• practice translation review with PISA and TIMSS items • discuss translation and translation review procedures • discuss the cognitive lab study		
Using 'GS.33b Catfish Performance Task_Template_CONFIDENTIAL.doc' file, translation review teams benchmark specific sections.	GS.33b	All
Country teams were to submit suggestions for adaptations for each PT	PK	All
Country teams, using drop down menus, were to select their choice of adaptation possibilities for each PT. Needed adobe acrobat to complete this task: • people names • people titles/surnames • location names • government positions • measurement units • proper nouns • financial units and cost • maps • graphs • keys	GS.33a	All
Using GS.34b Lake to River Performance Task_Template_CONFIDENTIAL translation review teams benchmark specific sections.	GS.34b	All
Country teams were to ensure construct equivalence across all versions of the PTs.	GS.36	All
Country teams were to ensure that level of difficulty was maintained across all versions of the PTs.	GS.36	All
Translations should read as if they were originally written in the country's language.	GS.36	All
Country teams can use the American Translators Association's (ATA) database or equivalent to find certified translators.	GS.36	All
Country teams should get translators who have lived within the country for a good part of their lives	GS.36	All
Country teams should get translators who attended school in the country.	GS.36	All
Country teams should get translators who are familiar with the local university/college system.	GS.36	All
Country teams should get translators who are familiar with dialects from different parts of the country.	GS.36	All
Country teams should get translators with experience translating:	GS.36	All

Task/Subtask	Doc	Cty
• tests or survey instruments • documents used in higher education • news and articles		
The full translation procedure was to be used with all primary documents: • documents give to students (PTs, scoring rubric, IT/Computer platform and interface language) • documents used to score performance	GS.36	All
Country teams need to hire three qualified translators.	**GS.36**	**All**
Translation/adaptation advisor works with 2 translators on reconciliation of the two independent translations.	GS.36	All
A third translator, the translation/adaptation advisor, and an assessment expert review translation after changes from cog labs for any modifications that may be needed: • same constructs • student interpretation is what was intended • consistent level of difficulty across languages	GS.36	All
Country teams must know what is a secondary document that used dual translation procedure: • mini-PT for tuning purposes • cog labs materials • scoring handbook charts • task administrator manual • scorer training materials (incl benchmark CLA Engl responses for calibration purposes)	GS.36	All
Country teams and translators must know the dual translation procedure: • two translators independently translate documents • the two translators and translation/adaptation advisor reconcile the two independent translations	GS.36	All
For the translation review process each country team needs to hire: • translation/adaptation advisor (acts as translation review leader) • assessment expert dedicated to AHELO • once certified professional translator (should not be one of the original translators)=independent translator 3	GS.36	All
Country teams should include additional personnel in the translation review team if possible:	GS.36	All

Task/Subtask	Doc	Cty
• lead scorer • linguist (prefer sociolinguist) • specialist in the country's language and literature NOTE: if budget was a problem CAE suggested getting help from a graduate student instead of a linguist and specialist.		
Translation review should be based on TTTE	GS.36	All
According to TTTE, a translation error can belong to multiple dimensions.	GS.36	All
Translation reviewers should be familiar with: • conceptual framework (GS.4) • scoring rubric (GS.35) • scoring handbook charts (GS. 39)	GS.36	All
Country teams must provide several documents: • hard copies of the translated version of the documents • electronic version of original to be projected • paragraph numbering for benchmarking • error coding forms • a master database for capturing agreed upon errors	GS.36	All
Translation reviewers will make comments and edits directly on the hard copies of the translated version of each document.	GS.36	All
Translation review leader is responsible for projecting the original after translation reviewers comment on and edit the translation.	GS.36	All
Translation team look for all dimension errors but each member focuses on the dimension(s) most closely related to their area of expertise.	GS.36	All
Using both language versions each reviewer records the errors they id on the coding form.	GS.36	All
After each reviewer has recorded errors, the entire translation review team discuss their findings.	GS.36	All
The translation review leader facilitates discussion when sharing coding form information until consensus is reached.	GS.36	All
Based on translation review team consensus, each reviewer updates his/her coding form to reflect decisions made by team.	GS.36	All
The translation review leader edits his/her copy of the translation based on review team consensus of errors.	GS.36	All
The translation review leader is responsible for entering errors identified by consensus into a master database.	GS.36	All

Task/Subtask	Doc	Cty
The translation review leader must collect all coding forms and hard copies of AHELO documents.	GS.36	All
Country teams were to review "GS.36 and GS.37 addendum" which includes 4 updates made to the Translation Guide and 1 update to the Cognitive Labs Guide.	GS.36/ GS.37 Addendum	All
GS.36 and GS.37 addendum—changes to the Translation Guide and Cog Labs Guide—was based on discussion that took place among CAE team members and during the 2010 CAE country site visits.	GS.36/ GS.37 Addendum	All
Update 1 for GS.36=Country teams were to read and implement new deadlines emailed on June 25, 2010.	GS.36/ GS.37 Addendum	All
Country teams received additional time (in Oct 2010) to finalize their PT translations.	GS.36/ GS.37 Addendum	All
Update 2 for GS.36=CAE updated the review process for the translated language going onto the internet testing platform.	GS.36/ GS.37 Addendum	All
Update 3 for GS.36=additional information is provided for the reason why the translation team needed to pay particular attention needed to be paid to Register.	GS.36/ GS.37 Addendum	All
Update 4 for GS.36=translation and translation review teams were to use each document as an analytical unit—not a sentence or paragraph.	GS.36/ GS.37 Addendum	All
Update 1 for GS.1=interviewer is not to model how to think aloud.	GS.36/ GS.37 Addendum	All
Documents for internet platform will undergo a specific translation process: • dual translation and reconciliation • send translation to CAE to be uploaded onto internet platform • NPM and team conduct a quality assurance review of internet platform • Changes will be made based on results from review • Country teams will ensure that HEIs conduct a small mini-pilot (~4-5 students and 1-3 HEI personnel) working through all elements of the internet platform (language, navigation, visuals, etc.) • Provide CAE with feedback after the mini pilot.	GS.36/ GS.37 Addendum	All

Task/Subtask	Doc	Cty
•		
NPM/team quality assurance review of internet platform with translation and reconciliation process includes examining language and the platform. Results from this review will determine modifications made.	GS.36/ GS.37 Addendum	All
Translation and translation review teams should examine each document in its entirety. (e.g., coding forms for translation review should be filled out looking at entire document first)	GS.36/ GS.37 Addendum	All
Interviewers should use the updated Appendix I the script for cog labs.	GS.36/ GS.37 Addendum	All
Project staff from each country will conduct the AHELO cog labs.	**GS.37**	**All**
Country teams must recruit volunteer students to participate in cog labs.	GS.37	All
Country teams must provide volunteer students participating in cog labs with paper copies of one of the PTs.	GS.37	All
Country teams must be trained to conduct the cog labs.	GS.37	All
Project staff will conduct cog labs, which have three stages: • training—telling students what to expect during the think aloud and training student to think aloud with small tasks • talk-aloud—students share what they are thinking; interviewer will take notes • follow-up interview—interviewer asks specific questions based on observations and general questions	GS.37	All
Country teams must provide cog lab interviewer with recording device.	GS.37	All
Cog lab interviewer will audio record students.	GS.37	All
Country teams must acquire student consent for audio recording cog labs.	GS.37	All
Project staff will listen to audio recordings and take additional notes on unintended challenges caused by translation.	GS.37	All
Project staff will identify ways in which translation of the tasks need to be improved.	GS.37	All
Country teams must recruit a large representative group of students to choose from for the cog labs. 6-10 students should participate for each PT.	GS.37	All
Country teams must budget time and money for cog labs.	GS.37	All
Country teams should take into account several factors when selecting a representative sample of students for the cog labs: • gender • socio-economic status	GS.37	All

Task/Subtask	Doc	Cty
• geographic areas		
• HEI types		
Interviewer (and anyone involved) should be familiar with both PTs.	GS.37	All
Country team should provide interviewer with:	GS.37	All
• talk aloud script (provided by CAE)		
• enough copies of the PTs		
• enough copies of follow-up questions form		
• recommended questions (provided by CAE)		
• labels for the recordings that will match the written notes		
• pens, pencils, scrap paper		
Project staff working on cog labs should practice completing the PTs	GS.37	All
• will help them properly interpret student verbalizations		
• help interpret student responses to follow-up questions		
• help id ways PTs can be improved		
Project staff need to setup the room for cog labs	GS.37	All
• no distractions		
• be private		
• maintained at room temperature		
• comfortable and working chair and table		
• appropriate lighting		
• working recording equipment with batteries		
• materials organized on the table		
Project staff should allow 90 minutes per student.	GS.37	All
During cog labs interviewer should:	GS.37	All
• put student at ease (use appropriate behavior for gender and age)		
• introduce himself and organization (do not set-up power dynamic)		
• explain the process (training, think aloud, follow-up questions)		
• state goal (not anything to do w/ transl) of finding out about student understanding of the task		
• give them opp to ask questions		
• give students opp to practice a think aloud (follow the full process)		

Task/Subtask	Doc	Cty
• give students actual PT (interviewer takes notes all along and reminds student to share thoughts if the student falls silent—and do not say anything else)		
• ask follow-up questions (use form to take notes)		
• ask general questions (use form to take notes)		
• collect all materials		
Project staff must become familiar with the cog lab script.	GS.37	All
Project staff must analyze cog lab results:	GS.37	All
• listen to audio recordings		
• review notes from follow-up questions form		
• review notes from general questions form		
Cog lab analysis includes:	GS.37	All
• having discussion between at least two people		
• identifying errors students		
• identifying struggles		
Project staff should know and determine if errors or struggles were due to:	GS.37	All
• translation altered the constructs		
• students interpreting translated task in ways not originally intended		
• level of difficulty changed		
• the language being different from PT that would have been created in country language		
Project staff must revise translations based on cog labs results:	GS.37	All
• keep copy of original task in English		
• keep copy of translated version of the PT (make notes on this copy)		
• use the 'modifications form' (provided by CAE) to note the exact location of issue and problem		
• use the 'modifications form' to note change being suggested and justification for it		
Country team must make appropriate changes.	GS.37	All
Country teams must translate and become familiar with mini PT.	GS.38	All
Country teams should become familiar with common response features for each PT.	GS.39	All
Country teams should become familiar with suggested motivation questions for the contextual survey.	GS.40	All

Task/Subtask	Doc	Cty
Country teams should review the U.S. CLA Internet Platform Tutorial (Confidential), which consists of seven screens	GS.41	All
Country teams should review the Internet Platform Instructions (CONFIDENTIAL)	GS.42	All
CAE will hire a translation company to translate remaining English text—associated with internet platform: • login screen • demographic profile form • welcome page text • orientation text • local survey • exit screens • other miscellaneous text	GS.43	All
Country teams will be responsible for reviewing translations completed by CAE hired translation company.	GS.43	All
Country teams will need to save images as .gif files and send to CAE electronically separately—for uploading to internet platform.	GS.43	All
Dec/Jan: Country teams receive ITS Word docs and .gif files from CAE for review and translations.	GS.43	All
Country teams must enter translated text into PT templates for the internet platform by third week of Dec.	GS.43	All
Country teams send CAE (by third week of Dec) pdf of: • PTs • Coding forms • Mini-PT • Scoring handbook charts • Interface instructions	GS.43	All
Country teams need to be able to use .CSV excel templates. They will place translated text into this format for internet platform upload.	GS.43	All
Country teams will be responsible for reconciling translations completed by CAE hired translation company.	GS.43	All
CAE will make changes to translations of ITS text based on country feedback.	GS.43	All
Country teams review the second iteration of ITS text—containing changes CAE made according to country team feedback—and give CAE any additional feedback.	GS.43	All
CAE uploads all final text onto internet platform.	GS.43	All

Task/Subtask	Doc	Cty
Dec 2010/Jan 2011 country teams will receive from CAE: • the proctor administration manual • the proctor training video	GS.44	All
CAE will be responsible for translation of proctor administration manual and proctor training video	GS.44	All
Country teams will be responsible for reviewing, reconciling, and making edits directly to the translated proctor administration manual documents: • pdf of initial translation • pdf of an adapted English version	GS.44	All
Country teams will be responsible for reviewing, reconciling, and making edits (using track changes feature in Word) directly to the translated pop-up text for the proctor training video: • .doc of initial translation • .doc of an English version	GS.44	All
CAE will make changes to the translated proctor administration manual documents and proctor training video documents to country teams for review.	GS.44	All
CAE had to collect findings and suggested modifications based on cog labs results from all countries and provide a collated set of modifications.	GS.44	All
Country teams confirm translated text for proctor administration manual documents and proctor training video documents—send any final changes back to CAE.	GS.44	All
Country teams needed to note updates in deadlines for AHELO activities (Oct 2010-May 2011) • Oct-Nov: finalize reconciled translation of PTs and rubrics • Oct-Nov: finalize dual translation of cog lab materials, mini PT, scoring handbook charts, Internet interface instructions • Oct-Nov: complete cog labs with at least 5 students per PT • Dec 2010: 1st week report to CAE cog labs findings including suggested modifications • Dec 2010: 2nd week modify PTs based on CAEs collated recommendations • Dec 2010: 1st week send CAE pdfs of final PTs, final coding forms, mini PT, scoring handbook charts, interface instructions • Jan/Feb 2011: country teams have 3 wks to enter and return to CAE completed docs with PT translations and revisions to initial translation of other text • Feb/March 2011: CAE will make changes to initial translations. And, country teams will have two weeks to confirm and return to CAE revised	GS. 45	All

Task/Subtask	Doc	Cty
translations. • Feb/March 2011: CAE will then send revised materials to NPM for review. • Feb/March 2011: country team will have 3 weeks to review and return materials • March/April 2011: CAE will make changes based on country team reviews • March/April 2011: country teams will have two weeks to confirm and return material to CAE • March/April 2011: country teams will conduct User Acceptance Testing (UAT) of internet platform…to be concluded by April/May		
June 2011-2012 countries work on final phase of AHELO: Implementation: • Recruit HEIs • Coordinate testing dates and sessions with HEIs • Train proctors (test administrators) with CAE help • Manage scoring process with CAE help • Collaborate with ACER, CAE, and others in planning for data collection • Participate in relevant OECD/ACER/CAE meetings • Conduct country specific data analyses—additional to what will be explored in AHELO	GS. 45	All
During recruitment of HEIs during final phase country teams will: • id a primary proctor (Institutional Coordinator) • support HEIs in student recruitment	GS. 45	All
During managing scoring process with CAE help country teams will: • id a chief reader (lead scorer) • recruit scorers • train scorers	GS. 45	All
Country teams were to encourage strongly, or require, all participating students to visit the website and complete the mini-PT.	GS.47[14]	All
During every review, CAE and TAT had to keep in mind comparability between the US PT and AHELO versions—allow for comparison with CAE's large data set of US HEIs.	GS.47	All
CAE must provide a sample response for the miniPT	GS.47	All

[14] GS.47: has overall notes from all countries as well as some information labeled specifically with one country…does this mean all countries voiced a concern unless labeled otherwise?

Task/Subtask	Doc	Cty
Country teams will translate sample response for the mini PT	GS.47	All
CAE provided a summary of changes made to PTs based on country feedback from cog labs.	GS.47	All
Country teams were to review the final CAE approved changes allowed to be made to PTs (and mini PT) based on cog labs.	GS.47	All
Country teams were to translate and integrate final CAE approved changes based on cog lab results.	GS.47	All
Once country teams translated and integrated final CAE approved changes based on cog lab results, they were to send pdfs of the PTs to cae.ahelo@gmail.com.	GS.47	All
Country teams must give correct permissions for the three types of platform uses: • student • proctor/administrator • scoring	GS.47	All
Country teams were to follow specific directions for reviewing and entering each document in each of the three platforms.	GS.48	All
GS.48 instructed country teams to cross-reference other documents for instructions on how to review and enter each of the different documents.	GS.48	All
There is a specific order in which PTs were to be uploaded onto the internet platform: • instructions and questions for PT1 • instructions and questions for PT2	GS.48	All
Country teams were to review and revise titles for documents that would be seen by students in a drop down menu that they would use to navigate through the documents.	GS.48	All
When typing onto the internet platform country teams were to keep exact formatting.	GS.48	All
To save a .gif onto the internet platform country teams were to access an instructional video online and follow the directions.	GS.48	All
Country teams will have to copy text from an Excel file to upload it onto the internet platform.	GS.48	All
CAE contracted a company to conduct external validation of the translations (final pdfs of PTs). This was not in the original plan.	GS.67	All
External translation validation included making sure that no inadvertent omission had been done resulting in issues with underlying constructs.	GS.67	All
The external translation company CAE hired to conduct external validation of the final translations were familiar with the PTs because they were working on platform text translation.	GS.67	All
EVENT: The external translation company found “minor issues” and provided	GS.67	All

Task/Subtask	Doc	Cty
suggestions for rewording.		
CAE sent suggested revisions provided by external translation company to respective country teams.	GS.67	All
Country teams were to review the revisions suggested by the external translation company.	GS.67	All
If country teams wanted to accept the changes made by the external translation company they were to follow specific steps via the internet platform.	GS.67	All
Ph1 of user acceptance: NPMs and other staff were to access online PTs by copying and pasting a link into a browser and enter a password.	GS.67	All
Ph1 of user acceptance: Country teams were to follow guidelines to review and accept the online versions of the PTs.	GS.67	All
User acceptance had two phases.	GS.67	All
Ph2 of user acceptance testing had two goals: • review all platform interfaces • review supplemental reference materials for AHELO	GS. 70	All
During Ph2 of user acceptance testing involved proctor, student, and scorer interfaces—each with its own URL and country-specific password provided by CAE.	GS. 70	All
During Ph2 country teams were to use a feedback log to note items or issues that require change.	GS. 70	All
The AHELO study consists of two main parts: • PT translation and adaptation • Field study that includes data collection, analysis, and reporting	GS.4	All
Translators were to translate document library for PTs: • letters • memos • summaries of research reports • newspaper articles • maps • charts • photographs • diagrams • tables • interview notes or transcripts	GS.4	All
CAE and OECD were to ascertain generalizability of the Generic Strand given	**GS.4**	**All**

Task/Subtask	Doc	Cty
that the CLA was created for US students and is to be used internationally		
Country teams must ensure that the tasks are meaningful in their countries by examining: • universality of PT theme • ease of translation based on PT's language complexity • ease of scoring	GS.4	All
CAE provided a useful reference for participating countries to use during the adaptation and translation of the PTs used in the AHELO study.	GS.4	All
CAE provided a reference to choosing PTs for adaptation.	GS.4	All
Country teams were to become familiar with PTs.	GS.4	All
Country teams were to become familiar with the PT scoring rubric.	GS.4	All
Country teams had a list of four key documents that informed the translation and adaptation procedures' conceptual framework.	GS.4	All
Country teams had to examine issues relevant to the adaptation process of PTs: • cultural differences • linguistic and cultural appropriateness (show evidence) • familiarity with computer-based assessment (show evidence) • context appropriateness (show evidence) • cognitive and linguistic equivalence (show judgmental evidence) • appropriateness of procedures (follow procedures)	GS.4	All
Country teams had to examine issues relevant to the adaptation process of computer administered AHELO tasks: • Technology: preparedness for the technology • Control: (control over comfort of test takers; control over test takers' practice w/ the PT format; generally, minimize cheating) • Security and Privacy: test materials, passwords and usernames, backing up data, collection and storage of personal data	GS.4	All
Country teams had to examine issues relevant to universal design:	GS.4	All
Country teams had to examine issues relevant to the test translation dimensions and error types.	GS.4	All
Country teams were to read and understand the ideas shown in four different pieces of literature contributing to adaption and translation conceptual framework.	GS.7-10	All
A TAT member visited a country and gave a presentation about: • the feasibility study and generic strand	GS.27	A

Task/Subtask	Doc	Cty
• framework and approaches to assessing learning		
• the CLA		
• practicing a PT		
• addressing validity		
• addressing reliability		
• value added		
• grading		
• internet platform		
• task format		
• response format		
• criterion sampling		
A TAT member visited a country and gave a presentation about:	GS.28	E
• the feasibility study and generic strand		
• framework and approaches to assessing learning		
• the CLA		
• practicing a PT		
• addressing validity		
• addressing reliability		
• value added		
• grading		
• internet platform		
• task format		
• response format		
• criterion sampling		
Country teams were encouraged to keep a diary to reflect on what is happening.	GS.26	All
Country teams were to use a collection of frameworks created by ACER:	GS.26	All
• Analysis Plan		
• Assessment Design Plan		
• Reporting Guidelines		
• Sampling Manual		
Per country team requests, CAE and TAT were to examine the adaptation of writing	GS.26	D

Task/Subtask	Doc	Cty
persuasiveness category in the rubric because: • it is not culturally sanctioned across all countries. • Persuasiveness is culturally constructed.		
Country teams were to monitor the potential impact of writing mechanics: • will differ between countries • may impact the amount of time it takes students to complete	GS.26	D; C
Country teams were to recruit students in the middle to end of their last year of college.	GS.26	All
Country teams will use guidance from ACER and CAE on student population sampling.	GS.26	All
Country teams will adhere to OECDs requirements for choosing HEIs.	GS.26	All
Scoring expert should work with a translator to help with scoring rubric adaptation.	GS.26	All
Countries will participate in training for cog labs.	GS.26	All
Country teams were to discuss implementation (e.g., funding) during the GNE meeting in March.	GS.26	All
Country teams were to use the ACER website to gain access to documents.	GS.26	All
Country teams were to copy ACER on any communication they have with CAE or TAT.	GS.26	All
Funding pg 7	GS.26	All
Country teams had to decide on how to handle budget: • continue to solicit assistance with funding • only continue to generic strand • only conduct 2 surveys	GS.26	All
OECD needed to secure funds for contractual commitment with ACER in order to begin work on implementation phase.	GS.26	All
13 country teams opted to participate in different strands: • 6 in generic skills strand • 3 in engineering strand • 4 in economic strand	GS.26	All
Country teams also needed to be aware of implementation logistics costs: • travel • attending meetings • translations, etc.	GS.26	All

Task/Subtask	Doc	Cty
When selecting PTs for the AHELO project country teams were to take into account: • that the PTs did not have to represent day-to-day activities • that the student population is graduating college students • keep in mind the strict instructions students are to follow • that the task/documents not emphasize one country • student motivational implications	GS.26	All
Country teams were vote on the two PTs that would be used for the full AHELO study.	GS.26	All
Country team members were to practice interacting with this type of performance task by responding to a release PT.	GS.26	All
Country teams were to become familiar with Bloom's Taxonomy.	GS.26	All
Country teams were to allow CAE to train all scorers.	GS.26	All
Country teams and CAE were to collect responses in local language to use for training scorers.	GS.26	All
To keep costs down, CAE would train scorers in their own countries.	GS.26	All
Country teams will have a run through for proctors as part of training process.	GS.26	All
CAE TAT advised country teams to have an IT person proctor.	GS.26	All
Country teams were to get participating students familiar with the CLA PT format.	GS.26	All
The contextual survey will capture information regarding student motivation.	GS.26	All
CAE will provide anchor papers that will help with inter-rater reliability	GS.26	All
Country teams were to decide on scoring system all countries will use—keeping in mind cost-effectiveness.	GS.26	All
Country teams are ultimate decision makers on final version following study guidelines.	GS.26	All
Rubrics were to reflect the nonlinear writing structure used in other countries	GS.26	C; D
Countries were to investigate IRB rules and follow them	GS.26	All
CAE had to adapt the US Administrator Manual for the internet platform.	GS.26	All
Countries are to conduct cognitive interviews: • with 10-15 students who vary in abilities • that will require the assessment expert/technical advisor collect response.	GS.26	All
Country teams will examine student responses from cognitive labs to understand if students are interpreting the PTs according to original aims.	GS.26	All
Country team members participating in the cognitive labs must go through training.	GS.26	All
Country team members participating in the cognitive labs must follow CAE TAT	GS.26	All

Task/Subtask	Doc	Cty
provided protocol.		
Country teams were to emphasize to students that participation in cognitive labs will have no consequences—particularly those countries in which students are not accustomed to presenting orally.	GS.26	All
Countries were to adapt names and geographic locations.	GS.26	All
During adaptation: • CAE/TAT will send countries word docs of the PTs. • Using 'track changes' country teams were to work on their PT word documents for adaptation. • By end of March the countries will send adapted PTs to TAT • The TAT will synthesize the recommendations from all 5 country teams into English • The TAT will send the adapted English versions of the PTs to country teams • Country teams will translate the adapted English versions of the PTs	GS.26	All
Adaptation was to begin during the NYC meeting.	GS.26	All
Country teams will gather information on student majors.	GS.26	All
CAE was to define what 'like' institutions are for future comparisons.	GS.26	All
Country teams chose two PTs during the NYC meeting.	GS.26	All
CAE and TAT will provide estimates of time commitments for scorers.	GS.26	All
CAE and TAT will provide information regarding the recruitment of scorers.	GS.26	All
Country teams were to consult the Administrator Manual for incentives on how to recruit students.	GS.26	All
80% of the funds for AHELO are not coming from countries but rather from other foundations.	GS.26	All
20% of funding coming from OECD or participating countries.	GS.26	All
CAE/TAT will provide sequence of events for pilots	GS.26	All
Country teams will discuss adaptation process at March meeting in Paris.	GS.26	All
In April CAE will synthesize countries' feedback on PTs from adaptation and create a single English version.	GS.26	All
In June-July country teams will conduct pilots—10 students per task.	GS.26	All
Selection of students for the pilots should be done in a way so as to: • not overwhelm any one institution • keep PTs safe from replication • have representation of all types of HEIs that will participate in the study	GS.26	All

Task/Subtask	Doc	Cty
Country teams and CAE will discuss PTs revised based on cog lab results.	GS.26	All
Country teams and CAE will discuss if additional cog labs need to be done after discussing the revised PTs.	GS.26	All
All students participating in cog labs will have to sign confidentiality agreements provided by CAE.	GS.26	All
TAT will create a video for training for cog labs that country teams were to use for training.	GS.26	All
Oct-Dec country teams will test out the internet platform in respective languages.	GS.26	All
Oct-Dec country teams will work closely with HEIs to finalize logistics.	GS.26	All
Oct-Dec country teams should begin looking at recruiting scorers.	GS.26	All
Country teams were to get overview of the faculty/classroom use of CLA.	GS.26	All
At the end of Milestone 1 (May 31, 2000), CAE will deliver a report to OECD.	GS.30	All
CAE was to provide OECD with a progress report on generic skills after the NYC meeting in February 2010.	GS.30	All
Country teams had to participate in teleconference calls in September of 2010 to review progress on the project and voice questions.	GS.46	All
The AHELO study was do begin no later than 2/1/2010	A-W-R	All
Country teams will agree to deliver the agreed upon version of the CLA via the Internet in a proctored environment on a platform approved by CAE	A-W-R	All
Countries will create a Country Team: • Country team rep to OECD • 1-2 assessment experts	A-W-R	All
Countries will create a translation team	**A-W-R**	**All**
Countries will create a translation review team.	**A-W-R**	**All**
Countries will identify professionals for the translation team	A-W-R	All
Countries will identify professionals for the translation review team.	A-W-R	All
Country teams will choose one language for translation.	A-W-R	All
Country teams will send two people to participate in two Paris meetings with funding provided by the funding of the project.	A-W-R	All
All team members must have access to software and computer/telecom equipment for conducting work using an agreed-upon software suite, most likely Microsoft Office and related products	A-W-R	All
CAE will adapt the two PTs and scoring rubrics now being used in the US.	A-W-R	All
Countries, with CAE, select Assessment Adaptation Group or AAG.	A-W-R	All

Task/Subtask	Doc	Cty
At NYC meeting AAG would be trained in: • adaptation • translation • translation review	A-W-R	All
Via internet meetings country teams will present suggested adaptations	A-W-R	All
Via internet meetings country teams will reach consensus on adaptations	A-W-R	All
CAE fully modifies two PTs (tasks, scoring rubrics and IT administration procedures) following agreed upon modifications at Internet Meetings	A-W-R	All
The on-line testing instructions will be translated early in the process (because these take longer to adapt onto the test delivery website).	A-W-R	All
In Internet Meetings country teams will: • Report, evaluate, and discuss findings on mini-pilots. • Discuss challenges of rough translations. • Finalize the two performance tasks. • Review, discuss and agree upon testing protocols with the PMSO contractor.	A-W-R	All
CAE and TAT to organize site visits to participating countries.	A-W-R	All
CAE expert in translation and translation review will visit, assist, train and guide Translation Teams and Translation Review Teams.	**A-W-R**	**All**
Technical Advisory Group (TAG) will provide a strong, diverse range of expert input regarding technical and higher education matters	1-2-10	All
The feasibility study was to be completed in 2012.	1-2-10	All
Several CAE support would comprise the team leading Module A: • principal investigator; • financial and legal support; • adaptation and translation; • work coordination; • coordination, technical support; • ask development/adaptation and CLA Faculty Academies; • testing operations consultant; and • task development/adaptation.	1-2-10	All
March 17 Paris meeting: county team representatives met face-to-face, review recommended country adaptations to the two performance tasks and present to Rich.	Prog Rept Mod A[15]	All
Originally, CAE asked for a second face-to-face meeting. • ***EVENT***: • budget didn't allow • telecommunication meeting proposed instead.	Prog Rept Mod A	All

[15] NEED TO GET: Please refer to GNE [2010]2 – Progress Report on Generic S

Prog Rept Mod A	All

kills Strand, presented at the March, 2010 GNE meeting, for a discussion of phases one and two

Task/Subtask	Doc	Cty
• All were attending Paris meeting so they took advantage of it.		
26 March country teams were to send CAE specific modifications in the form of track changes on each performance task document.	Prog Rept Mod A	All
CAE staff was to integrate track changes suggested modification sent by all countries and benchmark these recommendations against the original performance task(s) and between country recommendations.	Prog Rept Mod A	All
CAE fully modifies two PTs (tasks, scoring rubrics and IT instructions) following modifications agreed upon at telecom meetings and sends draft PTs for review, comment, and revision as they are completed.	Prog Rept Mod A	All
CAE contacted each country individually and explained the rationale behind why that adaptation would not be in the final English versions.	Prog Rept Mod A	All
Countries were to review each PT twice for adaptation. The first to suggest modifications. The second to make specific choices from possible adaptations and review PTs in general.	Prog Rept Mod A	All
May 11, 2010 country teams were to receive from CAE English versions of the performance tasks, performance task templates, the scoring rubric, and a translation guide.	Prog Rept Mod A	All
Countries needed to examine if students would understand the expectations involved in the PT.	Prog Rept Mod A	All
EVENT: Country teams were to review and adapt the mini PT.	Prog Rept Mod A	All
EVENT: Mini PT was to consist of: • an overview of what is a performance task • a shortened task scenario • a reduced document library (2 documents) • an answer sheet • an example answer, and • rationale as to why the example answer is considered a good one.	Prog Rept Mod A	All
Countries needed guidance on the goal for the cog labs as well as information on how to conduct them.	Prog Rept Mod A	All
Each Participating country's translation team creates first version in home language, circulates it to the respective translation review team. Feedback is provided and necessary modifications are made. This same cycle is repeated for the revised version. Once these two cycles are completed, the tasks are pilot tested.	Prog Rept Mod A	All
EVENT: The Country C team fall 2010 teleconference call was to be postponed until late October/early November.	Prog Rept Mod A	All
When applicable country teams should examine the following when selecting participating HEIs • Different types of institutions (e.g. universities and technical colleges); • Institutions which vary in terms of their research status;	12 Select	All

Task/Subtask	Doc	Cty
• Institutions which enrol the highest achieving students, and those which enrol less academically gifted students; • Institutions in different geographical areas; and • Institutions which enrol students with different characteristics (e.g. gender, ethnicity, religion, socio-economic status, first language or immigrant status, etc.).		
Country teams were to get individual and institution (multilevel) level engagement in the study from each HEI.	12 Select	All
Country teams were going to test in the following languages[16]: • Country A=Country A; Mod A • Country B=Country B (English optional second); Mod A • Country C=Arabic (English required second); Mod A • Country D=Spanish; Mods A, B, C • Country E=Norwegian (English optional second); Mod A	06 Strand	All
In November Paris meeting, national staff (National Project Managers (NPM), Lead Scorers (LS) and those responsible for test administration and sampling, if not the NPM) will receive training on: • national and international management; • sampling; • scoring; • test system; and • test administration.	Nov Train	All

[16] Are these languages across <u>all</u> strands in which they are participating?

Appendix G: Condensed list of tasks and subtasks included in translation and adaptation

Task/Subtask	Doc	Cty
Familiarize team with 16 contacts (2 CA, 2 CO, 12 NY)	**Annex B**	**All**
Country teams to be familiar with other technical advisors (Ron Hambleton, Ph. D.[17], Scott Elliot, Ph. D., other program managers/associates.	Annex D	All
Country teams to familiarize themselves with CAE staff and TAT backgrounds and responsibilities (Roger Benjamin, Ph. D., James Hundley, Rich Shavelson, Ph. D., Guillermo Solano-Flores, Ph. D., Amy Kurpius, Stephen Kelin, Ph. D., Marc Chun, Ph. D., Jeffrey Steedle, Ph. D.	Annex D	All
Conceptual framework for PT translation and translation review provided by CAE read by country teams.	**Intr ltr**	**All**
Familiarize team representatives with CAE provided conceptual framework for adaptation.	**Annex D**	**All**
Country teams had a list of four key documents that informed the translation and adaptation procedures' conceptual framework.	GS.4	All
Country teams were to read and understand the ideas shown in four different pieces of literature contributing to adaption and translation conceptual framework.	GS.7-10	All
Country teams were to become familiar with issues, designs, and technical guidelines for test translation and adaptation (Hambleton, 2005) • carefully choose test administrators • use appropriate item formats • control for speed effect • translators should be familiar with target group, their culture, test content, have some training in test development, and are most capable in test adaptation. • Choose judgmental designs appropriately • Choose appropriate data collection designs • Choose statistical analysis appropriately (differing curricula, cultural backgrounds, levels of motivation, socio-political factors) • Use appropriate ITC Guidelines for Test Adaptation	GS.7	All
Country teams were to become familiar and use applicable features of universal design (Thomson, Johnston, & Thurlow, 2002): • Design instruments so allow participation of widest range of students (and flexible enough to allow for changing student populations) • Precisely defined constructs • Accessible non-biased items • Amendable accommodations • Simple, clear, and intuitive instructions—and procedures • Maximum readability • Maximum legibility • Careful use of results	GS.10	All
Country teams were to become familiar with ITC guidelines (2005):	GS.8	All

[17] Document indicates, pg 8, that this function was now taken on by 'Consortium's technical advisory group. Is this ACER? Did Hambleton ever serve? Why the change?

Task/Subtask	Doc	Cty
• Consider the technological issues (CBT) and internet • Consider quality issues in CBT and internet testing • Provide appropriate level of control over CBT and Internet testing • Provide appropriate for security and safeguarding privacy		
Country teams were to become familiar with TTTE (Solano-Flores, 2008) • disconfirming evidence • multidimensionality of language • multidisciplinary-team approach • tension among error dimensions • attention to language usage, culture, and local curriculum	GS.9	All
Translation review should be based on TTTE	GS.36	All
According to TTTE, a translation error can belong to multiple dimensions.	GS.36	All
National committee should be able to provide input and feedback through regularly scheduled meetings (or some other alternative).	**GNE 19**	**All**
Contractors should schedule and coordinate meetings with NPMs.	GNE 19	All
Ask for each country team to provide availability for conference call to discuss future Feb 2010 meeting.	Intr ltr	All
Initial details provided for Feb 2010 meeting by CAE read by country teams.	Intr ltr	All
Country teams to prepare for the meeting in New York City (February 2010).	Annex D	All
Team representatives to participate in New York City meeting (February 2010).	Annex D	All
Via telecom meetings AAG member to present recommended modifications of CLA PTs.	Annex D	All
Via internet meetings country teams will present suggested adaptations	A-W-R	All
Via internet meetings country teams will reach consensus on adaptations	A-W-R	All
Via telecom meetings, based on CAE and Country Team evaluations and recommendations, gain consensus on the two PTs.	Annex D	All
c. Via telecom meetings w/ translation teams, CAE will provide support of translation process.	Annex D	All
c. Via telecom meetings w/ translation teams, CAE will provide share translations for review and finalization.	Annex D	All
CAE to participate in two AHELO GNE meetings in Paris.	Annex D	All
Mentions that in April 2009 in Paris AHELO GNE (Group of National Experts) agreed on proposed division of work by several entities: • Secretariat • Contractors • National experts and higher education institutions (HEIs) • GNEs The GNEs at that time also asked for clarification of: • the NPMs role. • Associated costs for countries • Associated costs for HEIs	GNE 19	All
NPMs will participate in five meetings throughout the study: 6. to become familiar and discuss project, assessment framework, sample items, and give national presentations. 7. discuss fieldwork procedures, sampling, national reports, data management and analysis systems.	GNE 19	All

Task/Subtask	Doc	Cty
8. Update on implementation updates, coder training, report on fieldwork. 9. Review fieldwork implementation and outcomes, review data and initial analysis, consider practical and scientific feasibility 10. Review and debrief on findings, results, and outcomes.		
Countries were to participate in tele or video conference call with CAE 9/27-9/29/11 to: • reflect on trans & adap translation process • discuss implementation phase expectations • ask any final questions • transitioning to CAE implementation team (from CAE adaptation team)	GS.1 Timeline	All
At NYC meeting country teams will: • become more familiar with PTs • select final two PTs • get an overview of the CLA proctor interface • get an overview of CLA PTs online interface • learn about security issues • learn about logistical issues	GS.5 Draft Mtg	All
Country team representatives were to attend the Generic Strands meeting in Paris (OECD) on March 17, 2010	GS.23	All
Country team member(s) participated in OECD, Paris AHELO Module A meeting from 9:00am-12:30pm.	GS.25	All
Country team members will participate in a two-day CAE representative site visit in June, July, or August 2010: • NPM • Assessment expert • Translators • Translation review team • Anyone else NPM thinks should be there (e.g., linguist, lead scorer)	GS.32	All
March 17 Paris meeting: county team representatives met face-to-face, review recommended country adaptations to the two performance tasks and present to Rich.	Prog Rept Mod A[18]	All
Country teams to select two suitable PTs that will be used in AHELO.	**Annex D**	**All**
Summary descriptions of 9 CLA tasks provided by CAE read by country teams.	Intr ltr	All
Country teams can add PT reducing criteria in NYC Feb 2010 meeting	GS.5 Crit Selection	All
Country teams are to use five criteria for reducing the number of PTs from nine to four to two.	GS.5 Crit Selection	All
Country teams are to become familiar with three classifications of performance tasks by nature of the task.	GS.5 Crit Selection	All
Country teams to become familiar with the three criteria to select the subset of PTs. (universality of PT theme; ease of translation [based on complexity of language in PT]; ease of scoring based on US experience w/ CLA).	Annex D	All

[18] NEED TO GET: Please refer to GNE [2010]2 – Progress Report on Generic S

Prog Rept Mod A	All

kills Strand, presented at the March, 2010 GNE meeting, for a discussion of phases one and two

Task/Subtask	Doc	Cty
Country team representatives to review a subset of at least nine CLA PTs.	Annex D	All
Recommendations and rationales for selecting from 4-5 subset PTs read by country teams at the NYC meeting.	Intr ltr	All
Country teams to select four suitable PTs—considered valid in an international context.	Annex D	All
Country teams to adapt each PT.		
Country teams need to know if students are used to proposing, challenging, or critiquing ideas, persons or institutions.	GS.13	All
Country teams needed to examine appropriateness of context and procedures.	GS.13	All
Country teams were to submit suggestions for adaptations for each PT	PK	All
Country teams, using drop down menus, were to select their choice of adaptation possibilities for each PT. Needed adobe acrobat to complete this task: • people names • people titles/surnames • location names • government positions • measurement units • proper nouns • financial units and cost • maps • graphs • keys	GS.33a	All
General instructions for CLA administration on Internet provided by CAE read by country teams.	Intr ltr	All
Country teams to read about CLA and its role in the larger context of assessment in higher education accountability.	**Annex C**	**All**
Country teams to understand the constructs being assessed in the CLA.	Annex C	All
Country teams to understand complexity of performance tasks: time required, cost, and scoring time requirements.	Annex C	All
Country teams to understand the impact of paperless administration.	Annex C	All
Country teams to become familiar with the controversial aspects of high-stakes testing (e.g., cheating, sanctions) to understand CLA's position on the issue.	Annex C	All
Country teams to understand the limitations of the CLA (needs measures for specific majors; measures of social, moral, and civic outcomes).	Annex C	All
During February 2010 meeting, country teams to become more familiar with CLA.	Annex D	All
Points out the role of contextual information in using test results for policy and practice-related decisions—to be read by country teams.	Intr ltr	All
Country teams were to understand: • performance task have three parts: task, response format, scoring system • CLA PT: holistic, complex, real-world task, students write a recommendation or decision, reach a conclusion or solve a problem supported with facts and evidence • CLA measures analytic reasoning and evaluation, problem solving, writing persuasiveness, and writing mechanics	GS.12	All
Country teams to get familiar with technical aspects of CLA.		

Task/Subtask	Doc	Cty
Explanation of complexities of generic strand (intended constructs) read by country teams.	Intr ltr	All
Mentions issues of validity in addition to those associated with cross-cultural appropriateness, and linguistic transferability—to be read by country teams.	Intr ltr	All
Country teams to understand the criterion sampling approach.	Annex C	All
Country teams to understand student operant responses and the context of performance tasks and multiple skills.	Annex C	All
Country teams to understand how matrix sampling can reduce testing time.	Annex C	All
Country teams to understand the impact of online rater scoring and calibration of the PT.	Annex C	All
Country teams to understand the accountability aspects of the CLA: signaling, benchmarking, value added focus	Annex C	All
Country teams to understand the issues associated with summative function of accountability: stakes, common set of indicators, incentive or punishment.	Annex C	All
Country teams to understand formative function of accountability at the student and school levels: diagnosing and providing feedback, monitoring change.	Annex C	All
Country teams to be aware of the importance of benchmarking or examining value added.	Annex C	All
Country teams to have general technical expertise.		
Country teams to be able to read tables, plot charts, and research abstracts.	Annex C	All
Country teams to understand natural language processing software and its relationship to reliability or validity.	Annex C	All
Countries must have access to software and computer/telecom equipment.	Annex D	All
Country teams to agree with the timely fashion in which activities must be completed.	**Annex D**	**All**
NPMs had to be available from Jan 2010 through Dec 2011.	GNE 19	All
2/19-3/12/2010: Country NPM/GNE and assessment experts begin recruiting" • translation team • translation review team	GS.1 Workplan	All
Country teams were to receive updated timeline (initial implementation expectations) from CAE by 7/28/10.	GS.1 Timeline	All
Country teams were to receive platform instructions for translation and addendums (GS.36/.37) from CAE by 7/28/10.	GS.1 Timeline	All
Countries were to send rough—or reconciled—translations of two PTs to CAE 8/15/10	GS.1 Timeline	All
Countries were to have translations of cognitive labs materials to CAE 8/15/10	GS.1 Timeline	All
Countries were to conduct cognitive labs and revise two PTs based on results between 8/15/10-10/1/10	GS.1 Timeline	All
Countries were to make final revisions to two PTs and instructions to CAE 10/1/-10/15/10 (based on teleconference and cog labs)	GS.1 Timeline	All
Countries were to complete final translations of performance tasks, scoring rubric, and instructions sent to CAE for review by 10/18/10	GS.1 Timeline	All
Final translations of mini PT (GS.38) and scoring handbook charts/response features (GS.39) sent to CAE for review by 12/1/10.	GS.1 Timeline	All

Task/Subtask	Doc	Cty
Country teams required personnel and financial resources.		

Country teams will need to determine:
• local staff rates
  • government staff
  • institution staff as test administrators
  • experts that take part
• translation; printing costs
• availability of premises or equipment needed
• if they will participate in the contextual dimension
• if HEIs need incentive
	GNE 19	All

AHELO GNEs were to take note of estimated implementation resource needs for their country.	GNE 19	All
Assessment representative requires 50% of full time for calendar year 2010 (continuous communication with both AHELO and national project staff and ensure highest quality standards and ensure proper implementation of AHELO procedures).	Annex E	All
Country teams will need staff: • NPM: 100-400 days • Research assistant: 100-400 days • Administrative assistant: 80-150 days • Translation/adaptation advisor: 0-30 days • Technical advisor: 10-30 days • Editorial support: 10-20 days	GNE 19	All
Country teams will need to cover costs (EUR): • Low: 2010=200K; 2011=200K; total: 400K • High: 2010=450K; 2011=450K; total: 900K	GNE 19	All
Country teams to modify the two suitable PTs for use in AHELO.	**Annex D**	**All**
Summary descriptions of 9 CLA tasks provided by CAE read by country teams.	Intr ltr	All
Recommendations and rationales for selecting from 4-5 subset PTs read by country teams at the NYC meeting.	Intr ltr	All
Country team representatives to review a subset of at least nine CLA PTs.	Annex D	All
Country teams to select four suitable PTs—considered valid in an international context.	Annex D	All
Country teams to select two suitable PTs that will be used in AHELO.	Annex D	All
Country teams to become familiar with the three criteria to select the subset of PTs. (universality of PT theme; ease of translation [based on complexity of language in PT]; ease of scoring based on US experience w/ CLA).	Annex D	All
Country teams are to become familiar with three classifications of performance tasks by nature of the task.	GS.5 Crit Selection	All
Country teams are to use five criteria for reducing the number of PTs from nine to four to two.	GS.5 Crit Selection	All
Country teams can add PT reducing criteria in NYC Feb 2010 meeting	GS.5 Crit Selection	All
Country teams to review contents of two selected/agreed upon PTs.	Annex D	All
After NYC meeting, country teams to create list of modifications for each PT to have each PT fit country context.	Annex D	All
CAE to fully modify two tasks following agreed upon modifications.	Annex D	All

Task/Subtask	Doc	Cty
CAE to fully modify scoring rubrics following agreed upon modifications.	Annex D	All
CAE to fully modify IT administration procedures following agreed upon modifications.	Annex D	All
Country teams must ensure that the tasks are meaningful in their countries by examining: • universality of PT theme • ease of translation based on PT's language complexity • ease of scoring	GS.4	All
Country assessment experts choose Assessment Adaptation Group (AAG).	**Annex D**	**All**
Each country team will have an assessment expert.		
Country teams each to nominate several persons with expertise in assessment who will represent country team in AHELO.	**Annex E**	**All**
One assessment representatives will be selected based on formal training, professional experience, and relevant set of technical qualifications.[19]	Annex E	All
Country teams to nominate people with required qualifications: • education and formal training (Ph.D. in psychometrics, statistics, social-science measurement or related field) • English proficiency (fluent in conversational and reading/writing English) • Professional experience (higher education, int'l or multicultural education; academic achievement assessment; assessment development—preferable constructed response tasks, computer simulations or computer-based training, test translation and test translation review	Annex E	All
Nominees for assessment representative will hopefully have desired qualifications: • experience coordinating the collection, management, and analysis of data for educational projects. • Ability to successfully work with multidisciplinary and/or multicultural teams • Record of technical publications in the areas of expertise	Annex E	All
Country teams were to establish a National Centre to provide appropriate infrastructure for managing key facets of the AHELO Feasibility Study.	**GNE 19**	**All**
When choosing NC site, each country team was to consider financial demands as well as the purpose and positioning of the AHELO feasibility study.	GNE 19	All
Country teams were to select an NC site that would allow each team to engage institutions in a scholarly and quality improvement perspective.	GNE 19	All
Country teams were to choose an NC site that would allow for efficient communication with OECD, government agencies, contractors, and HEIs.	GNE 19	All
Country teams were to staff the NC with people who had nuanced knowledge of the system, effective leadership capacity, --optimally—established relationships with opinion leaders, and sound technical footings.	GNE 19	All
Country teams were to staff the NC with a core staff to include: • NPM • Research assistant • Administrative assistant • Translation/adaptation advisor	GNE 19	All

[19] Unclear who does the selection of the assessment representative.

Task/Subtask	Doc	Cty
• Technical advisor • Editorial support		
Country teams will nominate the NPM.	**GNE 19**	**All**
NPMs should have several qualifications: • experience in planning, organizing, and conducting large-scale assessment • identify, select, and manage team of project staff • experience with successfully handling multiple tasks simultaneously • have excellent oral and written communication in local language and English—be able to represent country at international meetings • helpful if have previous work in the fields of higher education, educational assessment, and contextual surveys. • Helpful if familiar with data processing, survey quality control procedures, and data file structures.	GNE 19	All
If appropriate, the same person can serve as NPM and GNE.	GNE 19	All
Country team to learn about recruiting of test translation team.	**Annex D**	**All**
Translators were to translate the three PT components (task, response format, and scoring rubric) simultaneously.	**GS.13**	**All**
e. Each country team to select only one language for field test.[20]	Annex D	All
Country teams will choose one language for translation.	A-W-R	All
Translation reconciliation was to merge the two independent translations (done by translation/adaptation advisor and translators).	GS.13	All
Translators were to reconcile the three PT components (task, response format, and scoring rubric) simultaneously.	GS.13	All
An OECD designated agency was to verify the reconciled translation of all material against source versions to assure quality control.	GS.13	All
Translators were to translate document library for PTs: • letters • memos • summaries of research reports • newspaper articles • maps • charts • photographs • diagrams • tables • interview notes or transcripts	GS.4	All
Country teams had to examine issues relevant to the test translation dimensions and error types.	GS.4	All

[20] What is this field test referring to?

Task/Subtask	Doc	Cty
Country teams were to become familiar with and understand the translation process: • two independent translations • reconciliation of first two translations • conduct cognitive labs • make changes to PTs based on cognitive labs • conduct translation review • create final version of PTs based on AHELO • translation verification by OECD – designated agency • create final version of PTs	GS.14	All
Country teams are to conduct translation reviews to: • **examine disconfirming** • **focus on error, not appropriateness—assume that translation error is inevitable** • **examine tension among error dimensions** • **pay specific attention to language usage, culture, and local curriculum**	**GS.13**	**All**
Necessary training		
Familiarize team representatives with CAE provided procedures for training country members to adapt the PTs.	Annex D	All
AAG to train (by CAE) in task adaptation.	Annex D	All
AAG to train (by CAE) in translation process.	Annex D	All
AAG to train translation review process.	Annex D	All
Country teams must have hired and trained scorers close to the date when they will be doing actual scoring.	GS.23	All
Country teams must train proctors: • review the proctor guide • watch the online training video • be familiar with technical assistance • be familiar with PTs	GS.29	All
Country teams must be trained to conduct the cog labs.	GS.37	All
Countries will participate in training for cog labs.	GS.26	All
Country teams were to allow CAE to train all scorers.	GS.26	All
Country team members participating in the cognitive labs must go through training.	GS.26	All
TAT will create a video for training for cog labs that country teams were to use for training.	GS.26	All
At NYC meeting AAG would be trained in: • adaptation • translation • translation review	A-W-R	All
In November Paris meeting, national staff (National Project Managers (NPM), Lead Scorers (LS) and those responsible for test administration and sampling, if not the NPM) will receive training on: • national and international management; • sampling; • scoring; • test system; and	Nov Train	All

Task/Subtask	Doc	Cty
• test administration.		
Indication of information found in other documents—country teams to cross reference and read.	**Intr ltr**	**All**
Additional information about NPM responsibilities and timelines would be provided after contracts were signed with other entities.	GNE 19	All
Country teams were to review "GS.36 and GS.37 addendum" which includes 4 updates made to the Translation Guide and 1 update to the Cognitive Labs Guide.	GS.36/GS.37 Addendum	All
GS.36 and GS.37 addendum—changes to the Translation Guide and Cog Labs Guide—was based on discussion that took place among CAE team members and during the 2010 CAE country site visits.	GS.36/GS.37 Addendum	All
Update 1 for GS.36=Country teams were to read and implement new deadlines emailed on June 25, 2010.	GS.36/GS.37 Addendum	All
Update 2 for GS.36=CAE updated the review process for the translated language going onto the internet testing platform.	GS.36/GS.37 Addendum	All
Update 3 for GS.36=additional information is provided for the reason why the translation team needed to pay particular attention needed to be paid to Register.	GS.36/GS.37 Addendum	All
Update 4 for GS.36=translation and translation review teams were to use each document as an analytical unit—not a sentence or paragraph.	GS.36/GS.37 Addendum	All
Update 1 for GS.1=interviewer is not to model how to think aloud.	GS.36/GS.37 Addendum	All
GS.48 instructed country teams to cross-reference other documents for instructions on how to review and enter each of the different documents.	GS.48	All

Appendix H: Description matrices.

COUNTRY A	Evidence Type	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
Tasks	Timely communication	Expertise-measurement	Expertise-translation	Project management experience	Review opportunities	Training opportunities	Opportunities for progress documentation	Apt deadlines	User friendly materials	In-country support external to team	Support external to team outside country
Configure a team according to coordinating group's specifications.	(1) Country A and the US organizing agency shared timely communication about configuring the team.	(1) A team member for Country A was a national assessment expert and had academic background in measurement.	(1)The Country A team had translation experience.	(1) The NPM for Country A met all suggested qualifications.	(1) Country A team reviewed and chose persons for both positions--US organizing agency reviewed CVs afterwards	(1) Not all Country A team members attended all training; however, subsequent internal training took place.	(1) Country A NPM emailed the US organizing agency contact information and confidentiality agreements for each Country A team member	(1) Country A team was able to fill team positions in time for initial meetings and document review.	(1) Country A team did not express difficulty using documents provided for team configuration.	(1) Country A team had support from local government and academia to configure team.	(1) Country A received support from the international coordinating agency and US organizing agency when configuring the team.
Acquire funding for all steps of the process.	(1) Country A received timely communication about project costs	(1) Country A funded assessment expert for duration of project	(1) Country A funded translation team for duration of project	(1) Country A funded the NPM position for duration of project	(1) Country A had opportunities to review budget	(-1) Country A did not have training opportunities available for budget	(1) The US organizing agency collected data on fundin	(1) Country A only experienced difficulty with funding when	(1) Country A team did not express difficulty using documents	(1) Country A had financial support from government and acade	(1) With regard to funding activities, Country A received

COUNTRY A	Evidence Type	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
							g acquisition progress for Country A.	finalizing number of PTs for the study.	provided for acquiring funds.	mia.	support from the international and US organizing agencies.
Select test items based on specific criteria established by the coordinating group.	(1) there was timely communication for Country A to select PTs for the study.	(1) Country A used measurement expertise when selecting performance tasks for the study.	(1) Country A used translation expertise when selecting performance tasks for the study.	(1) Country A used project management experience when selecting performance tasks for the study.	(1) Country A participated in two reviews during PT selection process	(1) although there was no official training, Country A was able to learn and apply selection criteria provided through documentation	(1) Country A helped the US organizing agency's project manager document progress and report it to the international organizing agency.	(1) Country A team was able to complete all activities associated with final PT selection within the three weeks given.	(1) Country A team did not express difficulty using documents provided for PT selection	(-1) Country A did not receive additional in-country support external to the team	(1) Country A received support from other country teams, the international organizing agency, and US organizing agency
Acquire necessary technical infrastructure.	(1) Country A got timely communication regarding day-to-day communication,	(1) The Country A measurement expert had access to all of the technology necessary.	(1) The Country A translation experts had access to all of the technology necessary.	(1) Country A used project management experience when acquiring the technical infrastr	(1) Country A had opportunities to review acquisition of technical infrastructure.	(1) Country A participated in training and received documentation addressing	(1) The US organizing agency provided opportunities to document progre	(1) Country A team was able to acquire the technical infrastructure needed	(1) Country A required minimal clarification on documentation regarding	(1) government agencies and academia provided Country A with the necessary	(-1) When acquiring the necessary technical infrastructure Country A did not receive, or

COUNTRY A	Evidence Type	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	testing the computer interface, and implementing the assessment			ucture for the study.		technical infrastructure.	ss for Country A.	throughout the study.	acquisition of technical infrastructure.	technical infrastructure.	require, support from outside of the country.
Adapt test based on agreed upon cultural adaptation suggestions.	(1) There was timely communication for PT adaptation between the Country A team and US PI.	(1) Country A measurement expert participated in the PT adaptation process.	(-1) Country A did not include translation expertise during the adaptation process.	(-1) Although Country A participated in each step of adaptation process, they did not address all issues of adaptation.	(1) Country A participated in every review opportunity available during adaptation.	(1) Country A participated in training and received documentation addressing PT adaptation.	(1) Country A participated in every opportunity to document progress on task adaptation.	(1) Country A team was able to meet deadlines throughout the adaptation process.	(-1) During adaptation Country A did not address all of the topics discussed in the materials supplied.	(1) During adaptation Country A had in-country support external to the team.	(1) During adaptation Country A received support from the US organizing agency.
Hire translators possessing qualifications set by coordinating group.	(1) Country A team did not indicate any challenges with timely communication during hiring of translators.	(1) Country A translators had measurement expertise.	(1) although there was no certification, Country A's translators were professionals with appropriate experience	(1) Country A used project management experience when hiring translators for the study.	(-1) There was no review opportunityfor Country A during the hiring of translators.	(1) Country A participated in training and received documentation addressing hiring translators.	(1) Alhtough they were not officially planned activities, Country A participated in every opportunity	(-1) Deadlines caused Country A some challenges when hiring translators.	(1) Country A team did not express difficulty using documents provided for hiring translators.	(1) While hiring translators Country A had in-country support external to the team.	(1) When hiring translators Country A received support from the US organizing agency.

COUNTRY A	Evidence Type	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
							to document progress on hiring translators.				
Translate the assessment.	(1) There was timely communication during translation between the Country A team and US organizing agency.	(-1) Following criteria provided, Country A's translators did not have measurement expertise. However, they were able to work through challenges in scoring.	(1) Country A's translation team had the desired translation expertise.	(1) Country A's project management experience was helpful during the translation process.	(-1) There is no evidence of a review taking place during Country A's initial translation phase.	(1) Country A participated in training and received documentation addressing translation and reconciliation.	(1) The Country A team and US organizing agency documented team progress in translation and reconciliation.	(-1) Country A was not able to complete all translation activities prior to the site visit.	(1) Country A team did not express difficulty using documents provided for translation and reconciliation.	(1) Translators from Country A completed the translation process.	(1) Country A received support from the US organizing agency with the translation process.
Review translation and notes from translation process.	(1) There was timely communication during translation review between the Country A	(1) Country A's translation review team included measurement expertise.	(1) Country A's translation review team had the desired translation expertise.	(1) Country A's project management experience was helpful during the translation review	(1) Country A had opportunities to review their work as they completed the transla	(1) Country A participated in training and received documentation addressing	(1) The Country A translation review team had opportunities to document their	(-1) The US organizing agency did not provide the Country A team with a deadli	(1) Country A team did not express difficulty using documents provided for transla	(1) Country A had in-country support during the translation review process that	(1) Country A received support from the US organizing agency with the translation

COUNTRY A	Evidence Type	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	team and US organizing agency.			process.	tion review process.	translation review.	progress during the translation review process.	ne for the translation review process.	tion review.	was external to the team.	review process.
Translate ancillary materials as described by coordinating group.	(1) There was timely communication between the Country A team and US organizing agency during translation of ancillary materials.	(-1) Following criteria provided, Country A did not include measurement expertise during the translation of ancillary materials.	(1) Country A's translation review team had the desired translation expertise.	(1) Country A's project management experience was helpful during the dual translation process.	(-1) There is no evidence that Country A reviewed translation of ancillary materials.	(1) Country A participated in training and received documentation addressing ancillary material translation.	(1) The Country A team and US organizing agency documented team progress in translation and reconciliation of ancillary material.	(-1) The US organizing agency did not provide the Country A team with apt deadlines for translating all ancillary materials.	(1) Country A team did not express difficulty using documents provided for the dual translation of ancillary materials.	(1) Translators from Country A completed the dual translation process of ancillary materials.	(1) Country A received support from the US organizing agency with the dual translation process to be used with the ancillary materials.
Review translation of material for assessment implementation.	(1) Communication between the US organizing agency and Country A team addressing	(1) Country A's NPM, who had measurement experience, helped review external translations	(0) There is limited information about Country A's review of external translation of materi	(0) There is limited information about Country A's review of external translation of materi	(1) Country A had opportunities to review their work as they reviewed the external transla	(1) Country A participated in training and received documentation addressing transla	(1) The Country A team had opportunities to document their progress during the	(1) Country A did not express difficulty in completing the review in time to	(1) Country A team did not express difficulty using documents provided for reviewing	(0) There is limited information about Country A's support during the review of extern	(1) Country A received support from the US organizing agency while reviewing the external

COUNTRY A	Evidence Type	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	external translation of material for assessment implementation was timely.	of material for assessment implementation.	al for assessment implementation.	al for assessment implementation.	tion of material for assessment implementation.	tion review of materials for assessment implementation.	review of external translations.	upload the material onto the internet platform.	external translation of material for assessment implementation.	al translation of material for assessment implementation.	translation of material for assessment implementation.
Implement changes based on verification procedures prescribed by coordinating group.	(1) Communication between the US organizing agency and Country A team during verifiation was timely.	(1) Country A's NPM, who had measurement experience, helped review translations verification notes.	(0) There is limited information about Country A's review of translation verification suggestions.	(1) Country A's project management experience was helpful while reviewing translation verification information.	(0) Country A received limited guidelines on how to review translation verification suggestions.	(-1) Country A did not have training opportunities for reviewing translation verfication results.	(-1) Although there were opportunities to document Country A's progress, the US organizing agency did not keep documentation.	(-1) Although the US agency did not include this activity in the study's workplan, Country A completed the work within days.	(1) Country A team did not express difficulty using documents provided for implementing changes from translation verification.	(0) There is limited information about Country A's support in implementing changes from translation verification.	(1) Country A received support from the US organizing agency while implementing changes from translation verification.
Make agreed upon changes resulting from validation procedure established	(1) Communication between the US organizing agency and Country A team	(1) Country A's NPM, who had measurement experience, helped with the cogniti	(-1) The staff from Country A conducting the labs and making change	(1) Country A had project management experience while making changes	(1) Country A participated in two review opportunities.	(1) Country A participated in training and had access to training materi	(1) The Country A team had opportunities to document their progress	(1) Country A did not express difficulty in completing the cognitive lab	(1) Country A found materials for cognitive labs easy to use.	(1) Country A had in-country support during the validation procedure.	(1) Country A had external country support during the validation

COUNTRY A	Evidence Type	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
by coordinating group.	during validation procedures was timely.	ve labs.	s resulting from the process did not have expertise in translation.	based on results from validation procedure.		al addressing how to conduct and use notes from cognitive labs.	during the cognitive lab process.	activities on time.			procedure.
Test assessment implementation process for target population usability.	(-1) Country A did not test the assessment implementation process with their target population.	(-1) Country A did not test the assessment implementation process with their target population.	(-1) Country A did not test the assessment implementation process with their target population.	(-1) Country A did not test the assessment implementation process with their target population.	(-1) Country A did not test the assessment implementation process with their target population.	(-1) Country A did not test the assessment implementation process with their target population.	(-1) Country A did not test the assessment implementation process with their target population.	(-1) Country A did not test the assessment implementation process with their target population.	(-1) Country A did not test the assessment implementation process with their target population.	(-1) Country A did not test the assessment implementation process with their target population.	(-1) Country A did not test the assessment implementation process with their target population.
Provide students with an opportunity to become familiar with test format and expectations.	(1) Communication between the US organizing agency and Country A team about the mini-PT was timely.	(1) Country A's NPM, who had measurement experience, helped with the mini-PT.	(0) Information is unclear as to the translation expertise available while Country A was working on the	(1) Country A had project management experience while working on the mini-PT.	(1) Country A had opportunities for review when working with the mini-performance task.	(1) Country A participated in training and had access to training material addressing the mini-perfor	(1) Country A had opportunities to document their progress while working with the mini-PT.	(1) Country A did not express difficulty in completing the work for the mini-PT on time.	(1) Country A team did not express difficulty using documents provided for working with the mini-PT.	(1) Country A had in-country support during the work with the mini-PT.	(1) Country A had external country support while working with the mini-PT.

COUNTRY A	Evidence Type	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
			mini-PT.			mance task.					
Hire scorers according to coordinating group's specifications.	(1) Communication about scorers was timely between the US organizing agency and Country A.	(1) Country A scorers acquired measurement expertise through the study's training.	(-1) There is no evidence that Country A's scorers possessed translation expertise.	(1) Country A successfully used project management expertise while hiring scorers.	(1) Country A's scorers participated in training reviews.	(1) Country A team members and scorers took part in training.	(-1) There were no opportunities for Country A to document the progress of hiring scorers.	(1) Country A met all deadlines associated with scorers.	(1) Country A did not express finding the documents addressing scorer hiring challenging.	(0) There is limited information about who Country A hired for scoring.	(1) Country A had external country support while hiring scorers.
Attend kick-off meeting as well as in-person and phone meetings to discuss progress.	(1) Communication about meetings was timely between Country A and organizing agencies.	(1) Country A included team members with measurement expertise in telephone and in-person meetings.	(1) One meeting required a translation expert to participate; Country A was able to participate.	(1) Country A successfully used project management expertise to attend meetings.	(-1) Country A did not participate in any review opportunities while attending meetings.	(-1) There was no need or opportunity for Country A to train to attend meetings.	(1) Organizing and coordinating agencies documented information about meetings in which Country A participated.	(1) Country A did not indicate challenges with dates for planned meetings.	(1) Country A did not indicate that the material addressing meetings was challenging.	(-1) In-country support external to the team was not required or necessary for the Country A team.	(1) Country A received support with meetings from the US organizing agency and international coordinating agency.
Submit feedback on process and technical reports regardi	(-1) Communication between Country A and	(1) Country A included measurement expertise in	(1) Country A included translation expertise in	(1) Country A successfully used project management	(-1) There were no review opportunities while Countr	(-1) There was no need or opportunity for Countr	(1) The US organizing agency documented	(1) Country A did not indicate challenges	(-1) Country A provided feedback during meetin	(1) Country A had in-country support while providi	(-1) Country A did not require support external to

COUNTRY A	Evidence Type	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
ng progress.	organizing agencies regarding feedback was not always timely.	feedback provided to organizing and coordinating agencies.	their feedback.	expertise to provide feedback.	y A provided feedback.	y A to train to provide feedback.	Country A's feedback throughout the study.	with dates for providing feedback.	gs, conference calls, and via emails; no materials were needed.	ng feedback about the study.	the team and outside of the country to provide feedback.
Recruit institutions and students to participate in the assessment.	(1) Country A received timely communication from the organizing agencies regarding recuirtment of HEIs and students.	(0) There is limited information about the extent to which Country A's measurement expert was involved in HEI and student recruitment.	(-1) Translation expertise was not necessary while Country A recruited HEIs and students.	(0) There is limited information about the extent to which Country A's NPM was involved in HEI and student recruitment.	(0) There is no informaiton about opportunities for review while Country A recruited HEIs and students.	(1) Country A participated in training and received documentation addressing recruitment and sampling of HEIs and students.	(1) An international coordinating agency documented Country A's progress in sampling.	(0) It is unclear if Country A found the due dates for HEIs and student recruitment and sampling challenging.	(1) Country A did not indicate that the material addressing recruiting and sampling difficult to use.	(1) Country A had in-country support to recruit and sample HEIs and students.	(1) When recruiting HEIs and students Country A received support from an international coordinating agency.

COUNTRY B	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
Tasks	Timely communication	Expertise-measurement	Expertise-translation	Project management experience	Review opportunities	Training opportunities	Opportutnities for progress documentation	Apt deadlines	User friendly materials	In-country support external to team	Support external to team outside country
Configure a team according to coordinating group's specifications.	(-1) Country B had inconsistent communication with the US organizing agency about team members.	(1) A team member for Country B was a national assessment expert and had academic background in measurement.	(-1)The Country B team did not have translation experience.	(-1) two people took role in Country B; only one had qualifications stipulated. (-1) 2nd NPM: no CV; limited information available	(1) Country B reviewed and chose persons for both positions--CAE reviewed CVs afterwards.	(1) Country B team members participated in all training	(1) Country B emailed the US organizing agency contact information and confidentiality agreements for each Country B team member.	(-1) Country B was not able to fill team positions in time for initial meetings and document review	(1) Country B did not express difficulty using documents provided for team configuration.	(1) Country B had support from government and academia to configure team; however, internal issues delayed team configuration.	(1) Country B received support from the international coordinating agency and US organizing agency when configuring the team.
Acquire funding for all steps of the process.	(1) Country B received timely communication about project costs	(1) Country B funded assessment expert for duration of project	(1) Country B funded translation team for duration of project	(1) Country B funded the NPM position for duration of project	(1) Country B had opportunities to review budget	(-1) Country B did not have training opportunities available for budget	(1) The US organizing agency collected data on funding	(1) Country B only experienced difficulty with funding when finaliz	(1) Country B team did not express difficulty using documents provid	(1) Country B had financial support from government and academia.	(1) With regard to funding activities, Country B received suppor

COUNTRY B	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
							acquisition progress for Country B.	ing number of PTs for the study.	ed for acquiring funds.		t from the international and US organizing agencies.
Select test items based on specific criteria established by the coordinating group.	(1) there was timely communication for Country B to select PTs for the study.	(1) Country B used measurement expertise when selecting performance tasks for the study.	(1) Country B used translation expertise when selecting performance tasks for the study.	(1) Country B used project management experience when selecting performance tasks for the study.	(1) Country B participated in two reviews during PT selection process	(1) although there was no official training, Country B was able to learn and apply selection criteria provided through documentation	(1) Country B helped the US organizing agency's project manager document progress and report it to the international organizing agency.	(1) Country B team was able to complete all activities associated with final PT selection within the three weeks given.	(1) Country B team did not express difficulty using documents provided for PT selection	(-1) Country B did not receive additional in-country support external to the team	(1) Country B received support from other country teams, the international organizing agency, and US organizing agency
Acquire necessary technical infrastructure.	(1) Country B got timely communication regarding day-to-day communicati	(1) The Country B measurement expert had access to all of the technology necess	(1) The Country B translation experts had access to all of the technology necess	(1) Country B used project management experience when acquiring the technical	(1) Country B had opportunities to review acquisition of technical infrastructure.	(1) Country B participated in training and received documentation addres	(1) The US organizing agency provided opportunities to document	(1) Country B team was able to acquire the technical infrastructure neede	(1) Country B team did not express difficulty using documents addressing	(1) government agencies and academia provided Country B with the necess	(-1) When acquiring the necessary technical infrastructure Country B did not receive

COUNTRY B	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	on, testing the computer interface, and implementing the assessment	ary.	ary; however, they had some technical difficulties.	infrastructure for the study.		sing technical infrastructure.	progress for Country B.	d throughout the study.	acquisition of technical infrastructure.	ary technical infrastructure.	, or require, support from outside of the country.
Adapt test based on agreed upon cultural adaptation suggestions.	(1) There was timely communication for PT adaptation between the Country B team and US PI.	(1) Country B measurement expert participated in the PT adaptation process.	(-1) Country B did not include translation expertise during the adaptation process.	(-1) Although Country B participated in each step of adaptation process, they did not address all issues of adaptation.	(1) Country B participated in every review opportunity available during adaptation.	(1) Country B participated in training and received documentation addressing PT adaptation.	(1) Country B participated in every opportunity to document progress on task adaptation.	(1) Country B team was able to meet deadlines throughout the adaptation process.	(-1) During adaptation Country B did not address all of the topics discussed in the materials supplied.	(-1) During adaptation Country B did not have in-country support external to the team.	(1) During adaptation Country B received support from the US organizing agency.
Hire translators possessing qualifications set by coordinating group.	(1) Country B team did not indicate any challenges with timely communication during hiring of	(-1) Country B translators did not have measurement expertise.	(-1) Country B selected translation team members guided by their own connections rather	(1) Country B used project management experience when hiring translators for the study.	(-1) There was no review opportunity for Country B during the hiring of translators.	(1) Country B participated in training and received documentation addressing hiring transla	(1) Although they were not officially planned activities, Country B participated in	(-1) Deadlines caused Country B some challenges when hiring translators.	(1) Country B team did not express difficulty using documents provided for hiring translators.	(1) While hiring translators Country B had in-country support external to the team.	(1) When hiring translators Country B received support from the US organizing agency.

COUNTRY B	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	translators.		than their professional qualifications			tors.	every opportunity to document progress on hiring translators.				
Translate the assessment.	(-1) There was timely communication on the part of the US organizing agency but not always from the Country B team.	(-1) There is no evidence that Country B's translators had measurement expertise.	(-1) Country B's translation team did not possess the desired translation expertise.	(-1) Country B's project management experience was helpful during the translation process.	(-1) There is no evidence of a review taking place during Country B's initial translation phase.	(-1) Country B team members participated in training and received documentation addressing translation and reconciliation. However, the training for translators is unclear.	(1) The Country B team and US organizing agency documented team progress in translation and reconciliation.	(1) The Country B team was able to meet the original deadline for translation, the team did not need the revised deadline.	(1) Country B team did not express difficulty using documents provided for translation and reconciliation.	(1) Academics from Country B completed the translation process.	(1) Country B received support from the US organizing agency with the translation process.
Review translation and notes	(1) There was timely communicati	(1) Country B's translation review	(0) There is limited information	(0) There is limited information	(1) Country B had opportunities	(1) Country B participated in	(1) The Country B translation	(-1) The US organizing agenc	(1) Country B team did not expres	(1) Country B had in-country	(1) Country B received suppor

COUNTRY B	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
from translation process.	on during translation review between the Country B team and US organizing agency.	team included measurement expertise.	about Country B's translation review team's translation expertise.	about Country B's translation review team's project management expertise.	to review their work as they completed the translation review process.	training and received documentation addressing translation review.	review team had opportunities to document their progress during the translation review process.	y did not provide the Country B team with a deadline for the translation review process.	s difficulty using documents provided for translation review.	support during the translation review process that was external to the team.	t from the US organizing agency with the translation review process.
Translate ancillary materials as described by coordinating group.	(-1) There was timely communication on the part of the US agency during the dual translation but not from the Country B team.	(-1) Following criteria provided, Country B did not include measurement expertise during the translation of ancillary amterials.	(0) There is limited information about Country B's translation review team's translation expertise.	(-1) Country B was not able to manage the dual translation process or meet the deadlines.	(-1) There is no evidence that Country B reviewed translation of ancillary materials.	(1) Country B participated in training and received documentation addressing ancillary material translation.	(1) The Country B team and US organizing agency documented team progress in translation and reconciliation of ancillary material.	(-1) The US organizing agency did not provide the Country B team with apt deadlines for translating all ancillary materials.	(1) Country B team did not express difficulty using documents provided for the dual translation of ancillary materials.	(1) Academics from Country B completed the dual translation process of ancillary materials.	(1) Country B received support from the US organizing agency with the dual translation process to be used with the ancillary materials.
Review translation of materi	(1) Communication betwee	(0) There is no information	(0) There is limited inform	(0) There is limited inform	(1) Country B had opport	(1) Country B participated	(1) The Country B team	(1) Country B did not	(1) Country B team did not	(0) There is limited inform	(1) Country B received

COUNTRY B	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
al for assessment implementation.	n the US organizing agency and Country B team addressing external translation of material for assessment implementation was timely.	on the qualifications of the US translators who completed the translation of material for assessment implementation in Country B.	ation about Country B's review of external translation of material for assessment implementation.	ation about Country B's review of external translation of material for assessment implementation.	unities to review their work as they reviewed the external translation of material for assessment implementation.	in training and received documentation addressing translation review of materials for assessment implementation.	had opportunities to document their progress during the review of external translations.	express difficulty in completing the review in time to upload the material onto the internet platform.	express difficulty using documents provided for reviewing external translation of material for assessment implementation.	ation about Country B's support during the review of external translation of material for assessment implementation.	support from the US organizing agency while reviewing the external translation of material for assessment implementation.
Implement changes based on verification procedures prescribed by coordinating group.	(-1) There was timely communication on the part of the US agency during the translation verification but not from the Country B team.	(1) A Country B team member with measurement expertise helped review translations verification notes.	(0) There is limited information about Country B's review of translation verification suggestions.	(-1) Country B had difficulty managing the review of translation verification results.	(0) Country B received limited guidelines on how to review translation verification suggestions.	(-1) Country B did not have training opportunities for reviewing translation verfication results.	(1) Country B had opportunities to document their progress in implementing changes based on translation verification.	(-1) The US organizing agency did not include a deadline for implementing changes resulting from translation verification in the	(-1) Country B had difficulty working with material provided for implementing changes resulting from translation verification.	(0) There is limited information about Country B's support in implementing changes from translation verification.	(1) Country B received support from the US organizing agency while implementing changes from translation verification.

COUNTRY B	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
								study work plan.			
Make agreed upon changes resulting from validation procedure established by coordinating group.	(-1) Although communication by the US organizing agency was timely during validation procedures, the Country B team had challenges.	(0) There is limited information on Country B's staff who conducted the cognitive labs.	(-1) The staff from Country B conducting the labs and making changes resulting from the process did not have expertise in translation.	(0) There is limited information about project management experience on Country B's team while making changes resulting from verification.	(-1) Country B participated in one of two review opportunities.	(1) Country B participated in training and had access to training material addressing how to conduct and use notes from cognitive labs.	(1) The Country B team had opportunities to document their progress during the cognitive lab process.	(-1) Country B was not able to complete the cognitive lab activities on time.	(-1) Country B had difficulty working with material provided for conducting cognitive labs.	(1) Country B had in-country support during the validation procedure.	(1) Country B had external country support during the validation procedure.
Test assessment implementation process for target population usability.	(-1) Country B did not test the assessment implementation process with their target population.	(-1) Country B did not test the assessment implementation process with their target population.	(-1) Country B did not test the assessment implementation process with their target population.	(-1) Country B did not test the assessment implementation process with their target population.	(-1) Country B did not test the assessment implementation process with their target population.	(-1) Country B did not test the assessment implementation process with their target population.	(-1) Country B did not test the assessment implementation process with their target population.	(-1) Country B did not test the assessment implementation process with their target population.	(-1) Country B did not test the assessment implementation process with their target population.	(-1) Country B did not test the assessment implementation process with their target population.	(-1) Country B did not test the assessment implementation process with their target population.

COUNTRY B	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
Provide students with an opportunity to become familiar with test format and expectations.	(1) Communication between the US organizing agency and Country B team about the mini-PT was timely.	(1) A Country B team member with measurement expertise helped with the mini-PT.	(0) Information is unclear as to the translation expertise available while Country B was working on the mini-PT.	(1) Country B had project management experience while working on the mini-PT.	(1) Country B had opportunities for review when working with the mini-performance task.	(1) Country B participated in training and had access to training material addressing the mini-performance task.	(1) Country B had opportunities to document their progress while working with the mini-PT.	(-1) Country B was not able to complete the work for the mini-PT on time.	(1) Country B team did not express difficulty using documents provided for working with the mini-PT.	(0) There is limited information about Country B's support in working with the mini-PT.	(1) Country B had external country support while working with the mini-PT.
Hire scorers according to coordinating group's specifications.	(1) Communication about scorers was timely between the US organizing agency and Country B.	(1) Country B scorers acquired measurement expertise through the study's training.	(-1) There is no evidence that Country B's scorers possessed translation expertise.	(-1) Country B did not successfully use project management expertise while hiring scorers.	(-1) There is no evidence that Country B's scorers participated in training reviews.	(-1) Country B team members participated in training; however, there is no evidence that scorers completed training.	(-1) There were no opportunities for Country B to document the progress of hiring scorers.	(-1) There is no evidence that Country B met all deadlines associated with scorers.	(1) Country B did not express finding the documents addressing scorer hiring challenging.	(0) There is limited information about who Country B hired for scoring.	(1) Country B had external country support while hiring scorers.
Attend kick-off meeting as	(1) Communication about	(1) Country B included a	(-1) Country B did not includ	(1) Country B successfully	(-1) Country B did not partici	(-1) There was no need or	(1) Organizing and coordi	(1) Country B did not	(1) Country B did not indicat	(-1) In-country suppor	(1) Country B received

COUNTRY B	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
well as in-person and phone meetings to discuss progress.	meetings was timely between Country B and organizing agencies.	team member with measurement expertise in telephone and in-person meetings.	e translation expertise in any of the study's meetings.	used project management expertise to attend meetings.	pate in any review opportunities while attending meetings.	opportunity for Country B to train to attend meetings.	nating agencies documented information about meetings in which Country B participated.	indicate challenges with dates for planned meetings.	e that the material addressing meetings was challenging.	t external to the team was not required or necessary for the Country B team.	support with meetings from the US organizing agency and international coordinating agency.
Submit feedback on process and technical reports regarding progress.	(-1) Communication between Country B and organizing agencies regarding feedback was not always timely.	(1) Country B included measurement expertise in feedback provided to organizing and coordinating agencies during meetings.	(-1) Country B did not include translation expertise in feedback.	(1) Country B successfully used project management expertise to provide feedback.	(-1) There were no review opportunities while Country B provided feedback.	(-1) There was no need or opportunity for Country B to train to provide feedback.	(1) The US organizing agency documented Country B's feedback throughout the study.	(1) Country B did not indicate challenges with dates for providing feedback.	(-1) Country B provided feedback during meetings, conference calls, and via emails; no materials were needed.	(1) Country B had in-country support while providing feedback about the study.	(-1) Country B did not require support external to the team and outside of the country to provide feedback.
Recruit institutions and students to participate in the	(1) Country B received timely communication from	(0) There is limited information about the extent	(-1) Translation expertise was not necessary while Countr	(0) There is limited information about the extent to	(0) There is no informaiton about opportunities for review	(1) Country B participated in training and received	(1) An international coordinating agency documented Countr	(0) It is unclear if Country B found the due dates	(1) Country B did not indicate that the material addres	(1) Country B had in-country support to recruit and	(1) When recruiting HEIs and students Country B

COUNTRY B	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
assessment.	the organizing agencies regarding recuirtment of HEIs and students.	to which Country B's measurement expert was involved in HEI and student recruitment.	y B recruited HEIs and students.	which Country B's NPM was involved in HEI and student recruitment.	while Country B recruited HEIs and students.	documentation addressing recruitment and sampling of HEIs and students.	y B's progress in sampling.	for HEIs and student recruitment and sampling challenging.	sing recruiting and sampling difficult to use.	sample HEIs and students.	received support from an international coordinating agency.

COUNTRY C	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
Tasks	Timely communication	Expertise-measurement	Expertise-translation	Project management experience	Review opportunities	Training opportunities	Opportunities for progress documentation	Apt deadlines	User friendly materials	In-country support external to team	Support external to team outside country
Configure a team according to coordinating group's specifications.	(-1) Country C provided the US organizing agency limited communication about team members.	(-1) Neither of the two primary Country C team members had the required or desired measurement qualifications.	(1)The Country C team had translation experience.	(1) the Country C NPM met all suggested qualifications.	(1) Country C reviewed and chose persons for both positions--CAE did not review CVs afterwards	(-1) all Country C team members attended taining; no evidence of reading theoretical literature--Solano-Flores, Site, 2010	(1) Country C contact list and confidentiality agreements were emailed for each team member	(-1) Country C team was not able to fill team positions in time for initial meetings and document review	(1) Country C did not express difficulty using documents provided for team configuration	(1) Country C team had support from government and academia to configure team; however, assessment expert did not participate as originally planned	(1) Country C received support from the international coordinating agency and US organizing agency when configuring the team.
Acquire funding for all steps of the process.	(1) Country C received timely communication about project costs	(-1) no evidence Country C funded assessment expert for duration of project	(1) Country C funded translation team for duration of project	(1) Country C funded the NPM position for duration of project	(1) Country C had opportunities to review budget	(-1) Country C did not have training opportunities available for budget	(1) The US organizing agency collected data on funding	(1) Country C only experienced difficulty with funding when finaliz	(1) Country C team did not express difficulty using documents provid	(1) Country C had financial support from government and academia.	(1) With regard to funding activities, Country C received suppor

COUNTRY C	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
							acquisition progress for Country C.	ing number of PTs for the study.	ed for acquiring funds.		t from the international and US organizing agencies.
Select test items based on specific criteria established by the coordinating group.	(1) there was timely communication for Country C to select PTs for the study.	(1) Country C used measurement expertise when selecting performance tasks for the study.	(1) Country C used translation expertise when selecting performance tasks for the study.	(1) Country C used project management experience when selecting performance tasks for the study.	(1) Country C participated in two reviews during PT selection process	(1) although there was no official training, Country C was able to learn and apply selection criteria provided through documentation	(1) Country C helped the US organizing agency's project manager document progress and report it to the international organizing agency.	(1) Country C team was able to complete all activities associated with final PT selection within the three weeks given.	(1) Country C team did not express difficulty using documents provided for PT selection	(-1) Country C did not receive additional in-country support external to the team	(1) Country C received support from other country teams, the international organizing agency, and US organizing agency
Acquire necessary technical infrastructure.	(1) Country C got timely communication regarding day-to-day communicati	(0) There is no evidence that Country C had a measurement expert participating in the	(1) The Country C translation experts had access to all of the technology necess	(1) Country C used project management experience when acquiring the technical	(1) Country C had opportunities to review acquisition of technical infrastructure.	(1) Country C participated in training and received documentation addres	(1) The US organizing agency provided opportunities to document	(1) Country C team was able to acquire the technical infrastructure neede	(1) Country C team did not express difficulty using documents addressing	(1) government agencies and academia provided Country C with the necess	(-1) When acquiring the necessary technical infrastructure Country C did not receive

COUNTRY C	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	on, testing the computer interface, and implementing the assessment	study.	ary.	infrastructure for the study.		sing technical infrastructure.	progress for Country C.	d throughout the study.	acquisition of technical infrastructure.	ary technical infrastructure.	, or require, support from outside of the country.
Adapt test based on agreed upon cultural adaptation suggestions.	(-1) During PT adaptation, the USPI provided timely communication but the Country C team did not.	(0) There is no evidence that Country C had a measurement expert participating in the study.	(-1) Country C did not include translation expertise during the adaptation process.	(-1) Country C addressed most of the adaptation issues; however, they did not participate in each step of the adaptation process in a timely way.	(1) Country C participated in almost every review opportunity available during adaptation.	(1) Country C participated in training and received documentation addressing PT adaptation.	(1) Country C participated in every opportunity except for one to document progress on task adaptation.	(1) Country C team was able to meet all but one deadline throughout the adaptation process.	(-1) During adaptation Country C did not address all of the topics discussed in the materials supplied.	(-1) During adaptation Country C did not have in-country support external to the team.	(1) During adaptation Country C received support from the US organizing agency.
Hire translators possessing qualifications set by coordinating group.	(1) Country C team did not indicate any challenges with timely comm	(-1) Country C translators did not have measurement expertise.	(1) Although there was no certification process available in the	(1) Country C used project management experience when hiring transla	(-1) There was no review opportunityfor Country C during the hiring	(1) Country C participated in training and received documentatio	(1) Alhtough they were not officially planned activities,	(-1) Deadlines caused Country C some challenges when hiring transla	(1) Country C team did not express difficulty using documents	(1) While hiring translators Country C had in-country support	(1) When hiring translators Country C received support from the US

COUNTRY C	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	unication during hiring of translators.		country, Country C translators were professionals with appropriate experience	tors for the study.	of translators.	n addressing hiring translators.	Country C participated in every opportunity to document progress on hiring translators.	tors.	provided for hiring translators.	external to the team.	organizing agency.
Translate the assessment.	(-1) There was timely communication during translation on the part of the US organizing agency but not always from the Country C team.	(-1) Following criteria provided, Country C's translators did not have measurement expertise. However, they addressed challenges in scoring.	(1) Country C's translation team had the desired translation expertise.	(1) Country C's project management experience was helpful during the translation process.	(-1) There is no evidence of a review taking place during Country C's initial translation phase.	(1) Country C participated in training and received documentation addressing translation and reconciliation.	(1) The Country C team and US organizing agency documented team progress in translation and reconciliation.	(1) The Country C team completed translation activities by the site visit; however, the site visit took place in the fall 2010.	(1) Country C team did not express difficulty using documents provided for translation and reconciliation.	(1) Translators from Country C completed the translation process.	(1) Country C received support from the US organizing agency with the translation process.
Review translation and	(1) There was timely comm	(-1) There is no evidence that	(1) Country C's translation	(0) There is limited inform	(1) Country C had opport	(1) Country C participated	(1) The Country C transla	(-1) The US organizing	(1) Country C team did not	(1) Country C had in-countr	(1) Country C received

COUNTRY C	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
notes from translation process.	unication during translation review between the Country C team and US organizing agency.	Country C's translation review team included expertise in measurement.	review team had the desired translation expertise.	ation about Country C's translation review team's project management expertise.	unities to review their work as they completed the translation review process.	in training and received documentation addressing translation review.	tion review team had opportunities to document their progress during the translation review process.	agency did not provide the Country C team with a deadline for the translation review process.	express difficulty using documents provided for translation review.	y support during the translation review process that was external to the team.	support from the US organizing agency with the translation review process.
Translate ancillary materials as described by coordinating group.	(-1) There was timely communication on the part of the US agency during the dual translation but not from the Country C team.	(-1) Following criteria provided, Country C did not include measurement expertise during the translation of ancillary amterials.	(1) Country C's translation review team had the desired translation expertise.	(-1) It was difficult for Country C's project manager to deal with local religious holidays and schedule.	(-1) There is no evidence that Country C reviewed translation of ancillary materials.	(1) Country C participated in training and received documentation addressing ancillary material translation.	(1) The Country C team and US organizing agency documented team progress in translation of ancillary material.	(-1) The US organizing agency did not provide the Country C team with apt deadlines for translating all ancillary materials.	(1) Country C team did not express difficulty using documents provided for the dual translation of ancillary materials. However, the team did not apply the procedure as	(1) An academic from Country C, who had trasnlation expertise, completed the dual translation process of ancillary materials.	(1) Country C received support from the US organizing agency with the dual translation process to be used with the ancillary materials.

COUNTRY C	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
									planned.		
Review translation of material for assessment implementation.	(1) Communication between the US organizing agency and Country C team addressing external translation of material for assessment implementation was timely.	(0) There is no information on the qualifications of the US translators who completed the translation of material for assessment implementation in Country C.	(0) There is limited information about Country C's review of external translation of material for assessment implementation.	(0) There is limited information about Country C's review of external translation of material for assessment implementation.	(1) Country C had opportunities to review their work as they reviewed the external translation of material for assessment implementation.	(1) Country C participated in training and received documentation addressing translation review of materials for assessment implementation.	(1) The Country C team had opportunities to document their progress during the review of external translations.	(1) Country C did not express difficulty in completing the review in time to upload the material onto the internet platform.	(1) Country C team did not express difficulty using documents provided for reviewing external translation of material for assessment implementation.	(0) There is limited information about Country C's support during the review of external translation of material for assessment implementation.	(1) Country C received support from the US organizing agency while reviewing the external translation of material for assessment implementation.
Implement changes based on verification procedures prescribed by coordinating group.	(0) US communication during translation verification was timeley, however, there	(0) There is limited information on the qualifications of Country C team memb	(0) There is limited information about Country C's review of translation verification	(0) There is limited information about Country C's review of translation verification	(0) Country C received limited guidelines on how to review translation verification sugges	(-1) Country C did not have training opportunities for reviewing translation verfica	(-1) Although there were opportunities to document Country C's progress, the US	(-1) The US organizing agency did not include a deadline for implementing	(1) Country C team did not express difficulty using documents provided for imple	(0) There is limited information about Country C's support in implementing change	(1) Country C received support from the US organizing agency while implementing

COUNTRY C	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	is limited information on Country C's communication.	ers who reviewed translation verification results.	suggestions.	suggestions.	tions.	tion results.	organizing agency did not keep documentation.	changes resulting from translation verification in the study work plan.	menting changes from translation verification.	s from translation verification.	changes from translation verification.
Make agreed upon changes resulting from validation procedure established by coordinating group.	(-1) Although communication by the US organizing agency was timely during validation procedures, the Country C team had challenges.	(0) There is no information on Country C's staff who conducted the cognitive labs.	(0) There is no information on Country C's staff who conducted the cognitive labs.	(1) Country C had project management experience while making changes based on results from validation procedure.	(-1) Country C participated in one of two review opportunities.	(1) Country C participated in training and had access to training material addressing how to conduct and use notes from cognitive labs.	(1) The Country C team had opportunities to document their progress during the cognitive lab process.	(-1) Country C was not able to complete the cognitive lab activities on time.	(1) Country C found materials for cognitive labs easy to use.	(1) Country C had in-country support during the validation procedure.	(1) Country C had external country support during the validation procedure.
Test assessment implementation process for target popula	(-1) Country C did not test the assessment implementation	(-1) Country C did not test the assessment imple	(-1) Country C did not test the assessment implementation	(-1) Country C did not test the assessment implementation	(-1) Country C did not test the assessment implementation	(-1) Country C did not test the assessment implementation	(-1) Country C did not test the assessment imple	(-1) Country C did not test the assessment imple	(-1) Country C did not test the assessment implementation	(-1) Country C did not test the assessment implementation	(-1) Country C did not test the assessment implementation

COUNTRY C	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
tion usability.	process with their target population.	mentation process with their target population.	process with their target population.	process with their target population.	process with their target population.	process with their target population.	mentation process with their target population.	mentation process with their target population.	process with their target population.	process with their target population.	process with their target population.
Provide students with an opportunity to become familiar with test format and expectations.	(1) Communication between the US organizing agency and Country C team about the mini-PT was timely.	(0) There is no information on Country C's staff who worked on the mini-PT.	(0) Information is unclear as to the translation expertise available while Country C was working on the mini-PT.	(1) Country C had project management experience while working on the mini-PT.	(1) Country C had opportunities for review when working with the mini-performance task.	(1) Country C participated in training and had access to training material addressing the mini-performance task.	(1) Country C had opportunities to document their progress while working with the mini-PT.	(1) Country C did not express difficulty in completing the work for the mini-PT on time.	(1) Country C team did not express difficulty using documents provided for working with the mini-PT.	(1) Country C had in-country support during the work with the mini-PT.	(1) Country C had external country support while working with the mini-PT.
Hire scorers according to coordinating group's specifications.	(1) Communication about scorers was timely between the US organizing agency and Country C.	(1) Country C scorers acquired measurement expertise through the study's training.	(-1) There is no evidence that Country C's scorers possessed translation expertise.	(1) Country C successfully used project management expertise while hiring scorers.	(1) Country C's scorers participated in training reviews.	(1) Country C team members and scorers took part in training.	(-1) There were no opportunities for Country C to document the progress of hiring scorers.	(1) Country C met all deadlines associated with scorers.	(1) Country C did not express finding the documents addressing scorer hiring challenging.	(0) There is limited information about who Country C hired for scoring.	(1) Country C had external country support while hiring scorers.

COUNTRY C	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
Attend kick-off meeting as well as in-person and phone meetings to discuss progress.	(1) Communication about meetings was timely between Country C and organizing agencies.	(-1) There is no indication that Country C included measurement expertise in any meeting.	(-1) Country C did not include translation expertise in any of the study's meetings.	(1) Country C successfully used project management expertise to attend meetings.	(-1) Country C did not participate in any review opportunities while attending meetings.	(-1) There was no need or opportunity for Country C to train to attend meetings.	(1) Organizing and coordinating agencies documented information about meetings in which Country C participated.	(1) Country C did not indicate challenges with dates for planned meetings.	(1) Country C did not indicate that the material addressing meetings was challenging.	(-1) In-country support external to the team was not required or necessary for the Country C team.	(1) Country C received support with meetings from the US organizing agency and international coordinating agency.
Submit feedback on process and technical reports regarding progress.	(-1) Communication between Country C and organizing agencies regarding feedback was not always timely.	(-1) There is no indication that Country C included measurement expertise when providing feedback.	(1) Country C included translation expertise in their feedback.	(1) Country C successfully used project management expertise to provide feedback.	(-1) There were no review opportunities while Country C provided feedback.	(-1) There was no need or opportunity for Country C to train to provide feedback.	(1) The US organizing agency documented Country C's feedback throughout the study.	(1) Country C did not indicate challenges with dates for providing feedback.	(-1) Country C provided feedback during meetings, conference calls, and via emails; no materials were needed.	(1) Country C had in-country support while providing feedback about the study.	(-1) Country C did not require support external to the team and outside of the country to provide feedback.
Recruit institutions and student	(1) Country C received timely	(0) There is no evidence that Countr	(-1) Translation expertise was not	(0) There is limited information	(0) There is no informaiton about	(1) Country C participated in	(1) An international coordinating agenc	(0) It is unclear if Country C	(1) Country C did not indicate that	(1) Country C had in-country	(1) When recruiting HEIs and

COUNTRY C	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
s to participate in the assessment.	communication from the organizing agencies regarding recuirtment of HEIs and students.	y C had a measurement expert participating in the study.	necessary while Country C recruited HEIs and students.	about the extent to which Country C's NPM was involved in HEI and student recruitment.	opportunities for review while Country C recruited HEIs and students.	training and received documentation addressing recruitment and sampling of HEIs and students.	y documented Country C's progress in sampling.	found the due dates for HEIs and student recruitment and sampling challenging.	the material addressing recruiting and sampling difficult to use.	support to recruit and sample HEIs and students.	students Country C received support from an international coordinating agency.

COUNTRY D	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
Tasks	Timely communication	Expertise-measurement	Expertise-translation	Project management experience	Review opportunities	Training opportunities	Opportutnities for progress documentation	Apt deadlines	User friendly materials	In-country support external to team	Support external to team outside country
Configure a team according to coordinating group's specifications.	(1) Country D and the US organizing agency shared timely communication about configuring the team.	(1) A team member for Country D was national assessment expert and had academic background in measurement.	(1)The Country D team had translation experience.	(1) the Country D NPM met amost alll suggested qualifications: no assessment background, but, worked closely with the national assessment expert.	(1) Country D reviewed and chose persons for both positions--CAE reviewed CVs afterwards	(1) all Country D team members participated in all training	(1) contact list and confidentiality agreements were emailed for each Country D team member	(1) Country D was able to fill team positions in time for initial meetings and document review	(1) Country D did not express difficulty using documents provided for team configuration	(1) Country D had support from government and academia to configure team	(1) Country D received support from the international coordinating agency and US organizing agency when configuring the team.
Acquire funding for all steps of the process.	(1) Country D team received timely communication about project costs	(1) Country D funded assessment expert for duration of project	(1) Country D funded translation team for duration of project	(1) Country D funded the NPM position for duration of project	(1) Country D had opportunities to review budget	(-1) Country D did not have training opportunities available for budget	(1) The US organizing agency collected data on funding	(1) Country D only experienced difficulty with funding when finaliz	(1) Country D team did not express difficulty using documents provid	(1) Country D had financial support from government and academia.	(1) With regard to funding activities, Country D received suppor

COUNTRY D	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
							acquisition progress for Country D.	ing number of PTs for the study.	ed for acquiring funds.		t from the international and US organizing agencies.
Select test items based on specific criteria established by the coordinating group.	(1) there was timely communication for Country D to select PTs for the study.	(1) Country D used measurement expertise when selecting performance tasks for the study.	(-1) Country D team did not possess translation expertise when selecting performance tasks for the study.	(1) Country D used project management experience when selecting performance tasks for the study.	(1) Country D participated in two reviews during PT selection process	(1) although there was no official training, Country D was able to learn and apply selection criteria provided through documentation	(1) Country D helped the US organizing agency's project manager document progress and report it to the international organizing agency.	(1) Country D team was able to complete all activities associated with final PT selection within the three weeks given.	(1) Country D team did not express difficulty using documents provided for PT selection	(-1) Country D did not receive additional in-country support external to the team	(1) Country D received support from other country teams, the international organizing agency, and US organizing agency
Acquire necessary technical infrastructure.	(1) Country D got timely communication regarding day-to-day communicati	(1) The Country D measurement expert had access to all of the technology necess	(1) The Country D translation experts had access to all of the technology necess	(1) Country D used project management experience when acquiring the technical	(1) Country D had opportunities to review acquisition of technical infrastructure.	(1) Country D participated in training and received documentation addres	(1) The US organizing agency provided opportunities to document	(1) Country D team was able to acquire the technical infrastructure neede	(1) Country D required minimal clarification on documentation regardi	(1) government agencies and academia provided Country D with the necess	(-1) When acquiring the necessary technical infrastructure Country D did not receive

COUNTRY D	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	on, testing the computer interface, and implementing the assessment	ary.	ary.	infrastructure for the study.		sing technical infrastructure.	progress for Country D.	d throughout the study.	ng acquisition of technical infrastructure.	ary technical infrastructure.	, or require, support from outside of the country.
Adapt test based on agreed upon cultural adaptation suggestions.	(1) There was timely communication for PT adaptation between the Country D team and US PI.	(1) Country D measurement expert participated in the PT adaptation process.	(-1) Country D did not include translation expertise during the adaptation process.	(-1) Although Country D participated in each step of adaptation process, they did not address all issues of adaptation.	(1) Country D participated in every review opportunity available during adaptation.	(1) Country D participated in training and received documentation addressing PT adaptation.	(1) Country D participated in every opportunity to document progress on task adaptation.	(1) Country D team was able to meet deadlines throughout the adaptation process.	(-1) During adaptation Country D did not address all of the topics discussed in the materials supplied.	(-1) During adaptation Country D did not have in-country support external to the team.	(1) During adaptation Country D received support from the US organizing agency.
Hire translators possessing qualifications set by coordinating group.	(1) Country D team did not indicate any challenges with timely communication during hiring of	(1) A Country D translators had some experience with measurement tools.	(-1) Country D selected their translation team members guided by their own connections rather	(1) Country D used project management experience when hiring translators for the study.	(-1) There was no review opportunityfor Country D during the hiring of translators.	(1) Country D participated in training and received documentation addressing hiring transla	(1) Alhtough they were not officially planned activities, Country D participated in	(1) Deadlines did not cause Country D challenges when hiring translators.	(1) Country D team did not express difficulty using documents provided for hiring translators.	(1) While hiring translators Country D had in-country support external to the team.	(1) When hiring translators Country D received support from the US organizing agency.

COUNTRY D	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	translators.		than their professional qualifications			tors.	every opportunity to document progress on hiring translators.				
Translate the assessment.	(1) There was timely communication during translation between the Country D team and US organizing agency.	(-1) Following criteria provided, Country D's translators did not have measurement expertise. However, they addressed challenges in scoring.	(-1) Country D's translation team did not have the desired translation expertise.	(1) Country D's project management experience was helpful during the translation process.	(-1) There is no evidence of a review taking place during Country D's initial translation phase.	(1) Country D participated in training and received documentation addressing translation and reconciliation.	(1) The Country D team and US organizing agency documented team progress in translation and reconciliation.	(-1) The Country D translators did not complete all activities in time for the site visit.	(1) Country D team did not express difficulty using documents provided for translation and reconciliation.	(1) Translators from Country D completed the translation process.	(1) Country D received support from the US organizing agency with the translation process.
Review translation and notes from translation process.	(1) There was timely communication during translation review between the	(1) Country D's translation review team included measurement expertise.	(-1) Country D's translation review team did not have the desired translation	(0) There is limited information about Country D's translation review team's	(1) Country D had opportunities to review their work as they completed	(1) Country D participated in training and received documentation	(1) The Country D translation review team had opportunities to docum	(-1) The US organizing agency did not provide the Country D team	(1) Country D team did not express difficulty using documents provid	(-1) Country D did not seek in-country support for the translation	(1) Country D received support from the US organizing agency with the

COUNTRY D	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	Country D team and US organizing agency.		expertise.	project management expertise.	the translation review process.	addressing translation review.	ent their progress during the translation review process.	with a deadline for the translation review process.	ed for translation review.	review process that was external to the team.	translation review process.
Translate ancillary materials as described by coordinating group.	(1) There was timely communication between the Country D team and US organizing agency during translation of ancillary materials.	(-1) Following criteria provided, Country D did not include measurement expertise during the translation of ancillary amterials.	(-1) Country D's translation review team did not have the desired translation expertise.	(1) Country D's project management experience was helpful during the dual translation process.	(-1) There is no evidence that Country D reviewed translation of ancillary materials.	(1) Country D participated in training and received documentation addressing ancillary material translation.	(1) The Country D team and US organizing agency documented team progress in translation and reconciliation of ancillary material.	(-1) The US organizing agency did not provide the Country D team with apt deadlines for translating all ancillary materials.	(1) Country D team did not express difficulty using documents provided for the dual translation of ancillary materials.	(1) Translators from Country D completed the dual translation process of ancillary materials.	(1) Country D received support from the US organizing agency with the dual translation process to be used with the ancillary materials.
Review translation of material for assessment implementation.	(1) Communication between the US organizing agency and Countr	(1) Country D's assessment expert helped review external translations	(0) There is limited information about Country D's review of extern	(0) There is limited information about Country D's review of extern	(1) Country D had opportunities to review their work as they review	(1) Country D participated in training and received documentatio	(1) The Country D team had opportunities to document their	(1) Country D did not express difficulty in completing the	(1) Country D team did not express difficulty using documents	(0) There is limited information about Country D's support during	(1) Country D received support from the US organizing agency while

COUNTRY D	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	y D team addressing external translation of material for assessment implementation was timely.	of material for assessment implementation.	al translation of material for assessment implementation.	al translation of material for assessment implementation.	ed the external translation of material for assessment implementation.	n addressing translation review of materials for assessment implementation.	progress during the review of external translations.	review in time to upload the material onto the internet platform.	provided for reviewing external translation of material for assessment implementation.	the review of external translation of material for assessment implementation.	reviewing the external translation of material for assessment implementation.
Implement changes based on verification procedures prescribed by coordinating group.	(1) Communication between the US organizing agency and Country D team during verifiation was timely.	(1) A Country D team member with measurement expertise helped review translations verification notes.	(0) There is limited information about Country D's review of translation verification suggestions.	(1) Country D's project management experience was helpful while reviewing translation verification information.	(0) Country D received limited guidelines on how to review translation verification suggestions.	(-1) Country D did not have training opportunities for reviewing translation verfication results.	(1) Country D had opportunities to document their progress in implementing changes based on translation verification.	(-1) Although the US agency did not include this activity in the study's workplan, Country D completed the work quickly.	(-1) Country D had difficulty working with changes suggested from translation verification.	(0) There is limited information about Country D's support in implementing changes from translation verification.	(1) Country D received support from the US organizing agency while implementing changes from translation verification.
Make agreed upon changes resulting from validat	(1) Communication between the US organizing	(1) The Country D staff in charge of making	(-1) The staff from Country D conducting the	(1) Country D had project management experience	(-1) Country D participated in one of two review opport	(1) Country D participated in training and had	(1) The Country D team had opportunities to	(-1) Country D was not able to complete the cogniti	(-1) Country D had difficulty working with materi	(1) Country D had in-country support during	(1) Country D had external country suppor

COUNTRY D	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
ion procedure established by coordinating group.	agency and Country D team during validation procedures was timely.	changes based on the lab results possessed measurement expertise.	labs and making changes resulting from the process did not have expertise in translation.	while making changes based on results from validation procedure.	unities.	access to training material addressing how to conduct and use notes from cognitive labs.	document their progress during the cognitive lab process.	ve lab activities on time.	al provided for conducting cognitive labs.	the validation procedure.	t during the validation procedure.
Test assessment implementation process for target population usability.	(-1) Country D did not test the assessment implementation process with their target population.	(-1) Country D did not test the assessment implementation process with their target population.	(-1) Country D did not test the assessment implementation process with their target population.	(-1) Country D did not test the assessment implementation process with their target population.	(-1) Country D did not test the assessment implementation process with their target population.	(-1) Country D did not test the assessment implementation process with their target population.	(-1) Country D did not test the assessment implementation process with their target population.	(-1) Country D did not test the assessment implementation process with their target population.	(-1) Country D did not test the assessment implementation process with their target population.	(-1) Country D did not test the assessment implementation process with their target population.	(-1) Country D did not test the assessment implementation process with their target population.
Provide students with an opportunity to become familiar with test	(1) Communication between the US organizing agency and Country D	(1) A Country D team member with measurement expertise helped with the	(-1) The Country D team, which worked with the mini-PT, did not posses	(1) Country D had project management experience while working on the	(1) Country D had opportunities for review when working with the mini-	(1) Country D participated in training and had access to training	(1) Country D had opportunities to document their progress while	(1) Country D did not express difficulty in completing the work	(1) Country D team did not express difficulty using documents provid	(1) Country D had in-country support during the work with the	(1) Country D had external country support while working with the

COUNTRY D	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
format and expectations.	team about the mini-PT was timely.	mini-PT.	s translation expertise.	mini-PT.	performance task.	material addressing the mini-performance task.	working with the mini-PT.	for the mini-PT on time.	ed for working with the mini-PT.	mini-PT.	mini-PT.
Hire scorers according to coordinating group's specifications.	(1) Communication about scorers was timely between the US organizing agency and Country D.	(1) Country D scorers acquired measurement expertise through the study's training.	(-1) There is no evidence that Country D's scorers possessed translation expertise.	(1) Country D successfully used project management expertise while hiring scorers.	(1) Country D's scorers participated in training reviews.	(1) Country D team members and scorers took part in training.	(-1) There were no opportunities for Country D to document the progress of hiring scorers.	(1) Country D met all deadlines associated with scorers.	(1) Country D did not express finding the documents addressing scorer hiring challenging.	(0) There is limited information about who Country D hired for scoring.	(1) Country D had external country support while hiring scorers.
Attend kick-off meeting as well as in-person and phone meetings to discuss progress.	(1) Communication about meetings was timely between Country D and organizing agencies.	(1) Country D included a team member with measurement expertise in all telephone and some in-person meetings.	(-1) Country D did not include translation expertise in any of the study's meetings.	(1) Country D successfully used project management expertise to attend meetings.	(-1) Country D did not participate in any review opportunities while attending meetings.	(-1) There was no need or opportunity for Country D to train to attend meetings.	(1) Organizing and coordinating agencies documented information about meetings in which Country D participated.	(1) Country D did not indicate challenges with dates for planned meetings.	(1) Country D did not indicate that the material addressing meetings was challenging.	(-1) In-country support external to the team was not required or necessary for the Country D team.	(1) Country D received support with meetings from the US organizing agency and international coordinating agency.

COUNTRY D	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
Submit feedback on process and technical reports regarding progress.	(-1) Communication between Country D and organizing agencies regarding feedback was not always timely.	(1) Country D included measurement expertise in feedback provided to organizing and coordinating agencies.	(-1) Country D did not include translation expertise in feedback.	(1) Country D successfully used project management expertise to provide feedback.	(-1) There were no review opportunities while Country D provided feedback.	(-1) There was no need or opportunity for Country D to train to provide feedback.	(1) The US organizing agency documented Country D's feedback throughout the study.	(1) Country D did not indicate challenges with dates for providing feedback.	(-1) Country D provided feedback during meetings, conference calls, and via emails; no materials were needed.	(1) Country D had in-country support while providing feedback about the study.	(-1) Country D did not require support external to the team and outside of the country to provide feedback.
Recruit institutions and students to participate in the assessment.	(1) Country D received timely communication from the organizing agencies regarding recuirtment of HEIs and students.	(0) There is limited information about the extent to which Country D's measurement expert was involved in HEI and student recruitment.	(-1) Translation expertise was not necessary while Country D recruited HEIs and students.	(0) There is limited information about the extent to which Country D's NPM was involved in HEI and student recruitment.	(0) There is no informaiton about opportunities for review while Country D recruited HEIs and students.	(1) Country D participated in training and received documentation addressing recruitment and sampling of HEIs and students.	(1) An international coordinating agency documented Country D's progress in sampling.	(0) It is unclear if Country D found the due dates for HEIs and student recruitment and sampling challenging.	(1) Country D did not indicate that the material addressing recruiting and sampling difficult to use.	(1) Country D had in-country support to recruit and sample HEIs and students.	(1) When recruiting HEIs and students Country D received support from an international coordinating agency.

COUNTRY E	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
Tasks	Timely communication	Expertise-measurement	Expertise-translation	Project management experience	Review opportunities	Training opportunities	Opportunities to document progress	Apt deadlines	User friendly materials	In-country support external to team	Support external to team outside country
Configure a team according to coordinating group's specifications.	(1) Country E and the US organizing agency shared timely communication about configuring the team.	(1) A team member for Country E was a national assessment expert and had academic background in measurement.	(1)The Country E team had translation experience.	(1) the Country E NPM met all suggested qualifications.	(1) Country E reviewed and chose persons for both positions--CAE reviewed CVs afterwards	(1) all Country E team members participated in all training	(1) contact list and confidentiality agreements were emailed for each Country E team member	(1) Country E was able to fill team positions in time for initial meetings and document review	(1) Country E did not express difficulty using documents provided for team configuration	(1) Country E had support from government and academia to configure team	(1) Country E received support from the international coordinating agency and US organizing agency when configuring the team.
Acquire funding for all steps of the process.	(1) Country E team received timely communication about project costs	(1) Country E funded assessment expert for duration of project	(1) Country E funded translation team for duration of project	(1) Country E funded the NPM position for duration of project	(1) Country E had opportunities to review budget	(-1) Country E did not have training opportunities available for budget	(1) The US organizing agency collected data on funding acquisition progress for	(1) Country E only experienced difficulty with funding when finalizing number of PTs	(1) Country E team did not express difficulty using documents provided for acquiring funds.	(1) Country E had financial support from government and academia.	(1) With regard to funding activities, Country E received support from the international

COUNTRY E	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
							Country E.	for the study.			and US organizing agencies.
Select test items based on specific criteria established by the coordinating group.	(1) there was timely communication for Country E to select PTs for the study.	(1) Country E used measurement expertise when selecting performance tasks for the study.	(1) Country E used translation expertise when selecting performance tasks for the study.	(1) Country E used project management experience when selecting performance tasks for the study.	(1) Country E participated in two reviews during PT selection process	(1) although there was no official training, Country E was able to learn and apply selection criteria provided through documentation	(1) Country E helped the US organizing agency's project manager document progress and report it to the international organizing agency.	(1) Country E team was able to complete all activities associated with final PT selection within the three weeks given.	(1) Country E team did not express difficulty using documents provided for PT selection	(-1) Country E did not receive additional in-country support external to the team	(1) Country E received support from other country teams, the international organizing agency, and US organizing agency
Acquire necessary technical infrastructure.	(1) Country E got timely communication regarding day-to-day communication, testing the compu	(1) The Country E measurement expert had access to all of the technology necessary.	(1) The Country E translation experts had access to all of the technology necessary.	(1) Country E used project management experience when acquiring the technical infrastructure for the study.	(1) Country E had opportunities to review acquisition of technical infrastructure.	(1) Country Ey participated in training and received documentation addressing technical infrastr	(1) The US organizing agency provided opportunities to document progress for Country E.	(1) Country E team was able to acquire the technical infrastructure needed throughout the	(1) Country E team did not express difficulty using documents addressing acquisition of technical	(1) government agencies and academia provided Country E with the necessary technical infrastr	(-1) When acquiring the necessary technical infrastructure Country E did not receive, or require, suppor

COUNTRY E	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	ter interface, and implementing the assessment					ucture.		study.	infrastructure.	ucture.	t from outside of the country.
Adapt test based on agreed upon cultural adaptation suggestions.	(1) There was timely communication for PT adaptation between the Country E team and US PI.	(1) Country E measurement expert participated in the PT adaptation process.	(1) Country E included translation expertise during the adaptation process.	(-1) Although Country E participated in each step of adaptation process, they did not address all issues of adaptation.	(1) Country E participated in every review opportunity available during adaptation.	(1) Country E participated in training and received documentation addressing PT adaptation.	(1) Country E participated in every opportunity to document progress on task adaptation.	(1) Country E team was able to meet deadlines throughout the adaptation process.	(-1) During adaptation Country E did not address all of the topics discussed in the materials supplied.	(1) During adaptation Country E had in-country support external to the team.	(1) During adaptation Country E received support from the US organizing agency.
Hire translators possessing qualifications set by coordinating group.	(1) Country E team did not indicate any challenges with timely communication during hiring of translators.	(1) Country E translators had measurement expertise.	(1) although there was no certification, Country E translators were professionals with appropriate experience	(1) Country E used project management experience when hiring translators for the study.	(-1) There was no review opportunity for Country E during the hiring of translators.	(1) Country E participated in training and received documentation addressing hiring translators.	(1) Alhtough they were not officially planned activities, Country E participated in every opportunity to	(1) Deadlines did not cause Country E challenges when hiring translators.	(1) Country E team did not express difficulty using documents provided for hiring translators.	(1) While hiring translators Country E had in-country support external to the team.	(1) When hiring translators Country E received support from the US organizing agency.

COUNTRY E	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
							document progress on hiring translators.				
Translate the assessment.	(-1) There was timely communication during translation on the part of the US organizing agency but not always from the Country E team.	(1) Country E's measurement experts took part in translation process.	(1) Country E's translation team had the desired translation expertise.	(1) Country E's project management experience was helpful during the translation process.	(1) Country E included an unplanned and unrequired review during the initial phase of the translation process.	(1) Country E participated in training and received documentation addressing translation and reconciliation.	(1) The Country E team and US organizing agency documented team progress in translation and reconciliation.	(1) The Country E team was able to complete translation prior to the site visit.	(1) Country E team did not express difficulty using documents provided for translation and reconciliation.	(1) Translators from Country E completed the translation process.	(1) Country E received support from the US organizing agency with the translation process.
Review translation and notes from translation process.	(1) There was timely communication during translation review between the Country E team and US organi	(1) Country E's translation review team included measurement expertise.	(1) Country E's translation review team had the desired translation expertise.	(0) There is limited information about Country E's translation review team's project management expertise.	(1) Country E had opportunities to review their work as they completed the translation review process.	(1) Country E participated in training and received documentation addressing translation review.	(1) The Country E translation review team had opportunities to document their progress during the	(-1) The US organizing agency did not provide the Country E team with a deadline for the translation	(1) Country E team did not express difficulty using documents provided for translation review.	(1) Country E had in-country support during the translation review process that was external to the	(1) Country E received support from the US organizing agency with the translation review process.

COUNTRY E	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	zing agency.						translation review process.	review process.		team.	
Translate ancillary materials as described by coordinating group.	(1) There was timely communication between the Country E team and US organizing agency during translation of ancillary materials.	(1) Although the process did not demand it, Country E included measurement expertise during the translation of ancillary materials.	(1) Country E's translation review team had the desired translation expertise.	(1) Country E's project management experience was helpful during the dual translation process.	(1) Country E included an unplanned and unrequired review of ancillary material translation.	(1) Country E participated in training and received documentation addressing ancillary material translation.	(1) The Country E team and US organizing agency documented team progress in translation and reconciliation of ancillary material.	(-1) The US organizing agency did not provide the Country E team with apt deadlines for translating all ancillary materials.	(1) Country E team did not express difficulty using documents provided for the dual translation of ancillary materials.	(1) Translators from Country E completed the dual translation process of ancillary materials.	(1) Country E received support from the US organizing agency with the dual translation process to be used with the ancillary materials.
Review translation of material for assessment implementation.	(1) Communication between the US organizing agency and Country E team addressing external	(1) Although the process did not require it, Country E included measurement expertise during the	(1) Country E included translation expertise during the review of the external translation of materi	(0) There is limited information about Country E's review of external translation of material for assess	(1) Country E had opportunities to review their work as they reviewed the external translation of materi	(1) Country E participated in training and received documentation addressing translation review	(1) The Country E team had opportunities to document their progress during the review of	(1) Country E did not express difficulty in completing the review in time to upload the	(1) Country E team did not express difficulty using documents provided for reviewing external	(0) There is limited information about Country E's support during the review of external transla	(1) Country E received support from the US organizing agency while reviewing the external translation of

COUNTRY E	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	translation of material for assessment implementation was timely.	review of the external translation of material for assessment implementation.	al for assessment implementation.	ment implementation.	al for assessment implementation.	of materials for assessment implementation.	external translations.	material onto the internet platform.	translation of material for assessment implementation.	tion of material for assessment implementation.	material for assessment implementation.
Implement changes based on verification procedures prescribed by coordinating group.	(1) Communication between the US organizing agency and Country E team during verifiation was timely.	(1) A Country E team member with measurement expertise helped review translations verification notes.	(1) Country E included translation expertise during the review of translation verification results.	(1) Country E's project management experience was helpful while reviewing translation verification information.	(0) Country E received limited guidelines on how to review translation verification suggestions.	(-1) Country E did not have training opportunities for reviewing translation verfication results.	(-1) Although there were opportunities to document Country E's progress, the US organizing agency did not keep documentation.	(-1) Although the US agency did not include this activity in the study's workplan, Country E completed the work within days.	(-1) Country E had difficulty working with changes suggested from translation verification.	(0) There is limited information about Country E's support in implementing changes from translation verification.	(1) Country E received support from the US organizing agency while implementing changes from translation verification.
Make agreed upon changes resulting from validation procedure	(1) Communication between the US organizing agency and Countr	(1) The Country E staff in charge of making changes based	(1) The Country E staff in charge of making changes based	(1) Country E had project management experience while making	(-1) Country E participated in one of two review opportunities.	(1) Country E participated in training and had access to trainin	(1) The Country E team had opportunities to document their	(-1) Country E was not able to complete the cognitive lab activities on	(1) Country E found materials for cognitive labs easy to use.	(1) Country E had in-country support during the validation	(1) Country E had external country support during the

COUNTRY E	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
established by coordinating group.	y E team during validation procedures was timely.	on the lab results possessed measurement expertise.	on the lab results possessed translation expertise.	changes based on results from validation procedure.		g material addressing how to conduct and use notes from cognitive labs.	progress during the cognitive lab process.	time.		procedure.	validation procedure.
Test assessment implementation process for target population usability.	(-1) Country E did not test the assessment implementation process with their target population.	(-1) Country E did not test the assessment implementation process with their target population.	(-1) Country E did not test the assessment implementation process with their target population.	(-1) Country E did not test the assessment implementation process with their target population.	(-1) Country E did not test the assessment implementation process with their target population.	(-1) Country E did not test the assessment implementation process with their target population.	(-1) Country E did not test the assessment implementation process with their target population.	(-1) Country E did not test the assessment implementation process with their target population.	(-1) Country E did not test the assessment implementation process with their target population.	(-1) Country E did not test the assessment implementation process with their target population.	(-1) Country E did not test the assessment implementation process with their target population.
Provide students with an opportunity to become familiar with test format and expectations.	(1) Communication between the US organizing agency and Country E team about the mini-	(1) The Country E team with measurement expertise helped with the mini-PT.	(1) The Country E staff responsible for mini-PT work had expertise in translation.	(1) Country E had project management experience while working on the mini-PT.	(1) Country E had opportunities for review when working with the mini-performance task.	(1) Country E participated in training and had access to training material addressing	(1) Country E had opportunities to document their progress while working with the	(1) Country E did not express difficulty in completing the work for the mini-PT on time.	(1) Country E team did not express difficulty using documents provided for working with the	(1) Country E had in-country support during the work with the mini-PT.	(1) Country E had external country support while working with the mini-PT.

COUNTRY E	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
	PT was timely.					the mini-performance task.	mini-PT.		mini-PT.		
Hire scorers according to coordinating group's specifications.	(1) Communication about scorers was timely between the US organizing agency and Country E.	(1) Country E scorers acquired measurement expertise through the study's training.	(-1) There is no evidence that Country E's scorers possessed translation expertise.	(1) Country E successfully used project management expertise while hiring scorers.	(1) Country E's scorers participated in training reviews.	(1) Country E team members and scorers took part in training.	(-1) There were no opportunities for Country E to document the progress of hiring scorers.	(1) Country E met all deadlines associated with scorers.	(1) Country E did not express finding the documents addressing scorer hiring challenging.	(0) There is limited information about who Country E hired for scoring.	(1) Country E had external country support while hiring scorers.
Attend kick-off meeting as well as in-person and phone meetings to discuss progress.	(1) Communication about meetings was timely between Country E and organizing agencies.	(1) Country E included members with measurement expertise in telephone and in-person meetings.	(-1) Country E did not include translation expertise in any of the study's meetings.	(1) Country E successfully used project management expertise to attend meetings.	(-1) Country E did not participate in any review opportunities while attending meetings.	(-1) There was no need or opportunity for Country E to train to attend meetings.	(1) Organizing and coordinating agencies documented information about meetings in which Country E participated.	(1) Country E did not indicate challenges with dates for planned meetings.	(1) Country E did not indicate that the material addressing meetings was challenging.	(-1) In-country support external to the team was not required or necessary for the Country E team.	(1) Country E received support with meetings from the US organizing agency and international coordinating agency.
Submit feedback on process and technic	(-1) Communication between	(1) Country E included measur	(1) Country E included transla	(1) Country E successfully used	(-1) There were no review opport	(-1) There was no need or opport	(1) The US organizing agenc	(1) Country E did not indicat	(-1) Country E provided feedba	(1) Country E had in-country	(-1) Country E did not require suppor

COUNTRY E	Evidence Type 1 (key: 1=CEB; 1=DEB; 0=N/A)	Evidence Type 2	Evidence Type 3	Evidence Type 4	Evidence Type 5	Evidence Type 6	Evidence Type 7	Evidence Type 8	Evidence Type 9	Evidence Type 10	Evidence Type 11
al reports regarding progress.	Country E and organizing agencies regarding feedback was not always timely.	ement expertise in feedback provided to organizing and coordinating agencies.	tion expertise in their feedback.	project management expertise to provide feedback.	unities while Country E provided feedback.	unity for Country E to train to provide feedback.	y documented Country E's feedback throughout the study.	e challenges with dates for providing feedback.	ck during meetings, conference calls, and via emails; no materials were needed.	support while providing feedback about the study.	t external to the team and outside of the country to provide feedback.
Recruit institutions and students to participate in the assessment.	(1) Country E received timely communication from the organizing agencies regarding recuirtment of HEIs and students.	(0) There is limited information about the extent to which Country E's measurement expert was involved in HEI and student recruitment.	(-1) Translation expertise was not necessary while Country E recruited HEIs and students.	(0) There is limited information about the extent to which Country E's NPM was involved in HEI and student recruitment.	(0) There is no informaiton about opportunities for review while Country E recruited HEIs and students.	(1) Country E participated in training and received documentation addressing recruitment and sampling of HEIs and students.	(1) An international coordinating agency documented Country E's progress in sampling.	(0) It is unclear if Country E found the due dates for HEIs and student recruitment and sampling challenging.	(1) Country E did not indicate that the material addressing recruiting and sampling difficult to use.	(1) Country E had in-country support to recruit and sample HEIs and students.	(1) When recruiting HEIs and students Country E received support from an international coordinating agency.

Appendix I: Detailed qualitative data about each cell.

Task 1: Configure a team according to coordinating group's specifications.

Country A

There was timely communication on the part of the Country A NPM with regard to configuration of the team. Within one week of initial contact by the US organizing agency the Country A NPM was able to provide detailed information about the team and national centre (Ursi, personal communication, January 12, 2010). At the beginning of 2009, months prior to initial contact, the Country A Ministry of Education authorized the Country A Institute for Education Research to be responsible for the AHELO project (Ursi, personal communication, January 12, 2010). Through this agreement, the NPM, the assessment expert, and a research assistant were chosen to work on AHELO and a curriculum vitae or biography was submitted for each. The NPM also provided the US organizing agency with a contact list with all necessary information for each team member (Ursi, personal communication, January 12, 2010). Throughout the project the NPM asked clarifying questions about qualifications for additional staff, such as translators, and provided logistical documents (e.g., confidentiality agreements) for new hires (Ursi, personal communication, April 22, 2010).

Country A was able to include a measurement expert on their national AHELO team who had obtained a doctorate in statistics (CAE, GS.11). The person was a senior researcher at the Country A Institute for Educational Research (FIER), which is part of a research university in Country A. At the time of the project the measurement expert had almost 20 years of experience teaching in the Department of Mathematics and Statistics at the same university. The measurement expert had also served as a researcher for the Centre of Excellence for Study of Variation, Contacts, and Change in English in the Department of Languages at the same higher education institution. The person had expertise in linear models, mixed and multilevel models, multivariate statistics, and survey statistics. In addition, the person demonstrated fluency in speaking as well as reading and writing English.

The national project manager for Country A was able to help with performance task translation and adaptation. The national project manager had experience with higher education assessments that took place across diverse European countries (Ursin, 2010). The Country A NPM contributed to each PTs adaptation process and submitted the final adaptation suggestions and final modified PTs (Ursin, personal communication, June 8, 2010).

The national project manager for the Country A team had all of the qualifications stipulated by the US organizing agency. The NPM was a senior researcher at the Country A Institute for Educational Research (CAE, 2010, GS.11). The Country A NPM had an earned masters and doctorate in education from a research university in Country A (Ursi, 2010, CV). The manager had successfully led five research projects, some of which overlapped. The projects included topics such as evaluating teacher competencies, internal quality assurance systems in Country A universities, the impact of the Bologna Process, and research group work. Having

published and presented finding from numerous papers the NPM was fluent in reading, writing, and speaking English as well as Country A, Swedish, and German.

In Country A, the Country A Ministry of Education (MinEdu) authorized the Country A Institute for Educational Research (FINHEEC) to be the body responsible for the AHELO project. As a result a national coordination group was appointed by the ministry of education consisting of seven members from higher education institutions, student unions, FINHEEC, and MinEdu (Ursi, personal communication, January 12, 2010). The national coordination group selected the national project manager creating good opportunities for review. It is important to note, however, that Country A provided the assessment expert's curriculum vitae to the US organizing agency for review (Ursi, personal communication, January 12, 2010).

During the discussion of the theoretical framework for the AHELO translation and adaptation process the Country A team was able to participate actively. They understood the level of detail and amount of work necessary; as a result they asked about the various ways that different documents could be handled during the translation process (CAE, GS.26, 2010). The Country A team members also asked about the technical requirements to administer the assessment (CAE, GS.26, 2010). The entire Country A team participated in the training that took place in New York City (CAE, GS.26, 2010). Of the core country team members, only the national project manager from the Country A team attended the training (Solano-Flores, G., Shavelson, R., & Chia, M., 2010, Visit). However, it became clear that the national project manager subsequently trained the another member of the team as the person was able to address aspects of the translation review and verification process during communication throughout the study (Hyytinen, personal communication, September 23, 2011).

The Country A national project manager collaborated well with the US project manager to document the addition of each team member. The national project manager emailed the project manager for the US organizing agency a contact list including all members of the Country A team (Ursin, personal communication, January 12, 2010). The Country A NPM also emailed a signed confidentiality agreement for each of the team members (Ursin, personal communication, January 27, 2010).

Country A was able to contract all team members prior to the initial conference call and meeting in New York City (Ursin, personal communication, January 12, 2010). Country A did not indicate any difficulty in having formed their team in time for the meetings.

The NPM for Country A acknowledge receipt of the information regarding team configuration and did not express experiencing any challenges with them (Ursin, personal communication, January, 12, 2010). The country NPM also provided country team member documents, as requested via documents, without any confusion (Ursin, personal communication, January 27, 2010).

Country A was able to find team members from various national agencies and universities. During the year prior to beginning the AHELO study, the Country A Ministry of Education (MinEdu) authorized that personnel be made available from the Country A Institute

for Educational Research (FIER) of the University of Jyväskyla. FIER then coordinated with the Helsinki University Centre for Research and Development of Higher Education and the Country A Higher Education Evaluation Council (FINHEEC) (Ursin, personal communication, January 12, 2010). The AHELO national team also received support from seven members from higher education institutions and student unions) (Ursin, personal communication, January 12, 2010). As a result the country had all of the support necessary to create the AHELO team in Country A.

The international organizing agency informed members of Country A Ministry of Education about the performance assessment as well as initial steps in forming a national team for the study (Ursin, personal communication, January 11, 2010). As a result, the international organizing agency was key during the initial stages of the study (Ursin, personal communication, January 11, 2010). The support from the international organizing agency helped country universities and agencies support the creation of a national team (Solano-Flores & Chia, 2010, Interview). The US organizing agency sent an initial email and held a conference call with Country A's NPM at the beginning of the project. These efforts gave country representatives an opportunity to ask questions about the creation of the team (Shavelson, personal communication, January 11, 2010).

Country B

Communication with the team for Country B regarding the configuration of the AHELO staff was sporadic. Two weeks after initial communication from the US organizing agency the Country B team responded to the email with names and contact information for three of the four team members (Choi, personal communication, February 1, 2010). One person was the assessment expert; the other the interim NPM. The biographical sketches and one curriculum vitae were submitted one day later (Choi, personal communication, February 2, 2010). Although the Country B team had some additions to the team throughout the process and emailed their contact information, their CVs were not provided.

One of the primary team members on the Country B team met the required and desired qualifications (CAE, GS.11). Having obtained a doctoral degree in social research methodology at a research institution of higher education in the United States, the Country B measurement expert was a professor in the department of education for a national university in Seoul. The expert specialized in statistical research design, statistical computer programming, and development of achievement tests. The person had led several government-funded research projects including the development of a collegiate higher order thinking ability scale, a national entrance examination, and the national system of school evaluation.

The team for Country B did not appear to have expertise in translation or assessment across diverse linguistic or cultural groups. Team members participated in the translation and adaptation process. However, the lack of translation expertise was evident in the initial translations created (Solano-Flores, Visit, 2010). The national project manager was responsible for compiling all information and final versions of the tasks (Young, personal communication, February 8, 2011).

Country B did not have one national project manager for the entire process—instead each of the two managed the project during different times of the translation and adaptation process. The first national project manager was a senior researcher at the Country B Educational Development Institute, which was the center for AHELO (CAE, 2010. GS.11). The NPM had experience working for the Country B Council for University Education. The person's research focus was on quality assurance in higher education and public funding of higher education. Having completed a doctorate at a research university in the United States, the first NPM was fluent in reading, writing, and speaking English. The second NPM also obtained an advanced degree in education (personal communication, January 18, 2010). At the time of the project the second NPM had several publications and worked as a researcher for higher education research projects through the project laboratory in Country B. The second NPM was also fluent in reading, writing, and speaking English. It is not clear that either of the two researchers who served as NPM had experience leading a study.

The process for filling both positions in Country B included several organizations and allowed for internal and external review of team members. Filling the position of national assessment expert was completed at the country level. The Country B Educational Development Institute (KEDI), which housed the national Center of the AHELO project in Country B, handled initial team organization (CAE, 2010, GS.11). The interim national project manager provided the project manager for the US organizing agency with each team member's curriculum vitae after they were chosen by KEDI (Choi, personal communication, February 1, 2010). However, the project manager for the US organizing agency reviewed the person's curriculum vitae and agreed that the person met all of the qualifications (Shavelson, personal communication, January 18, 2010). The US organizing agency did not have an opportunity to review the permanent NPM's background.

The Country B team members were able to incorporate information from some of the documents in the discussion of the translation and adaptation process. The Country B team brought up the challenges with the amount of adaptation required for some aspects of the performance tasks (CAE, GS.26, 2010). The team also addressed challenges involved due to register and organization of information (CAE, GS.26, 2010). Country B also opened a discussion about student familiarity with the test format and implications on test performance (CAE, GS.26, 2010). The entire Country B team was able to attend the translation review and verification training conducted during the site visit to their country and complete the practice exercises (Solano-Flores, 2010, Visit). However, during the training it became clear that the Country B team had not implemented the translation procedures on which the team had been trained during the initial meeting in New York City (Solano-Flores, 2010, Visit).

The country team in Country B helped the US project manager document the addition of each team member. One of the Country B team members emailed a signed confidentiality agreement for each of the three team members (Choi, personal communication, February 1, 2010). The team member also emailed a list with contact information for each of the team

members (Choi, personal communication, February 1, 2010). There was also email communication providing contact information for a new person who was replacing an original team member (Choi, personal communication, January 14, 2010). This person also signed a confidentiality agreement.

Country B experienced some difficulty in finalizing their team. Although the team had chosen the assessment expert, they would not be able to fill permanently the position of national project manager in time for initial meetings (Choi, personal communication, January 14, 2010). An interim national project manager participated in the conference call (Choi, personal communication, January 21, 2010) and attended the initial meeting in New York City (CAE, GS.1, 2010).

The team members for Country B did not express any difficulty understanding and following the documents provided to guide team configuration. It was clear that challenges in team creation were internal to within the country (Choi, personal communication, January 14, 2010). The team was also able to provide the documentation as stated in the document (Choi, personal communication, February 1, 2010).

Country B experienced in-country support when configuring the team that was to work on the study. The Ministry of Education pledged support for Country B's participation in the AHELO study (CAE, GS.26, February 17, 2010). As a result, several country organizations and universities provided staff for the study. The national project manager worked for the Country B Educational Development Institute (KEDI) (CAE, GS.11, 2010). The assessment expert was a professor in the department of education at the Seoul National University in Country B (CAE, GS.11, 2010). It is important to note, however, that due to issues—not made public—in organizing the national team there was a delay in naming the national project manager (Choi, personal communication, January 14, 2010).

The international organizing agency informed members of Country B Ministry of Education about the AHELO study (Choi, personal communication, February 1, 2010). Given the support of the organizing international organizing agency professionals from a local research institute located within a research university provided staff for the study (Solano-Flores & Chia, 2010, Interview). The US organizing agency sent an initial email and held a conference call with Country B's NPM at the beginning of the project. These efforts gave country representatives an opportunity to ask questions about the creation of the team (Shavelson, personal communication, January 11, 2010).

Country C

Throughout the project, communication with the Country C country team regarding configuration of the AHELO group was challenging. The week after an initial attempt at communicating with the Country C contact the project manager by the US organizing agency, another attempt was made (Shavelson, personal communication, January 13, 2013). At that point the Country C NPM responded with some information. There was some biographical

information shared for the team's NPM and one other contact. However, there was no information shared about a team assessment expert.

Although the Country C team had experience with research, there did not appear to be a member of the national project team with the required and desired qualifications for the measurement expert. The Country C national project manager for AHELO was a recognized distinguished researcher in engineering and was a certified educational institution auditor for the Ministry of Higher Education (CAE, GS.11). Another member of the team was overseeing a study regarding program accreditation and specialized in developing and teaching courses in cultural and women's studies (CAE, GS.11). The team did not have a person with a background in statistics, psychometrics, or in designing assessments. However, both possessed a high degree of fluency in English.

Country C team members were able to help throughout the translation and adaptation process. The Country C team had experience with higher education across different groups within the country and across countries (CAE, GS.11, 2010). One team member had experience developing and teaching courses in cultural studies (CAE, GS.11, 2010). The national project manager was responsible for compiling all information and final versions of the tasks (Al-Atiqi, personal communication, January 1, 2011).

Country C had to share the NPM responsibilities between two people. However, the first NPM fully completed the translation and adaptation process; the second NPM worked on pre-implementation and implementation responsibilities. The first NPM had extensive experience in research and higher education accreditation and practices. Having completed advanced degrees in the United States, including a PhD in chemical engineering, the NPM was fluent in reading, writing, and speaking in English (CAE, 2010, GS.11). The NPM showed experience managing multiple projects at various institutions of higher learning simultaneously, demonstrating knowledge of the higher education system in Country C and project management success (Solano-Flores, 2010, Meeting).

Country C's team configuration process included a number of diverse organizations with limited opportunity for review outside the country. The Country C national team included representatives from each of the six participating higher education institutions. Each person was experienced in quality assurance procedures at the university level (Solano-Flores, 2010, Interview). In addition, the Private University Council headed AHELO. The Private University Council chose the national project manager. Although the Country C NPM provided the US project manager with their biographic sketch, the US did not receive a curriculum vitae for the assessment expert.

The team members from Country C requested a second copy of the documents (CAE, GS.26, 2010). Although there was a brief discussion of these documents during the initial meeting, the team members were to review the documents on their own. However, the two team members were able to participate in the adaptation training and translation overview training that took place during the initial meeting in New York City (CAE, GS.26, 2010). During the

translation review and verification training conducted in Country C not all members of the country team were present; only the national project manager and a representative from one of the universities were present from the team (Solano-Flores, 2010, Visit).

The Country C team provided the information needed to track the progress of team configuration. The US organizing agency had a signed confidentiality agreement on file for each of the members of the Country C team (Keeley, personal communication, September, 13, 2011). The US agency also had contact information for each member of the Country C team (The Country C team was able to communicate changes to the national team and included contact information (Al-Atiqi, personal communication, January 7, 2011).

Country C had difficulty meeting the deadlines for gathering the essential team members. The team was had decided on the national project manager to the first planned conference call; however, they were not able to do the same for the national assessment expert position (Choi, personal communication, January 14, 2010; CAE, GS.11, 2010). However, it is important to note that the assessment expert was not as heavily involved in the project as those from the other countries.

The Country C NPM did not share information indicating that the documents provided for configuring the national team were not user friendly. The NPM acknowledged receipt of the documents and provided confidentiality agreements and biographical sketch (Al-Atiqui, personal communication, January 13, 2010).

Country C obtained support from various agencies and universities when creating the national team for the study. The national project manager was a professor at Country C University (CAE, GS.11, 2010). The NPM was also the Secretary General of Council for Private Universities as well as a Certified Educational Institution Auditor for the Ministry of Higher Education in Country C (CAE, GS.11, 2010). Another member of the team was the Assistant Dean for Accreditation for Curriculum and Assessment and a professor at the University of Country C (AUK) (CAE, GS.11, 2010). Despite the support from several organizations, the Country C team did not have an assessment expert participate as originally suggested by the study.

The international organizing agency informed members of Country C Ministry of Education about the assessment as well as initial steps needed to form a national team for the study (CAE, GS.26, 2010). As a result, the international organizing agency was key during the initial stages of the study (Solano-Flores, 2010, Visit). The support from the international organizing agency helped country universities and agencies support the creation of a national team by providing staff (Solano-Flores & Chia, 2010, Interview). The US organizing agency sent an initial email and held a conference call with Country C's NPM at the beginning of the project. These efforts gave country representatives an opportunity to ask questions about the creation of the team (Shavelson, personal communication, January 11, 2010).

Country D

The communication with Country D was constant and timely. The three core members of the Country D team shared responsibility in communicating with the project manager for the US organizing agency. Each provided biographical sketches containing information about their academic and professional backgrounds (CAE, 2010, GS.11). There was no record of follow-up email asking for information about the team members.

The measurement expert on the Country D team met the required and desired qualifications (CAE, GS.11). The person obtained a doctorate in statistics and focused on statistic simulation, experimental design, and statistical analysis of data. The measurement expert taught pure and applied mathematics and statistics as a research university in Country D and, at the time of the AHELO project, had authored or co-authored five scientific articles and a book. The measurement expert had successfully led three research projects and collaborated on an additional two. The person was fluent in reading, writing, and speaking in English.

The Country D team members had experience that helped them work through the translation and adaptation process. Among them they had experience in working on diverse public policy issues, bilingual education, and formal education in archaeology (CAE, GS.11, 2010). They participated in the actual translation and adaption activities (Solano-Flores, Visit, 2010; Solano-Flores & Chia, Interview, 2010). The team member with the most experience in cultural and linguistic diversity was responsible for compiling all information and final versions of the tasks (Urrea, personal communication, January 21, 2011). However, none of the team members had specific experience or expertise in translation.

The Country D team was able to maintain one national project manager for the entirety of the project. Professionally, the NPM had several years of experience simultaneously working as the coordinator of innovation for undergraduate programs and masters in technologies for learning (CAE, 2010, GS.11). The NPM served as consultant for diverse government agency research projects with a focus on strengthening public policies related to civil society and sustainable development. The NPM had published several scholarly articles, book chapters, and books. The NPM had obtained a masters degree and a doctoral degree in a field akin to urban development and sustainability (Solano-Flores, 2010, Meeting). The NPM was fluent in reading, speaking, and writing English. The NPM worked very closely with the assessment expert who was a key member of national team.

Country D team members provided the project manager for the US organizing agency with academic and professional background information for each of the Country D team members. However, the information was provided after the national project manager and assessment expert had already been chosen for AHELO (Rosas Chavez, personal communication, January 14, 2010). The AHELO national center was housed at the University of Guadalajara and the staff for the national team were all faculty and staff at the university (CAE, 2010, GS.11).

Team members for the Country D team were able to incorporate information from the documents in the discussion of the translation and adaptation process (CAE, GS.26, 2010). The

team was aware of issues related to different dialects used throughout the country, but, explained that this was less of an issue for the age group participating in the study (CAE, GS.26, 2010). The team also brought up issues involved in syntax and discourse with respect to the type of writing the assessment required (CAE, GS.26, 2010). All core members of the Country D team participated in the adaptation and translation training that took place at the initial meeting in New York City (CAE, GS.26, 2010). During the training in Country D all three team members were present for the training (Solano-Flores, 2010, Visit).

The Country D team helped track team configuration progress. A Country D team member emailed a confidentiality agreement for each of the Country D team's members (Rosas Chavez, personal communication, January 25, 2010). The US organizing agency also had a list of contact information for each of the Country D team member (Keeley, personal communication, September 13, 2011).

Country D successfully met the deadline for filling the team positions. The national project manager and assessment expert were able to participate in the initial conference call (Rosas Chavez, personal communication, January 14, 2010). In addition, the country team was able to attend the initial in-person meeting that took place in New York City (CAE, GS.26, 2010).

The Country D team found the documents with guidelines on country team configuration user friendly. They stated that the document was helpful to have (Solano-Flores, 2010, Visit). Also, the Country D team was able to provide confidentiality agreements and biographical sketches (Rosas Chaves, personal communication, January 25, 2010).

Country D configured its team with the support of several organizations. The Ministry of Education showed initial and continual support for the study (CAE, GS.26, 2010). All of the team members, including the national project manager and assessment expert, worked at the University of Guadalajara (CAE, GS.11, 2010). The national project manager was chose first by the ministry; the NPM in turn extended official offers to the other team members (Solano-Flores and Chia, 2010, Interview).

The international organizing agency worked with members of Country D Ministry of Education to include the country in the study and to complete initial steps in forming a national team (Solano-Flores & Chia, 2010, Interview). The support from the international organizing agency helped create the national team with staff from a major research university in the country (Solano-Flores & Chia, 2010, Interview). The US organizing agency sent an initial email and held a conference call with Country D NPM at the beginning of the project. These efforts gave country representatives an opportunity to ask questions about the creation of the team (Shavelson, personal communication, January 11, 2010).

Country E

The team from Country E provided information about their team in a timely and detailed manner. Within 24 hours of the project manager for the US team having initiated communication, the NPM for Country E had responded with information about team members

(Opheim, personal communication, January 13, 2010). The immediate response from Country E included the curriculum vitae for the team's assessment experts and the NPM. Any changes in team staff was also shared immediately (Opheim, personal communication, April 18, 2011).

Country E was able to hire a member of the Country E country assessment expert team (CAE, GS.11). Country E's measurement expert was fluent in English and had obtained a Ph.D. in statistics. The person was professor in the Department of Teacher Education and School Development at a research university in Oslo. In addition, the measurement expert has worked as senior advisor for the country's Ministry of Education and Research and, at the time of the project, had been involved in the implementation of the PISA study in Country E for over ten years. The expert also had several publications related to large scale international achievement studies and was author of a widely used teacher education textbook.

The Country E team members had extensive experience with translation procedures used in international comparison studies. Some team members worked on translation of PISA and TIMSS (CAE, GS.11, 2010). The team directly helped with task adaptation, translation review, and final versions of the performance tasks (Solano-Flores, Visit, 2010; Solano-Flores & Chia, Interview, 2010; Opheim, personal communication, January 25, 2010).

The Country E national project manager was head of research at the Country E Institute for Studies in Innovation, Research and Education (CAE, 2010, GS.11), which is evidence of the person's ability to manage multiple projects simultaneously. The person's research interests, as illustrated by list of publications, included education equity, education policy, and student finance. The NPM has worked closely with OECD on other international comparison studies such as PISA (Solano-Flores, G. & Chia, M., 2010, Interview). In 2004, the NPM prepared the Country E background report for the OECD project "Thematic Review of Equity in Education (CAE, GS.11, 2010). The NPM was fluent in reading, writing, and speaking English.

Country E's process for building their national project team was clearly structured at the local level, yet, there was some opportunity for review. The Director General of the Country E Ministry of Education and Research informed the project manager for the US organizing agency of the ministry's staffing results (Opheim, personal communication, January 13, 2010). The national project manager provided the US organizing agency's project manager with a CV for each of the Country E team members (Opheim, personal communication, January 13, 2010). It is important to note, however, that background information was not provided with addition of new members (Roe, personal communication, April 28, 2010).

The Country E team was able to discuss some aspects of the literature addressing the theoretical foundation of the translation and adaptation process used in the study. Country E asked for suggestions on how to familiarize students with the specific test format used in order to minimize differences caused by format unfamiliarity (CAE, GS.26, 2010). Also, the team asked about the possibility of not translating certain documents given the amount of work required in the AHELO process (CAE, GS.26, 2010). The team members actively participated in the adaptation and translation training conducted at the initial meeting in New York City (CAE,

GS.26, 2010). Finally, during the site visit to Country E all of the core country team members participated in the translation review training that took place (Solano-Flores, G., Shavelson, R., & Chia, M., 2010, Visit).

The team for Country E helped track the progress of team configuration. The national project manager for the Country E team provided the confidentiality agreements signed by each of the team members (Opheim, personal communication, January 26, 2010). The original national project manager for Country E also notified the US organizing agency of changes in the team (Opheim, personal communication, April 18, 2011; Roe, personal communication, April 28, 2010). The Country E team also provided a list of contact information for all team members (Opheim, personal communication, January 26, 2010).

Country E was able to fill the two important positions on the country team within the amount of time given. The national project manager and assessment expert had been hired prior to the initial conference call (Opheim, personal communication, January 13, 2010). In addition, the NPM and assessment expert both attended the New York City meeting (CAE, GS.26, 2010).

The team for Country E found the documents created to help configure teams clear. The NPM for Country E expressed that there was clarity in team member requirements (Opheim, personal communication, January 13, 2010). They were able to provide a curriculum vitae and confidentiality agreement for each team member as requested in the documents (Opheim, personal communication, January 26, 2010).

Country E was able to configure their team through the support of several organizations found within the country. During the summer of 2009 the Ministry of Education and Research in Country E sent out information about the AHELO study to University of Oslo (CAE, GS.26, 2010). Specifically the Department of Education and Teacher Research responded positively by providing staff for the project (CAE, GS.26, 2010).

The international organizing agency informed members of Country E's Ministry of Education about AHELO participation as well as initial steps in creating a national team for the study (Opheim, personal communication, January 13, 2010). As a result, the international organizing agency was key during the initial stages of the study (Opheim, personal communication, January 13, 2010). The support from the international organizing agency helped country universities and agencies support the creation of a national team (Solano-Flores & Chia, 2010, Interview). The US organizing agency sent an initial email and held a conference call with Country E's NPM at the beginning of the project. These efforts gave country representatives an opportunity to ask questions about the creation of the team (Shavelson, personal communication, January 11, 2010).

Task 2: Acquire funding for all steps of the process.

Timely communication-- ***Across all countries.***

International test comparison studies can prove costly and time consuming. Given the requirement for staffing and task demanded in this particular study, it was important that each country receive timely communication of information that could affect the funding for each step

of the process: pre-planning and organization, meeting in New York, adaptation, and translation (CAE, Work, 2010).

There was timely communication with all countries regarding the organization of the project. Countries had received information from the international organizing agency the year before the project was to begin. This allowed the ministries of education to communicate staffing, infrastructure, and logistical needs with research institutes and universities (country, personal communications).

The communication about financial responsibilities for the initial meeting in New York City was also timely. The project manager for the US organizing agency set an email to each country team representative at the beginning of January containing information about the dates and agenda for the initial meeting (Shavelson, personal communication, January 9, 2010). The US project manager also emailed a document containing specific information about costs for attending the meeting: options for travel within the city as well as hotel options (CAE, Appendix F, 2010). The US project manager sent out the information over one month prior to the meeting taking place; there is no record of countries finding the information untimely.

Another major step in the project was the adaptation of the performance tasks prior to actual translation. The US organizing agency emailed general information about the requirements for this process to all country representatives (CAE, Initial, 2010). Countries received more specific information about the task requirements at the beginning of January, when country teams were initiating work on budgets (CAE, Appendix D, 2010). Early in January country teams also received a work plan with specific dates for completing adaptations tasks (CAE, GS.1, 2010).

At the beginning of January, country teams also received some detailed information about the resources needed, as well as an overview of the tasks required, to complete the translation process (CAE, Appendix D, 2010). The document included the number of people required throughout the process, the different materials involved, and some of the coordination needed. Finally, in early January country teams also received a work plan with specific dates for completing translation tasks (CAE, GS.1, 2010). Although there were some initial questions about translation costs expressed via emails and at the meeting in New York City, staff from the US organizing agency was able to provide answers within 24-48 hours.

Country A

Country A paid for an assessment expert to be present throughout the entire study. The assessment expert attended the meeting in New York City (CAE, GS.26, 2010). The assessment expert also provided feedback throughout the process. The assessment expert participated in task adaptation (Solano-Flores & Chia, 2010, Interview). The NPM also reported that, as part of the team, the assessment expert would be giving feedback until the completion of the translation process (Shavelson, 2010, Conference).

The two independent translators working with Country A were present during the site visit (Solano-Flores, Report, 2010). However, there was no third translator. Instead, the Country

A NPM acted as the translation reviewer (Solano-Flores, Report, 2010). It is important to note, however, that the NPM discussed difficulty in finding qualified translators but did not attribute challenges as being related to funding for the positions (Solano-Flores & Chia, Interview, 2010).

Country A did not express a challenge for funding the national project manager position. The NPM was one of the first positions on the team filled (Ursin, personal communication, January 11, 2010). The country funded the position so that the Country E national project manager participated in and completed all tasks that were a part of the official NPM responsibilities (CAE, 2010, Progress).

During the initial meeting in New York, the Country A team discussed that the Ministry of Education and the research institute were financially responsible for country's participation in the study (CAE, GS.26, 2010). To be able to report on budget accurately, the Country A NPM asked for additional information about tasks and staffing (Ursin, personal communication, April 22, 2010).

Country A did not have any official opportunities available for training on the acquisition of fund for the project (CAE, 2010, Appendix D).

Country A provided the US organizing agency with information for several reports that included updates on progress in the study. The team did not share any challenges with funding when sharing information for a milestone report (CAE, 2010). There was also no report of challenges in funding during the progress report (CAE, 2011). In addition, reflection on the project, which included input from US staff, did not include information on challenges to funding (CAE, 2010).

Originally, the study called for the use of three performance tasks. However, due to lack of available funding at the beginning of the study Country A could only work with two performance tasks (OECD, 2010).

The Country A team received the documentation addressing funding needs. The NPM for Country E shared thoughts about costs for scoring and cognitive labs (CAE, GS.26, 2010). The country NPM also asked for clarification on some of the costs. Specifically, the teams were unclear as to which costs were paid for by the institutions of higher learning who would be participating in the study (CAE, GS.26, 2010).

The great deal of financial support that Country A provided facilitated the country's participation in the project. The Country A Ministry of Education (MinEdu) and the Country A Institute for Educational Research (FIER) contributed with personnel for the country team, national centre, and technical infrastructure. (Ursin, personal communication, January 12, 2010). The AHELO national team also received support from seven members from higher education institutions and student unions) (Ursin, personal communication, January 12, 2010). Also, higher education institutions provided students and technology for assessment implementation. Fundamentally, these were all ways that the country financially supported the project.

In order to participate in the AHELO feasibility study each country had to pay the international organizing agency. Country A received financial support with the fee from the

international organizing agency. The international organizing agency supported Country A by providing twenty percent of the total participation fee (Shavelson, personal communication, March 21, 2012). The US organizing agency also helped by providing resources that would have cost Country A a great deal of money. The US organizing agency provided training on adaptation, translation, and scoring (CAE, 2010, On-Site; ACER, 2011, November). The US organizing agency also provided training on conducting and analyzing results from cognitive labs (CAE, GS.37, 2010).

Country B

Country B funded the assessment expert position during the entire translation and adaptation phase of the study. The assessment expert was present during the initial meeting in New York City (CAE, GS.26, 2010). The team also paid for the expert to participate in the site visit (Solano-Flores, 2010, Visit). The team also had the assessment expert on staff for the debriefing conference call that took place at the end of the study (Shavelson, 2010, Conference).

The Country B team had a difficult experience finding translators with the qualifications stipulated by the US organizing agency; however, team members did not voice funding as being an issue (Solano-Flores & Chia, Interview, 2010). During the site visit the two translators who conducted independent translations and the translator responsible for translation review were present (Solano-Flores, Site, 2010).

Country B had challenges finding a permanent national project manager. However, the difficulty was not related to financial challenges (Choi, personal communication, January 14, 2010). The Country B team was able to fund an acting NPM for the entire process (CAE, 2010, Progress).

While a discussion regarding financial responsibility arose during the initial meeting in New York, the Country B team explained that funds for the project were provided by the government and government research institute (CAE, GS.26, 2010). The Country B NPM also shared that they had conversations about finances with funding organizations (Shavelson, personal communication, January 26, 2010).

Country B did not have any official opportunities available for training on the acquisition of fund for the project (CAE, 2010, Appendix D).

The US organizing agency solicited information from the Country B team members regarding their progress on the study. When providing information for a milestone report the Country B team did not indicate that they had any issues with funding (CAE, 2010). There was also no indication of difficulty in finding enough funds during the progress report (CAE, 2011). Also, US staff members did not share knowledge of challenges that countries faced in acquiring funding throughout the translation and adaptation process (CAE, 2010).

During the planning of the AHELO study there was support of using three performance tasks with all countries. However, lack of available funding to cover the costs resulted in Country B only using two performance tasks (OECD, 2010).

The team members for Country B received documentation addressing funding needs and costs. During a conversation about conducting cognitive labs the Country B team brought up balancing the rewards of additional student participants with the costs (CAE, GS.26, 2010). It was clear that challenges in team creation were internal to within the country (Choi, personal communication, January 14, 2010). The team was also able to provide the documentation as stated in the document (Choi, personal communication, February 1, 2010).

Country B benefited from in-country financial support and as a result they were able to participate in the AHELO feasibility study. The Country B Educational Development Institute (KEDI) provided staff—specifically the national project manager was an employee of the institute who was allowed to work on AHELO (CAE, GS.11, 2010). The Seoul National University in Country B also financially supported the project by lending a staff member for the project. The assessment expert was a professor in the department of education at the Seoul National University in Country B (CAE, GS.11, 2010). In addition, by housing the country's national centre for the project KEDI financially supported the project by provided space and the technical infrastructure that would help country participation (CAE, GS.11, 2010). Finally, institutions of higher learning provided students, space, and the technology necessary for assessment implementation.

Country B received support from organizations outside of the country that helped the team acquire funding necessary to project participation. Each country was to pay the international organizing agency a fee to participate in the study. The international organization assisted Country B by providing twenty percent of the fee (Shavelson, personal communication, March 21, 2012). Additionally, the US organizing agency provided support with funding. The US organizing agency provided training on adaptation, translation, and scoring (CAE, 2010, On-Site; ACER, 2011, November). The US organizing agency also provided training on conducting and analyzing results from cognitive labs (CAE, GS.37, 2010; Solano-Flores, Visit, 2010).

Country C

There was little documented participation throughout the project by the assessment expert for Country C. The assessment expert did not participate in any of the scheduled meetings—including the initial meeting in New York City (CAE, GS.26, 2010). The assessment expert also did not participate in the site visit training held in the country (Solano-Flores, Report, 2010). Finally, the assessment expert was not part of the conference call to share experiences throughout the implementation of adaptation and translation (Shavelson, Call, 2010).

Country C did not express lack of funding for translators. Country C had several people with a translation background attend the site visit. There were two translators who completed independent translations present during the site visit (Solano-Flores, Site, 2010). Additionally, the person in charge of translation review was also present (Solano-Flores, Site, 2010).

Country C did not express any challenges with funding the national project manager position for the study. The Country C NPM position was budgeted and the person participated in

all of the meetings (CAE, GS.26, 2010; Solano-Flores, 2010, Visit; Shavelson, 2010, Conference).

During a conversation around project funding that took place at the initial meeting in New York, the Country C NPM explained that the government was providing financially (CAE, GS.26, 2010). as part of a 5-year accreditation plan (CAE, GS.26, 2010). During the site meeting it became clear that the Country B NPM had regular communication with other members of the larger national group involved in the project (Solano-Flores & Chia, 2010, Interview).

Country C did not have any official opportunities available for training on the acquisition of fund for the project (CAE, 2010, Appendix D).

Country C did not indicate any difficulty with funding during any of the documentation opportunities. The team did not share any challenges with funding when sharing information for a milestone report (CAE, 2010). The Country C team did not report any challenges in funding during the progress report (CAE, 2011). In addition, in the reflections report US staff did not indicate that country teams had shared that they faced difficulties in acquiring necessary funding (CAE, 2011).

When Country C agreed to participate in the AHELO feasibility study the project planned on administering three performance tasks to all countries. In the end, only two performance tasks were used because the funding available at the beginning of the study would not cover the costs of three (OECD, 2010).

There is evidence that the Country C team reviewed documentation provided that addressed funding and costs. The Country C NPM provided options about training scorers so as to decrease costs for which he understood the country was responsible (CAE, GS.26, 2010).

Country C provided a great amount of financial resources needed for the country to participate in the AHELO feasibility study. The Secretary General of Council for Private Universities, the Ministry of Higher Education in Country C, and the University of Country C provided staff for the national team (CAE, GS.11, 2010). Each staff member was also able to use their offices and the technical infrastructure available at the institutions in which they originally worked. In addition, the higher education institutions provided students, facilities, and the technical infrastructure necessary for testing.

The international and US organizing agencies provided Country C with support when funding various activities throughout the study. In order to participate in the study each country was to pay a fee to the international organizing agency. The international organizing agency provided Country C with twenty percent of the fee (Shavelson, personal communication, March 21, 2012). Also, the US organizing agency provided support with issues around funding. The US organizing agency provided training on adaptation, translation, and scoring (CAE, 2010, On-Site; ACER, 2011, November). The US organizing agency also provided training on conducting and analyzing results from cognitive labs (CAE, GS.37, 2010; Solano-Flores, Visit, 2010).

Country D

Country D funded the assessment expert throughout the entire AHELO project. The assessment expert was present at the New York City initial meeting (CAE, GS.26, 2010). The team also budgeted for the assessment expert to participate in the site visit (Solano-Flores, 2010, Visit). The assessment expert also participated in the conference call to debrief about the translation and adaptation process (Shavelson, 2010, Conference).

There was no evidence that Country D was unable to fund the translation expert positions. Two translators were present during the site visit (Solano-Flores, Site, 2010). One of the core Country D team members was responsible for translation review; this person also participated in the site visit (Solano-Flores, Site, 2010).

Country D did not share any information pertaining to challenges due to funding for the national project manager position. The Country D NPM participated in conference calls and in-person meetings (CAE, GS.26, 2010; Solano-Flores, 2010, Visit; Shavelson, 2010, Conference).

While participating in a conversation about funding that took place during the initial meeting in New York, Country D explained that the government was providing funding for the project (CAE, GS.26, 2010). Country D had budget reviews when major phases of the study were completed (Urrea, personal communication, July 27, 2010). The Country D team also had to review budget when the exchange rate between the euro, dollar, and Country D peso fluctuated (Urrea, personal communication, September 28, 2010).

Country D did not have any official opportunities available for training on the acquisition of fund for the project (CAE, 2010, Appendix D).

Country D provided the US organizing agency with information for several reports. The team members did not indicate that they experienced challenges with funding when reporting on progress for a milestone report (CAE, 2010). The Country D team did not report challenges in acquiring funding during the progress report (CAE, 2011). In addition, the Country D team members did not indicate challenges in funding acquisition when speaking with US staff (CAE, 2011).

Country D had agreed to participate in the AHELO feasibility study and administer three performance tasks to their student population. When it became clear that funding would not be available at the beginning of the project to do so, the team chose to administer only two performance tasks as part of the general skills module (OECD, 2010). In addition, it was important that deadlines for funding and payment keep in mind the exchange rate between the euro and a country's local currency (Urrea, personal communication, September 28, 2010).

There is evidence that the Country D team understood and was able to apply information from the documents regarding the acquisition of funds. Team members provided suggestions on how to share information while minimizing costs (CAE, GS.26, 2010). The team also discussed available technology that could allow meeting with teams and organizing agencies without travel (CAE, GS.26, 2010).

Country D supported project participation by providing resources that would otherwise have cost a great deal of money. The Ministry of Education showed initial and continual support

for the study (CAE, GS.26, 2010). The ministry provided some of the funds that the international organizing agency required of each country to participate (Urrea, personal communication, September 28, 2010). All of the team members, including the national project manager and assessment expert, worked at the University of Guadalajara (CAE, GS.11, 2010). The country team members were able to use the technical infrastructure in place as well as their office and conference room spaces at the University of Guadalajara for the project. In addition, several higher education institutions provided students, technical infrastructure, and the space necessary to implement the assessment.

Country D received support with acquiring funding from organizations outside of the country. The international and US organizing agencies helped Country D in several ways. To participate in the AHELO study Country D needed to pay the international organizing agency a fee. The international organizing agency provided twenty percent of the fee (Shavelson, personal communication, March 21, 2012). The US organizing agency also helped Country D with areas requiring funding. The US organizing agency provided training on adaptation, translation, and scoring (CAE, 2010, On-Site; ACER, 2011, November). The US organizing agency also provided training on conducting and analyzing results from cognitive labs (CAE, GS.37, 2010; Solano-Flores, Visit, 2010).

Country E

Country E had their respective assessment expert attend the initial project meeting in New York City (CAE, GS.26, 2010). These experts remained on the project through to the final stages of translation. In fact, the Country E NPM argued that it was important to continue working in team and have communication even during stages after test translation and adaptation (Shavelson, 2010, Conference).

Country E did not have difficulty funding its translation team. During the site visit the translators were presented (Solano-Flores & Shavelson, Site, 2010). The translator responsible for the translation review was also present (Solano-Flores & Shavelson, Site, 2010). The team did not express concern about lack of funding for translators.

Country E was able to fund the national project manager position for the duration of the project. There were two people who took on the responsibilities of NPM for Country E. Each person hired as the Country E NPM participated in her respective conference calls and in-person meetings (CAE, GS.26, 2010; Solano-Flores, 2010, Visit; Shavelson, 2010, Conference).

During the initial New York meeting, Country E explained to the US organizing agency, the international organizing agency, and the other country teams that the Ministry of Education and Research were financially responsible for the project (CAE, GS.26, 2010). The Country E NPM had constant communication about project progress with the ministry (Opheim, personal communication, January 13, 2010).

Country E did not have any official opportunities available for training on the acquisition of fund for the project (CAE, 2010, Appendix D).

Throughout the process the US organizing agency solicited information from Country E about their progress in the study—including on their acquisition of funding. The team gave no indication of difficulty with funding when reporting progress for a milestone report (CAE, 2010). The team from Country E did not indicate facing challenges in acquiring funding during the progress report (CAE, 2011). Lastly, the US staff did not include reports by the Country E team that they had difficulty acquiring necessary funds in their reflection report (CAE, 2011).

At the beginning of the AHELO feasibility study, organizing agencies and countries thought that they would be able to work with three performance tasks. However, the inability to acquire sufficient funding by the beginning of the project resulted in Country E working with only two performance tasks throughout the study (OECD, 2010).

There is evidence that Country E team members received enough information regarding costs through the documents provided and understood it. One of the team members recalled that there were costs for which each country was responsible—such as travel, meetings, and actual translation procedures (CAE, GS.26, 2010). Another team member addressed the question of cost-effectiveness during a discussion about double scoring (CAE, GS.26, 2010).

Country E was able to configure their team through the support of several organizations found within the country. During the summer of 2009 the Ministry of Education and Research in Country E sent out information about the AHELO study to University of Oslo and committed to paying some of the fee that the international organizing agency required for participation (CAE, GS.26, 2010). Specifically the Department of Education and Teacher Research responded positively by providing staff for the project (CAE, GS.26, 2010). The staff was able to use their offices, conference areas, and the technical infrastructure in place. In addition, higher education institutions supported the country by providing students, technical infrastructure, and facilities needed to carry out the assessment.

The international and US organizing agencies supported Country E with activities requiring the acquisition of funds. There was a fee associated with study participation that was to be paid to the international organizing agency. Country E received twenty percent of the funding needed to participate from the international organizing agency (Shavelson, personal communication, March 21, 2012). The US organizing agency also supported Country E with areas requiring funding. he US organizing agency provided training on adaptation, translation, and scoring (CAE, 2010, On-Site; ACER, 2011, November). The US organizing agency also provided training on conducting and analyzing results from cognitive labs (CAE, GS.37, 2010; Solano-Flores, Visit, 2010).

Task 3: Select test items based on specific criteria established by the coordinating group.

Country A

The Country A team had three weeks to learn the criteria for selecting performance tasks, select their top four performance tasks, and prepare their reasonings for the meeting in New York City. Three weeks prior to the New York City meeting, the project manager for the US organizing agency emailed the criteria for performance task selection to each member of the

Country A team (Kurpius & Shavelson, personal communication, January 29, 2010). During the New York City meeting Country A team members were able to use specific criteria to explain why they chose certain performance assessment (CAE, GS.26, 2010). The Country A team indicated that they had enough time to complete these activities.

The Country A team used assessment measurement expertise when selecting the performance tasks for the AHELO study. During the initial project meeting in New York City the Country A team assessment expert and NPM discussed the difficulty with securing construct validity and cultural validity for some of the performance tasks (CAE, GS.26, 2010). They also discussed the varying degrees of task 'universality' (CAE, GS.26, 2010).

The Country A team applied their knowledge of translation while selecting the performance tasks for the AHELO study. The team discussed the ease of translating a performance task that contained mainly visuals (CAE, GS.26, 2010). They also shared concerns about the difficulty of translating highly technical documents, numerous documents, or documents that were long (CAE, GS.26, 2010). Finally, the country questioned the translation of material that had copyright protections (CAE, GS.26, 2010). However, they did not include translation expertise in the process.

The national project manager for Country A used the team's project management experience while selecting performance tasks to be used in the AHELO study. The NPM scheduled a meeting in Country A—prior to the New York City meeting—to discuss which tasks would be best for the study (CAE, GS.26, 2010). The NPM shared that the team had addressed all of the selection criteria during that internal meeting (CAE, GS.26, 2010). The national NPM also differentiated between performance tasks that were ideal, acceptable, and not acceptable for the Country A student population so as to make it easier when all countries were to select the final two PTs (CAE, GS.26, 2010).

Country A team members participated in two reviews that were conducted during selection of performance tasks for the study. Both reviews took place during the initial meeting in New York City. Country A team members responded to other country's suggestions about performance tasks: indicating when the context was not suitable for their students or when they felt there would be complex translation requirements (CAE, GS.26, 2010). The next review took place after the country teams voted for the final performance task. After all votes were tallied Country A expressed that they agreed with the task chosen (CAE, GS.26, 2010).

Country A received documentation with criteria that they could use while selecting performance tasks for the study. They were able to review the criteria, discuss it as a team, and apply it (CAE, GS.26, 2010).

The Country A team reported their process of performance task selection during the New York City initial meeting. The Country A national project manger shared that the team met as a group in Country A to discuss the criteria for PT selection (CAE, GS.26, 2010). Furthermore, the team shared detailed explanations as to why they favored certain performance tasks over others

for their students. The project manager for the US organizing agency documented the country team's progress and reported it to the international organizing agency.

Country A did not express any challenge in the amount of time afforded to select performance tasks for the study. The team had three weeks from the time that they received documentation with selection criteria to the New York City meeting when they would collaborate with other teams to select the study's PT (Kurpius & Shavelson, personal communication, January 29, 2010). The team reviewed selection criteria, examined nine performance tasks, and, ultimately, helped select the final PTs that would be used in the AHELO study (CAE, GS.26, 2010).

Country A received the document containing the criteria for performance selection that all countries were to follow. Country A team members were able to use the material effectively (CAE, GS.26, 2010). During the meeting in New York City, team members were able to discuss specific criteria as it applied to each performance task (CAE, GS.26, 2010). Country A team members did not express encountering any difficulty working with the document.

Country A team members did not require support external to the team within the country. In Country A the team was able to discuss the criteria for performance task selection among themselves (CAE, GS.26, 2010).

During performance task selection Country A benefited from support provided by other country teams and the international organizing agency. During the initial meeting in New York city Country A was able to get input from each of the other participating countries regarding performance tasks that could serve all student populations (CAE, GS.26, 2010). The representative from the international organizing agency also provided insight gained from other international test comparison practices (CAE, GS.26, 2010). In addition, the US organizing agency provided guidance throughout the process (CAE, GS.5, 2010).

Country B

Three weeks prior to the New York City meeting, the project manager for the US organizing agency emailed the criteria for performance task selection to each member of the Country B team (Kurpius & Shavelson, personal communication, January 29, 2010). During the New York City meeting the Country B team suggested two performance tasks (CAE, GS.26, 2010). The Country B team indicated that they had enough time to complete these activities.

Country B's measurement expertise helped during the selection of performance tasks that would be used in the AHELO study. The New York City meeting provided opportunities to discuss the performance task selection. The Country B team was concerned with the measurement error that could occur due to unfamiliarity of the task type (CAE, GS.26, 2010). They also shared their concern regarding the type of questions used and the impact that this could have on test performance (CAE, GS.26, 2010).

The Country B team's knowledge of translation helped during the selection of performance tasks that would be used in the AHELO study. The New York City meeting provided opportunities to discuss the performance task selection. The Country B team explained

that they preferred the tasks that had simpler language and straightforward questions because they were easier to translate (CAE, GS.26, 2010). The team also differentiated between translating questions, historical documents, and captions (CAE, GS.26, 2010). However, they did not include translation expertise in the process.

Country B's project management experience helped during the selection of the performance tasks that were to be used for the AHELO study (CAE, GS.26, 2010). The team was eager to compromise with the needs of other countries and only expressed concern over using performance tasks that could prove unfair to Country B students participating in the study (CAE, GS.26, 2010). Also, this was the team that provided three alternatives to a particular PT that proved to bring forth different levels of interest by other country teams (CAE, GS.26, 2010).

The Country B team members were active participants in the two reviews that took place during performance task selection. Countries were given the opportunity to review the initial performance tasks that were suggested for the study. Country B provided feedback on the task format, type of questions, and task information context—making clear which tasks were and were not appropriate for their students (CAE, GS.26, 2010). Once country teams voted for the final performance tasks used, the group had an opportunity to review the findings. The Country B team voiced concerns about cultural differences in some of the PTs, which psychometricians were able address to the team's satisfaction (CAE, GS.26, 2010).

The Country B team members received documentation with criteria that they could use while selecting performance tasks for the study. They were able to review the criteria, discuss it as a team, and apply it (CAE, GS.26, 2010).

The Country B team members were able to share their performance task selection process with the project manager for the US organizing agency during the initial project meeting in New York City. The team provided explanations for choosing certain performance tasks over others (CAE, GS.26, 2010). The team responded to other country teams' questions (CAE, GS.26, 2010). The project manager for the US organizing agency reported the team's process to the international organizing agency (CAE, GS.26, 2010).

Country B had three weeks do review the performance task selection criteria, examine nine performance tasks, apply the criteria for selecting PTs for their student population, and help decide on the final two PTs for the study (Kurpius & Shavelson, personal communication, January 29, 2010). The team successfully completed all of the activities on time. The Country B team members did not indicate any challenge with the deadline for performance task selection.

Country B received documentation with specific criteria to be used during performance task selection from the project manager for the US organizing agency (Kurpius & Shavelson, personal communication, January 29, 2010). Country B team members were able to apply the information in the document when selecting the performance tasks that would be best suited for their students and again when selecting the final PTs (CAE, GS.26, 2010). The team did not indicated having difficulty working with the document.

Country B team members did not need support from within the country and external to the team. The team was able to apply the criteria for performance task selection among themselves accurately (CAE, GS.26, 2010).

The Country B team was able to get support from other country teams, a representative of the international organizing agency, and the project manager for the US organizing agency. During the New York City meeting the Country B team members exchanged suggestions and information while deciding which performance tasks would be best for all student groups (CAE, GS.26, 2010). The representative from the international organization shared best practices from other international test comparisons (CAE, GS. 26, 2010). Finally, the Country B team used PT selection criteria provided by the US organizing agency.

Country C

At the end of January 2010, the project manager for the US organizing agency emailed the criteria for performance task selection to each member of the Country C team (Kurpius & Shavelson, personal communication, January 29, 2010). During the New York City meeting Country C team members shared the performance tasks that they felt were best suited for their students based on the selection criteria provided (CAE, GS.26, 2010). The Country C team indicated that they had enough time to complete these activities.

The Country C team applied the knowledge of measurement provided through documents by the US organizing agency to performance task selection for the AHELO study. The team addressed areas for potential test bias by gender (CAE, GS.26, 2010). The team also questioned the cultural validity of some of the performance tasks' context and subject matter (CAE, GS.26, 2010). The team assured that certain tasks would be more fair because of the cultural habits with which students would be familiar (CAE, GS.26, 2010). However, there was no evidence of a measurement expert being part of the team.

The Country C team applied their knowledge of translation to performance task selection for the AHELO study. The team expressed a preference for performance tasks that contained less text and a greater number of visuals (CAE, GS.26, 2010). The team also discussed the difference in difficulty of translating different types of documents (CAE, GS.26, 2010). However, they did not include translation expertise in the process.

The Country C team used its project management experience when selecting the performance tasks that would be used in the AHELO study. The national project manager brought up the long-term implications that the translation demands could have on measurement results (CAE, GS.26, 2010). The team NPM also helped find compromise among the country teams during the meeting in New York City (CAE, GS.26, 2010).

Country C was able to actively participate in both reviews that took place during performance task selection. After all countries gave their preferences for performance tasks Country C provided feedback. The team explained which of the performance tasks worked best culturally for their students (CAE, GS.26, 2010). The team also connected certain aspects of the PTs with the adaptation and translation work required and potential for bias through those

processes (CAE, GS.26, 2010). After country teams had the opportunity to vote for the final performance tasks, Country C participated in the review discussion. Country C team members were curious about the impact of certain PT context and skills on students across different majors (CAE, GS.26, 2010). The team also voiced concern about the amount of testing time available for certain performance tasks (CAE, GS.26, 2010). Finally, the team also explained how some performance tasks were better culturally suited for their students (CAE, GS.26, 2010).

The project manager for the US organizing agency sent Country C team members documentation containing criteria that should be applied when selecting tasks for the entire study. During the New York City meeting, it was clear that the Country C team members were able to apply the criteria (CAE, GS.26, 2010).

The Country C team members reported their performance task selection process to the project manager for the US organizing agency. During the New York City meeting, the Country C national project manager provided details on how the team applied the PT selection criteria provided (CAE, GS.26, 2010). The team was able to respond to questions about their chosen performance tasks (CAE, GS.26, 2010). The US organizing agency's project manager shared the information that they documented with regard to PT selection with the international organizing agency.

The Country C team members did not indicate that the deadline for selecting the final performance tasks for the study was problematic. The team was to review the documentation containing criteria for performance task selection, review nine performance tasks, select the performance tasks best suited for their students, and help the other country teams select the final PTs for the study (Kurpius & Shavelson, personal communication, January 29, 2010). The Country C team completed all of the activities within the time allotted.

The project manager for the US organizing agency provided the document with explicit criteria types along with an explanation for each (Kurpius & Shavelson, personal communication, January 29, 2010; CAE, GS.5, 2010). Country C team members accurately applied the selection criteria when discussing potential PTs with other countries. The Country C team members did not indicate that they had difficulty working with the document containing the criteria for performance task selection.

Country C team members did not need support from within the country but external to the team. When the team discussed the performance tasks they felt were best for their students it was clear that they had been able to apply the criteria for PT selection within the team (CAE, GS.26, 2010).

Country C team members were able to collaborate with other country teams when choosing the performance tasks for the study. During the initial meeting in New York City Country B got feedback from the other country teams so that the performance tasks selected would work best with all students (CAE, GS.26, 2010). The Country B team also took into account information that a representative from the international organizing agency shared concerning practices done in other cross-national comparison studies. The Country B team also

received guidance from the US organizing agency through documentation that they provided with criteria for test selection, which the Country B team used (CAE, GS.26, 2010).

Country D

The project manager for the US organizing agency sent all members of the Country D team the performance task selection criteria (Kurpius & Shavelson, personal communication, January 29, 2010). The team members shared their choices of performance tasks for the study and included reasons from the selection criteria (CAE, GS.26, 2010). None of the team members indicated that they needed additional time for these activities.

The Country D team's assessment expertise proved important during the performance task selection process. The team shared concerns with the cultural validity of some of the tasks due to the unfamiliar context to which students would be exposed (CAE, GS.26, 2010). The team also discussed the potential impact of the question types and test format used (CAE, GS.26, 2010). The team also discussed the possible impact of text and visuals on test performance (CAE, GS.26, 2010).

The Country D team's knowledge about translation proved helpful during the performance task selection process. The team shared concerns with translation of complex or numerous documents (CAE, GS.26, 2010). The team also addressed the impact that the language complexity included in documents would have on translation (CAE, GS.26, 2010). The team also discussed the possible impact of text and visuals on ease of translation (CAE, GS.26, 2010). However, they did not include translation expertise in the process.

The Country D team used its project management experience while choosing the performance tasks that would be used in the AHELO study. The team had met prior to the New York meeting to discuss the performance tasks best suited for the study (CAE, GS.26, 2010). The team carefully presented reasons for choosing certain performance tasks over others (CAE, GS.26, 2010). The team constantly reminded all participants present at the New York City meeting that they needed to apply the selection criteria within the context of the student age (CAE, GS.26, 2010).

Country D participated in both review opportunities during the performance task selection process. The country team members were active during the review discussion that that took place after each country had suggested certain performance tasks for the study. The team members were able to clearly explain the relationship between certain types of documents and performance tasks and the country's type of discourse and syntax (CAE, GS.26, 2010). The team also felt it was important to choose performance tasks that could more clearly measure the intended constructs across countries (CAE, GS.26, 2010). The team members also shared their ideas after the country teams voted for the final performance tasks. The team explained the how certain performance tasks were better contextual matches for their students (CAE, GS.26, 2010). The team also shared how certain types of tasks were better suited for translation into their language—even providing an example of the number of words in the original version versus the

number in the translated version (CAE, GS.26, 2010). They also shared which questions would be better for their students (CAE, GS.26, 2010).

There were no official training opportunities on task selection procedures. However, the project manager representing the US organizing agency sent the Country D team members documentation explaining the criteria that should be used when selecting performance task for the study. The Country D team was able to use the criteria while selecting the study's performance tasks during the New York City meeting (CAE, GS.26, 2010).

The Country D team members were able to participate in the initial New York City meeting during which the performance selection was documented. The team members shared that they team members had met prior to the New York City meeting and reported their strategies for choosing certain performance tasks (CAE, GS.26, 2010). During the discussion with other country teams and the project manager for the US organizing agency the Country D team members were forthcoming with questions and further explanations supporting their choices for PTs (CAE, GS.26, 2010). The project manager for the US organizing agency documented all of the information and reported it to the international organizing agency.

Each country team had approximately three weeks to select the two performance tasks that would be used in the AHELO study (Kurpius & Shavelson, personal communication, January 29, 2010). Country D was able to complete all activities associated with PT selection in the amount of time allotted. The team did not express any difficulty in meeting the PT selection deadline. The team met to discuss the criteria and select the best PTs for their students and subsequently helped choose the final PTs for the study (CAE, GS.26, 2010).

The project manager for the US organizing agency sent the Country D team a document with criteria for performance task selection (Kurpius & Shavelson, personal communication, January 29, 2010). Country Ds team members were able to complete all activities associated with PT selection for their students as well as for the full project by following the information contained in the document (CAE, GS.26, 2010). The team did not express anything challenging about working with the criteria or the document.

Country D team members did not require support external to the team but from within the country. The team members met and were able to discuss the criteria and the performance tasks that were best for their students (CAE, GS.26, 2010). The team members accurately applied the criteria.

During the performance task selection process the Country D team received support from the other participating countries, the international organizing agency, and the US organizing agency. The Country D team received information from the other country teams that helped them reach a consensus on which tasks would be best for the study (CAE, GS.26, 2010). The Country D team also got support from a representative of the international agency who shared practices used for the same activity in other international test studies (CAE, GS.26, 2010). The Country D team also received support from the US organizing agency, who provided the criteria for task selection that the Country D team used (CAE, GS.5, 2010).

Country E

Country E received an email with the performance task selection criteria at the end of January 2010 (Kurious & Shavelson, personal communication, January 29. 2010). During the New York meeting team members from Country E shared discussions that they had prior to the meeting based on the criteria (CAE, GS.26, 2010). The team gave their suggestions for performance tasks that should be used in the AHELO study (CAE, GS.26, 2010). The team indicated that they had enough time to review the selection criteria and choose performance tasks for the study.

The Country E team used their measurement expertise during their selection of performance tasks to be used for the AHELO study. The team was sensitive to the constructs and attributes required to do well on the test (CAE, GS.26, 2010). The team also shared the tasks that would be more suitable with their student culture and avoid bias (CAE, GS.26, 2010). The team also discussed various questioning formats and the impact that this could have on scoring and student performance (CAE, GS.26, 2010).

The Country E team used their translation expertise during their selection of performance tasks to be used for the AHELO study. The team discussed the complexity of the number of documents in some of the performance tasks and how this could impact the translation process (CAE, GS.26, 2010). The team also addressed the impact of the number of visuals included in a performance task could have on the translation process (CAE, GS.26, 2010). The team also brought up the difference in translating different types of documents—specifically using the example of poetry (CAE, GS.26, 2010). The team members involved in the process had extensive experience with translation of other international assessments.

The Country E team's project management experience facilitated choosing the performance tasks that would be used during the AHELO study. The team clearly differentiated between tasks that they preferred, tasks with which they did not have serious objections, and tasks that would not be appropriate for their student population (CAE, GS.26, 2010). Since the team had met in Country E to discuss which performance tasks would be best for the study they were able to present their views in an organized manner (CAE, GS.26, 2010).

Country E was active during the reviews that took place during performance task selection for the study. The Country E team members shared their reactions to other teams' suggestions for performance tasks. The team contributed to the discussion by pointing out the impact that suggested performance tasks could have on scoring and measurement error (CAE, GS.26, 2010). The team was also curious about ways that translations for certain PTs could provide respondents with clues (CAE, GS.26, 2010). The Country E team also voiced their ideas after all teams voted for the study's final performance tasks. The team expressed their preference for questions included in specific performance tasks as well as the amount of text involved in others (CAE, GS.26, 2010). This team was also curious about the relationship between student major, student educational background, and student plans for the future and student performance on specific tasks (CAE, GS.26, 2010).

There was no official training offered on how to complete the selection for the study's performance tasks. Instead, the project manager for the US organizing agency provided documentation with a list of criteria for selection. Country E was able to review the documentation and successfully apply the criteria provided during the task selection process that took place in New York City (CAE, GS.26, 2010).

The Country E team was able to report their progress on performance task selection process during the New York City meeting. The Country E team members had discussed the selection criteria and possible performance tasks prior to meeting with all of the other countries; they reported their finding during the meeting (CAE, GS.26, 2010). In addition, the team members were able to respond to questions and ask other country teams for clarification (CAE, GS.26, 2010). The project manager for the US organizing agency recorded all of the information and shared it with the international organizing agency.

Country E team members did not express any difficulty in completing all of the activities associated with final performance task selection within the time afforded. The team was able to review PT selection criteria, study nine PTs, choose the PTs best suited for their students, and help the other countries decide on the final PTs for the AHELO study (Kurpius & Shavelson, personal communication, January 29, 2010).

At the end of January the project manager for the US organizing agency sent Country E team members a document that included the criteria for selecting performance tasks (Kurpius & Shavelson, personal communication, January 29, 2010). Country E team members were able to review the information and accurately use it when they met as a group to select the PTs most appropriate for their student population (CAE, GS.26, 2010). They were also able to apply the information from the documents during the PT selection discussion with other country teams (CAE, GS.26, 2010).

Country E team members were able to complete the performance task selection process without the assistance from people external to the team but within the country. The team was comfortable applying the criteria for selecting PTs that were best suited for their students (CAE, GS.26, 2010).

During the performance task activity, Country E received support from the other participating countries, the international organizing agency, and the US organizing agency. The project manager for the US organizing agency provided the performance selection criteria that the Country E team applied (CAE, GS.5, 2010). The other participating teams provided explanations for their selection of tasks, which helped Country E and the other countries reach consensus and choose the final PTs (CAE, GS.26, 2010). A representative of the international organizing agency provided support by sharing best practices used for the same activity in other international test comparisons (CAE, GS.26, 2010).

Task 4: Acquire necessary technical infrastructure.

Country A

The US organizing agency provided information regarding technical requirements for implementation during the meeting in New York City that the Country A team attended (CAE, GS26, 2010). The discussion after the assessment internet platform presentation included information about the number of computers that may be necessary, internet accessibility, and trouble shooting potential technical problems (CAE, GS.26, 2010). The teams also learned about technology tools that could help with their international communication (CAE, GS.26, 2010). In addition, the technology program manager for the US organizing agency shared qualifications that assessment proctors should have to facilitate assessment administration (CAE, GS.26, 2010). The US organizing agency also provided specific information about computer needs for reviewing the testing platform (CAE, personal communication, November 1, 2010; Keeley, personal communication, June 6, 2011). Country A did not express concern about timely communication regarding technical needs.

The Country A measurement expert had all of the technology required to contribute his expertise. The expert was part of the performance task review process (Solano-Flores & Chia, Interview, 2010). The expert was able to access the performance tasks for review online (Keeley, personal communication, June 8, 2011). The measurement expert was also able to work with Microsoft office programs to provide feedback on the performance assessment (Keeley, personal communication, June 8, 2011).

The Country A team was able to provide translation experts with the technology they needed to provide feedback throughout the project. The members with translation expertise were able to access, work on, and provide feedback on performance tasks using word, adobe, and excel (Kurpius, personal communication, June 29, 2010). They were able to use the internet to communicate via email and phone (Chia, Rubric, 2010). The team did not indicate any challenge to acquiring or accessing technology needed.

The national project manager for Country A was able to successfully manage the technical infrastructure for the team. The team was able to participate in conference calls via telephone and Skype (Shavelson, Conference, 2010; Chia, Rubric, 2011). The NPM also made sure that team members had the software and hardware to contribute to the project and complete required tasks (Shavelson, Conference, 2010). Even when there was a technical preference, such as using word instead of a pdf document, the national project manager for Country A was able to communicate with the US organizing agency and help the team (Kurpius, personal communication, June 9, 2010).

Country A team members participated in the person and over the telephone reviews that were conducted. The Country A team attended the New York City meeting and indicated that they had the technology in place to complete tasks assigned up to that point (CAE, GS.26, 2010). Country A team members were also able to participate in two project meetings that took place during OECD GA meetings. The team did not indicate in either meeting that they were experiencing difficulty (Kurpius, personal communication, March 17, 2010; Tremblay, personal communication, February 18, 2011). The measurement expert working with the US organizing

agency was able to visit the team and there is no indication of difficulty in putting the technical infrastructure necessary (CAE, Milestone, 2010; Solano-Flores, Visit, 2010; Solano-Flores & Chia, Interview, 2010). The team was also able to participate in conference calls during which they did not report challenges to acquiring the technical infrastructure needed to participate in the project (Shavelson, Conference, 2010; Ursi, personal communication, January 12, 2010).

Country A team members participated in the training offered in New York City that addressed technical needs for assessment administration. The training included information about test online format, proctor's technical needs, technology for test security, handling potential technical problems, and additional online training (CAE, GS.26, 2010). The US agency also provided documentation and support for user testing (Keeley, personal communication, May 2, 2011; Keeley, personal communication, June 6, 2011) and working with graphics that are part of the test (Keeley, personal communication, June 8, 2011).

Country A provided feedback for translation verification and were able to test the proctor, student, and scorer interfaces. They were also able to review the online platform.

Country A did not express any challenge in the amount of time afforded to acquire the necessary technical infrastructure. The team was able to participate in conference call, email, and use documents in word and pdf formats (Ursin, personal communication, January 12, 2010). The team was also able to work with Excel documents throughout the study (Ursin, 2010, Labs). The national project manager was able to conduct, record, and report out on cognitive labs in time (Ursin, 2010, Labs).

Country A received documents containing information about the necessary technical infrastructure (Kurpius & Shavelson, personal communication, January 19, 2010). The team was aware of the need to have internet access, working email, conference call capabilities, the ability to conduct and record cognitive labs, and the ability to work with word, excel and pdf documents (CAE, 2010, GS.1; CAE, 2010, Annex D). Country A needed clarification on the technology required for the internet platform: type of web browser required and any specific software needed (CAE, personal communication, November 1, 2010).

Country A provided the technical infrastructure required for the country's participation in the project. The national team consisted of employees who worked for Country A Ministry of Education (MinEdu) and the Country A Institute for Educational Research (FIER) (Ursin, personal communication, January 12, 2010). As a result the team used the organizations' technical infrastructure to work on the project. (Ursin, personal communication, January 12, 2010). In addition, higher education institutions provided technology for assessment implementation.

During the New York City meeting Country A team members received training on hiring translators from the US organizing agency (CAE, 2010, GS.26). The Country A team also received electronic copies of materials addressing translator qualifications from the US organizing agency (CAE, 2010, GS.13; CAE, 2010, GS.4). Finally, the US agency's PI sent

Country A team members an email containing a list of translator qualifications (Kurpius & Shavelson, personal communication, April 21, 2010).

Country A team members did not require support from outside of the country to put in place the technical infrastructure that the US and international organizing agencies required.

Country B

The Country B team was able to gather information about technical requirements for the project during the initial meeting in New York City. The team was able to participate in the assessment internet platform presentation. The team members heard information about the number of computers that may be necessary, internet accessibility, and trouble shooting potential technical problems (CAE, GS.26, 2010). The Country B team also learned about technology tools that could help with their international communication (CAE, GS.26, 2010). In addition, the technology program manager for the US organizing agency shared qualifications that assessment proctors should have to facilitate assessment administration (CAE, GS.26, 2010). The US organizing agency also provided specific information about computer needs for reviewing the testing platform (CAE, personal communication, November 1, 2010; Keeley, personal communication, June 6, 2011). Country B team members did not express concern about timely communication regarding technical needs.

The measurement expert for Country B had access to all of the technology needed to provide feedback throughout the process. The measurement expert was able to access the performance task through email and using computer software (Solano-Flores, Visit, 2010). The expert also had all of the technology needed to review the performance tasks online and provide feedback (Kelley, personal communication, June 8, 2011).

Country B's translation experts were able to access the performance tasks throughout the study by using appropriate technology. The translation experts were able to open word documents and provide feedback using specific software: word, excel, and adobe acrobat (Young, personal communication, March 26, 2010). They also had access to the internet and email (Solano-Flores, Visit, 2010). However, the Country B team experienced some initial difficulty when trying to open attachments that were password protected (Young, January 25, 2011). The team did not indicate any challenge to acquiring or accessing technology needed.

Country B's national project manager helped the team acquire the technical infrastructure needed throughout the study. The team was able to participate in conference calls (Shavelson, Conference, 2010). The team was also able to work with word documents, pdf files, and Excel worksheets (Young, personal communication, February 8, 2011). In addition, when there was a technical challenge with accessing documents, the national project manager quickly contacted the US organizing agency to rectify it (Young, personal communication, January 25, 2011).

The Country B team members were active participants in the reviews that provide information regarding technical infrastructure. The Country B team attended the New York City meeting and indicated that they had the technology in place because they were able to complete tasks that had been assigned (CAE, GS.26, 2010). Country B team members also participated in

two project meetings that took place during OECD GA meetings. There was no indication that the team was experiencing difficulty (Kurpius, personal communication, March 17, 2010; Tremblay, personal communication, February 18, 2011). The measurement expert working with the US organizing agency was able to visit the team and there is no indication of difficulty in putting the technical infrastructure necessary (CAE, Milestone, 2010; Solano-Flores, Visit, 2010; Solano-Flores & Chia, Interview, 2010). The Country B team participated in conference calls and did not report challenges to acquiring the technical infrastructure needed to participate in the project (Shavelson, Conference, 2010; Choi, personal communication, January 21, 2010).

The Country B team members received documentation addressing technical aspects of the project and participated in the training offered during the New York City meeting. The training included information about test online format, proctor's technical needs, technology for test security, handling potential technical problems, and additional online training (CAE, GS.26, 2010). The US agency also provided documentation and support for user testing (Keeley, personal communication, May 2, 2011; Keeley, personal communication, June 6, 2011) and working with graphics that are part of the test (Keeley, personal communication, June 8, 2011).

Country B provided feedback for translation verification and were able to test the proctor, student, and scorer interfaces. They were also able to review the online platform.

Country B did not indicate that they had any difficulty in acquiring the necessary technical infrastructure within the amount of time allotted. Via email communication, team members were able to access and work with word and pdf documents and participate in conference calls (Choi, personal communication, January 21, 2010). Throughout the study the team was also able to work with Excel documents (Keeley, personal communication, June 6, 2011). Additionally, the team was able to record the cognitive labs that they conducted (Kurpius, personal communication, February 8, 2011).

Country B received documents containing information about the necessary technical infrastructure (Kurpius & Shavelson, personal communication, January 19, 2010). The team was aware of the need to have internet access, working email, conference call capabilities, the ability to conduct and record cognitive labs, and the ability to work with word, excel and pdf documents (CAE, 2010, GS.1; CAE, 2010, Annex D). The team did not give any indication that the documents addressing the acquisition of the technical infrastructure necessary for the project were challenging.

Country B benefited from in-country technical support and as a result they were able to participate in the AHELO feasibility study. The Country B Educational Development Institute (KEDI) provided staff who was, subsequently, allowed to use the institute's technology while working on the project (CAE, GS.11, 2010). The Seoul National University in Country B also provided a staff member and the university's technical infrastructure for the project (CAE, GS.11, 2010). Finally, institutions of higher learning provided students, space, and the technology necessary for assessment implementation.

Country B team members did not require support from outside of the country to put in place the technical infrastructure that the US and international organizing agencies required.

Country C

During the initial meeting in New York City the Country C team was able to gather information about technical needs for the project. After a presentation on the assessment internet platform the Country C team asked questions and heard information on the number of computers that may be necessary, internet accessibility, and trouble shooting potential technical problems (CAE, GS.26, 2010). The teams also learned about technology tools that could facilitate international communication (CAE, GS.26, 2010). In addition, the technology program manager for the US organizing agency shared qualifications for assessment proctors (CAE, GS.26, 2010). The US organizing agency also provided specific information about computer needs for reviewing the testing platform (CAE, personal communication, November 1, 2010; Keeley, personal communication, June 6, 2011). Country C did not express concern about timely communication regarding technical needs.

Country C was to gather measurement expertise throughout the process. However, there was no evidence of an assessment expert giving direct feedback to Country C's national project manager or the project manager for the US organizing agency.

The Country C team was able to provide its translation experts with the technology that they needed in order to participate in the study. Those with translation expertise had access to the internet, email, and software—such as word and adobe acrobat—to open and work on the performance tasks (Al-Atiqi, personal communication, January 1, 2011). The team did not indicate any challenge to acquiring or accessing technology needed.

Country C's national project manager made sure to acquire all of the technical infrastructure necessary to participate in the study. The team was able to use word and pdf documents (Al-Atiqi, personal communication, January 1, 2011). The team was also able to acquire the technology to participate in conference calls (Shavelson, Conference 2010). When there was a technical difficulty accessing a document the national project manager contacted the US organizing agency to rectify the situation (Al-Rashed, personal communication, February 13, 2011).

Country C was able to actively participate in both reviews that would include information about technical infrastructure. The Country C team attended the New York City meeting and indicated that they had the technology in place to complete tasks assigned up to that point (CAE, GS.26, 2010). The national project manager for Country C team members were also able to participate in two project meetings that took place during OECD GA meetings. The team did not indicate in either meeting that they were experiencing difficulty (Kurpius, personal communication, March 17, 2010; Tremblay, personal communication, February 18, 2011). The measurement expert working with the US organizing agency was able to visit the team and there is no indication of difficulty in putting the technical infrastructure necessary (CAE, Milestone, 2010; Solano-Flores, Visit, 2010; Solano-Flores & Chia, Interview, 2010). The team was also

able to participate in conference calls during and did not report challenges in acquiring the technical infrastructure needed to participate in the project (Shavelson, Conference, 2010).

Country C participated in the training offered by the US organizing agency during the New York City meeting. The training included information about test online format, proctor's technical needs, technology for test security, handling potential technical problems, and additional online training (CAE, GS.26, 2010). The US agency also provided documentation and support for user testing (Keeley, personal communication, May 2, 2011; Keeley, personal communication, June 6, 2011) and working with graphics that are part of the test (Keeley, personal communication, June 8, 2011).

Country C provided feedback for translation verification and were able to test the proctor, student, and scorer interfaces. They were also able to review the online platform.

The Country C team members did not indicate that the deadline for acquiring the technical infrastructure needed for the study was problematic. The team was able to participate in conference calls (Shavelson, personal communication, January 14, 2010). During the calls the national project manager discussed word and pdf documents that the US organizing agency sent via email (Shavelson, personal communication, January 14, 2010). In addition, the Country C team was able to work with Excel documents (Keeley, personal communication, June 6, 2011). Finally, the Country C team was able to record each of the cognitive labs that they conducted (CAE, 2010, Module).

Country C received documents containing information about the necessary technical infrastructure (Kurpius & Shavelson, personal communication, January 19, 2010). The team was aware of the need to have internet access, working email, conference call capabilities, the ability to conduct and record cognitive labs, and the ability to work with word, excel and pdf documents (CAE, 2010, GS.1; CAE, 2010, Annex D). The team did not give any indication that the technical infrastructure documentation was challenging.

Country C provided a great amount of technical resources needed for the country to participate in the AHELO feasibility study. The Secretary General of Council for Private Universities, the Ministry of Higher Education in Country C, and the University of Country C provided staff and as a result technical infrastructure for the national team (CAE, GS.11, 2010). In addition, the higher education institutions provided students, facilities, and the technical infrastructure necessary for testing.

Country C team members did not require support from outside of the country to put in place the technical infrastructure that the US and international organizing agencies required.

Country D

The Country D team attended the initial meeting in New York City and was able to gather information about the technical needs for the project. The Country D team participated in the assessment internet platform presentation and discussion that followed. During the discussion the team gathered information about the number of computers that may be necessary, internet accessibility, and trouble shooting potential technical problems (CAE, GS.26, 2010). The

Country D team also gathered information about technology that could facilitate international communication (CAE, GS.26, 2010). In addition, the US organizing agency shared qualifications that assessment proctors should have to facilitate assessment administration (CAE, GS.26, 2010). The US organizing agency also provided specific information about computer needs for reviewing the testing platform (CAE, personal communication, November 1, 2010; Keeley, personal communication, June 6, 2011). Country D did not express concern about timely communication regarding technical needs.

The measurement expert for Country D had access to all of the technology needed to provide information throughout the project. The measurement expert was able to provide feedback on the performance task translations (Solano-Flores, Visit, 2010). Also, the Country D team's measurement expert accessed the performance tasks online by using the authorization hyperlinks and passwords provided via email (Sanchez, personal communication, 2011).

The Country D team provided members of their team with the technology necessary for them to participate in the study. Translators were able to use the internet to communicate and access documents pivotal to study (Chia, Rubric, 2011). The translation experts also had access to the software required to open, work with, and provide feedback on documents: word, adobe acrobat, and excel (Chia, Rubric, 2011; Urrea, personal communication, September 20, 2010). The team did not indicate any challenge to acquiring or accessing technology needed.

The Country D team's national project manager put in place the technical infrastructure necessary to project participation. The team acquired technology to participate in conference calls as well as video conferencing (Shavelson, Conference, 2010; Chia, Rubric, 2011). The team also used word, adobe acrobat, and Excel throughout the study (Sanchez-Gomez, personal communication, January 21, 2011; Urrea, personal communication, September 27, 2010).

Country D participated in all review opportunities that included information about technical infrastructure. The Country D team attended the New York City meeting and did not indicate challenges in acquiring necessary technical infrastructure (CAE, GS.26, 2010). The Country D team members were also able to participate in two project meetings that took place during OECD GA meetings. The team did not indicate in either meeting that they were experiencing difficulty (Kurpius, personal communication, March 17, 2010; Tremblay, personal communication, February 18, 2011). The measurement expert working with the US organizing agency was able to visit the team and there is no indication of difficulty in putting the technical infrastructure necessary (CAE, Milestone, 2010; Solano-Flores, Visit, 2010; Solano-Flores & Chia, Interview, 2010). The team was also able to participate in conference calls during which they did not report challenges to acquiring the technical infrastructure needed to participate in the project (Shavelson, Conference, 2010; Urrea, personal communication, July 26, 2010).

The US organizing agency provided formal training and supporting documentation to help country teams acquire the technical infrastructure needed for the study. The Country D team members participated in the training, which took place during the New York City meeting. The training included information about test online format, proctor's technical needs, technology for

test security, handling potential technical problems, and additional online training (CAE, GS.26, 2010). The US agency also provided documentation and support for user testing (Keeley, personal communication, May 2, 2011; Keeley, personal communication, June 6, 2011) and working with graphics that are part of the test (Keeley, personal communication, June 8, 2011).

Country D provided feedback for translation verification and were able to test the proctor, student, and scorer interfaces. They were also able to review the online platform.

There were different deadlines for acquiring the technical infrastructure necessary for study participation. The Country D team did not indicate that they experienced any difficulty meeting the deadlines. The team members were able to participate in conference calls (Rosas Chavez, personal communication, January 14, 2010). During the call team members were able to discuss word and pdf documents that they had received via email (Rosas Chavez, personal communication, January 14, 2010). The Country D team was also able to work with Excel files throughout the process (Urrea, personal communication, September 27, 2010). Lastly, the team was able to record the cognitive labs conducted (Country D, 2010, Respuestas).

Country D received documents containing information about the necessary technical infrastructure (Kurpius & Shavelson, personal communication, January 19, 2010). The team was aware of the need to have internet access, working email, conference call capabilities, the ability to conduct and record cognitive labs, and the ability to work with word, excel and pdf documents (CAE, 2010, GS.1; CAE, 2010, Annex D). Country D only had one follow-up question about internet security (Keeley, personal communication, April 28, 2011).

Country D supported project participation by providing the technical infrastructure demanded to participate in the study. The Ministry of Education showed initial and continual support for the study (CAE, GS.26, 2010). All of the team members, including the national project manager and assessment expert, worked at the University of Guadalajara (CAE, GS.11, 2010). The country team members were able to use the technical infrastructure in place for the project. In addition, several higher education institutions provided students, technical infrastructure, and the space necessary to implement the assessment.

Country D team members did not require support from outside of the country to put in place the technical infrastructure that the US and international organizing agencies required.

Country E

The team for Country E was able to acquire information about the technical needs required for in the project. The team was present during the assessment internet platform presentation and subsequent discussion. They gathered information about the number of computers that may be necessary, internet accessibility, and trouble shooting potential technical problems (CAE, GS.26, 2010). The Country E team also learned about technology that could help with their international communication (CAE, GS.26, 2010). In addition, the technology program manager for the US organizing agency shared qualifications that assessment proctors should have to facilitate assessment administration (CAE, GS.26, 2010). The US organizing agency also provided specific information about computer needs for reviewing the testing

platform (CAE, personal communication, November 1, 2010; Keeley, personal communication, June 6, 2011). Country E did not express concern about timely communication regarding technical needs.

The Country E team provided all of the technology for the measurement expert to provide feedback throughout the process. The expert was able to access performance tasks and provide feedback by using office software, email, and the internet (Solano-Flores, Interview, 2010). The measurement expert for Country E was also able to access the online performance tasks and provide feedback (Keeley, personal communication, June 8, 2011).

The team from Country E provided their translation experts with the necessary technology for them to participate in the study. They were able to use the internet and email to communicate and access performance tasks and other documents important to the process (Opheim, personal communication, January 21, 2011). The team also had access to important software that allowed them to work on the performance tasks (Solano-Flores, Visit, 2010). The team did not indicate any challenge to acquiring or accessing technology needed.

Country E's national project manager put in place the technical infrastructure needed to participate in the project. The NPM was able to participate in conference calls (Shavelson, Conference, 2010). The team also had access to software, email and the internet. As a result, the team was able to complete tasks that required that technology (Opheim, personal communication, January 21, 2011).

Country E was active during the reviews that could provide information regarding the acquisition of the needed technical infrastructure. The Country E team attended the New York City meeting and indicated that they had the technology in place to complete tasks assigned up to that point (CAE, GS.26, 2010). Country E team members were also able to participate in two project meetings that took place during OECD GA meetings. The team did not indicate in either meeting that they were experiencing difficulty (Kurpius, personal communication, March 17, 2010; Tremblay, personal communication, February 18, 2011). The measurement expert working with the US organizing agency was able to visit the team and there is no indication of difficulty in putting the technical infrastructure necessary (CAE, Milestone, 2010; Solano-Flores, Visit, 2010; Solano-Flores & Chia, Interview, 2010). The team was also able to participate in conference calls during which they did not report challenges to acquiring the technical infrastructure needed to participate in the project (Shavelson, Conference, 2010; Opheim, personal communication, January 13, 2010).

The Country E team participated in the training that US organizing provided during the initial meeting that took place in New York City. The training included information about test online format, proctor's technical needs, technology for test security, handling potential technical problems, and additional online training (CAE, GS.26, 2010). The US agency also provided documentation and support for user testing (Keeley, personal communication, May 2, 2011; Keeley, personal communication, June 6, 2011) and working with graphics that are part of the test (Keeley, personal communication, June 8, 2011).

Country E provided feedback for translation verification and were able to test the proctor, student, and scorer interfaces. They were also able to review the online platform.

Country E team members did not express any difficulty in completing all of the activities associated with acquiring the technical infrastructure needed for the study. The team was able to participate in conference calls (Opheim, personal communication, January 13, 2010). Also, the team was able to work with word and pdf files (Opheim, personal communication, January 13, 2010). In addition, the team worked with Excel files throughout the project (Keeley, personal communication, June 6, 2011). Finally, the team was able to record the cognitive labs that they conducted (CAE, personal communication, January 20, 2011).

Country E received documents containing information about the necessary technical infrastructure (Kurpius & Shavelson, personal communication, January 19, 2010). The team was aware of the need to have internet access, working email, conference call capabilities, the ability to conduct and record cognitive labs, and the ability to work with word, excel and pdf documents (CAE, 2010, GS.1; CAE, 2010, Annex D). The team did not give any indication that documentation addressing technical infrastructure necessary for the project was challenging.

Country E was able to acquire the necessary technical infrastructure through the support of several organizations found within the country. Specifically the Department of Education and Teacher Research responded positively by providing staff for the project (CAE, GS.26, 2010). The staff was able to use the technical infrastructure in place while working on the AHELO study. In addition, higher education institutions supported the country by providing students, technical infrastructure, and facilities needed to carry out the assessment.

Country E team members did not require support from outside of the country to put in place the technical infrastructure that the US and international organizing agencies required.

Task 5: Adapt test based on agreed upon cultural adaptation suggestions.

Country A

The Country A team did not indicate any challenge in communication. Team members were able to discuss specific adaptation needs during the initial meeting in New York (CAE, GS.26, 2010). Communication worked well and the team met the mid-March deadline for submission of suggested modifications (Ursin, personal communication, March 25, 2010). Communication was also timely for the US organizing agency to modify PTs and send to countries for review, comment and completion (CAE, personal communication, April 19, 2010; Ursin, personal communication, April 28, 2010). The organizing agency was able to send the final version of the English PTs to Country A in time by the May deadline (CAE, personal communication, May 11, 2010).

The Country A measurement expert participated in the adaptation process of the two performance tasks used in the study (Solano-Flores & Chia, Interview, 2010).

The Country A team had members who were familiar with assessment and evaluation programs (CAE, GS.11, 2010). However, the team did not include translation expertise and did not hire translation experts during the adaptation process. This may have contributed to the fact

that the modifications suggested during the official adaptation phase were minor (CAE, 2010, Progress). Therefore, translators adapted the performance tasks during the translation process (Solano-Flores & Chia, 2010, Interview).

The national project manager for Country A was able to coordinate the multiple steps involved in the adaptation process within the time allotted (see B7-D7 for documents). However, the team did not address all potential challenges to task appropriateness during the adaptation process. The country only focused on surnames, name of places, and people titles (Solano-Flores & Chia, 2010, Interview).

The Country A team participated in all of the review opportunities. Team members were able to discuss specific adaptation needs during the initial meeting in New York (CAE, GS.26, 2010). The team submitted their modification suggestions to the US PI for review by the third week in March (Ursin, personal communication, March 25, 2010). Upon review reviewing the US organizing agency's comments and accepted modifications the Country A team was able to submit their final adapted PTs in the middle of April 2010 (CAE, personal communication, April 19, 2010; Ursin, personal communication, April 28, 2010). The organizing agency was able to send the final version of the English PTs to Country A in May for final review (CAE, personal communication, May 11, 2010).

In January 2010, the Country A team received the conceptual framework explaining the goals, importance, and an overview of the challenges associated with test adaptation (Shavelson & Kurpius, personal communication, January 29, 2010; CAE, 2010, GS.4). The following month, the US organizing agency sent the Country A team members four documents related to test adaptation: Adapting Educational and Psychological Tests for Cross-Cultural Assessment (2005) by R. Hambleton, P. F. Merenda, C. D. Spielberger (Eds.) (GS.7); International Guidelines on Computer-Based and Internet Delivered Testing (2005) (GS.8); Theory of Test Translation Error (2009) by G. Solano-Flores, E. Backhoff, L. A. Contreras-Niño (GS.9); Universal Design Applied to Large Scale Assessments (2002) by S. J. Thompson, C. J. Johnstone, M. L. Thurlow (GS.10) (Shavelson & Kurpius, personal communication, February 9, 2010). In addition, the team participated in the training offered in New York City that addressed details of test adaptation. A psychometrician specializing in test translation and adaptation explained and provided hands on practice exercises regarding issues of context, discourse, dialects, register, graphical representations, computer administration, cultural appropriateness, cognitive and linguistic equivalence, and format familiarity (CAE, 2010, GS.13). The training also stressed the importance of simultaneously and iteratively examining the task, scoring rubric, and response format (CAE, 2010, GS.13).

The national project manager for Country A provided information at each opportunity to document progress relevant to the country team. Team members were able to provide feedback on adaptation documentation and process during the initial meeting in New York (CAE, GS.26, 2010). In March 2010 the team submitted to the US PI a list of suggested modifications for documentation (Ursin, personal communication, March 25, 2010). In addition, Country A

provided the US PI with comments and changes to both PTs based on the list of suggested modifications (Ursin, personal communication, April 28, 2010).

Country A met all of the deadlines scheduled throughout the performance task adaptation process. Team members had read the provided documentation addressing the importance of and process for adapation initial meeting in New York (CAE, GS.26, 2010). The team met the mid-March deadline for submission of suggested modifications (Ursin, personal communication, March 25, 2010). The team was also able to complete the task adaptation by the April deadline (CAE, personal communication, April 19, 2010; Ursin, personal communication, April 28, 2010).

Country A received documents containing information about the performance task adaptation. However, the Country A team did not follow all of the articles' guidelines closely. The team did not address issues with language, scoring rubrics, or computer administration.

During the official adaptation of the performance tasks Country A had support from professionals outside of the team. The translation team was brought in prior to the translation stage to help with the adaptation process (Solano & Chia, 2010, Interview).

Country A team members received support from the US organizing agencies during performance task adaptation. The organization provided documentation that included guidance for the adaptation process (Kurpius & Shavelson, personal communication, January 9, 2010). The US organizing agency also created a compilation of all modifications suggested by all countries that would not have a negative impact on the intended constructs (CAE, personal communication, April 19, 2010). The agency also provided final English versions of the performance task that incorporated all accepted adaptations (CAE, personal communication, May 11, 2010).

Country B

The Country B team did not indicate any challenge in communication. Team members were able to discuss specific adaptation needs during the initial meeting in New York (CAE, GS.26, 2010). The communication for submission of possible task modifications in March 2010 was also timely (Young, personal communication, March 26, 2010). Communication was also timely for the US organizing agency to modify PTs and send them to countries for review, comment and complete (CAE, personal communication, April 19, 2010; Young, personal communication, April 28, 2010). The organizing agency was able to send the final version of the English PTs to Country B in time by the May deadline (CAE, personal communication, May 11, 2010).

The measurement expert for Country B had access to all of the technology needed to provide feedback throughout the process. The measurement expert was able to access the performance task through email and using computer software (Solano-Flores, Visit, 2010).

The Country B team had expertise in measurement and research. However, the team did not include members with translation expertise (CAE, GS.11, 2010). The team implemented knowledge about factors to consider during adaptation that were shared during the initial meeting

in New York City (Solano-Flores & Chia, 2010, Interview). This could explain why the adaptation modifications suggested by Country B were minor and superficial and why additional adaptation occurred during the translation process (CAE, 2010, Progress).

Country B's national project manager was able to coordinate the multiple steps involved in the adaptation process within the time allotted (see B7-D7 for documents). However, the team did not address all potential challenges to task appropriateness during the adaptation process. The country team members involved in adaptation focused on naming conventions—both for people and cities (Solano-Flores & Chia, 2010, Interview). They also worked on adapting the use of titles keeping in mind placing it within the Country B hierarchy context as well as historical contexts (Solano-Flores & Chia, 2010, Interview).

The Country B participated in every review opportunity available during the adaptation process. Team members reviewed specific adaptation needs during the initial meeting in New York (CAE, GS.26, 2010). The team also submitted possible task modifications to the US organizing agency in March 2010 for review (Young, personal communication, March 26, 2010). The US organizing agency modified the PTs and send them to countries for review in April 2010 and within two weeks the Country B team emailed back their final adapted PTs for review (CAE, personal communication, April 19, 2010; Young, personal communication, April 28, 2010). The organizing agency was able to send the final version of the English PTs to Country B for review in May as well (CAE, personal communication, May 11, 2010).

The Country B team received the conceptual framework explaining the goals, importance, and an overview of the challenges associated with test adaptation in January 2010 (Shavelson & Kurpius, personal communication, January 29, 2010; CAE, 2010, GS.4). In February, the US organizing agency sent the Country B team members four documents related to test adaptation: Adapting Educational and Psychological Tests for Cross-Cultural Assessment (2005) by R. Hambleton, P. F. Merenda, C. D. Spielberger (Eds.) (GS.7); International Guidelines on Computer-Based and Internet Delivered Testing (2005) (GS.8); Theory of Test Translation Error (2009) by G. Solano-Flores, E. Backhoff, L. A. Contreras-Niño (GS.9); Universal Design Applied to Large Scale Assessments (2002) by S. J. Thompson, C. J. Johnstone, M. L. Thurlow (GS.10) (Shavelson & Kurpius, personal communication, February 9, 2010). In addition, the team participated in the training offered in New York City that addressed details of test adaptation. A psychometrician specializing in test translation and adaptation explained and provided hands on practice exercises regarding issues of context, discourse, dialects, register, graphical representations, computer administration, cultural appropriateness, cognitive and linguistic equivalence, and format familiarity (CAE, 2010, GS.13). The training also stressed the importance of simultaneously and iteratively examining the task, scoring rubric, and response format (CAE, 2010, GS.13).

Country B provided feedback for performance task adaptation that was documented by the US organizing agency. The team was able to provide the US PI information about adaptation during the initial meeting in New York (CAE, GS.26, 2010). Also, the team provided feedback

on their progress in March 2010 (Young, personal communication, March 26, 2010). In April 2010, the team was also able to email their comments on the list of modifications that the US organizing agency had compiled as well as their final modification suggestions (Young, personal communication, April 28, 2010).

The Country B team completed all tasks associated with performance task adaptation within the allotted amount of time. Team members read all material associated with adaptation by the initial meeting in New York (CAE, GS.26, 2010). The team also submitted all task modifications suggestions in March 2010 (Young, personal communication, March 26, 2010). In addition, the Country B team reviewed the US PI's list of acceptable modifications and completed the task adaptation process in April 2010 (CAE, personal communication, April 19, 2010; Young, personal communication, April 28, 2010).

The US organizing agency sent the Country B team four published articles and one document that the agency's staff created. All material addressed important issues requiring attention during the adaptation phase. Despite having access to this information, the Country B team did not use all of the tools suggested in the documents. The team did not address the scoring rubric, language usage, or the attention to disconfirming evidence.

During the official adaptation stage of the study, the Country B team worked on its own. The team did not obtain assistance from anyone in their country who was not already on their team. All five members of the Country B team met to discuss modification suggestions (Solano-Flores & Chia, 2010, Interview).

Country B team members received support from the US organizing agencies during performance task adaptation. The organization provided documentation that included guidance for the adaptation process (Kurpius & Shavelson, personal communication, January 9, 2010). The US organizing agency also created a compilation of all modifications suggested by all countries that would not have a negative impact on the intended constructs (CAE, personal communication, April 19, 2010). The agency also provided final English versions of the performance task that incorporated all accepted adaptations (CAE, personal communication, May 11, 2010).

Country C

The Country C team did not indicate any challenge in communication. Team members participated in the specific adaptation needs during the initial meeting in New York (CAE, GS.26, 2010). Communication was also timely for the US organizing agency to collect modification suggestions from Country C as well as modify PTs and send them to countries for review, comment and completion (CAE, personal communication, April 19, 2010). However, there is no evidence that the Country C team completed the last feedback activity for task adaptation. The organizing agency was able to send the final version of the English PTs to Country C in time by the May deadline (CAE, personal communication, May 11, 2010).

Country C was to gather measurement expertise throughout the process. However, there was no evidence of an assessment expert giving direct feedback to Country C's national project manager or the project manager for the US organizing agency.

The Country C team had extensive experience with research and issues regarding higher education (CAE, GS.11, 2010). However, the team did not include expertise in translation. Instead the team included input from representatives of each participating university (Solano-Flores & Chia, 2010, Interview). The representatives came from diverse background (e.g., literature, engineering, compute science) and applied the core principal of the theory of test translation error: actively seek components of the performance tasks that would not be appropriate to students (Solano-Flores & Chia, 2010, Interview). The Country C performance tasks included minor modifications (CAE, 2010, Progress) but the national project manager stated that the team had made more serious suggestions for adaptation that were not included (Solano-Flores & Chia, 2010, Interview) and had concerns with some components of the rubric (Solano-Flores, 2010, Visit).

There is not enough evidence indicating that the team for Country C was able to coordinate the multiple steps involved in the adaptation process within the time allotted (see B7-D7 for documents). There were some points in the process during which Country C did not offer suggestions. However, the team seemed to focus on various aspects of adaptation. Guided by the theory of test translation error (TTTE) the Country C team examined if the contexts were culturally appropriate (Solano-Flores & Chia, 2010, Interview). Using paragraphs as a unit of analysis the team critically reviewed the use of idiomatic expressions, names, proper titles, proper scientific terminology, and geographic locations (Solano-Flores & Chia, 2010, Interview). Working with American schools, the team also was aware of issues of language and dialect (Solano-Flores & Chia, 2010, Interview).

The Country C team was not able to participate in every review opportunity that took place during the adaptation process. Team members reviewed specific adaptation needs during the initial meeting in New York (CAE, GS.26, 2010). The team was also able to share their modification suggestions with the organizing agency's PI during the third week of April 2010 (CAE, personal communication, April 19, 2010). However, there is no evidence that the Country C team completed the last feedback activity for review during task adaptation. The organizing agency was able to send the final version of the English PTs to Country C in May for their review (CAE, personal communication, May 11, 2010).

In January 2010, the US organizing agency's principal investigator emailed the Country C team the conceptual framework explaining the goals, importance, and an overview of the challenges associated with test adaptation (Shavelson & Kurpius, personal communication, January 29, 2010; CAE, 2010, GS.4). The following month, the US organizing agency sent Country C team members four documents related to test adaptation: Adapting Educational and Psychological Tests for Cross-Cultural Assessment (2005) by R. Hambleton, P. F. Merenda, C. D. Spielberger (Eds.) (GS.7); International Guidelines on Computer-Based and Internet

Delivered Testing (2005) (GS.8); Theory of Test Translation Error (2009) by G. Solano-Flores, E. Backhoff, L. A. Contreras-Niño (GS.9); Universal Design Applied to Large Scale Assessments (2002) by S. J. Thompson, C. J. Johnstone, M. L. Thurlow (GS.10) (Shavelson & Kurpius, personal communication, February 9, 2010). In addition, the team participated in the training offered in New York City that addressed details of test adaptation. A psychometrician specializing in test translation and adaptation explained and provided hands on practice exercises regarding issues of context, discourse, dialects, register, graphical representations, computer administration, cultural appropriateness, cognitive and linguistic equivalence, and format familiarity (CAE, 2010, GS.13). The training also stressed the importance of simultaneously and iteratively examining the task, scoring rubric, and response format (CAE, 2010, GS.13).

Country C provided feedback regarding performance task adaptation progress. Team members gave feedback on the documentation and process planned for adaptation during the initial meeting in New York (CAE, GS.26, 2010). The US organizing agency's PI was also able to collect modification suggestions from Country C (CAE, personal communication, April 19, 2010). However, there is no evidence that the Country C team participated in the end of April opportunity to document progress on adaptation.

The Country C team had a challenge meeting a deadline during the adaptation process. Team members were able to read all of the adaptation material that the US organizing agency provided prior to the initial meeting in New York (CAE, GS.26, 2010). The Country C team also provided the US PI with modification suggestions (CAE, personal communication, April 19, 2010). However, there is no evidence that the Country C team completed the last feedback activity for task adaptation.

The Country C team received all of the documents addressing task adaptation that the US organizing agency made available. The team's modification suggestions did not incorporate all of the issues addressed in the literature. The suggestions did not include issues with the scoring rubric, the technological issues, or maximum readability.

The Country C team was able to work on its own during the performance task adaptation activities. The team did not require support from people within the country who were not already on the team. The team consisted of one representative from each participating institution, two administrators, and the national project manager (Solano & Chia, 2010, Interview).

Country C team members received support from the US organizing agencies during performance task adaptation. The organization provided documentation that included guidance for the adaptation process (Kurpius & Shavelson, personal communication, January 9, 2010). The US organizing agency also created a compilation of all modifications suggested by all countries that would not have a negative impact on the intended constructs (CAE, personal communication, April 19, 2010). The agency also provided final English versions of the performance task that incorporated all accepted adaptations (CAE, personal communication, May 11, 2010).

Country D

The Country D team did not indicate any challenge in communication. Team members were able to discuss specific adaptation needs during the initial meeting in New York (CAE, GS.26, 2010). The PI for the US study was able to collect modification suggestions from Country D in time for the US organizing agency to modify PTs and send to them to countries for review, comment and completion (CAE, personal communication, April 19, 2010; Sanchez-Gomez, personal communication, April 21, 2010). The organizing agency was able to send the final version of the English PTs to Country D in time by the May deadline (CAE, personal communication, May 11, 2010).

The measurement expert for Country D was involved in the adaptation process of both performance tasks used in the AHELO study. The measurement expert was able to provide feedback on the performance task adaptation to the other country team members (Solano-Flores, Interview, 2010).

Although the Country D team included members who were knowledgeable regarding measurement, project management, and culture, it did not include translation expertise. One member of the team had a background in archaeology and two considered themselves bicultural (Solano-Flores & Chia, 2010, Interview). Although the team members were aware of the impact that certain aspects of performance tasks would have on future translation procedures there was little evidence of their knowledge helping the adaptation process (Solano-Flores & Chia, 2010, Interview). The modifications that arose through the adaptation process were minor and superficial (CAE, 2010, Progress).

The team for Country D was able to coordinate the multiple steps involved in the adaptation process within the time allotted (see B7-D7 for documents). However, the team did not address all potential challenges to task appropriateness during the adaptation process. The country focused on people's names, terms for geographic locations, and the impact that these may have on a student's cognition (Solano-Flores & Chia, 2010, Interview). The team seemed to focus more on the constructs addressed in each performance task more than the issues of adaptation (Solano-Flores & Chia, 2010, Interview).

The Country D team was able to participate in every review opportunity available during the adaptation process. Team members were able to review specific adaptation needs during the initial meeting in New York (CAE, GS.26, 2010). The PI for the US study was able to review modification suggestions from Country D by the third week in April 2010 and two days later the Country D team responded with notes from their review of that work (CAE, personal communication, April 19, 2010; Sanchez-Gomez, personal communication, April 21, 2010). The organizing agency was able to send the final version of the English PTs to Country D for review in May 2010 (CAE, personal communication, May 11, 2010).

In January 2010, the Country D team received the conceptual framework explaining the goals, importance, and an overview of the challenges associated with test adaptation (Shavelson & Kurpius, personal communication, January 29, 2010; CAE, 2010, GS.4). The following month, the US organizing agency sent Country D team members four documents related to test

adaptation: Adapting Educational and Psychological Tests for Cross-Cultural Assessment (2005) by R. Hambleton, P. F. Merenda, C. D. Spielberger (Eds.) (GS.7); International Guidelines on Computer-Based and Internet Delivered Testing (2005) (GS.8); Theory of Test Translation Error (2009) by G. Solano-Flores, E. Backhoff, L. A. Contreras-Niño (GS.9); Universal Design Applied to Large Scale Assessments (2002) by S. J. Thompson, C. J. Johnstone, M. L. Thurlow (GS.10) (Shavelson & Kurpius, personal communication, February 9, 2010). In addition, the team participated in the training offered in New York City that addressed details of test adaptation. A psychometrician specializing in test translation and adaptation explained and provided hands on practice exercises regarding issues of context, discourse, dialects, register, graphical representations, computer administration, cultural appropriateness, cognitive and linguistic equivalence, and format familiarity (CAE, 2010, GS.13). The training also stressed the importance of simultaneously and iteratively examining the task, scoring rubric, and response format (CAE, 2010, GS.13).

Country D contributed to each opportunity available to document task adaptation progress. Team members gave feedback on adaptation documentation and process captured by the US organizing agency during the initial meeting in New York (CAE, GS.26, 2010). The Country D team provided the US PI modification suggestions and follow-up comments on the list of acceptable modifications that the US organizing agency compiled (CAE, personal communication, April 19, 2010; Sanchez-Gomez, personal communication, April 21, 2010).

The Country D team was able to meet the deadlines created for the performance task adaptation process. Team members read all of the documents that the US organizing agency provided explaining the reasons for and the process of adaptation prior to the initial meeting in New York (CAE, GS.26, 2010). The team was able to provide the US PI with suggestions for modification by the end of April 2010 and applied the final US modification list to their final task adaptations (CAE, personal communication, April 19, 2010; Sanchez-Gomez, personal communication, April 21, 2010).

The US organizing agency provided the Country D team with four documents that provided assistance for task adaptation. The team was to use guidelines included in the document that the US organizing agency created and four academic articles. During the adaptation process the team's modification suggestions did not incorporate information from all of the documents. The team did not address some contextual issues, the scoring rubric, or the differences across the original and new students populations.

During the official adaptation process, the Country D team did not require support from anyone in the country who was not already on the team. The team felt that their bicultural, academic, and professional experiences would provide enough insight throughout the process (Solano-Flores & Chia, 2010, Interview).

Country D team members received support from the US organizing agencies during performance task adaptation. The organization provided documentation that included guidance for the adaptation process (Kurpius & Shavelson, personal communication, January 9, 2010).

The US organizing agency also created a compilation of all modifications suggested by all countries that would not have a negative impact on the intended constructs (CAE, personal communication, April 19, 2010). The agency also provided final English versions of the performance task that incorporated all accepted adaptations (CAE, personal communication, May 11, 2010).

Country E

The Country E team did not indicate any challenge in communication. Team members were able to discuss specific adaptation needs during the initial meeting in New York (CAE, GS.26, 2010). The PI for the US study was able to collect modification suggestions from Country E in time for the US organizing agency to modify PTs and send to them to countries for review, comment and completion (CAE, personal communication, April 19, 2010; Throndsen, personal communication, April 28, 2010). The organizing agency was able to send the final version of the English PTs to Country E in time by the May deadline (CAE, personal communication, May 11, 2010).

The Country E team had several team members with measurement expertise. The measurement experts participated in the performance task adaptation process (Solano-Flores, Interview, 2010).

The team from Country E had extensive experience in translation of international assessments, which they incorporated during the adaptation process (CAE, 2010, GS.11). One team member was a linguist who focused on reading education (Solano-Flores & Chia, 2010, Interview). In addition, during the adaptation process the Country E team collaborated with a translation company (Solano-Flores & Chia, 2010, Interview).

The national project manager for Country E was able to coordinate the multiple steps involved in the adaptation process within the time allotted (see B7-D7 for documents). However, the team did not address all potential challenges to task appropriateness during the adaptation process. The country focused on names and limited use of titles within the Country E context; the team admitted that they offered very few and minor modifications (Solano-Flores & Chia, 2010, Interview).

The Country E team participated in all of the review activities that took place during the adaptation process. Team members reviewed specific adaptation needs during the initial meeting in New York (CAE, GS.26, 2010). The PI for the US study was able to review modification suggestions from Country D in April 2010 (CAE, personal communication, April 19, 2010). Country E was able to review the US work and email comments for review within a couple of days (Throndsen, personal communication, April 28, 2010). The organizing agency was able to send the final version of the English PTs to Country E for review in May 2010 (CAE, personal communication, May 11, 2010).

In January 2010, the Country E team received the conceptual framework explaining the goals, importance, and an overview of the challenges associated with test adaptation (Shavelson & Kurpius, personal communication, January 29, 2010; CAE, 2010, GS.4). In February 2010,

the US organizing agency sent the Country E team members four documents related to test adaptation: Adapting Educational and Psychological Tests for Cross-Cultural Assessment (2005) by R. Hambleton, P. F. Merenda, C. D. Spielberger (Eds.) (GS.7); International Guidelines on Computer-Based and Internet Delivered Testing (2005) (GS.8); Theory of Test Translation Error (2009) by G. Solano-Flores, E. Backhoff, L. A. Contreras-Niño (GS.9); Universal Design Applied to Large Scale Assessments (2002) by S. J. Thompson, C. J. Johnstone, M. L. Thurlow (GS.10) (Shavelson & Kurpius, personal communication, February 9, 2010). In addition, the team participated in the training offered in New York City that addressed details of test adaptation. A psychometrician specializing in test translation and adaptation explained and provided hands on practice exercises regarding issues of context, discourse, dialects, register, graphical representations, computer administration, cultural appropriateness, cognitive and linguistic equivalence, and format familiarity (CAE, 2010, GS.13). The training also stressed the importance of simultaneously and iteratively examining the task, scoring rubric, and response format (CAE, 2010, GS.13).

Country E provided feedback during each opportunity to document task adaptation progress. The US PI collected feedback on adaptation documentation and planned process from the Country E team during the initial meeting in New York (CAE, GS.26, 2010). The PI for the US study was also able to document modification suggestions from Country E and send the team a compiled list of approved modifications for review (CAE, personal communication, April 19, 2010). The team provided comments on the US PI's modifications and provided information on their progress (Throndsen, personal communication, April 28, 2010).

The Country E team was able to work on task adaptation within the deadlines provided by the organizing agencies. Prior to the initial meeting in New York Country E team members had read documents explaining the adaptation process (CAE, GS.26, 2010). Country E team members provided the US PI with suggestions for modifications, reviewed the US PI's list of acceptable modifications, and used the list to complete their task adaptations (CAE, personal communication, April 19, 2010; Throndsen, personal communication, April 28, 2010).

Country E received five documents aimed to help country teams throughout the adaptation process. The US organizing agency created one document and provided four well-respected published articles. The Country E team did incorporate all of the suggestions included in the documents. The team did not study the differences between the original and new student populations. The team also did not address the issues associated with computer administration or the response format.

While adapting the performance tasks, the Country E team obtained support from people inside of Country E but who were external to the team. Initially only members of the team worked on the adaptation; the assessment experts began the process (Solano-Flores & Chia, 2010, Interview). The team also incorporated information from discussions that the team had while attending the New York City meeting (Solano-Flores & Chia, 2010, Interview). Finally,

the Country E team also included support from the translation company that would continue working with the team throughout the study (Solano-Flores & Chia, 2010, Interview).

Country E team members received support from the US organizing agencies during performance task adaptation. The organization provided documentation that included guidance for the adaptation process (Kurpius & Shavelson, personal communication, January 9, 2010). The US organizing agency also created a compilation of all modifications suggested by all countries that would not have a negative impact on the intended constructs (CAE, personal communication, April 19, 2010). The agency also provided final English versions of the performance task that incorporated all accepted adaptations (CAE, personal communication, May 11, 2010).

Task 6: Hire translators possessing qualifications set by coordinating group.

Country A

The US organizing agency provided information on translator qualifications during the initial meeting in New York City (CAE, GS.13, 2010) and again via email (CAE, personal communication, April 21, 2010). In addition, the agency shared specific deadlines about hiring translators (CAE, 2010, GS.1). The Country A team recruited and hired translators and did not indicate any challenges with regard to timeliness of the communication.

The Country A team was able to hire translators with extensive academic training and translation experience (Solano-Flores & Chia, Interview, 2010). In addition, the translators who worked with the Country A team had experience with academic test translation (CAE, 2010, Module).

At the time of the project Country A did not have a national or regional professional association that certified translators but the team was able to find a company to complete the work (Solano-Flores & Chia, 2010, Interview). As a result, the national project manager focused on working with professional translation companies that had translators with many years of experience working with different texts. These translators had formal training in translation studies which includes the study of language, linguistics, Country A, cultural studies, and communication. The translation company that the national project manager chose to work with verifies translators' degrees and requires that each translator complete a test. The translators hired for this project had graduated from the department of translation studies in a Country A university (Solano-Flores, Shavelson, & Chia, 2010, Meeting).

The national project manager for Country A was able to coordinate the multiple activities involved in hiring translators for the study. The national project manager was able to acquire funds for the translators and navigate the university's budget system (Solano-Flores & Chia, 2010, Interview). The national project manager was also able to hire translators and provide logistic and training documentation (Solano-Flores & Chia, 2010, Interview).

The US organizing agency provided the Country A team with qualifications that they were to use when hiring translators for the study (CAE, 2010, GS.13). However, there was no review opportunity for the translator hiring process.

Country A team members were able to take advantage of the training that the US organizing agency provided regarding hiring of translators. The team attended the training on translator qualifications that took place during the initial New York City meeting (CAE, 2010, GS.26). The US organizing agency conducted the training and made the training material available to the Country A team (CAE, 2010, GS.13; CAE, 2010, GS.4). In addition, the US organizing agency's PI sent Country A team members an email containing translator qualifications (Kurpius & Shavelson, personal communication, April 21, 2010).

Although there were no opportunities to document progress on hiring translators included in the study timeline and work plan, the US organizing agency collected information on the process after the fact. General notes were taken during a US agency's staff's site visit to Country A (Solano-Flores, 2010, Visit) and during an interview with Country A's translation team (Solano-Flores & Chia, 2010, Interview). The US organizing agency included information from these sources in official progress reports (CAE, 2010, Module A; CAE, Milestone; CAE, 2010, GS.30).

The US organizing agency provided Country A with initial information on translator qualifications during the initial meeting in New York City (CAE, GS.13, 2010) and again via email (CAE, personal communication, April 21, 2010). The Country A team recruited and hired translators. However, they indicated that it was challenging for them to do so because of the time of year. The language center at the university could not supply the resources needed (Solano-Flores & Chia, 2010, Interview). In addition, the amount of time necessary to gather all of the required approvals through the university system for the funds was challenging to do during that time of year (Solano-Flores & Chia, 2010, Interview).

Early in the study Country A received documents containing information about hiring translators from the US organizing agency (CAE, 2010, GS.4; CAE, 2010, GS.13). Furthermore, the US agency emailed the team specific qualifications for translators (Kurpius & Shavelson, personal communication, April 21, 2010). The team was able to discuss the qualifications for translators (Solano-Flores & Chia, 2010, Interview) and apply the information when hiring (CAE, 2010, Module A). There is no evidence that the Country A team found using the documents challenging.

The Country A team had support from within the country that was external to the team. At the time of the project Country A did not have a national or regional professional association that certified translators but the team was able to find a company to complete the work (Solano-Flores & Chia, 2010, Interview). The national project manager gathered information from university professionals and, ultimately, worked with a professional translation company located in Country A (Solano-Flores & Chia, 2010, Interview).

During the New York City meeting Country A team members received training on hiring translators from the US organizing agency (CAE, 2010, GS.26). The Country A team also received electronic copies of materials addressing translator qualifications from the US organizing agency (CAE, 2010, GS.13; CAE, 2010, GS.4). Finally, the US agency's PI sent Country A team members an email containing a list of translator qualifications (Kurpius & Shavelson, personal communication, April 21, 2010).

Country B

The Country B team received information about translator requirement during the meeting in New York City (CAE, GS.13, 2010) and again through an email that the US coordinating agency sent (CAE, personal communication, April 21, 2010). Country B did not indicate any issues with communication timeliness.

The Country B team was able to hire people to act as members of the translation team. Although the translators had experience within their academic fields of study, none of the people hired had measurement expertise (Solano-Flores, 2010, Visit). However, the team's measurement expert participated in the translation process (Solano-Flores & Chia, 2010, Interview).

The Country B team initially approached a specialist who worked with the PISA process about translation services. This contact's recommendations contradicted the recommendations made by the US organizing agency (Solano-Flores & Chia, 2010). This person advised that hiring a person who only specializes in translation could cause problems due to lack of education or assessment backgrounds. As a result, the national team focused on finding someone with a doctorate, and favored a doctorate from a US institution of higher learning, and not a certification. As a result the two translators were professors in a department similar to linguistics departments in the United States (Solano-Flores & Chia, 2010).

Country B's national project manager was able to coordinate the multiple steps involved in hiring translators for the AHELO study. The team contacted academics, researchers, and professional translators and was able to hire staff for translation (Solano-Flores & Chia, 2010, Interview). The team was also able to fund the translation team (Solano-Flores & Chia, 2010, Interview). Finally, the team provided the translation team with the logistical and training documents (Solano-Flores & Chia, 2010, Interview).

The US organizing agency provided the Country B team with qualifications that they were to use when hiring translators for the study (CAE, 2010, GS.13). However, there was no review opportunity for the translator hiring process.

Country B team members participated in the training opportunities addressing hiring translators that the US organizing agency made available. Team members attended the training that took place during the initial New York City meeting (CAE, 2010, GS.26). The US organizing agency also made the materials from the training session available to the Country B team (CAE, 2010, GS.13; CAE, 2010, GS.4). In addition, the PI for the US organizing agency

emailed the Country B team the translator qualifications (Kurpius & Shavelson, personal communication, April 21, 2010).

The US organizing agency did not include official opportunities to document progress on hiring translators included in the study timeline and work plan. However, the US agency collected information on the Country B team's process after the hiring was completed—during the translation process. General notes were taken during a US agency's staff's site visit to Country B (Solano-Flores, 2010, Visit) and during an interview with Country B's translation team (Solano-Flores & Chia, 2010, Interview). The US organizing agency included information from these sources in official progress reports (CAE, 2010, Module A; CAE, Milestone; CAE, 2010, GS.30).

The Country B team received information about translator requirement during the meeting in New York City (CAE, GS.13, 2010) and again through an email that the US coordinating agency sent (CAE, personal communication, April 21, 2010). The team experienced some difficulty when hiring translators because of the due dates. Translators with the most pertinent qualifications were working on another international assessment (Solano-Flores & Chia, 2010, Interview).

The US organizing agency provided the Country B team with documents addressing the hiring of translators for the study. The agency emailed the documents and made them available online (CAE, 2010, GS.4; CAE, 2010, GS.13). In addition, the US agency provided a succinct list of translator qualifications in an email (Kurpius & Shavelson, personal communication, April 21, 2010). The team was able to discuss the qualifications and their connection to the translation process (Solano-Flores & Chia, 2010, Interview). At no point did the Country B team indicate that the documents were difficult to follow.

The Country B team had support from within the country that was external to the team. The Country B team approached a locals specialist who worked with the PISA process about translation services, made further inquiries among academics, finally two translators who were professors in a department similar to linguistics departments in the United States (Solano-Flores & Chia, 2010).

Country B team members attended the training that the US organizing agency offered during the initial New York City meeting (CAE, 2010, GS.26). The US organizing agency also made the materials including information about hiring translators available to the Country B team (CAE, 2010, GS.13; CAE, 2010, GS.4). In addition, the PI for the US organizing agency emailed the Country B team a succinct list of translator qualifications (Kurpius & Shavelson, personal communication, April 21, 2010).

Country C

The US organizing agency provided the Country C team with information addressing translator qualifications weeks before the team had to hire the translators (CAE, 2010, GS.13; CAE, personal communication, April 21, 2010; CAE, 2010, GS.1). At no point did the Country

C team share that they experienced challenges due to communication during the translator hiring process.

Following guidelines that the US organizing agency provided, Country C was able to hire translators with strong academic and professional backgrounds (Solano-Flores, 2010, Visit). However, there is no evidence that any of the translators on the Country C team had measurement expertise.

The Country C team experienced challenges with regard to meeting all of the qualifications suggested by the US organizing entity. Country C does not have a professional translation certification process (Solano-Flores & Chia, 2010, Interview). Initially, the national project manager and committee members representing different institutions of higher learning in Country C approached academics in the translation and linguistics departments. Unfortunately, due to the timeframe and previous commitments these academics were not able to participate. The Country C team was able to seek assistance from academics specializing in translation. The team was able to find professional translators with experience working with complex documents (Solano-Flores & Chia, 2010, Interview). The translation reviewers were highly qualified (Solano-Flores & Chia, 2010, Meeting). One had a formal background in translation and works as a professional translator. This translator was finishing graduate studies in the translation masters program at a university (Solano-Flores, 2010, Meeting). The second translator had an earned doctorate in linguistics (Solano-Flores, 2010, Meeting).

The Country C national project manager went through a rigorous process to hire translators for the study. The NPM was able to acquire funding for the translation team (Solano-Flores & Chia, 2010, Interview). The NPM worked through three rounds of translator selection until the NPM was able to hire the translators best suited for the study (Solano-Flores & Chia, Interview, 2010). The NPM also provided the translators with logistic and training materials (Solano-Flores & Chia, 2010, Interview).

The US organizing agency provided the Country C team with qualifications that they were to use when hiring translators for the study (CAE, 2010, GS.13). However, there was no review opportunity for the translator hiring process.

Country C team members participated in the training opportunities addressing hiring translators that the US organizing agency made available. Team members attended the training that took place during the initial New York City meeting (CAE, 2010, GS.26). The US organizing agency also made the materials from the training session available to the Country C team (CAE, 2010, GS.13; CAE, 2010, GS.4). In addition, the PI for the US organizing agency emailed the Country C team the translator qualifications (Kurpius & Shavelson, personal communication, April 21, 2010).

The study's timeline and work plan that the US organizing agency created did not include official opportunities to document each country's progress while hiring translators. However, the Country C team provided information about the process later in the study. The Country C team participated in an interview conducted by the US organizing agency (Solano-Flores & Chia,

2010, Interview). In addition. the US agency's staff who visited the country provided notes on the team's progress, which included information on the translators (Solano-Flores, 2010, Visit). The US organizing agency included information from these sources in official progress reports (CAE, 2010, Module A; CAE, Milestone; CAE, 2010, GS.30).

The US organizing agency provided initial information on translator qualifications during the initial meeting in New York City (CAE, GS.13, 2010) and again via email (CAE, personal communication, April 21, 2010). The Country C team was able to hire translators but shared that there it was somewhat challenging in the amount of time allotted. The team had to conduct three rounds of searching because the initial prospects did not have the qualifications stipulated in the documents (Solano-Flores & Chia, 2010, Interview). During the third round the team sought assistance from academics who had connections within the translation field, which demanded additional time (Solano-Flores & Chia, 2010, Interview).

At the beginning of the study Country C received documents containing information about hiring translators from the US organizing agency (CAE, 2010, GS.4; CAE, 2010, GS.13). In addition, the US agency emailed the Country C team a succinct list of specific qualifications for translators (Kurpius & Shavelson, personal communication, April 21, 2010). The team was able to discuss the qualifications for translators and reasons for the qualifications (Solano-Flores & Chia, 2010, Interview). There is no evidence that the Country C team found using the documents challenging.

The Country C team had in-country support that was external to the team. The Country C national project manager and committee members representing different institutions of higher learning in Country C approached academics in the translation and linguistics departments (Solano-Flores & Chia, 2010, Interview). The Country C team was able to find assistance from academics specializing in translation and, ultimately, was able to find professional translators (Solano-Flores & Chia, 2010, Interview).

During the New York City meeting Country C team members participated in training addressing hiring translators from the US organizing agency (CAE, 2010, GS.26). The Country C team also had access to electronic copies of materials addressing translator qualifications from the US organizing agency (CAE, 2010, GS.13; CAE, 2010, GS.4). Finally, as a reminder, the US agency's PI sent Country C team members an email containing a list of translator qualifications (Kurpius & Shavelson, personal communication, April 21, 2010).

Country D

The Country D team received information about translator qualification at several points. In addition to documentation at the beginning of the study, the US organizing agency also provided information immediately preceding the translation stage of the study (CAE, 2010, GS.13; CAE, 2010, GS.1; CAE, 2010, GS.26). The team did not give any indication that they experperineced trouble with timely communication.

The Country D team was able to hire translators for the study. One of the translators had experience with test translation for a law school in Country D (Chia, 2011, Rubric). In addition, the measurement expert for Country D was involved in the translation process.

During selection of translators for the project, the Country D team placed emphasis on finding translators who were bicultural. The core members of the national team wanted to find people who could dominate both cultures (Solano-Flores & Chia, 2010, Interview). The second criteria on which the team focused was that the person have academic training in education because of the evaluation and assessment component of the project. One translator's academic background was in social sciences with a major in economics and a minor in Latin American studies. A second translator was a systems engineer with a master's degree in education. A core team member, with an academic background in archaeology and a master's degree in education with a focus on curriculum and instruction served as a third translator. Both translators usually live in Guadalajara (Solano-Flores, 2010, Meeting).

The team for Country D was able to coordinate the multiple steps involved in hiring translators for the AHELO feasibility study. The team provided the translators with the documentation that they needed to complete the hiring process: logistical and training material (Solano-Flores & Chia, 2010, Interview). The team considered the qualifications that the US organizing agency provided for translators and hired the people they felt were best suited for the study (Solano-Flores & Chia, 2010, Interview). However, the staff hired did not possess the translation expertise included in the guidelines that the US organizing agency provided (CAE, 2010, GS.13).

The US organizing agency provided the Country D team with qualifications that they were to use when hiring translators for the study (CAE, 2010, GS.13). However, there was no review opportunity for the translator hiring process.

The US organizing agency provided all country teams training opportunities about hiring translators. Country D team members attended the training that the US organizing agency conducted during the initial meeting in New York City (CAE, 2010, GS.26). The team was able to access training material used during the training online (CAE, 2010, GS.13; CAE, 2010, GS.4). In addition, the US organizing agency sent Country D team members a reminder email containing translator qualifications (Kurpius & Shavelson, personal communication, April 21, 2010).

The US organizing agency created a timeline and work plan for all of the countries participating in the study. Neither document included opportunities to document individual country progress during the hiring of translators. However, Country D provided information on the process during the translation process. The US organizing agency conducted an interview during which information about hiring translators was gathered (Solano-Flores & Chia, 2010, Interview). Also, the US agency's staff who visited Country D shared notes about the experience with the translation team (Solano-Flores, 2010, Visit). The US organizing agency included

information from these sources in official progress reports (CAE, 2010, Module A; CAE, Milestone; CAE, 2010, GS.30).

The Country D team had several weeks to hire translators once the US organizing agency provided information on translator qualifications. The Country D team indicated that the deadlines did not create any challenge in hiring translators. The team had found and hired translators before they had even completed the adaptation process (Solano-Flores, 2010, Interview).

The US organizing agency provided the Country D team members with documents addressing hiring of translators for the study (CAE, 2010, GS.4; CAE, 2010, GS.13). Additionally, the US agency emailed the team specific qualifications for translators (Kurpius & Shavelson, personal communication, April 21, 2010). During the site visit the team was able to discuss the qualifications for translators and how they used them during the hiring process (Solano-Flores & Chia, 2010, Interview). There is no evidence that the Country D team found using the documents challenging.

During selection of translators for the project, the Country D team found in-country support that was external to the team. The team members looked to people with whom they had worked in the past (Solano-Flores & Chia, 2010, Interview).

Country D team members attended a training session that the US organizing agency conducted during the initial meeting in New York City (CAE, 2010, GS.26). The US organizing agency made training materials and other related documents available electronically (CAE, 2010, GS.13; CAE, 2010, GS.4). In addition, the US organizing agency sent Country D team members an email containing a succinct list of translator qualifications (Kurpius & Shavelson, personal communication, April 21, 2010).

Country E

Early in the study the US communicated hiring qualifications for translators to the Country E team. In addition to documentation at the beginning of the study, the US organizing agency also provided information immediately preceding the translation stage of the study (CAE, 2010, GS.13; CAE, 2010, GS.1; CAE, 2010, GS.26). The Country E team did not express challenges to the hiring of translators due to communication timeliness.

Country E focused on working with academics who had extensive experience in translation and administration of other international assessments, such as PISA and TIMSS (Solano-Flores and Chia, 2010, Interview). One person was a linguist working originally in reading education. Another translation team member had a background in math and science. The third had experience in special needs and general education. These three team members also had extensive experience in assessment and considered themselves experts in that field. The team also worked with a private translation company who was hired by the university in charge of the project (Solano-Flores, G & Chia, 2010, Meeting).

The Country E team was able to hire translators with extensive translation experience. The translators and several team members who participated in the translation process had measurement expertise (Solano-Flores, Interview, 2010).

The national project manager for Country E was able to ensure that the national team included people with expertise that would allow them to serve as translators for the project (Solano-Flores & Chia, 2010, Interview).

The US organizing agency provided the Country E team with qualifications that they were to use when hiring translators for the study (CAE, 2010, GS.13). However, there was no review opportunity for the translator hiring process.

Country E team members were given several training opportunities that addressed hiring translators for the study. The team attended training that the US organizing agency conducted during the initial meeting in New York City (CAE, 2010, GS.26). In addition, the US organizing agency made material used during the training available to Country E team members online and via email (CAE, 2010, GS.13; CAE, 2010, GS.4). Finally, the US organizing agency emailed the Country E team a reminder of qualifications that they should look for when hiring translators (Kurpius & Shavelson, personal communication, April 21, 2010).

The timeline and work plan that the US organizing agency created for the study did not include official opportunities to document county progress on hiring translators. However, the US organizing agency collected data on the translator hiring process during translation. The US organizing agency conducted an interview with Country E's translation team during which they shared information about the hiring process (Solano-Flores & Chia, 2010, Interview). In addition, the US agency's staff who visited Country E shared notes that included information about the translation team (Solano-Flores, 2010, Visit). The US organizing agency included information from these sources in official progress reports (CAE, 2010, Module A; CAE, Milestone; CAE, 2010, GS.30).

The US organizing agency provided information regarding qualifications for translators early in the study. There were several weeks from the time that the Country E team received documentation on hiring translators to the time that they had to have the team in place. The team stated that they did not experience any challenges hiring translators (Solano-Flores, 2010, Interview).

Early in the study Country E received documents containing information about hiring translators from the US organizing agency (CAE, 2010, GS.4; CAE, 2010, GS.13). Furthermore, the US agency emailed the team specific qualifications for translators (Kurpius & Shavelson, personal communication, April 21, 2010). The team was able to discuss the qualifications for translators (Solano-Flores & Chia, 2010, Interview) and apply the information when hiring (CAE, 2010, Module A). There is no evidence that the Country A team found using the documents challenging.

Country E team members had support from professionals who were in the country but not part of the team. The team worked with a private translation company who was hired by the university in charge of the project (Solano-Flores, G & Chia, 2010, Interview).

The US organization supported Country E team members for hiring translators. The team attended training that the US organizing agency conducted during the initial meeting in New York City (CAE, 2010, GS.26). In addition, the US organizing agency made material used during the training available to Country E team members online and via email (CAE, 2010, GS.13; CAE, 2010, GS.4). Finally, the US organizing agency emailed the Country E team a reminder of qualifications that they should look for when hiring translators (Kurpius & Shavelson, personal communication, April 21, 2010).

Task 7: Translate the assessment.

Country A

Throughout the translation process, country teams relied on constant communication with the US organizing agency's PI for guidance and support. The PI organized communication between countries as well. Communication was always initiated via email but teleconferences also took place. There is no indication that anyone on the Country A team experienced a challenge with timely communication. Furthermore, the PI asked the Country A NPM for feedback on communication procedures at the end of the translation and adaptation process. The NPM stated that communication made it easy to following what was happening throughout the study—despite having very few in-person meetings (Shavelson & Kurpius, 2010, End). The NPM said the communication process itself was not challenging (Shavelson & Kurpius, 2010, End). The NPM also expressed that the teleconferences to discuss progress worked well (Shavelson & Kurpius, 2010, End). The team responded to requests for communication throughout the process in a timely manner. In fact, the Country A team responded to emails immediately.

The translators working with the Country A team familiarized themselves with the tasks' constructs, scoring rubric, and scoring handbook (Chia, 2011, Rubric). Specifically, they examined the applicability of the point scale system used in the rubric to Country A raters and students and the terminology that differentiated the score categories (Chia, 2011, Rubric). The translators also examined each score category, which represented each construct being measured, for appropriateness (Chia, 2011, Rubric). Finally, within each construct, the translators looked at the familiarity that Country A participants would have with each attribute as defined in the scoring rubric (Chia, 2011, Rubric). However, during an interview one of the translators shared that, although the team did not suggest changes at the time of translation, further reflection led him to believe that Country A students would not be aware that all attributes would hold equal weight in the scoring process (Chia, 2011, Rubric). Furthermore, one of the attributes was not something that would normally be emphasized to or expected from Country A students (Chia, 2011, Rubric). Finally, the translator noted that in Country A the scale usually has less point

differentiation for each construct in the scale but insisted this would be challenging if scorer training took place (Chia, 2011, Rubric).

The professional translation company that the Country A NPM contracted provided translators with many years of experience working with different texts. The person from the translation team responsible for the AHELO study was a professional translator (Solano-Flores, 2010, Visit). The person expressed some knowledge about translating rubrics and grading scales—including the format usually used in Country A's education system (Chia, 2011, Rubric). The same translator demonstrated knowledge of working with higher education documents (Chia, 2011, Rubric). There was no information about the type, or amount of, translation experience that the other translators possessed.

The team from Country A was able to complete the translation phase of the study, which included translation reconciliation. The national project manager provided the translators with all of the required documentation for process (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010, Interview). The translations were completed on schedule and there is no evidence that the translators were not given enough time to complete them. Although the NPM managed the logistics, one of the translators managed the reconciliation process (Solano-Flores & Chia, 2010, Interview).

After completing two independent translations Country A's translation team followed the steps for translation reconciliation (Solano-Flores, 2010, Visit). The team members appeared to be very open in their discussions about the translation process and translation errors (Solano-Flores, 2010, Visit). However, there is no evidence that a review of the translations took place during independent translations or translation reconciliation activities. An extensive review of translations was scheduled for a separate step in the process (CAE, 2010, GS.13; CAE, 2010, GS.4).

Country A participated in the training that the US organizing agency conducted during the New York City meeting (CAE, 2010, GS.26). The training included information on the translation process, which stressed the importance of two translators creating two independent versions of each of the performance tasks (CAE, 2010, GS.26; CAE, 2010, GS.13). The in-person training also addressed the reconciliation process that would result in one translation per performance task (CAE, 2010, GS.26; CAE, 2010, GS.13). The US organizing agency made the training material and supplemental material available to the Country A team online (CAE, 2010, GS.13; CAE, 2010, GS.4). Country A provided their translators with the material and an explanation on implementing the translation activities. The translators were able to explain how they implemented each step while working on the translation and translation reconciliation procedures (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010, Interview, Chia, 2011, Rubric).

There were several opportunities to document Country A's progress during performance task translation and reconciliation. First, the two translators created individual translations and provided the documents to the Country A team (Chia, 2011, Rubric). In addition, for each

performance task the translators created one reconciled version and submitted the work to the team (Solano-Flores, 2010, Visit). Finally, the US organizing agency documented Country A's progress in translation in several reports to the international organizing agency (CAE, 2010, Milestone; CAE, 2010, Module).

The US organizing agency provided Country A with clear deadlines for the translation process (CAE, GS.1, 2010). Originally, the teams were to complete the initial translation prior to the expert site visits (CAE, 2010, GS.1). However, the deadline was extended to August 15, 2010 regardless of the scheduled site visit date (CAE, 2010, GS.1). When the site visit to Country A took place the first week of July 2010 the two independent translations had been completed but the translation team had not completed the reconciliation process (Solano-Flores, 2010, Visit). The revised deadline worked well for the Country A team.

The US organizing agency provided the Country A team with several documents addressing the translation process (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). Throughout the documents, the agency presented the information in several ways: paragraphs, lists, and flowchart. The Country A translators did not indicate that they experienced challenges when working with the documents.

Members of the Country A team could not translate and reconcile the two performance tasks on their own. However, the Country A team was able to find a company in the country to complete the translation work (Solano-Flores & Chia, 2010, Interview). The Country A team did not require additional assistance from outside the team within the country.

The Country A team required assistance outside of the country to complete the translation and reconciliation process for the AHELO feasibility study. During the New York City meeting Country A team members received training on the translation and reconciliation process from the US organizing agency (CAE, 2010, GS.26). Another US agency staff members helped the Country A national project manager and translators with some of the reconciliation procedures (Solano-Flores, 2010, Visit). The Country A team also received electronic copies of materials addressing translation and verification from the US organizing agency (CAE, 2010, GS.13; CAE, 2010, GS.4).

Country B

Throughout the translation process Country B team members relied on constant communication with the US organizing agency's PI for guidance and support. The PI organized communication between countries as well. Communication was always initiated via email with few conference calls also taking place. The PI asked Country B team members for feedback on communication procedures. The team indicated that communication was easy and helpful (Shavelson & Kurpius, 2010, End). There were a few occasions when the US agency's staff needed to send the Country B team follow-up emails due to a lack of response (e.g., Shavelson, personal communication, January 12, 2010).

The Country B team produced translations of the scoring materials for the performance tasks. However, there is no evidence of measurement expertise present during the translation

portion of the study for Country B. Furthermore, the Country B team was unable to participate in a follow-up interview that was to address specifically aspects of scoring for each performance task.

As professors in a university in Country B, two of the translators were familiar with measurement practices and student experiences in higher education (Solano-Flores & Chia, 2010, Interview). However, there is no information about the translators' prior experience in translation. There is no evidence that the translators had experience translating documents for higher education or with test translation.

The Country B team completed the translation and translation reconciliation process for the AHELO feasibility study. The national project manager provided all translation team members with the documents requiring translation and provided a special Country B reviewer, with experiencing in publishing, for the reconciliation process (Solano-Flores & Chia, 2010, Interview). The national project manager also participated in the reconciliation process (Solano-Flores & Chia, 2010, Interview). However, the project manager was not able to ensure that the translation team follow the specified guidelines created by the US organizing agency and failed to meet deadlines (Solano-Flores, 2010, Visit; Solano-Flores, 2010, Cross).

As called for in the study's translation process, the Country B translators finished two independent translations of each of the two performance tasks (CAE, 2010, GS.13; CAE, 2010, GS.4). The translation team also worked together to complete the translation reconciliation process and produce one translated version of each performance task (Solano-Flores, 2010, Visit). However, there is no indication that the team conducted a review during this process. On the contrary, there is evidence that communication between persons of different academic standing was not fluid (Solano-Flores, 2010, Visit).

The US organizing team provided training addressing the translation process during the initial meeting in New York City (CAE, 2010, GS.26). Country B team members participated in the training, which addressed independent translations and translation reconciliation (CAE, 2010, GS.26). In addition, the US organizing agency placed the training material and supplemental documents available to the Country B team online (CAE, 2010, GS.13; CAE, 2010, GS.4). The Country B team members made the training and supplemental documents available to the translators but relied on the translators' expertise during translation (Solano-Flores, 2010, Visit).

The Country B team was able to document progress during the translation and translation reconciliation process. The two translators created individual translations and provided them to the Country B team (Solano-Flores & Chia, 2010, Interview). Once the translators reconciled the translations they provided the translation of each performance task to the Country B team (Solano-Flores, 2010, Visit). The US organizing agency documented Country B's progress in translation in several reports to the international organizing agency (CAE, 2010, Milestone; CAE, 2010, Module).

The AHELO feasibility study included deadlines for the initial translation, including the reconciliation process (CAE, GS.1, 2010). Originally, the teams were to complete the translation

and reconciliation activities prior to the expert site visits (CAE, 2010, GS.1). However, the deadline was extended to August 15, 2010 regardless of site visit date (CAE, 2010, GS.1). When the site visit to Country B took place in June 2010 the two independent translations and reconciliation had been completed (Solano-Flores, 2010, Visit). The team was able to accomplish the process within the amount of time originally provided.

Country B team members had access to electronic versions of several documents addressing the translation and translation reconciliation process online (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). Subsequently, team members were able to provide their translators with the material (Solano-Flores, 2010, Visit). The documents presented the translation information in different formats: paragraph, list, and flowchart. Neither the translation team nor the Country B team shared any concerns or difficulties with the documents.

The Country B team was able to find translators from in the country to complete the translation work (Solano-Flores & Chia, 2010, Interview). The Country B team did not require additional assistance from outside the team within the country.

The Country B team found support for the translation and reconciliation process outside of the country. Country B team members attended the training addressing translation and reconciliation that the US organizing agency offered during the initial New York City meeting (CAE, 2010, GS.26). The US organizing agency also made the materials including information about translation and reconciliation available to the Country B team (CAE, 2010, GS.13; CAE, 2010, GS.4) who then made them available to translators.

Country C

Throughout the translation and adaptation process the Country C team relied on constant communication with the US organizing agency's PI for guidance and support. The PI organized communication between countries as well. Communication took place mainly via email with a few conference calls as well. During a call with the Country C team, the US PI asked for feedback on communication procedures. Team members stated that the US organizing agency had encouraged communication (Shavelson & Kurpius, 2010, End). There were very few occasions when the US agency's staff needed to send the Country C team follow-up emails due to a lack of response (e.g., Shavelson, personal communication, January 12, 2010).

The translators who worked with the Country C team became familiar with the scoring rubric, scoring handbook, and the intended constructs associated with each performance task. During the site visit one of the translators expressed concern with some of the constructs and attributes emphasized in the scoring rubric (Solano-Flores, 2010, Visit). During an interview that took place after the translated performance tasks had been submitted by the Country C NPM, a member of the Country C team expressed some concern with the scoring rubric. The team member expressed that some of the subtle differences throughout the point scale did not work well in Arabic and that some of the categories should be collapsed (Chia, personal communication, October 15, 2010).

The translators who worked with the Country C team had the desired qualifications that the US organizing agency suggested. The translators shared questions about localization and shared a different way to interpret the linguistic term 'register' based on past experience (Solano-Flores, 2010, Visit). In addition, as professors in local universities two of the translators were familiar with higher education documents. However, there is no additional information about the amount of experience conducting professional translation or the type of texts with which the translators had experience.

The Country C team completed the translation of the performance tasks for AHELO. The national project manager provided the translation team members with all of the documents requiring translation—and did so in a timely manner (Solano-Flores & Chia, 2010, Interview). Although Country C's NPM did not participate in the reconciliation process, he organized the translation team so that there was a clear leader during that activity (Solano-Flores & Chia, 2010, Interview).

As per the translation process stipulated by the US organizing agency, the Country C team asked two translators to independently translate each of the two performance tasks (CAE, 2010, GS.4; CAE, 2010, GS.13). The translators then proceeded to reconcile the two independent translations into one (Solano-Flores, 2010, Visit). At the time, however, there was no evidence that the reconciliation process included a review.

Two Country C team members participated in the translation and translation reconciliation training that the US organizing agency provided during the initial meeting in New York (CAE, 2010, GS.26). In addition, the team had access to the training materials and additional documents addressing the translation process online (CAE, 2010, GS.4; CAE, 2010, GS.13; CAE, 2010, 26). The team provided the translators with the training and supplemental documents and they were able to apply the process successfully during the translation and translation reconciliation activities (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010, Interview).

The country team and US organizing agency documented Country C's progress in performance task translation and reconciliation. The two translators created individual translations and provided them to the Country C team (Solano-Flores & Chia, 2010, Interview). In addition, the translators created one reconciled version of each performance task and submitted the work to the team (Solano-Flores, 2010, Visit). The US organizing agency documented Country C's progress in translation in several reports to the international organizing agency (CAE, 2010, Milestone; CAE, 2010, Module).

The US organizing agency provided country teams with deadlines for the translation process, which included two independent translations and translation reconciliation (CAE, GS.1, 2010; CAE, 2010, GS.13). Originally, the teams were to complete the initial translation prior to the expert site visits (CAE, 2010, GS.1). However, the deadline was extended to August 15, 2010 regardless of site visit date (CAE, 2010, GS.1). Due to religious holidays the site visit to Country C did not take place until September 2010. At the time of the Country C visit the two

independent translations and translation reconciliation were completed (Solano-Flores, 2010, Visit). Although the Country C team was able to complete the translation process, the team was not able to complete the activities by the deadline stipulated.

At the beginning of the AHELO study, the US organizing agency provided the Country C team with documents addressing the translation process (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). The documents presented information in paragraph, list, and graphic forms. The Country C team provided these documents to the translation team and they were able to apply the information (Solano-Flores, 2010, Visit). The team did not indicate a challenge with the documents.

The Country C team was able to find a company in the country to complete the translation work (Solano-Flores & Chia, 2010, Interview). During the translation process the Country C team did not require additional assistance from outside the team within the country.

Country C acquired help for the translation and reconciliation process from outside of the country. During the New York City meeting Country C team members participated in training addressing translation and reconciliation conducted by the US organizing agency (CAE, 2010, GS.26). The Country C team also had access to electronic copies of materials addressing translation and reconciliation, which were created and placed online by the US organizing agency (CAE, 2010, GS.13; CAE, 2010, GS.4).

Country D

Throughout the translation process Country D team members relied on constant communication with the US organizing agency's PI for guidance and support. The PI organized communication between countries as well. Communication was always initiated via email; however, there were also a few conference calls. There is no indication that anyone on the Country D team experienced a challenge with timely communication. Furthermore, during a conference call the PI asked the Country D team for feedback on communication procedures. The team indicated that communication was timely and helpful (Shavelson & Kurpius, 2010, End). Like the US agency's PI and other staff, the country team responded to requests for communication throughout the process in a timely manner.

The translators who worked with the Country D team familiarized themselves with the scoring rubric, scoring handbook, and the intended constructs measured within each performance task. At the time of translation, the team did not indicate that any changes needed to occur with regard to the scoring or intended constructs. However, during a conference call with one of the translators the person shared potential challenges with the rubric and constructs. The translator explained that the use of rubrics was not widespread in all departments throughout the universities (Chia, 2011, Interview). Furthermore, the majority of rubrics used in Country D contain a different point scale (Chia, 2011, Interview). However, the translator felt that with careful training scorers would be able to apply the rubric (Chia, 2011, Interview). The translator's greater concern dealt with one of the constructs included in the rubric. The translator explained that students would be expecting to be graded differently and that students would need

to see the rubric in advance to understand what they were being graded on (Chia, 2011, Interview). Furthermore, there were two attributes that the translator felt were not normally assessed in Country D and could cause students to be at a disadvantage (Chia, 20110, Interview).

Documents provide limited information about the previous translation experience of only one translator. The translator had experience translating documents for the law school at a local university in Country D (Chia, 2011, Rubric). However, there is limited information about the extent of experience that the translator possessed. In addition, there is no information about the other translator's experience.

The team for Country D was able to manage the performance task translation and translation reconciliation process of the feasibility study. The Country D team provided the translators with the documents requiring translation well in advance of the due date (Solano-Flores & Chia, 2010, Interview). The team also provided the translators with assistance during the translation reconciliation process (Solano-Flores & Chia, 2010, Interview).

Following the process instituted by the US organizing agency the Country D team asked that each of the two translators create independent translations for each of the two performance tasks (Solano-Flores, 2010, Visit). The two translators then reconciled the two translations into one version (Chia, 2011, Rubric). There was no indication that a review of the work occurred at this time. However, the communication regarding the translation process and translation errors appeared open and fruitful among the translators (Solano-Flores, 2010, Visit).

The US organizing agency provided the Country D team with training opportunities addressing the translation and reconciliation procedures. The Country D team attended the training that the US organizing agency conducted during the initial meeting in New York City (CAE, 2010, GS.26). The team also accessed the training materials used during the training—and supplemental documents—online (CAE, 2010, GS.13; CAE, 2010, GS.4). The Country D team members gave the translators all of the materials available online and provided guidance on the process. The translators were able to discuss and implement the translation process as well as the reconciliation step (Solano-Flores & Chia, 2010, Interview; Chia, 2011, Rubric).

The Country D team had two opportunities to document their progress in performance task translation and reconciliation. First, the two translators created individual translations and provided them to the Country D team (Chia, 2011, Rubric). Second, for each performance task the translators created one reconciled version and submitted the work to the team (Solano-Flores, 2010, Visit). In addition, the US organizing agency documented Country D's progress in translation in several reports to the international organizing agency (CAE, 2010, Milestone; CAE, 2010, Module).

At the beginning of the study, the US organizing agency provided Country D with a list of deadlines for important activities (CAE, 2010, GS. 1). The US agency was forced to revise the deadline for the translation process (CAE, 2010, GS.1). Originally, the teams were to complete the initial translation prior to the expert site visits (CAE, 2010, GS.1). However, the deadline was extended to August 15, 2010 regardless of site visit date (CAE, 2010, GS.1). When the site visit

to Country D took place the last week of July 2010 the two independent translations had been completed but the team had not completed the reconciliation process (Solano-Flores, 2010, Visit). The Country D team needed to have the original deadline extended.

The Country D team and translators had access to documents addressing the translation process. The US organizing agency created these documents and made them available electronically. The material presented the information in paragraph and list format as well as in a flowchart (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). The team stated that the documents were helpful (Chia, 2011, rubric; Solano-Flores & Chia, 2010, Interview) and did not indicate that they found any part of the material challenging.

The Country D team was able to find translators from Country D to complete the translation work (Solano-Flores & Chia, 2010, Interview). The Country D team did not require additional assistance from outside the team within the country during the translation and reconciliation process.

The US organizing agency helped Country D with the translation and reconciliation process. Country D team members attended a training session that the US organizing agency conducted during the initial meeting in New York City that addressed translation and reconciliation (CAE, 2010, GS.26). The US organizing agency made training materials and other related documents available electronically (CAE, 2010, GS.13; CAE, 2010, GS.4). In addition, the US organizing agency representative who completed the site visit helped translators with the reconciliation process (Solano-Flores, 2010, Visit).

Country E

The US organizing agency's PI provided Country E with guidance and support thoughout the translation process. The PI organized communication between countries as well. Communication was always initiated via email but included a few conference calls. There is no indication that anyone on the Country E team found the US agency's communication untimely. In addition, the PI asked Country E team members for feedback on communication procedures at the end of the translation process. The NPM indicated that the communication process itself was not challenging but was helpful (Shavelson & Kurpius, 2010, End). The country team responded to requests for communication throughout the process in a timely manner. There were very few occasions when the US agency's staff needed to send the Country E team follow-up emails due to a lack of response (e.g., Shavelson, personal communication, January 13, 2010).

Several of the Country E team members who participated in the translation process had measurement expertise (Solano-Flores, Interview, 2010). The team familiarized themselves with the scoring rubric, scoring handbook, and intended constructs. The team did not indicate any challenges with the scoring system for either PT. Lastly, the Country E team was not able to participate in a follow-up interview that was to address translation of scoring material.

The translators working with the Country E team had extensive experience in the field of translation. The translators had experience working with the translation of PISA and TIMSS

items (Solano-Flores and Chia, 2010, Interview). The team also had a great deal of experience working with educational documents—particularly in reading, mathematics, science, and special education (Solano-Flores, G & Chia, 2010, Interview). However, there was no information indicating that the team had experience with higher education material.

The Country E team completed the translation and translation reconciliation process for the feasibility study. The national project manager provided all members of the translation team with the documents requiring translation (Solano-Flores & Chia, 2010, Interview). Several members of the Country A team participated in the translation reconciliation process (Solano-Flores & Chia, 2010, Interview). The team did not express challenges with time or demands.

The Country E team conducted a translation and translation reconciliation process as required by the US organizing agency (CAE, 2010, GS.13, CAE, 2010, GS.4). Two translators each created a version of the two performance tasks; a third translator reconciled the two independent translations (Solano-Flores & Chia, 2010, Interview). In addition, the three translators worked together after the initial independent translations to improve the translations as much as possible (Solano-Flores & Chia, 2010, Interview). Therefore, Country E included a review process during this stage of the study without it being planned as part of the requirements listed by the US organizing agency.

The US organizing agency provided training and guidance for the translation and translation reconciliation process. The Country E team attended training that the US organizing agency conducted during the initial meeting in New York City (CAE, 2010, GS.26). In addition, the US organizing agency made material used during the training and supplemental documents available to Country E team members online (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.26). The Country E team provided translators with all documents addressing translation and reconciliation and guided them throughout the process (Solano-Flores, 2010, Visit).

Country E's progress in performance task translation and reconciliation was documented by the team and the US organizing agency. The translators submitted one reconciled version of each of the two performance tasks to the country team (Solano-Flores, 2010, Visit). Also, the US organizing agency documented Country E's translation progress in several reports to the international organizing agency (CAE, 2010, Milestone; CAE, 2010, Module).

The US organizing agency provided clear deadlines for the study. However, adjustments were made for the translation process (CAE, GS.1, 2010). Originally, each country team was to complete the initial translation—including the reconciliation process—prior to the expert site visit (CAE, 2010, GS.1). However, the deadline was extended to August 15, 2010 regardless of the site visit date (CAE, 2010, GS.1). The US agency's representative visited Country E in June 2010 (CAE, 2010, Visit). At the time of the site visit the two independent translations and the reconciliation process had been completed (Solano-Flores, 2010, Visit). The Country E translation team was able to work within the original time allotted for the process and did not require an extension of the deadline.

The US organizing agency created several documents addressing the translation process, which included reconciliation. The Country E team and translators had access to electronic versions of the document online. The material presented the information in paragraph and list format as well as in a flowchart (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). The team explained that they were able to work with the documents during translation (Solano-Flores & Chia, 2010, Interview) and did not indicate that they found any part of the material addressing translation challenging.

The Country E team was able to work with Country E translators to complete the translation work (Solano-Flores & Chia, 2010, Interview). The Country E team did not require additional assistance from outside the team within the country.

The US organization supported Country E team members with translation and reconciliation procedures. The team attended training that the US organizing agency conducted during the initial meeting in New York City (CAE, 2010, GS.26). In addition, the US organizing agency made material used during the training available to Country E team members online and via email (CAE, 2010, GS.13; CAE, 2010, GS.4).

Task 8: Review translation and notes from translation process.

Country A

At no point throughout the process did the Country A team indicate any challenge in communication about translation review. The US organizing agency emailed Country A information about deadlines for translation review during early communication (Kurpius & Shavelson, personal communication, January 19, 2010). During the initial meeting in New York City, in February 2010, the US agency's staff was able to answer team members' questions about the review process (CAE, GS.26, 2010). Subsequent informational exchanges took place related to the translation review process (Kurpius & Shavelson, personal communication, April 27, 2010; Kurpius & Shavelson, personal communication, May 11, 2010). A great deal of communication also took place to prepare for the translation review training scheduled to take place during the site visit (Ursin, personal communication, May 26, 2010; Kurpius, personal communication, June 10, 2010).

Using the AHELO guidelines Country A reviewed of the translations for each performance task, the scoring rubric, and the computer platform and interface language (CAE, 2010, GS.36). Using the TTTE, the team included team members with expertise in different fields, including measurement (CAE, 2010, GS.9). The country's national project manager, who had expertise in educational assessment, served as the measurement expert for the translation team (Ursin, personal communication, January 12, 2010).

The Country A national team was able to implement the translation review process as described by the TTTE (CAE, 2010, GS.36). The Country A translation team included translation expertise during the translation review process (Solano-Flores & Chia, 2010, Interview). The national project manager hired a third translator from a translation company to act as the translation expert during the review (Ursin, personal communication, August 19,

2010). However, there is no information about the amount of experience the person had as a professional translator. In addition, there is limited information about the team's translation adviser.

Although there is information about the assessment and translation expertise included in the translation review team, there is limited information on the process itself. The national project manager, who served on the translation review team, had proven to be an involved skilled facilitator during the translation review training (Solano-Flores, 2010, Visit). In addition, one of the translators involved in the review process shared that the translation review process had gone smoothly and focused on the TTTE (Chia, 2011, Rubric).

The US organizing agency provided the Country A team with instructions for the translation review process (CAE, 2010, GS.13). They also provided Country A with the coding form, which was to be used during the review (CAE, 2010, GS.36). Using these documents the Country A translation review team was able to review their translation review work.

Country A participated in the translation review training that the US organizing agency conducted during the initial group New York City meeting (CAE, 2010, GS.26). The training included information on the translation review process, which stressed the importance of multidisciplinary group discussion and collaboration when using the error dimensions provided (CAE, 2010, GS.26; CAE, 2010, GS.13). The training also stressed that reviewers should be actively seeking disconfirming evidence (CAE, 2010, GS.36). The US organizing agency made the training material and supplemental material available to the Country A team online, including the coding form the team would use to during the review process (CAE, 2010, GS.13; CAE, 2010, GS.4). During the site visit, Country A translation review team members learned about the process, reviewed key components of the TTTE, and were able to practice the process with the guidance from the US agency staff (Solano-Flores, 2010, Visit; Chia, 2011, Rubric).

The translation review process included constant opportunities to document progress. As part of the translation review documents the Country A review team had a list of ten error dimensions with their definitions and a coding form (CAE, 2010, GS.4). As part of the translation review process each member of Country A's review team filled out a coding form noting the identified error and the dimension(s) impacted (CAE, 2010, GS.36). The completed coding forms served as a means to document the progress that the Country A team made during the translation review process.

The US organizing agency did not provide Country A with a clear deadline for the translation review process (CAE, 2010, GS.1). The step is not addressed in the study's work plan (CAE, 2010, GS.1).

The US organizing agency provided Country A with material addressing the translation review of performance tasks, scoring rubric, and computer interface language. The Country A translation review team had access to electronic copies of the material used during the training, as well as supplemental documents, online (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.9; CAE, 2010, GS.36). The documents presented information in narrative, table, and bullet

form and explained the qualifications for the translation review team, the translation error dimensions, and the translation review process (CAE, 2010, GS.36). The material also included a coding form that teams would use to monitor and document their progress. Despite difficulty with the unit of analysis used for reviewing the translated documents, the Country A national project manager expressed that material was easy to follow (Solano-Flores, 2010, Visit; Shavelson & Kurpius, 2010, Conference).

The Country A national team was able to acquire in-country support that was external to the team when implementing the translation review process as described by the TTTE (CAE, 2010, GS.36). The Country A team was able to hire staff from a professional Country A translation company (Solano-Flores & Chia, 2010, Interview). The national project manager hired a third translator from a translation company to act as the translation expert during the review (Ursin, personal communication, August 19, 2010).

The Country A team required assistance outside of the country to complete the translation review process for the AHELO feasibility study. During the New York City meeting Country A team members received training on the translation review process from the US organizing agency (CAE, 2010, GS.26). Another US agency staff members helped the Country A national project manager and translators by training them on the process (Solano-Flores, 2010, Visit). The Country A team also received electronic copies of materials addressing translation review from the US organizing agency (CAE, 2010, GS.13; CAE, 2010, GS.4).

Country B

During initial communication, the US organizing agency emailed Country B team members information about deadlines for—and activities involve in—the translation review process (Kurpius & Shavelson, personal communication, January 19, 2010). In addition, at the initial meeting in New York City, in February 2010, the US agency's staff answered team members' questions about the review process (CAE, GS.26, 2010). The US PI and staff communicated additional information about the process and progress (Kurpius & Shavelson, personal communication, April 27, 2010; Kurpius & Shavelson, personal communication, May 11, 2010). A great deal of communication also took place to prepare for the translation review training scheduled to take place during the site visit (Ursin, personal communication, May 26, 2010; Kurpius, personal communication, June 10, 2010). At no point throughout the process did Country B team members indicate any challenge in communication about translation review. Country B's communication with the US agency's staff dealing with translation review was also timely (Solano-Flores, 2010, Comparison).

The Country B team completed the translation review process for the two performance tasks, the scoring rubric, and the computer platform and interface language (CAE, 2010, GS.36). As indicated by the TTTE, the Country B team included experts from different fields in the review process (CAE, 2010, GS.9). The Country B team included a measurement expert who participated in the process (Choi, personal communication, January 14, 2010).

The study's translation review process was different from that often used in cross-national studies. Based on the TTTE the translation review team should have translation expertise that is separate from the experts who worked on the actual translation and a translation advisor (CAE, 2010, GS.36). Although the Country B team was able to complete the internal translation review process, there is limited information about the team's implementation of the TTTE's multidisciplinary approach or attention to translator qualifications. In fact, there is no documentation addressing the people involved in the translation review process or their backgrounds.

Documents from the AHELO study provide information about the assessment and translation expertise included in Country B's translation review team (Solano-Flores & Chia, 2010, Interview; Solano-Flores, 2010, Visit). However, there was no information on Country B's implementation of the translation review process.

The US organizing agency provided the Country B team with material that would allow the translation team to review their work as they reviewed the translations. The agency made instructions for the translation review process available online (CAE, 2010, GS.13; CAE, 2010, GS.36). The agency also provided Country B with the coding form, which each reviewer used during the review process (CAE, 2010, GS.36). Using these documents the Country B translation review team was able to review their translation review work.

The US organizing agency provided several translation review training opportunities. Team members from Country B participated in the translation review training that the US organizing agency conducted during the initial group New York City meeting (CAE, 2010, GS.26). The training included information on the translation review process and stressed the importance of discussion between reviewers representing different disciplines (CAE, 2010, GS.26; CAE, 2010, GS.13). The training also emphasized that reviewers should be actively seeking disconfirming evidence (CAE, 2010, GS.36). The US organizing agency made the material available to the Country B team online, including the coding form the team would use to during the review process (CAE, 2010, GS.13; CAE, 2010, GS.4). During the site visit, Country B the translation review team learned about the process, reviewed key components of the TTTE, and were able to practice the review process with guidance from the US agency staff (Solano-Flores, 2010, Visit). However, given the social structure, it was difficult to communicate directly with the translation team. The team would speak to the NPM and the NPM reported back to TAT (Solano-Flores, 2010, Visit).

There were constant opportunities to document progress during the translation review process. As part of the material the US organizing agency provided for translation review the Country B review team received a list of ten error dimensions with their definitions and a coding form (CAE, 2010, GS.4). As part of the translation review process each member of Country B's review team filled out a coding form noting the identified error and the dimension(s) that each error impacted (CAE, 2010, GS.36). The completed coding forms served as a means to document the progress that the Country B team made during the translation review process.

The AHELO feasibility study did not include a deadline for the translation review process in the work plan that it provided to the Country B team members (CAE, GS.1, 2010).

The US organizing agency provided Country B with material addressing the translation review of performance tasks, scoring rubric, and computer interface language. The Country B translation review team had access to electronic copies of the material used during the training, as well as supplemental documents, online (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.9; CAE, 2010, GS.36). The documents presented information in narrative, table, and bullet form and explained the qualifications for the translation review team, the translation error dimensions, and the translation review process (CAE, 2010, GS.36). The material also included a coding form that teams would use to monitor and document their progress. During an end of study conference call, the Country B team expressed the material was easy to follow (Shavelson & Kurpius, 2010, Conference).

The Country B national team was able to acquire in-country support that was external to the team when implementing the translation review process (CAE, 2010, GS.36). The Country B team hired staff from local universities (Solano-Flores & Chia, 2010, Interview).

The Country B team found support for the translation review process outside of the country. Country B team members attended the training addressing translation review that the US organizing agency offered during the initial New York City meeting (CAE, 2010, GS.26). The US organizing agency also made the materials including information about translation review available to the Country B team (CAE, 2010, GS.13; CAE, 2010, GS.4) who then made them available to translators.

Country C

Communication between Country C and the US organizing agency was timely during the translation review process. During initial communication, the US organizing agency emailed team members information about the activities involved in the review process and their deadlines (Kurpius & Shavelson, personal communication, January 19, 2010). In addition, at the initial meeting in New York City, in February 2010, the US agency's staff and Country C's team members discussed the review process (CAE, GS.26, 2010). The US agency communicated additional information about the process and progress via email (Kurpius & Shavelson, personal communication, April 27, 2010; Kurpius & Shavelson, personal communication, May 11, 2010). The Country C team did not give any indication that communication about translation review had been challenging. Communication dealing with translation review on the part of Country C was also timely.

Following guidelines that the US organizing agency provided, Country C was able to review the translations of the two performance tasks, the scoring rubric, and the computer platform and interface language (CAE, 2010, GS.36). However, there is no evidence that any of the members of the translation review team were experts in measurement.

Based on the process on the TTTE, the translation review process for the AHELO study required a third translator—a translator who had not been part of the initial translation process

(CAE, 2010, GS.36). The person should have had enough experience to serve as a translation advisor (CAE, 2010, GS.36). The third translator for the Country C team was highly qualified and experienced. The person was a linguist and professor specializing in languages; the person was the coordinator of the graduate program in translation at a Country C team university (Solano-Flores & Chia, 2010, Interview).

The Country C translation review team included the translation expertise that the review process demanded (Solano-Flores, 2010, Interview). However, there is limited information on the level of project management expertise involved in Country C's implementation of the translation review process.

Each country had access to translation review documents online (CAE, 2010, GS.4; CAE, 2010, GS.13; CAE, 2010, GS.36). The US organizing agency provided the Country B translation review team with step-by-step instructions for completing the translation review process (CAE, 2010, GS.13; CAE, 2010, GS.36). They also provided Country C's reviewers with the coding form, which was to be used during the review (CAE, 2010, GS.36). Using these documents the Country C translation review team was able to review their translation review work.

The Country C translation review team participated in several translation review training opportunities offered by the US organizing agency. Country C participated in the New York City translation review training that the US organizing agency conducted (CAE, 2010, GS.26). The training addressed each step of the translation review process and stressed the importance of multidisciplinary group discussion while using the error dimensions provided (CAE, 2010, GS.26; CAE, 2010, GS.13). Reviewing the TTTE, the training also emphasized that reviewers should be actively seeking disconfirming evidence (CAE, 2010, GS.36; CAE, 2010, GS.9). The US organizing agency made the training material and supplemental material available to the Country C team online, including the coding form the team would use to during the review process (CAE, 2010, GS.13; CAE, 2010, GS.4). During the site visit, the Country C translation review team was able to practice the process with the guidance from the US agency staff (Solano-Flores, 2010, Visit).

The translation review process incorporated constant opportunities for Country C to document progress. As part of the translation review documents the US organizing agency sent Country C a list of ten error dimensions with their definitions and a coding form (CAE, 2010, GS.4). As part of the translation review process each member of Country C's review team used coding forms to note the errors each member identified and the dimension(s) that each error impacted (CAE, 2010, GS.36). The completed coding forms served as a means to document the progress that the Country C team made during the translation review process.

The US organizing agency did not provide the Country C translation review team with deadlines for the translation review process (CAE, 2010, GS.1). The study's work plan did not address the translation review process (CAE, 2010, GS.1).

The US organizing agency provided Country C with material addressing the translation review of performance tasks, scoring rubric, and computer interface language. The Country C

translation review team had access to electronic copies of the material used during the training, as well as supplemental documents, online (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.9; CAE, 2010, GS.36). The documents presented information in narrative, table, and bullet form and explained the qualifications for the translation review team, the translation error dimensions, and the translation review process (CAE, 2010, GS.36). The material also included a coding form that teams would use to monitor and document their progress. Despite difficulty with the unit of analysis used for reviewing the translated documents, the Country C national project manager expressed that material was easy to follow (Solano-Flores, 2010, Visit; Shavelson & Kurpius, 2010, Conference).

Based on the process on the TTTE, the translation review process for the AHELO study required a third translator—a translator who had not been part of the initial translation process (CAE, 2010, GS.36). The Country C team professionals associated with local universities (Solano-Flores & Chia, 2010, Interview).

Country C acquired help for the translation review process from outside of the country. During the New York City meeting Country C team members participated in training addressing translation review conducted by the US organizing agency (CAE, 2010, GS.26). The Country C team also had access to electronic copies of materials addressing translation and reconciliation, which were created and placed online by the US organizing agency (CAE, 2010, GS.13; CAE, 2010, GS.4). Finally, the US organizing agency provided detailed information and training on the translation review process during the country site visit (Solano-Flores, 2010, Visit).

Country D

Throughout the translation review process, communication between Country D and the US organizing agency was timely. Early in the project, the US organizing agency emailed team members information about the activities involved in the review process along with deadlines for key steps (Kurpius & Shavelson, personal communication, January 19, 2010). In addition, at the initial meeting in New York City, in February 2010, Country D team members were able to gather clarifying information regarding the review process from the US agency's staff (CAE, GS.26, 2010). The US agency also communicated additional information about the process via email (Kurpius & Shavelson, personal communication, April 27, 2010; Kurpius & Shavelson, personal communication, May 11, 2010). The Country D team did not give any indication that communication about translation review had been untimely. The Country D team members were also timely in their communication (Solano-Flores, 2010, Comparison).

The Country D team completed a review of the translations of the two performance tasks, the scoring rubric, and the computer platform and interface language (CAE, 2010, GS.36). The instructions for the translation review process that the US organizing agency provided followed the TTTE (CAE, 2010, GS.9). The process required team members with expertise in different fields, including measurement (CAE, 2010, GS.36). The Country D team had a measurement expert who was an integral part of study. The measurement expert worked on the translation review process (Chia, 2011, Rubric).

The study's translation review process followed the TTTE's emphasis on a multidisciplinary approach when examining a translation with a critical eye (CAE, 2010, GS.9). The review was to include an experienced translator who was not involved in the initial translation process and who could serve as a translation adviser (CAE, 2010, GS.36). The Country D team did not follow the US organizing agency's guidelines for the translation review personnel. The person who took on the responsibilities of the third translator for the translation review did not have a background as a professional translator (Solano-Flores & Chia, 2010, Interview). In addition, although the person did not complete either of the two independent translations they were involved in the translation reconciliation process (Solano-Flores, 2010, Visit).

Documents from the study provided information on the amount of measurement and translation expertise included in Country D's translation review team (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010, Interview). However, there is limited information on Country D's implementation of the translation review process itself.

The US organizing agency provided the Country D team with detailed instructions for that guided them on the translation review process (CAE, 2010, GS.13; CAE, 2010, GS.36). They also provided Country D team members with the translation review coding form, which was to be used during the review (CAE, 2010, GS.36). Using these documents the Country D translation review team was able to review their translation review work.

The US organizing agency provided several translation review training opportunities. Country D participated in the translation review training that the US organizing agency conducted during the initial group New York City meeting (CAE, 2010, GS.26). The training included information on the translation review process and stressed the importance of multidisciplinary group discussion when applying the error dimensions provided (CAE, 2010, GS.26; CAE, 2010, GS.13). The training also reviewed the TTTE and stressed that reviewers should be actively seeking disconfirming evidence (CAE, 2010, GS.36). The US organizing agency made the training material and supplemental material available to the Country D team online, including the coding form the team would use to during the review process (CAE, 2010, GS.13; CAE, 2010, GS.4). In addition, during the site visit, Country D translation review team members were able to practice the process with the guidance from the US agency staff (Solano-Flores, 2010, Visit; Chia, 2011, Rubric).

The translation review process used for the AHELO feasibility study included constant opportunities to document progress. The US organizing agency provided the Country D review team with a list of ten error dimensions and their definitions as well as a coding form (CAE, 2010, GS.4). As part of the translation review process each member of Country D's review team used coding forms to take note of the identified errors and the dimensions each error impacted (CAE, 2010, GS.36). The completed coding forms served as a means to document the progress that the Country D team made during the translation review process.

At the beginning of the study, the US organizing agency provided Country D with a work plan with deadlines for important project milestones (CAE, 2010, GS. 1). The work plan did not address the translation review process (CAE, 2010, GS.1).

The US organizing agency provided Country D with material addressing the translation review of performance tasks, scoring rubric, and computer interface language. The Country D translation review team had access to electronic copies of the material used during the training, as well as supplemental documents, online (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.9; CAE, 2010, GS.36). The documents presented information in narrative, table, and bullet form and explained the qualifications for the translation review team, the translation error dimensions, and the translation review process (CAE, 2010, GS.36). The material also included a coding form that teams would use to monitor and document their progress. Despite difficulty with the unit of analysis used for reviewing the translated documents, the Country D national project manager expressed that material was easy to follow (Solano-Flores, 2010, Visit; Shavelson & Kurpius, 2010, Conference).

The Country D national team did not need to seek support for the translation review process from outside of the team (Solano-Flores & Chia, 2010, Interview).

The US organizing agency helped Country D with the translation review process. Country D team members attended a training session that the US organizing agency conducted during the initial meeting in New York City that addressed translation review (CAE, 2010, GS.26). The US organizing agency made training materials and other related documents available electronically (CAE, 2010, GS.13; CAE, 2010, GS.4). In addition, the US organizing agency representative who completed the site visit helped translators with the review process (Solano-Flores, 2010, Visit).

Country E

The Country E team did not indicate that communication with the US organizing agency was untimely. The US organizing agency sent information about the translation review process—and deadlines involved—at the beginning of the study (Kurpius & Shavelson, personal communication, January 19, 2010). In addition, Country E team members and the US agency's staff were able to discuss the translation process (CAE, GS.26, 2010). The US agency also communicated additional information about the process via email (Kurpius & Shavelson, personal communication, April 27, 2010; Kurpius & Shavelson, personal communication, May 11, 2010). The Country E team did not give any indication that communication about translation review had been untimely. The Country E team members were also timely in their communication (Solano-Flores, 2010, Comparison).

The translation review process for the AHELO study used the TTTE requiring a translation team that included expertise in different fields (CAE, 2010, GS.9). To review the translations for the two performance tasks, the scoring rubric, and the computer platform and interface language the translation review team needed to include measurement expertise (CAE, 2010, GS.36). The Country E translation team included translators who possessed extensive

translation experience (Solano-Flores & Chia, Interview, 2010). In addition, several team members who participated in the translation review process had measurement expertise (Solano-Flores & Chia, Interview, 2010).

The Country E national team was able to implement the translation review process as described by the TTTE and as required by the US organizing agency (CAE, 2010, GS.36). The Country E translation review team included translation expertise (Solano-Flores & Chia, 2010, Interview). All of the translators involved in the translation and translation review activities came from a professional translation company (Solano-Flores & Chia, 2010, Interview).

Although there is information about the assessment and translation expertise included in the translation review team (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010, Interview) there is limited information on Country E's translation review process.

Country E had access to important documents addressing the translation review process. The US organizing agency provided the Country E translation review team with detailed instructions for the translation review process (CAE, 2010, GS.13; CAE, 2010, GS.36). They also provided Country E team members with the translation review coding form, which was to be used during the review (CAE, 2010, GS.36). Using these documents the Country E translation review team was able to review their translation review work.

Country E participated in several training opportunities addressing translation review that the US organizing agency provided. Country E participated in the translation review training that the US organizing agency conducted during the initial group New York City meeting (CAE, 2010, GS.26). The training included information on the translation review process and stressed the importance of multidisciplinary group discussion when using the error dimensions provided for the process (CAE, 2010, GS.26; CAE, 2010, GS.13). The training also underscored that reviewers should be actively seeking disconfirming evidence (CAE, 2010, GS.36). The US organizing agency made the training material and supplemental material available to the Country E team online, including the coding form the team would use to during the review process (CAE, 2010, GS.13; CAE, 2010, GS.4). In addition, the Country E translation review team reviewed key components of the TTTE, and were able to practice the process during the site visit (Solano-Flores, 2010, Visit).

The AHELO feasibility study used a translation review process that included constant opportunities to document progress. As part of the translation review documents that the US organizing agency provided the Country E review team there was a list of ten error dimensions, their definitions, and a coding form (CAE, 2010, GS.4). As part of the translation review process each member of Country E's review team filled out a coding form noting each identified error and the dimension(s) impacted (CAE, 2010, GS.36). The completed coding forms served as a means to document the progress that the Country E team made during the translation review process.

The US organizing agency provided the Country E team with a work plan containing deadlines for important project milestones (CAE, 2010, GS.1). The work plan did not include a

deadline by when the Country E translation review team had to complete the translation review process.

The US organizing agency provided Country E with material addressing the translation review of performance tasks, scoring rubric, and computer interface language. The Country E translation review team had access to electronic copies of the material used during the training, as well as supplemental documents, online (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.9; CAE, 2010, GS.36). The documents presented information in narrative, table, and bullet form and explained the qualifications for the translation review team, the translation error dimensions, and the translation review process (CAE, 2010, GS.36). The material also included a coding form that teams would use to monitor and document their progress. Despite difficulty with the unit of analysis used for reviewing the translated documents, the Country E national project manager expressed that material was easy to follow (Solano-Flores, 2010, Visit; Shavelson & Kurpius, 2010, Conference).

The Country E national team was able to gather in-country support from outside of the team (Solano-Flores & Chia, 2010, Interview). All of the translators involved in the translation and translation review activities came from a professional translation company located in Country E (Solano-Flores & Chia, 2010, Interview).

The US organization supported Country E team members with translation review procedures. The team attended training that the US organizing agency conducted during the initial meeting in New York City (CAE, 2010, GS.26). In addition, the US organizing agency made material used during the training available to Country E team members online and via email (CAE, 2010, GS.13; CAE, 2010, GS.4).

Task 9: Translate ancillary materials as described by coordinating group.

Country A

The US organizing agency sent Country A team members material addressing dual translation with a list of ancillary materials and due dates (CAE, 2010, GS.31; CAE, personal communication, May 11, 2010). Since it would require less financial and human resources, the US organizing agency chose to implement the dual translation process with ancillary materials: the mini performance task, cognitive interview materials, scoring handbook charts, the administrator manual, and scorer training materials (CAE, 2010, GS.31). The agency also sent the Country A team an update on due dates (CAE, 2010, GS.36 and GS.37; CAE, personal communication, July 28, 2010). The Country A NPM also communicated progress on the dual translation process in a timely manner (Ursin, personal communication, August 10, 2010).

The translators who worked with the Country A team worked with a professional translation company and had impressive academic credentials (Solano-Flores & Chia, 2010, Interview). However, as per the guidelines provided by the US organizing agency, the Country A team did not include measurement expertise in the dual translation process.

The Country A national project manager was able to hire translators for the AHELO study through a professional agency (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010,

Interview). The translation team members responsible for the dual translation process had experience translating educational assessment documents and material from higher education (Chia, 2011, Rubric).

The team from Country A was able to complete the dual translation and reconciliation phase process. The national project manager provided the translators with all of the required documentation to complete the process (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010, Interview). The translations were done on schedule and there is no evidence that the translators were not given enough time to complete them (Ursin, personal communication, September 20, 2012). Although the NPM managed the logistics, one of the translators managed the reconciliation process (Solano-Flores & Chia, 2010, Interview).

After completing two independent translations of all material, Country A's translation team followed the steps for translation reconciliation (Solano-Flores, 2010, Visit). All team members contributed to open discussions about the translation process while applying the ten translation error dimensions from the TTTE (Solano-Flores, 2010, Visit). However, the team did not share information indicating that they reviewed the translations.

Country A participated in the training that the US organizing agency conducted during the New York City meeting and provided the training material to country teams online (CAE, 2010, GS.26; CAE, 2010, GS.4; CAE, 2010, GS, 13). The training stressed the importance of having two translators create two independent versions of each ancillary document (CAE, 2010, GS.26; CAE, 2010, GS.13). The training also included the reconciliation process (CAE, 2010, GS.26; CAE, 2010, GS.13). Country A provided their translators with the training and supplemental material. As a result, the translators were able to explain how they implemented each step of the translation process (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010, Interview, Chia, 2011, Rubric).

There were several opportunities to document Country A's progress during performance task translation and reconciliation. The two translators provided the Country A team the individual translations and the reconciled version (Chia, 2011, Rubric; Solano-Flores, 2010, Visit). In addition, the US organizing agency documented Country A's ancillary material translation progress in several reports turned in to the international organizing agency (CAE, 2010, Milestone; CAE, 2010, Module).

Country teams were to translate several ancillary materials for the AHELO study: a mini performance task, cognitive lab materials, scoring handbook charts, administrator manual, and scorer training materials. Country A was to complete mini PT and cognitive labs materials translation by August 15, 2010 (CAE, 2010, GS.1). The scoring handbook charts were to be translated by December 1, 2010 (CAE, 2010, GS.1). There was no specific due date provided for translation of the administrator manual—which was to be completed by an external translation company hired by the US organizing agency—and scorer training materials (CAE, 2010, GS.1). By August 10, 2010 Country A had completed the translation of cognitive lab materials and scoring handbook charts; the mini PT translation was in progress (Ursin, personal

communication, August 10, 2010). Due to a minor technical error in the original English mini PT Country A completed their translation slightly after the due date (Ursin, personal communication, September 7, 2010).

The US organizing agency provided the Country A team with several documents addressing the dual translation process that was applied to the ancillary materials (CAE, 2010, GS.1; CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). Throughout the documents, the agency presented the information in several ways: paragraphs, lists, and flowchart. The Country A translators did not indicate that they experienced challenges when working with the documents.

Members of the Country A team could not translate and reconcile all of the ancillary materials on their own. The Country A team contracted a translation company located within the country to complete the dual translation work (Solano-Flores & Chia, 2010, Interview). The Country A team did not require additional assistance from outside the team within the country.

The Country A team required assistance outside of the country to complete the dual translation and reconciliation process for the AHELO study's ancillary materials. During the New York City meeting Country A team members received training on the dual translation and reconciliation process from the US organizing agency (CAE, 2010, GS.26). The Country A team also received electronic copies of materials addressing translation and verification from the US organizing agency (CAE, 2010, GS.13; CAE, 2010, GS.4).

Country B

The US organizing agency created the dual translation process for ancillary materials: the mini performance task, cognitive interview materials, scoring handbook charts, the administrator manual, and scorer training materials (CAE, 2010, GS.31). The US organizing agency provided the Country B team with information about the dual translation process throughout the study. The US organizing agency sent Country B team members electronic versions of material addressing dual translation containing a list of ancillary materials and due dates (CAE, 2010, GS.31; CAE, personal communication, May 11, 2010). In addition, the agency sent the Country B team updated due dates (CAE, 2010, GS.36 and GS.37; CAE, personal communication, July 28, 2010). Initially, the Country B team communicated about ancillary document translation in a timely manner (Young, personal communication, June 4, 2010). However, during the actual dual translation process there was limited information on the Country B's team part.

The Country B team implemented the dual translation process with the study's ancillary materials. As stipulated by the translation guidelines created by the US organizing agency for the AHELO study, there is no evidence of measurement expertise present during the dual translation portion of the study for Country B.

The Country B team hired translators to complete the dual translation process for ancillary material (CAE, GS.11, 2010). However, although the translation team members were well-respected academics, there is no indication that they possessed the expertise in translation as

suggested in the US organizing agency's translation documentation (Solano-Flores & Chia, 2010, Interview).

The national project manager provided all translation team members with the documents for dual translation (Solano-Flores & Chia, 2010, Interview). The national project manager also participated in the reconciliation process (Solano-Flores & Chia, 2010, Interview). The translation team did not initially follow the translation guidelines and was late in turning in the translations.

As called for in the study's translation process, the Country B translators finished two independent translations of all ancillary materials needed for the study (CAE, 2010, GS.13; CAE, 2010, GS.4). The translation team also worked together to complete the translation reconciliation process and produce one translated version of documents (Solano-Flores, 2010, Visit). However, there is no indication that the team conducted a review during this process. There is evidence that the team was not able to speak freely about the translation errors evident from their translation process (Solano-Flores, 2010, Visit).

The US organizing team trained Country B team members on translating ancillary material during the initial meeting in New York City and made training material available online (CAE, 2010, GS.26; CAE, 2010, GS.13; CAE, 2010, GS.4). The training addressed independent translations and translation reconciliation (CAE, 2010, GS.26). Although Country B team members made the training and supplemental documents available to the translators, they relied on the translators' expertise during translation (Solano-Flores, 2010, Visit).

The two Country B translators created individual translations of ancillary material and provided them to the Country B team (Solano-Flores & Chia, 2010, Interview). Once the translators reconciled the translations and provided all material to the Country B team (Solano-Flores, 2010, Visit). The US organizing agency documented Country B's progress in translation in several reports submitted to the international organizing agency (CAE, 2010, Milestone; CAE, 2010, Module).

Country teams were to translate several ancillary materials for the AHELO study: a mini performance task, cognitive lab materials, scoring handbook charts, administrator manual, and scorer training materials. Country B was to complete mini PT and cognitive labs materials translation by August 15, 2010 (CAE, 2010, GS.1). The scoring handbook charts were to be translated by December 1, 2010 (CAE, 2010, GS.1). There was no specific due date provided for translation of the administrator manual—that was to be translated by an external translation company hired by the US organizing agency—and scorer training materials (CAE, 2010, GS.1). There was very limited information from Country B regarding the deadlines and progress regarding the dual translation of ancillary materials.

Country B team members had access to electronic versions of several documents addressing the dual translation and translation reconciliation process of ancillary materials online (CAE, 2010, GS.1; CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). Subsequently, team members were able to provide their translators

with the material (Solano-Flores, 2010, Visit). The documents presented the translation information in different formats: paragraph, list, and flowchart. Neither the translation team nor the Country B team shared any concerns or difficulties with the documents.

The Country B team hired academics from within the country to complete the dual translation work of ancillary materials (Solano-Flores & Chia, 2010, Interview). The Country B team did not require additional assistance from outside the team within the country.

The Country B team found support for the dual translation and reconciliation process outside of the country. Country B team members attended the training addressing dual translation and reconciliation that the US organizing agency offered during the initial New York City meeting (CAE, 2010, GS.26). The US organizing agency also made the materials including information about translation of ancillary materials available to the Country B team (CAE, 2010, GS.13; CAE, 2010, GS.4) who then made them available to translators.

Country C

Throughout the study, the US organizing agency provided the Country C team information about the dual translation process. The US organizing agency chose to implement the less resources demanding dual translation process for ancillary materials: the mini performance task, cognitive interview materials, scoring handbook charts, the administrator manual, and scorer training materials (CAE, 2010, GS.31). The US organizing agency sent the Country C team electronic versions of material addressing dual translation containing a list of ancillary materials and due dates and made them available online (CAE, 2010, GS.31; CAE, personal communication, May 11, 2010). In addition, the agency sent the Country C team updated due dates for the dual translation (CAE, 2010, GS.36 and GS.37; CAE, personal communication, July 28, 2010). Although the US organizing agency communicated in a timely manner, there is no evidence that the translation team from Country C was able to do the same.

The translators who worked with the Country C team completed the dual translation process with ancillary study materials. Following the dual translation guidelines the US organizing agency, the Country C team did not include measurement expertise during the process.

The Country C translation team hired professional translators to complete the full and dual translation process (Solano-Flores & Chia, 2010, Interview). The Country C team's translators had all served as translation professionals and two of the three were professors of language and translation at local universities (Solano-Flores & Chia, 2010, Interview).

The Country C team completed the dual translation ancillary material for AHELO. The national project manager provided the translation team members with all of the documents necessary to complete the dual translation process—and did so in a timely manner (Solano-Flores & Chia, 2010, Interview). Also, the national project manager was not able to manage difficulties in the completing the translations within the stipulated deadlines because of local religious holidays (Solano-Flores & Chia, 2010, Interview).

The translation process stipulated by the US organizing agency required that two translators independently translate all of the ancillary materials and then reconcile them into one version of each document (CAE, 2010, GS.4; CAE, 2010, GS.13). However, the translator in charge of reviewing reconciled performance tasks translated the ancillary materials (Solano-Flores, 2010, Visit). No evidence indicated that a review of the ancillary material translation occurred.

Two Country C team members participated in the translation and translation reconciliation training of ancillary material that the US organizing agency provided during the initial meeting in New York (CAE, 2010, GS.26). In addition, the team had access to the translation training materials and additional documents online (CAE, 2010, GS.4; CAE, 2010, GS.13; CAE, 2010, 26). After the Country C team trained the translators, the translation team was able to apply the process successfully during the translation and translation reconciliation of ancillary material (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010, Interview).

The country team and US organizing agency documented Country C's progress in performance task translation and reconciliation. One translator translated ancillary material and provided the work to the Country C team (Solano-Flores & Chia, 2010, Interview; Solano-Flores, 2010, Visit). The US organizing agency documented Country C's progress in translation in several reports that they provided to the international organizing agency (CAE, 2010, Milestone; CAE, 2010, Module).

Country teams were to translate several ancillary materials for the AHELO study: a mini performance task, cognitive lab materials, scoring handbook charts, administrator manual, and scorer training materials. Country C was to complete mini PT and cognitive labs materials translation by August 15, 2010 (CAE, 2010, GS.1). The Country C team submitted their translation of the mini performance task on January 1, 2011 (Al-Atiqi, personal communication, January 1, 2011). The scoring handbook charts were to be translated by December 1, 2010 (CAE, 2010, GS.1). However, there is a lack of information regarding the team's progress on this task. There was no specific due date provided for translation of the administrator manual—that was to be translated by an external translation company hired by the US organizing agency—and scorer training materials (CAE, 2010, GS.1).

At the beginning of the AHELO study, the US organizing agency provided the Country C team with documents addressing the translation process (CAE, 2010, GS.1; CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). The documents presented information in paragraph, list, and graphic forms. The Country C team provided these documents to the translation team and they were able to apply the information (Solano-Flores, 2010, Visit). The team did not indicate a challenge with the documents. However, the team did not implement the dual translation process as required. Only one translator on his own completed the dual translation of documents (Solano-Flores, 2010, Visit).

The Country C team was able to hire translation experts to complete the dual translation work of ancillary materials (Solano-Flores & Chia, 2010, Interview). During the translation

process the Country C team did not require additional assistance from outside the team within the country.

Country C acquired help for the dual translation process to be applied with ancillary materials translation from outside of the country. During the New York City meeting Country C team members participated in training addressing dual translation conducted by the US organizing agency (CAE, 2010, GS.26). The Country C team also had access to electronic copies of materials addressing dual translation, which were created and placed online by the US organizing agency (CAE, 2010, GS.13; CAE, 2010, GS.4).

Country D

Throughout the study, the Country D team received information about the dual translation process from the US organizing agency. The agency chose to have the countries use the less costly dual translation process for ancillary materials: the mini performance task, cognitive interview materials, scoring handbook charts, the administrator manual, and scorer training materials (CAE, 2010, GS.31). The US organizing agency sent the Country D team electronic versions of material addressing dual translation containing a list of ancillary materials and due dates and made them available online (CAE, 2010, GS.31; CAE, personal communication, May 11, 2010). In addition, the agency communicated with the Country D team about updated due dates for the dual translation (CAE, 2010, GS.36 and GS.37; CAE, personal communication, July 28, 2010). Just as with the US organizing agency's communication, the Country D team emailed about translation of the ancillary material in a timely manner (Urrea, personal communication, June 1, 2010).

The translators who worked with the Country D team completed the dual translation process on the study's ancillary material. The US organizing agency's guidelines for the dual translation did not require measurement expertise. Therefore, the team did not include a measurement expert during dual translation.

The Country D team hired independent translators to complete the full and dual translation procedures. However, although the two translators were bicultural and had experience within the education field there is no evidence that the translators were certified or worked with professional agencies. One translator mentioned performing translation for a local law school but did include details about the type of amount of translation completed (Solano-Flores & Chia, 2010, Interview; Chia, 2011, Rubric).

The team for Country D was able to manage the performance task translation and translation reconciliation process of the feasibility study. The Country D team provided the translators with the documents requiring translation well in advance of the due date (Solano-Flores & Chia, 2010, Interview). The team also provided the translators with assistance during the translation reconciliation process (Solano-Flores & Chia, 2010, Interview). Translations were turned in to the US organizing agency without any negative comment regarding the management of the process (Urrea, personal communication, September 22, 2010).

Following the process instituted by the US organizing agency the Country D team asked that each of the two translators create independent translations of all ancillary materials (Solano-Flores, 2010, Visit). The two translators then reconciled the translations of each document into one version (Chia, 2011, Rubric). It does not appear that a review of the work took place.

After attending training on translating ancillary material offered by the US organizing agency, the Country D team gave the translators all of the materials available online and provided guidance on the process. The translators were able to discuss and implement the translation process when translating ancillary material (Solano-Flores & Chia, 2010, Interview; Chia, 2011, Rubric).

The Country D team had two opportunities to document their progress of translating ancillary material. First, the two translators created individual translations and provided them to the Country D team—which the team subsequently submitted to the US organizing agency (Chia, 2011, Rubric). Second, for each ancillary document the translators created one reconciled version and submitted the work to the team (Solano-Flores, 2010, Visit). In addition, the US organizing agency documented Country D's translation progress in several reports submitted to the international organizing agency (CAE, 2010, Milestone; CAE, 2010, Module).

Country teams were to translate several ancillary materials for the AHELO study: a mini performance task, cognitive lab materials, scoring handbook charts, administrator manual, and scorer training materials. Country D was to complete mini PT and cognitive labs materials translation by August 15, 2010 (CAE, 2010, GS.1). In early September the Country D team was still working on the translation of the mini performance task (Urrea, personal communication, September 1, 2010). The team submitted the translated mini PT on September 22, 2010 (Urrea, personal communication, September 22, 2010). Another version of the translated mini PT was submitted December 15, 2010 and a final translation on January 21, 2011 (Urrea, personal communication, December 15, 2010; Urrea, personal communication, January 21, 2011). The team submitted cognitive labs materials on November 4, 2010 (Urrea, personal communication, November 4, 2010). The scoring handbook charts were to be translated by December 1, 2010 (CAE, 2010, GS.1); however, there is no information on Country D's progress with this task. Also, no information could be found regarding Country D's translation of scorer training materials. There was no deadline for the translation of the administrator manual, which was to be translated by an external translation company hired by the US organizing agency (CAE, 2010, GS.31).

The Country D team and translators had access to documents addressing the dual translation process of ancillary materials. The US organizing agency created these documents and made them available electronically online. The material presented the information in paragraph and list format as well as in a flowchart (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). The team stated that the documents were helpful (Chia, 2011, rubric; Solano-Flores & Chia, 2010, Interview) and did not indicate that they found any part of the material challenging.

The Country D team hired translators from Country D to complete the dual translation of ancillary materials (Solano-Flores & Chia, 2010, Interview). The Country D team did not require additional assistance from outside the team within the country during the translation and reconciliation process.

The US organizing agency helped Country D with the dual translation process to be applied with ancillary materials. Country D team members attended a training session that the US organizing agency conducted during the initial meeting in New York City that addressed dual translation and reconciliation (CAE, 2010, GS.26). The US organizing agency made training materials and other related documents available electronically (CAE, 2010, GS.13; CAE, 2010, GS.4).

Country E

Throughout the study, the US organizing agency provided Country E team members information about the dual translation process. Due to its less costly process, the US agency chose to have the countries use the dual translation process for ancillary materials: the mini performance task, cognitive interview materials, scoring handbook charts, the administrator manual, and scorer training materials (CAE, 2010, GS.31). The US organizing agency sent the Country E team material addressing dual translation, including a list of ancillary materials and due dates (CAE, 2010, GS.31; CAE, personal communication, May 11, 2010). In addition, the agency communicated with the Country E team about updated due dates for the dual translation (CAE, 2010, GS.36 and GS.37; CAE, personal communication, July 28, 2010). Just as with the US organizing agency's communication, the Country E team emailed about translation of the ancillary material in a timely manner (Opheim, personal communication, June 14, 2010; Turmo, personal communication, June 3, 2010).

Several of the Country E team members who participated in the translation process had measurement expertise (Solano-Flores, Interview, 2010). Therefore, although the US organization did not require it, Country E included measurement experts in their dual translation process.

The team from Country E hired very experienced translators. The translation team included extensive experience in translation of international assessments, which they incorporated during the adaptation process (CAE, 2010, GS.11). Also, one team member was a linguist who focused on reading education (Solano-Flores & Chia, 2010, Interview).

The Country E team completed the translation and translation reconciliation process for the feasibility study. The national project manager provided all members of the translation team with the documents requiring translation (Solano-Flores & Chia, 2010, Interview). Several members of the Country A team participated in the translation reconciliation process (Solano-Flores & Chia, 2010, Interview). The team submitted translations and did not indicate any challenge with the management of the process (Opheim, personal communication, August 26, 2010).

The Country E team conducted a translation and translation reconciliation process of ancillary materials as required by the US organizing agency (CAE, 2010, GS.13, CAE, 2010, GS.4). In addition, the three translators worked together after the initial independent translations to improve the translations as much as possible (Solano-Flores & Chia, 2010, Interview). As a result, Country E included a review process during this stage of the study without it being part of the requirements listed by the US organizing agency.

The US organizing agency provided training and guidance for the translation of ancillary material. The Country E team attended training that the US organizing agency conducted during the initial meeting in New York City (CAE, 2010, GS.26). In addition, the US organizing agency made material used during the training and supplemental documents available to Country E team members online (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.26). The Country E team provided translators with all documents addressing translation and reconciliation and guided them throughout the process (Solano-Flores, 2010, Visit).

Country E's progress in translation and reconciliation of ancillary material was documented by the team and the US organizing agency. The translators submitted one reconciled version of all ancillary material to the country team (Solano-Flores, 2010, Visit). In addition, the US organizing agency documented Country E's progress in translation in several reports to the international organizing agency (CAE, 2010, Milestone; CAE, 2010, Module).

Country teams were to translate several ancillary materials for the AHELO study: a mini performance task, cognitive lab materials, scoring handbook charts, administrator manual, and scorer training materials. Country E was to complete mini PT and cognitive labs materials translation by August 15, 2010 (CAE, 2010, GS.1). The Country E team submitted final translation of the mini PT on January 21, 2011 (Opheim, personal communication, January 21, 2010). Country E submitted the initial translation of cognitive labs materials on August 26, 2010 and the final translation on December 28, 2010 (Opheim, personal communication, August 26, 2010; Opheim, personal communication, December 28, 2010). The scoring handbook charts were to be translated by December 1, 2010 (CAE, 2010, GS.1). There was no specific due date provided for translation of the administrator manual,—that was to be translated by an external translation company hired by the US organizing agency—scorer training materials , or scoring charts (CAE, 2010, GS.1).

The US organizing agency created several documents addressing the dual translation process of ancillary materials, which included reconciliation. The Country E team and translators had access to electronic versions of the document online. The material presented the information in paragraph and list format as well as in a flowchart (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). The team explained that they were able to work with the documents during translation (Solano-Flores & Chia, 2010, Interview) and did not indicate that they found any part of the material addressing translation challenging.

The Country E team was able to work with Country E translators to complete the dual translation work required for the ancillary materials (Solano-Flores & Chia, 2010, Interview).

The translators worked with members of the country team on this process (Solano-Flores & Chia, 2010, Interview). The Country E team did not require additional assistance from outside the team within the country.

The US organization supported Country E team members with the dual translation procedures that they were to follow when working with ancillary materials. The team attended training that the US organizing agency conducted during the initial meeting in New York City (CAE, 2010, GS.26). In addition, the US organizing agency made material used during the training available to Country E team members online and via email (CAE, 2010, GS.13; CAE, 2010, GS.4).

Task 10: Review translation of material for assessment implementation.

Country A

The US organizing agency hired an external translation company to translate material associated with assessment implementation: the administrator manual, proctor interface, student interface, and scorer interface into each of the five languages (Keeley, personal communication, June 8, 2011; CAE, 2010, GS.31). The US organizing agency sent Country A team members material addressing external translation procedures (CAE, 2010, GS.44). The US organizing agency communicated information about company and country team responsibilities and due dates and responsibilities to the Country A national project manager (Keeley, personal communication, March 4, 2011; CAE, 2010, GS.44). There is no evidence that the communication between Country A and the US organizing agency regarding external translation was untimely.

To alleviate some of the translation work demanded in the AHELO study Country A agreed that an external agency should help with translation of material for assessment implementation (Solano-Flores & Chia, 2010, Interview). The US organizing agency hired translators working for a translation company located in the United States to complete the work. However, there is no information as to the translators' qualifications and experience in the field. Furthermore, the guidelines provided by the US organizing agency did not demand that a measurement expert be included in the dual translation process (CAE, 2010, GS.13; CAE, 2010, GS.4). It was important that the Country A team review the external translators' work (Keeley, personal communication, April 26, 2011). The national project manager for Country A, who had measurement expertise, participated in the review of externally translated materials for assessment implementation (Ursin, personal communication, June 7, 2011).

The dual translation process used with material for assessment implementation required two independent translations and translation reconciliation (CAE, 2010, GS.13; CAE, 2010, GS.43). Therefore, it was important that the Country A team include translation expertise when reviewing the translations completed by the external company. There is no evidence that Country A included translation experts during the review.

Country A agreed that external translators located in the US could work on material for assessment implementation. The team also committed to reviewing the translations completed by

the external translators. Reviews were completed and communicated to the US organizing agency (Ursin, personal communication, June 7, 2011). However, there were no guidelines provided for this review process. In addition, there is little information about the process that the Country A team implemented to review the external translations.

Country A agreed to review the translation of material used for assessment implementation completed by an external translation team located in the United States (CAE, 2010, GS.31). The US organizing agency provided Country A with the translation review coding form, which was to be used during the review (CAE, 2010, GS.36). Using the coding forms, the Country A translation review team was able to review the external translation review work.

Country A participated in the training that the US organizing agency conducted during the New York City meeting (CAE, 2010, GS.26). The training included information on translation review and translation reconciliation (CAE, 2010, GS.26; CAE, 2010, GS.13). The US organizing agency made the training material and supplemental material available to the Country A team online (CAE, 2010, GS.13; CAE, 2010, GS.4). The Country A team was responsible for reviewing and reconciling the translations of material for assessment implementation completed by external translators hired by the US organizing agency (CAE, 2010, GS.43).

The translation review of material for assessment implementation included constant opportunities to document progress. The Country A team was to use a list of ten error dimensions and a coding form as they reviewed the external translations (CAE, 2010, GS.4; CAE, 2010, GS.43). Each member of Country A's review team filled out a coding form noting the identified error and the dimension(s) impacted (CAE, 2010, GS.36). The completed coding forms served as a means to document the team's progress while reviewing the translations of materials for assessment implementation, which were completed by a US translation agency hired by the US organizing agency.

The US organizing agency did not provide Country A with a clear deadline to complete the review of external translation of material for assessment implementation (CAE, 2010, GS.1). However, the Country A team needed to review the external translations in time to upload the material onto the internet platform (Keeley, personal communication, April 26, 2011). Country A team members did not indicate the deadlines were challenging.

The US organizing agency provided Country A team members with documents that they could use when reviewing the translations of material for assessment implementation (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). The Country A translation review team was to use translation error dimensions, step-by-step instructions, and coding forms throughout the process (CAE, 2010, GS.4; CAE, 2010, GS.13; CAE, 2010, GS.36; CAE, 2010, GS.9). Throughout the documents, the agency presented the information in several ways: paragraphs, lists, and flowchart. The Country A translators did not indicate that they experienced challenges when working with the documents.

During the review of their translation work, the Country A national team was able to acquire in-country support that was external to the team when implementing the translation review process as described by the TTTE (CAE, 2010, GS.36; CAE, 2010, GS.43). However, there is no information about the staff used to review the external translation of material for assessment implementation.

The Country A team required assistance outside of the country to complete the translation review process for the AHELO feasibility study. The Country A team was able to use the training and documents that the US organizing agency provided for full translation review when reviewing external translation (CAE, 2010, GS.43; CAE, 2010, GS.13).

Country B

The US organizing agency contracted a US translation company to translate the administrator manual, proctor interface, student interface, and scorer interface into each of the five languages (Keeley, personal communication, June 8, 2011). The US organizing agency created and shared material that addressed external translation procedures with the Country B team (CAE, 2010, GS.44). The US organizing agency communicated information about company and country team responsibilities and due dates and responsibilities to the Country B team (Keeley, personal communication, March 4, 2011; CAE, 2010, GS.44). There is no evidence that the communication between Country B and the US organizing agency regarding external translation was untimely.

Due to the amount and length of documents that were a part of the AHELO feasibility study that needed to be translated, the translation work for AHELO was extensive. To help with the translation work demands Country B agreed that an external agency should help with translation of material for assessment implementation. The US organizing agency hired translators through a translation company located in the United States to complete the translation of material for assessment implementation (CAE, 2010, GS.44). However, there is no information as to the translators' qualifications and experience in the field. Furthermore, the guidelines provided by the US organizing agency did not demand that a measurement expert be involved in the dual translation process (CAE, 2010, GS.13; CAE, 2010, GS.4). It was important that the Country B team review the external translators' work (Keeley, personal communication, April 26, 2011). However, there is no information about the Country B team's review of the external translation.

The US organizing agency hired translators to complete the dual translation process for material used in assessment implementation for Country B (CAE, GS.44, 2010; CAE, 2010, GS.43). It was important that the Country B team include translation experts when reviewing the work completed by external translators. There is no information about the Country B team's review of the external translation of material for assessment implementation.

Country B team members understood that hiring external translators to translate material for assessment implementation would be helpful. The Country B team also committed to reviewing the translations completed by the external translators located in the United States. The

Country B team completed the reviews (Keeley, personal communication, April 26, 2011). However, there were no guidelines provided for this review process. In addition, there is little information about the process that the Country B team implemented to review the external translations.

Country B agreed to review the translation of material used for assessment implementation completed by an external translation team located in the United States (CAE, 2010, GS.31, CAE, 2010, GS.43). The US organizing agency provided the Country B team with material that would allow the translation team to review their work as they reviewed the translations (CAE, 2010, GS.43). The agency also provided Country B with the coding form, which each reviewer used during the review process (CAE, 2010, GS.36). Using these documents the Country B translation review team was able to review the external translation review work.

The US organizing team provided training addressing translation review and translation reconciliation during the initial meeting in New York City (CAE, 2010, GS.26). Country B team members participated in the training (CAE, 2010, GS.26). In addition, the US organizing agency placed the training material and supplemental documents available to the Country B team online (CAE, 2010, GS.13; CAE, 2010, GS.4). Country B team members were to apply information from the training when reviewing the translation of material for assessment implementation completed by translators located in the United States that the US organizing agency contracted (CAE, 2010, GS.43).

There were constant opportunities to document progress during the review of external translations completed with material for assessment implementation. As part of the material the US organizing agency provided for the review, Country B team members received a list of ten error dimensions with their definitions and a coding form (CAE, 2010, GS.4). As part of the review process each member of Country B's review team filled out a coding form noting the identified error and the dimension(s) that each error impacted (CAE, 2010, GS.36). The completed coding forms served as a means to document Country B's progress while reviewing the external translations of material for assessment implementation.

The AHELO feasibility study did not include a specific deadline for Country B team members with regard to reviewing external translations (CAE, GS.1, 2010). However, the Country B team was to complete the review of translations of material for assessment implementation in time to upload the documents onto the internet platform (Keeley, personal communication, April 26, 2011). There is no record of Country B team members indicating that the amount of time for reviewing the external work was challenging.

The Country B team used documents provided by the US organizing agency that they could use when reviewing the translations of material for assessment implementation completed by an external review company (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). The Country B translation review team used step-by-step instructions, translation error dimensions, and coding forms throughout the process (CAE,

2010, GS.4; CAE, 2010, GS.13; CAE, 2010, GS.36; CAE, 2010, GS.9). Throughout the documents, the US organizing agency presented the information in several ways: paragraphs, lists, and flowchart. Country B team members did not indicate that they experienced challenges when working with the documents.

The Country B national team was able to acquire in-country support that was external to the team when implementing the translation review process with their own translation work (CAE, 2010, GS.36; Solano-Flores & Chia, 2010, Interview). However, there is no information about the people that the Country B team hired to review the external translation of material for assessment implementation.

The Country B team required assistance outside of the country to complete the translation review process with materials for assessment implementation. The Country B team was able to use the training and documents that the US organizing agency provided for full translation review when reviewing external translation work (CAE, 2010, GS.43; CAE, 2010, GS.13).

Country C

The US organizing agency contracted a US translation company to translate assessment implementation material. The US company completed translation for the administrator manual, proctor interface, student interface, and scorer interface into each of the five languages (Keeley, personal communication, June 8, 2011). The US organizing agency created and shared material that addressed external translation procedures with the Country C team (CAE, 2010, GS.44). The US organizing agency communicated information about company and country team responsibilities and due dates and responsibilities to the Country C team (Keeley, personal communication, March 4, 2011; CAE, 2010, GS.44). There is no evidence that the communication between Country C and the US organizing agency regarding external translation was untimely.

The translation process for the AHELO study was demanding because of the number of documents associated with each performance task and the text contained within each. All countries and the US organizing agency agreed that the translation of material for assessment implementation required only the dual translation process. Furthermore, to alleviate some of the translation work Country C agreed that an external agency should help with translation of material for assessment implementation (CAE, 2010, GS.44). The US organizing agency hired translators working for a translation company located in the United States to complete the work. However, there is no information as to the translators' qualifications and experience in the field. Furthermore, the guidelines provided by the US organizing agency did not demand that a measurement expert be involved in the dual translation process (CAE, 2010, GS.13; CAE, 2010, GS.4). However, there is no information about the Country B team's review of the external translation.

The US organizing agency hired translators to complete the dual translation process for material used in assessment implementation for Country C (CAE, GS.44, 2010; CAE, 2010, GS.43). It was important that the Country B team include translation experts when reviewing the

work completed by external translators. There is no information about the Country B team's review of the external translation of material for assessment implementation.

The US organizing agency and Country C team members agreed that hiring external translators to translate material for assessment implementation would be helpful. The Country C team committed to reviewing the translations completed by the translators located in the United States that the US organizing agency contracted. The Country C team completed the reviews (Keeley, personal communication, April 26, 2011). However, the US organizing agency did not provide guidelines specifically for this review process. In addition, there is little information about the process that the Country C team implemented to review the external translations.

Country C agreed to review the translation of material used for assessment implementation completed by an external translation team located in the United States (CAE, 2010, GS.31). The US organizing agency provided the Country B translation review team with the translation review coding form, which was to be used during the review (CAE, 2010, GS.36). Using these documents the Country C translation review team was able to review the external translation review work.

Two Country C team members participated in the translation review and translation reconciliation training that the US organizing agency provided during the initial meeting in New York (CAE, 2010, GS.26). In addition, the team had access to the training materials and additional documents addressing the translation process online (CAE, 2010, GS.4; CAE, 2010, GS.13; CAE, 2010, 26). Country C team members were to apply information from the training when reviewing the translation of material for assessment implementation completed by translators located in the United States that the US organizing agency contracted (CAE, 2010, GS.43).

The translation review of external translations incorporated constant opportunities for Country C to document progress. As part of the translation review documents the US organizing agency sent Country C a list of ten error dimensions with their definitions and a coding form (CAE, 2010, GS.4). While reviewing the translation of material for assessment implementation Country C's team used coding forms to note the errors each member identified and the dimension(s) that each error impacted (CAE, 2010, GS.36). The completed coding forms served as a means to document Country C's translation review progress.

The US organizing agency did not provide the Country C translation review team with specific deadlines for the translation review of materials for assessment implementation (CAE, 2010, GS.1). However, Country C team members understood that they were to complete the review of external translations in time to upload the documents onto the internet platform (Keeley, personal communication, April 26, 2011). The team did not indicate having difficulty completing the review in the amount of time provided.

The US organizing agency provided Country C team members documents that they were to use while reviewing the translations of material for assessment implementation completed by an external review company (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE,

2010, GS.31; CAE, 2010, GS.36). The Country C team had access to step-by-step instructions, translation error dimensions, and coding forms while reviewing the external translations (CAE, 2010, GS.4; CAE, 2010, GS.13; CAE, 2010, GS.36; CAE, 2010, GS.9). Throughout the documents, the US organizing agency presented the information in several ways: paragraphs, lists, and flowchart. Country C team members did not indicate that they experienced challenges when working with the documents.

While reviewing their own translation work for the AHELO study the Country C team hired a third translator—a translator who had not been part of the initial translation process (CAE, 2010, GS.36). The team followed the TTTE. However, there was limited information regarding the Country C team's review process for the external translation of materials for assessment implementation.

Country C team members required assistance outside of the country to complete the translation review process with materials for assessment implementation. The Country C team was able to use the training and documents that the US organizing agency provided for full translation review when reviewing external translation work (CAE, 2010, GS.43; CAE, 2010, GS.13).

Country D

The US organizing agency contracted a translation company located in the United States to translate the administrator manual, proctor interface, student interface, and scorer interface into each of the five languages (Keeley, personal communication, June 8, 2011). The US organizing agency created and shared material that addressed external translation procedures with the Country D team (CAE, 2010, GS.44). The US organizing agency communicated information about company and country team responsibilities and due dates and responsibilities to the Country D team (Keeley, personal communication, March 4, 2011; CAE, 2010, GS.44). There is no evidence that the communication between Country D and the US organizing agency regarding external translation was untimely.

The translation process for the AHELO study was demanding because of the amount of text associated with each performance task. The US organizing agency's guidelines stated that the translation of material for assessment implementation required only the dual translation process. Furthermore, Country D agreed that an external agency should help with translation of material for assessment implementation (CAE, 2010, GS.44). The US organizing agency hired translators working for a translation company located in the United States to complete the work. However, there is no information as to the translators' qualifications and experience in the field. Furthermore, the guidelines provided by the US organizing agency did not demand that a measurement expert be involved in the dual translation process (CAE, 2010, GS.13; CAE, 2010, GS.4). It was important that the Country D team review the external translators' work (Keeley, personal communication, April 26, 2011). The assessment expert for the Country D team participated in the review of externally translated materials for assessment implementation (Rosas, personal communication, June 6, 2011).

The US organizing agency hired translators to complete the dual translation process for material used in assessment implementation for Country D (CAE, GS.44, 2010; CAE, 2010, GS.43). It was important that the Country D team include translation experts when reviewing the work completed by external translators. However, there is no information about the Country D team's review of the external translation of material for assessment implementation.

The Country D team agreed that the US organizing agency could hire external translators located in the United States to translate material for assessment implementation. Also, the Country D team committed to reviewing the translations of the material completed by the external translators. The Country D team completed the reviews (Keeley, personal communication, April 26, 2011; Rosas, personal communication, June 6, 2011). However, there were no guidelines provided for this review process. In addition, there is little information about the process that the Country D team implemented to review the external translations.

Country D agreed to review the translation of material used for assessment implementation completed by an external translation team located in the United States (CAE, 2010, GS.31). The S organizing agency provided Country D team members with the translation review coding form, which was to be used during the review of materials for assessment implementation (CAE, 2010, GS.36; CAE, 2010, GS.43). Using these documents the Country D translation review team was able to review their translation review work.

The US organizing agency provided the Country D team with training opportunities addressing the translation review and reconciliation procedures. The Country D team attended the training that the US organizing agency conducted during the initial meeting in New York City (CAE, 2010, GS.26). The team also accessed the training materials used during the training—and supplemental documents—online (CAE, 2010, GS.13; CAE, 2010, GS.4). Country D team members were to apply information from the training when reviewing the translation of material for assessment implementation completed by translators located in the United States that the US organizing agency contracted (CAE, 2010, GS.43).

The translation review process used to review external translations included constant opportunities to document progress. The US organizing agency provided the Country D review team with a list of ten error dimensions and their definitions as well as a coding form (CAE, 2010, GS.4). As part of reviewing the translation of material for assessment implementation each member of Country D's review team used coding forms to take note of the identified errors and the dimensions each error impacted (CAE, 2010, GS.36; CAE, 2010, GS.43). The completed coding forms served as a means to document the progress that the Country D team made during the translation review process.

At the beginning of the study, the US organizing agency provided Country D with a work plan with deadlines for important project milestones (CAE, 2010, GS. 1). The work plan did not address the translation review process (CAE, 2010, GS.1). However, the US organizing agency made clear that the Country D team was to complete the review of external translation work on material for assessment implementation in time to upload the documents onto the internet

platform (Keeley, personal communication, April 26, 2011). Country D team members did not express that the amount of time given to complete the task as difficult.

While reviewing translations of material for assessment implementation completed by external translators, Country D team members used documents provided by the US organizing agency (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). The Country D translation review team used step-by-step instructions, translation error dimensions, and coding forms throughout the process (CAE, 2010, GS.4; CAE, 2010, GS.13; CAE, 2010, GS.36; CAE, 2010, GS.9). The US organizing agency presented the information in several ways: paragraphs, lists, and flowchart. Country D team members did not indicate that they experienced challenges when working with the documents.

The Country D national team did not need to seek support for the translation review process of their work from outside of the team (Solano-Flores & Chia, 2010, Interview). However, it is not clear how the Country D team completed the review of external translation of materials for assessment implementation.

The Country D team required assistance outside of the country to complete the translation review process with materials for assessment implementation. The Country D team was able to use the training and documents that the US organizing agency provided for full translation review when reviewing external translation work (CAE, 2010, GS.43; CAE, 2010, GS.13).

Country E

The US organizing agency contracted an external translation company to translate the administrator manual, proctor interface, student interface, and scorer interface into each of the five languages (Keeley, personal communication, June 8, 2011). The US organizing agency created and shared material that addressed external translation procedures with the Country E team (CAE, 2010, GS.44). The US organizing agency communicated information about the US translation company and country team responsibilities and due dates and responsibilities to the Country E team (Keeley, personal communication, March 4, 2011; CAE, 2010, GS.44). There is no evidence that the communication between Country E and the US organizing agency regarding external translation was untimely.

Due to the amount of text associated with each performance task the translation process for the study was resource intensive. The US organizing agency required that material for assessment implementation only go through the dual translation process. In addition, to alleviate some of the translation work Country E agreed that an external agency should help with translation of that material (CAE, 2010, GS.44). The US organizing agency hired translators working for a translation company located in the United States to complete the work. The guidelines provided by the US organizing agency did not demand that a measurement expert be involved in the dual translation process (CAE, 2010, GS.13; CAE, 2010, GS.4). However, the Country E team's measurement experts helped review the external translation of material for assessment implementation (Opheim, personal communication, March 4, 2011).

The US organizing agency hired translators to complete the dual translation process for material used in assessment implementation for Country B (CAE, GS.44, 2010; CAE, 2010, GS.43). It was important that the Country B team include translation experts when reviewing the work completed by external translators. The Country E team, which included members with expertise in translation, helped review the external translation of material for assessment implementation (Opheim, personal communication, March 4, 2011).

Country E team members agreed with the US organizing agency that hiring external translators to translate material for assessment implementation would be helpful. The Country E team also committed to reviewing the translations completed by the external translators who were located in the United States. The Country E team completed the reviews (Keeley, personal communication, April 26, 2011). However, there were no guidelines provided for this review process. In addition, there is little information about the process that the Country E team implemented to review the external translations.

Country E agreed to review the translation of material used for assessment implementation completed by an external translation team located in the United States (CAE, 2010, GS.31). The US organizing agency provided the Country E translation review team with detailed instructions for the translation review process (CAE, 2010, GS.13; CAE, 2010, GS.36). They also provided Country E team members with the translation review coding form, which was to be used during the review (CAE, 2010, GS.36). Using these documents the Country E translation review team was able to review their translation review work.

The US organizing agency provided training and guidance for the translation review and translation reconciliation process. The Country E team attended training that the US organizing agency conducted during the initial meeting in New York City (CAE, 2010, GS.26). In addition, the US organizing agency made material used during the training and supplemental documents available to Country E team members online (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.26). Country E team members were to apply information from the training when reviewing the translation of material for assessment implementation completed by translators located in the United States that the US organizing agency contracted (CAE, 2010, GS.43).

The review process in place for external translation included constant opportunities to document progress. As part of the translation review documents that the US organizing agency provided the Country E review team there was a list of ten error dimensions, their definitions, and a coding form (CAE, 2010, GS.4; CAE, 2010, GS.43). Each member of Country E's review team filled out a coding form noting each identified error and the dimension(s) impacted (CAE, 2010, GS.36). The completed coding forms served as a means to document Country E's progress when reviewing the external translation of material for assessment implementation.

The US organizing agency provided the Country E team with a work plan containing deadlines for important project milestones (CAE, 2010, GS.1). The work plan did not include a deadline by when the Country E translation review team had to complete the review of external translation work. The team had to complete the review of materials for assessment

implementation in time to upload the documents onto the internet platform (Keeley, personal communication, April 26, 2011). The team did not communicate difficulty with the timeframe given for the activity.

The Country E team used documents provided by the US organizing agency while reviewing the external translations of material for assessment implementation completed by an external review company (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). Country E team members used step-by-step instructions, translation error dimensions, and coding forms throughout the review process (CAE, 2010, GS.4; CAE, 2010, GS.13; CAE, 2010, GS.36; CAE, 2010, GS.9). The US organizing agency presented the information included in the documents in several ways: paragraphs, lists, and flowchart. Country E team members did not indicate that they experienced challenges when working with the documents.

During the review of their work, the Country E national team was able to gather in-country support from outside of the team (Solano-Flores & Chia, 2010, Interview). However, there was no information regarding Country E's review process of the external translation of materials for assessment implementation.

The US organization supported Country E team members with translation review procedures. The Country E team required assistance outside of the country to complete the translation review process with materials for assessment implementation. The Country E team was able to use the training and documents that the US organizing agency provided for full translation review when reviewing external translation work (CAE, 2010, GS.43; CAE, 2010, GS.13).

Task 11: Implement changes based on verification procedures prescribed by coordinating group.

Country A

The US organizing agency communicated results from the external translation verification process with the Country A national project manager (Keeley, personal communication, January 20, 2012). The US organizing agency requested Country A's feedback on the verification notes and that the team implement changes with which they agreed (Keeley, personal communication, March 29, 2011). Country A team members were to complete this activity during their user acceptance testing of the internet platform (CAE, 2010, GS.45; Keeley, personal communication, March 29, 2011). Neither the Country A team nor the US organizing agency indicated experiencing challenges with communication timeliness.

The US organizing agency communicated results from the external translation verification process with the Country A national project manager (Keeley, personal communication, January 20, 2012). Country A's national project manager, who had expertise in educational assessment, helped examine the translation verification documents (Ursin, personal communication, December 10, 2011).

Communication from Country A's national project manager indicates that he helped examine the translation verification documents (Ursin, personal communication, December 10, 2011). However, there is no information about other members of the Country A team who may have assisted with this process.

The Country A team's national project manager helped the team complete the review of the translation verification process (Ursin, personal communication, December 10, 2011). The team provided feedback on the verification notes and made any changes necessary indicating good project management.

The Country A team received limited information about the process that they were to implement when reviewing translation verification results (CAE, 2011, Validation). There was no specific mention of reviewing their work as country teams examined results from the translation verification process.

The Country A team received limited information about the process that they were to implement when reviewing translation verification results (CAE, 2011, Validation). There was no training available for this activity in the study.

Country A had two opportunities to document their progress when implementing changes to their translations based on feedback from the translation verification results (CAE, 2011, Validation). Although Country A completed this activity, the US organizing agency did not keep documentation of the progress.

The US organizing agency did not provide Country A with a clear deadline for implementing changes based on the translation verification procedures (CAE, 2010, GS.1). This activity was not included in the work plan's timeline that the US organizing agency provided Country A (CAE, 2010, GS.1). However, Country A was able to complete the activity within the amount of time provided.

The Country A team did not have an opportunity to participate in official training addressing the implementation of changes resulting from translation verification (CAE, 2011, Validation). The US organizing agency provided the Country A team with a spreadsheet that included the suggested changes from the verification translators, electronic copies of the Country A's translations, and an email explaining what to do with the material (Geisinger, personal communication, January 21, 2012). There were very minor and few changes suggested by the translator who completed the verification process.

The US organizing agency provided Country A with limited information about the steps that country teams were to implement when reviewing translation verification results (CAE, 2011, Validation). In addition, the US organizing agency did not specify who was to complete the work or the qualifications necessary. The Country A national project manager helped review information from translation verification. However, it is not clear who else—if anyone—assisted in the process.

The US organizing agency provided Country A team members with assistance while implementing changes from translation verification process. The agency provided information from external translators, documents to complete changes, and helped manage the process.

Country B

The US organizing agency contracted independent translators to verify Country B's translations and communicated results from the verification process with the Country B national project manager (Keeley, personal communication, September 28, 2012). The US organizing agency requested Country B's feedback on the verification notes and that the team implement changes with which they agreed (Keeley, personal communication, March 29, 2011). Country B team members were to complete this activity during their user acceptance testing of the internet platform (CAE, 2010, GS.45; Keeley, personal communication, March 29, 2011). Although the US organizing agency sent timely emails, the Country B team members were not able to respond in turn (Keeley, personal communication, January 20, 2012).

The US organizing agency emailed notes from the external translation verification process to the Country B national project manager (Keeley, personal communication, January 20, 2012). One of the Country B team members was a measurement expert and helped examine the translation verification documents (Young, personal communication, January 22, 2012).

The US organizing agency emailed results from an external translation verification process to the Country B national project manager (Keeley, personal communication, January 20, 2012). The Country B team was able to review the work. (Young, personal communication, January 22, 2012). However, it is not clear who completed the review for the team. Therefore, it is not certain that those who completed the review possessed translation expertise.

There was some difficulty with Country B's review of the translation verification work (Geisinger, personal communication, January 20, 2012; CAE & Burros, 2012, Report). There was a lack of successful project management demonstrated during this activity.

The Country B team received limited information about the process that they were to implement when reviewing translation verification results (CAE, 2011, Validation). There was no specific mention of reviewing their work as country teams examined results from the translation verification process.

The Country B team received limited information about the process that they were to implement when reviewing translation verification results (CAE, 2011, Validation). There was no training available for this activity in the study.

Country B had two opportunities to document their progress when implementing changes to their translations based on feedback from the translation verification results (CAE, 2011, Validation). Country B completed this process and the progress was documented (Keeley, personal communication, September 28, 2011).

The US organizing agency did not include a deadline for implementing changes resulting from the translation verification procedures in the project timeline that they provided the Country B team (CAE, 2010, GS.1).

The US organizing agency did not provide Country B team members training addressing how to implement changes resulting from translation verification (CAE, 2011, Validation). Instead, the US organizing agency provided material to help complete the work. The agency provided the Country B team with a spreadsheet that included the suggested changes from the verification translators, electronic copies of the Country B's translations, and an email explaining what to do with the material (Geisinger, personal communication, January 21, 2012). However, the Country B team had difficulty working with the materials provided and implementing the changes suggested from translation verification.

Material that the US organizing agency gave the Country B team did not include specifics on how to implement changes resulting from translation verification (CAE, 2011, Validation). In addition, the US organizing agency did not specify who was to complete the work or the qualifications necessary. It is unclear who—if anyone—in addition to the Country B team members worked on this process.

The US organizing agency provided the Country B team with assistance while implementing changes from translation verification process. The agency provided information from external translators, documents to complete changes, and helped manage the process.

Country C

To verify Country C's translations, the US organizing agency contracted independent translators to examine the final documents and communicated results from the verification process with the Country C's national project manager (Keeley, personal communication, September 28, 2012). The US organizing agency requested Country C's feedback on the verification notes and that the team implement changes with which they agreed (Keeley, personal communication, March 29, 2011). Country C team members were to complete this activity during their user acceptance testing of the internet platform (CAE, 2010, GS.45; Keeley, personal communication, March 29, 2011). Although the US organizing agency sent timely emails, there is very limited information about Country C's response.

To verify Country C's translations, the US organizing agency contracted independent translators to examine the final documents and communicated results from the verification process with the Country C's national project manager (Keeley, personal communication, September 28, 2012). There is limited information available regarding Country C's work on translation verification.

The US organizing agency contracted independent translators to examine the final documents and communicated results from the verification process with the Country C's national project manager (Keeley, personal communication, September 28, 2012). However, there is limited information available regarding Country C's work on translation verification. As a result, it is unclear if the reviewers had translation expertise.

The Country C team's national project manager helped the team complete the review of the translation verification process (Ursin, personal communication, December 10, 2011).

However, there is very limited information addressing Country C's review of the translation verification suggestions.

The Country C team received limited information about the process that they were to implement when reviewing translation verification results (CAE, 2011, Validation). There was no specific mention of reviewing their work as country teams examined results from the translation verification process.

The Country C team received limited information about the process that they were to implement when reviewing translation verification results (CAE, 2011, Validation). There was no training available for this activity in the study.

Country C had two opportunities to document their progress when implementing changes to their translations based on feedback from the translation verification results (CAE, 2011, Validation). Although Country C completed this activity, the US organizing agency did not keep documentation of the progress.

The study timeline that the US organizing agency provided Country C team members did not include a deadline for implementing changes resulting from the translation verification process (CAE, 2010, GS.1).

The US organizing agency did not provide Country C team members training addressing how to implement changes resulting from translation verification (CAE, 2011, Validation). However, the US organizing agency provided material to help complete the work. The agency provided the Country C team with a spreadsheet that included the suggested changes from the verification translators, electronic copies of the Country C's translations, and an email explaining what to do with the material (Geisinger, personal communication, January 21, 2012).

The US organizing agency provided Country C limited information about the steps that country teams were to implement when reviewing translation verification results (CAE, 2011, Validation). In addition, the US organizing agency did not specify who was to complete the work or the qualifications necessary. It is unclear who worked on the process with or for the Country C team.

The US organizing agency provided the Country C national project manager with assistance while implementing changes from translation verification process. The agency provided information from external translators, documents to complete changes, and helped manage the process.

Country D

The US organizing agency communicated results from the external translation verification process with the Country D national project manager (Keeley, personal communication, August 22, 2011). The US organizing agency requested that Country D provide feedback on the verification notes and implement changes with which they agreed during user acceptance testing of the internet platform (Keeley, personal communication, March 29, 2011; CAE, 2010, GS.45; Keeley, personal communication, March 29, 2011). Neither the Country D

team nor the US organizing agency indicated experiencing challenges with communication timeliness.

The US organizing agency communicated results from the external translation verification process with the Country D national project manager (Keeley, personal communication, August 22, 2011). The measurement expert who was a member of the Country D team helped examine and comment on the translation verification notes.

The US organizing agency communicated results from the external translation verification process with the Country D national project manager (Keeley, personal communication, August 22, 2011). Although the Country D team completed the activity, it is not clear that those who reviewed the verification work had expertise in translation.

Country D team members collaborated to review notes from the translation verification process (Solano-Flores, personal communication, August 22, 2011). The team provided feedback on the verification notes and made any changes necessary.

Country D team members received limited information about the process that they were to implement when reviewing translation verification results (CAE, 2011, Validation). There was no specific mention of reviewing their work as country teams examined results from the translation verification process.

Country D team members received limited information about the process that they were to implement when reviewing translation verification results (CAE, 2011, Validation). There was no training available for this activity in the study.

Country D had two opportunities to document their progress when implementing changes to their translations based on feedback from the translation verification results (CAE, 2011, Validation). Country D completed this process and the progress was documented (Solano-Flores, personal communication, August 22, 2011).

At the beginning of the study, the US organizing agency provided Country D with a work plan that included deadlines for important project milestones (CAE, 2010, GS. 1). The work plan did not include a deadline for implementing changes resulting from the translation verification process (CAE, 2010, GS.1). However, Country D was able to complete the activity within the amount of time provided.

The US organizing agency provided material to help complete the work. The agency provided the Country B team with a spreadsheet that included the suggested changes from the verification translators, electronic copies of the Country B's translations, and an email explaining what to do with the material (Geisinger, personal communication, January 21, 2012). It is important to note that the Country D team did not agree with most of the suggestions made by the translator who completed the verification process. It appears that although the translator knew Spanish, the person was not familiar with the standard local dialect used in Country D. Most of the suggestions made as a result of the verification process did not apply to Country D; the suggestions contradicted the normal discourse, syntax, and semantics used in Country D.

The US organizing agency provided Country D limited information about the steps that country teams were to implement when reviewing translation verification results (CAE, 2011, Validation). In addition, the US organizing agency did not specify who was to complete the work or the qualifications necessary. The Country D team reviewed information from translation verification. However, it is not clear who else—if anyone—assisted in the process.

The US organizing agency provided Country D team members with assistance while implementing changes from translation verification process. The agency provided information from external translators, documents to complete changes, and helped manage the process.

Country E

The US organizing agency contracted an independent translation company to verify Country E's translations (Keeley, personal communication, August 22, 2011). The US organizing agency requested that Country E provide feedback on the verification notes and implement changes with which they agreed during user acceptance testing of the internet platform (Keeley, personal communication, March 29, 2011; CAE, 2010, GS.45; Keeley, personal communication, March 29, 2011). Neither the Country E team nor the US organizing agency indicated experiencing challenges with communication timeliness.

To ensure that there were no issues with underlying constructs, the US organizing agency contracted an independent translation company to verify Country E's translations (Keeley, personal communication, August 22, 2011). The Country E team included members with extensive experience in educational measurement who participated in the review of the translation verification results.

The US organizing agency provided the Country E team with results from translation verification work completed by an independent translation company for review (Keeley, personal communication, August 22, 2011). The Country E team included members with extensive experience in translation who participated in the review of the translation verification results.

Country E's team members collaborated to review notes from the translation verification process (Keeley, personal communication, January 20, 2012). The team provided feedback on the verification notes and made any changes necessary.

Country E's team members received limited information about the process that they were to implement when reviewing translation verification results (CAE, 2011, Validation). There was no specific mention of reviewing their work as country teams examined results from the translation verification process.

Country E's team members received limited information about the process that they were to implement when reviewing translation verification results (CAE, 2011, Validation). There was no training available for this activity in the study.

Country E had two opportunities to document their progress when implementing changes to their translations based on feedback from the translation verification results (CAE, 2011, Validation). Although Country E completed this activity, the US organizing agency did not keep documentation of the progress.

The US organizing agency provided the Country E team with a work plan containing deadlines for important project milestones (CAE, 2010, GS.1). The work plan did not include a deadline by when the Country E had to implement changes resulting from the translation verification process (CAE, 2010, GS.1). However, Country E was able to complete the activity within the amount of time provided.

The US organizing agency did not provide Country E team members training addressing how to implement changes resulting from translation verification (CAE, 2011, Validation). However, the US organizing agency provided material to help complete the work. The agency provided the Country E team with a spreadsheet that included the suggested changes from the verification translators, electronic copies of the Country E's translations, and an email explaining what to do with the material (Geisinger, personal communication, January 21, 2012). It is important to note that the Country E team did not agree with most of the changes suggested by the verification translator.

The US organizing agency provided Country E with limited information about the steps that country teams were to implement when reviewing translation verification results (CAE, 2011, Validation). In addition, the US organizing agency did not specify who was to complete the work or the qualifications necessary. The Country E national project manager helped review information from translation verification. However, it is not clear who else—if anyone—assisted in the process.

The US organizing agency provided the Country E national project manager assistance while implementing changes from translation verification process. The agency provided information from external translators, documents to complete changes, and helped manage the process.

Task 12: Make agreed upon changes resulting from validation procedure established by coordinating group.

Country A

During the initial New York City meeting, country teams asked the US organizing agency for training and protocol on how to conduct cognitive interviews (CAE, 2010, GS.26). The US organizing agency met the requests prior to Country A completing the full translation process. During the site visit, country teams watched a video and received training on how to conduct the labs (CAE, 2010, GS.26; CAE, 2010, GS.37). Country A team members completed the cognitive labs and provided results to the US agency (Ursin, personal communication, September 24, 2010). After a conference call addressing all of the findings and subsequent suggestions the team made necessary changes (Shavelson & Kurpius, personal communication, September 22, 2010; CAE, 2010, Module). Communication between the US and Country A teams was timely.

The national project manager and a doctoral student on the Country A team conducted and recorded the cognitive labs (Shavelson & Kurpius, 2010, End). Both team members were knowledgeable of the performance tasks (CAE, 2010, GS.26). In addition, the national project

manager possessed extensive experience in assessment (Ursin, 2010, CV). Furthermore, the doctoral student was studying issues of validity related to performance across different types of institutions of higher learning (Shavelson, 2010, End).

The national project manager and a doctoral student on the Country A team conducted and recorded the cognitive labs (Shavelson & Kurpius, 2010, End). Both team members were knowledgeable of the performance tasks (CAE, 2010, GS.26). In addition, the national project manager and student had experience in assessment (Ursin, 2010, CV; Shavelson, 2010, End). However, there is no evidence that the interviewers had experience with translation.

The national project manager on the Country A team was responsible for conducting cognitive labs and making changes based on the results (Shavelson & Kurpius, 2010, End). Country A's NPM had extensive experience managing projects (Ursin, 2010, CV; Shavelson, 2010, End).

The team for Country A participated in two opportunities to review potential changes resulting from the verification procedure. The national project manager provided the US organizing agency with results from the cognitive labs (Shavelson, 2010, End). In addition, the NPM and PI for the US organizing agency reviewed results via a conference call (Shavelson & Kurpius, 2010, End).

During the site visit by the US organizing agency's staff, Country A's national project manager participated in training addressing cognitive labs (Solano-Flores, 2010, Visit). The team watched a video that the US agency's staff created specifically for this project (Solano-Flores, 2010, Visit). In addition, the US agency made all of the training material available to the Country A team online (CAE, 2010, GS.37).

The AHELO study's protocol for cognitive labs included several opportunities for interviewers to document progress. Country A's interviewers noted peculiarities and student explanations about them during the student's talk aloud on a form that the US organizing agency provided (CAE, 2010, GS.37). In addition, the interviewers used another form to note student responses to five general questions that interviewers were to ask each student (CAE, 2010, GS.37). Furthermore, the interviewers audio and video recorded the cognitive labs (CAE, 2010, GS.37). Finally, the team provided the US organizing agency with notes from the cognitive labs (Kurpius, personal communication, June 9, 2010).

The study's timeline provided by the US organizing agency indicated that country teams were to conduct cognitive labs and make revisions to their translations based on their results between August 15 and October 1, 2010 (CAE, 2010, GS.1). The US PI scheduled a call for September 2010 with Country A's NPM that included discussing cognitive lab progress on the agenda (Shavelson & Kurpius, 2010, End). During the call, Country A's NPM explained that the team had completed the labs and discussed results (Shavelson & Kurpius, 2010, End).

The Country A team had access to information regarding the cognitive labs in different formats. The US organizing agency provided a video modeling how to conduct the labs (CAE, 2010, GS.37). The Country A team also received an electronic copy of a script and data

collection forms that they were to use (CAE, 2010, GS.37). Finally, the US agency provided a narrative of how to use the forms and script when conducting the interviews (CAE, 2010, GS.37). The team's national project manager explained that the materials were helpful and easy to use (Shavelson & Kurpius, 2010, End).

To complete cognitive labs as a means to validate translations, Country A acquired support from people within Country A who were not part of the team. First, Country A was able to coordinate with higher education institutions (HEIs) to identify students (Shavelson & Kurpius, 2010, End). In addition, twelve students attending Country A HEIs agreed to participate in the labs (Shavelson & Kurpius, 2010, End). Country A team members conducted and recorded all of the cognitive labs (Shavelson & Kurpius, 2010, End).

The US organizing agency supported Country A team members in their efforts to conduct cognitive labs as a means to validate the translation results. Country A's NPM participated in training the US agency conducted that addressed how to conduct labs for the AHELO study, record data from the interviews, and use the information within the context of translation validation (CAE, 2010, GS.26; Solano-Flores, 2010, Visit; CAE, 2010, GS.37; Shavelson & Kurpius, 2010, End).

Country B

During the initial New York City meeting, Country B team members asked the US organizing agency for training and protocol on how to conduct cognitive interviews (CAE, 2010, GS.26). The US organizing agency met the requests during the site visit to Country B (Solano-Flores, 2010, Visit). During the site visit, Country B team members watched a video and received training on how to conduct the labs (CAE, 2010, GS.26; CAE, 2010, GS.37). Country B team members completed the cognitive labs and provided results to the US agency in time to discuss them during a conference call (Shavelson & Kurpius, 2010, September 22, 2010). However, the communication on Country B's part was not timely (Kurpius, personal communication, December 3, 2010).

The Country B team was able to complete the cognitive labs (CAE, 2010, Module). One of the team members who conducted the labs was knowledge of the performance tasks (CAE, 2010, GS.26). However, the team member did not have experience with assessment or educational measurement. Furthermore, there is no information available concerning the other staff involved in carrying out the labs.

The Country B team was able to complete the cognitive labs (CAE, 2010, Module). One of the team members who conducted the labs was knowledge of the performance tasks (CAE, 2010, GS.26). However, none of the team members had experience with translation.

The Country B team was able to complete the cognitive labs with guidance from the national project manager (CAE, 2010, Module). There was limited information about the NPM's experience with project management (CAE, 2010, GS.26).

Making changes due to results from the verification procedure for translation included two review procedures. Country B was able to participate in one. The team submitted potential

changes to the US country team (CAE, 2010, GS.46). However, the team had not completed the cognitive labs in time to discuss them during a scheduled conference call (Shavelson & Kurpius, 2010, End).

Country B team members participated in training addressing how to conduct and analyze data collected during cognitive labs (CAE, 2010, GS.37). During the site visit, team members watched a training video, reviewed protocol, and discussed the purpose of the labs (Solano-Flores, 2010, Site). The Country B team learned how to use forms that the US organizing agency provided to note any components of their translations that students found confusing, unusual, or demanding unintended cognitive processes interfering with student performance (CAE, 2010, GS.37). In addition, the US organizing agency provided electronic copies of all training materials to the Country B team.

The project's cognitive labs protocol included several opportunities for the Country B team to document progress. To note peculiarities and student explanations about them during the student's talk aloud the US organizing agency provided the Country B team with a specific form (CAE, 2010, GS.37). The interviewers used another form to note student responses to five general questions that interviewers were to ask each student (CAE, 2010, GS.37). Furthermore, the interviewers recorded the cognitive labs (CAE, 2010, GS.37). Finally, the team provided the US organizing agency with notes from the cognitive labs (CAE, 2010, GS.47).

The study's timeline provided by the US organizing agency indicated that country teams were to conduct cognitive labs and make revisions to their translations based on their results between August 15 and October 1, 2010 (CAE, 2010, GS.1). The US PI scheduled a call for September 2010 with Country B's NPM that included discussing cognitive lab progress on the agenda (Shavelson & Kurpius, 2010, End). During the call, Country B's team explained that they had not begun the cognitive labs (Shavelson & Kurpius, 2010, End). They initially asked if they could postpone the activity until late December but understood that it was not possible (Shavelson & Kurpius, 2010, End). Country B explained that they were still unsure about the process and asked for additional assistance (Shavelson & Kurpius, 2010, End).

The US organizing agency provided the Country B team with information regarding the cognitive labs in different formats. The US organizing agency created a video modeling how to conduct the labs (CAE, 2010, GS.37). The Country B team also received an electronic copy of a script and data collection forms that they were to use (CAE, 2010, GS.37). Finally, the US agency provided a narrative of how to use the forms and script when conducting the interviews (CAE, 2010, GS.37). However, Country B's NPM expressed that despite the material provided the team was confused about process and required additional assistance to complete the task (Shavelson & Kurpius, 2010, End).

To complete cognitive labs for the translation validation the Country B team needed support from people in the country who were not members of the country team. First, Country B needed to coordinate with higher education institutions (HEIs) to identify students who could

help (Shavelson & Kurpius, 2010, End). Then Country B needed to get students to agree to participate in the process (Shavelson & Kurpius, 2010, End).

To complete cognitive labs as a means to validate translation results, Country B team members acquired support from the US organizing agency. The US agency provided Country B team members training on how to conduct the labs for the AHELO study, record data from the interviews, and use the information within the context of translation validation (CAE, 2010, GS.26; Solano-Flores, 2010, Visit; CAE, 2010, GS.37; Shavelson & Kurpius, 2010, End).

Country C

While attending the initial New York City meeting, Country C team members asked the US organizing agency for training and protocol on how to conduct cognitive interviews (CAE, 2010, GS.26). The US organizing agency met the requests prior to the site visit to Country C. During the site visit, the Country C team watched a training video and received training material on how to conduct the labs and analyze results (CAE, 2010, GS.26; CAE, 2010, GS.37). Country C completed the cognitive labs and provided results to the US agency (CAE, 2010, Module). Communication, particularly the conference call, was delayed because of religious holidays and Country C school schedules during the summer.

The Country C team was able to complete the cognitive labs (CAE, 2010, Module). The second national project manager for the team reported on the cognitive labs but did not provide information as to the staff that completed them (CAE, 2010, Module). It is unclear if any of the interviewers possessed measurement expertise.

The Country C team was able to complete the cognitive labs (CAE, 2010, Module). The second national project manager for the team reported on the cognitive labs but did not provide information as to the staff that completed them (CAE, 2010, Module). It is unclear if any of the interviewers possessed translation expertise.

The Country C team was able to complete the cognitive labs (CAE, 2010, Module). The national project manager for the team worked on conducting the labs and making (CAE, 2010, Module). The NPM had extensive experience with project management.

The process of making changes as a result of translation verification included two opportunities for review. The Country C team submitted their findings to the US organizing agency for review. However, the Country C team had not completed the cognitive labs in time to discuss them during a planned conference call (Shavelson & Kurpius, 2010, End).

The US organizing agency trained Country C team members on how to conduct cognitive labs and how to use the notes taken during each to modify their translations (CAE, 2010, GS.37). Team members watched a training video, reviewed cognitive labs protocol, and discussed the purpose of the labs (Solano-Flores, 2010, Site). The US organizing agency provided Country C team members with forms on which they could note any components of the translations that students found confusing, unusual, or demanding unintended cognitive processes interfering with student performance (CAE, 2010, GS.37). Country C team members had access to electronic copies of all training materials.

The AHELO study's protocol for cognitive labs included several opportunities for interviewers to document progress. Country C's interviewers noted any peculiarities during the student's talk aloud on a form that the US organizing agency provided along with student explanations about them (CAE, 2010, GS.37). In addition, the interviewers used another form to note student responses to five general questions that interviewers were to ask each student (CAE, 2010, GS.37). Furthermore, the interviewers recorded the cognitive labs (CAE, 2010, GS.37). Finally, the Country C team provided the US organizing agency with notes from the cognitive labs (CAE, 2010, GS.47).

The US organizing agency provided the Country C team with the study's timeline, which indicated that country teams were to conduct cognitive labs and make revisions to their translations based on lab results between August 15 and October 1, 2010 (CAE, 2010, GS.1). The US PI and Country C NPM scheduled a conference call for October 2010; the call included discussing cognitive lab progress (Shavelson & Kurpius, 2010, End). During the call, Country C's NPM explained that although they had a team in place to conduct the labs they had not yet begun the interviews (Shavelson & Kurpius, 2010, End).

The US organizing agency provided the Country C team with material on cognitive labs. The US organizing agency provided a video modeling how to conduct the labs (CAE, 2010, GS.37). The Country C team also received an electronic copy of a script and data collection forms that they were to use (CAE, 2010, GS.37). Finally, the US agency provided a narrative of how to use the forms and script when conducting the interviews (CAE, 2010, GS.37). The Country C national project manager indicated that the materials were helpful in training the interviewers and carrying out the work (Shavelson & Kurpius, 2010, End).

To complete cognitive labs as a means to validate translations, Country C acquired support from people within Country C who were not part of the team. Country C was able to coordinate with higher education institutions (HEIs) to identify students (Shavelson & Kurpius, 2010, End). In addition, Country C team members recruited students who agreed to participate in the labs (Shavelson & Kurpius, 2010, End).

The US organizing agency supported Country C team members in their efforts to conduct cognitive labs as a means to validate the translation results. Country C's team members participated in training the US agency conducted that addressed how to conduct labs for the AHELO study, record data from the interviews, and use the information within the context of translation validation (CAE, 2010, GS.26; Solano-Flores, 2010, Visit; CAE, 2010, GS.37; Shavelson & Kurpius, 2010, End).

Country D

During the initial New York City meeting, Country D team members expressed interest in training that the US organizing agency could offer on how to conduct cognitive interviews (CAE, 2010, GS.26). The US organizing agency met the request prior to Country D having completed the full translation process. During the site visit, country teams watched a video and received training on how to conduct the labs and analyze results (CAE, 2010, GS.26; CAE, 2010,

GS.37). Country D team members completed the cognitive labs and provided results to the US agency (CAE, 2010, Module). After a conference call addressing all of the findings and subsequent suggestions the team made necessary changes (Shavelson & Kurpius, personal communication, September 23, 2010; CAE, 2010, Module). Communication between the US and Country D teams was timely.

The Country D team completed the cognitive labs (CAE, 2010, Module). However, there is no information on the staff that the team trained to conduct the interviews. One of the team members in charge of making changes based on the lab results possessed expertise in measurement and was also familiar with the performance tasks.

The Country D team completed the cognitive labs (CAE, 2010, Module). However, there is no information on the staff that the team trained to conduct the interviews. The team members in charge of making changes based on the lab results did not have expertise in translation.

The Country D team completed the cognitive labs (CAE, 2010, Module). The national project manager for the team was responsible for conducting the labs and making changes to according to their results. The NPM had experience with project management.

The AHELO study included two opportunities for review during the process of making changes based on translation verification. Country D was able to participate in one. The team submitted potential changes to the US country team (CAE, 2010, GS.46). However, the team had not completed the cognitive labs in time to discuss them during a scheduled conference call (Shavelson & Kurpius, 2010, End).

The Country D team participated in training on how to conduct cognitive labs and how to use the notes taken during each to make changes to translations (CAE, 2010, GS.37). During the site visit, the team watched a training video, reviewed protocol, and discussed the purpose of the labs (Solano-Flores, 2010, Site). The team also learned how to use forms that the US organizing agency provided to note any components of their translations that students found confusing, unusual, or demanding unintended cognitive processes interfering with student performance (CAE, 2010, GS.37). The team used these notes to suggest changes to the PTs. The US organizing agency provided the team with electronic copies of all training materials.

The AHELO study's protocol for cognitive labs included several opportunities for interviewers to document their progress. Country D's interviewers noted any peculiarities during the student's talk aloud on a form that the US organizing agency provided (CAE, 2010, GS.37). Interviewers used the same form to write explanations students provided about any of the peculiarities the interviewer noted (CAE, 2010, GS.37). In addition, the interviewers used another form to note student responses to five general questions that interviewers were to ask each student (CAE, 2010, GS.37). Furthermore, the interviewers recorded the cognitive labs (CAE, 2010, GS.37). Finally, the Country D team provided the US organizing agency with notes from the cognitive labs (CAE, 2010, GS.47).

The study's timeline provided by the US organizing agency indicated that country teams were to conduct cognitive labs and make revisions to their translations based their results

between August 15 and October 1, 2010 (CAE, 2010, GS.1). The US PI scheduled a call for September 2010 with Country D's team members that included discussing cognitive lab progress on the agenda (Shavelson & Kurpius, 2010, End). During the call, Country D's team explained that they had not begun the cognitive labs (Shavelson & Kurpius, 2010, End).

The US organizing agency used different formats to provide the Country D team with information on cognitive labs. The US organizing agency provided a video modeling how to conduct the labs (CAE, 2010, GS.37). The Country D team also received an electronic copy of a script and data collection forms that they were to use (CAE, 2010, GS.37). Finally, the US agency provided a narrative of how to use the forms and script when conducting the interviews (CAE, 2010, GS.37). During a conference call it became clear that the team was unclear about the process. The team shared that they thought they had to conduct a full pilot of the assessment and not a cognitive lab (Shavelson & Kurpius, 2010, End).

To complete cognitive labs as a means to validate translations, Country D had to acquire support from people within Country D who were not part of the team. Country D was able to coordinate with higher education institutions (HEIs) to identify students who could participate in the cognitive labs process (Shavelson & Kurpius, 2010, End). Then Country D identified students attending Country D HEIs who agreed to assist with the labs (Shavelson & Kurpius, 2010, End). Also, in addition to one team member Country D hired additional staff to conduct and record all of the cognitive labs (Shavelson & Kurpius, 2010, End).

To complete cognitive labs as a means to validate translations, Country D had to acquire support from people outside of the country. The US organizing agency supported Country D team members in their efforts to conduct cognitive labs as a means to validate the translation results. Team members participated in training the US agency conducted that addressed how to conduct labs for the AHELO study, record data from the interviews, and use the information within the context of translation validation (CAE, 2010, GS.26; Solano-Flores, 2010, Visit; CAE, 2010, GS.37; Shavelson & Kurpius, 2010, End).

Country E

During the initial New York City meeting, country teams asked the US organizing agency for training on how to conduct cognitive interviews (CAE, 2010, GS.26). Country E in particular asked for protocol that they were to follow (CAE, 2010, GS.26). During the site visit, Country E team members watched a video and received training material addressing how to conduct the labs (CAE, 2010, GS.26; CAE, 2010, GS.37). Country E team members completed the cognitive labs and provided results to the US agency (Opheim, personal communication, December 28, 2010). After a conference call addressing all of the findings and subsequent suggestions the team made necessary changes (Shavelson & Kurpius, personal communication, September 22, 2010; CAE, 2010, Module). Communication between the US and Country E teams was timely.

The Country E team conducted cognitive labs and provided feedback based on results (CAE, 2010, Module). There is evidence that the team had members with previous experience

conducting the labs, were familiar with the performance tasks, and had experience with measurement (Solano-Flores & Shavelson, 2010, Visit).

The Country E team conducted cognitive labs and provided feedback based on results (CAE, 2010, Module). There is evidence that the team had members with previous experience translating international assessments (Solano-Flores & Shavelson, 2010, Visit).

The Country E team completed all steps involved in the cognitive labs (CAE, 2010, Module). The NPM for the team was responsible for the work. The NPM had extensive experience with project management.

Making changes due to results from the validation procedure for translation included two review procedures. Country E was able to participate in one. The team submitted potential changes to the US country team (CAE, 2010, GS.46). However, the team had not completed the cognitive labs in time to discuss them during a scheduled conference call (Shavelson & Kurpius, 2010, End).

Some members of the Country E team had extensive experience conducting cognitive labs. However, it was important that they received training on the AHELO protocol. During the site visit, Country E team members watched a training video, reviewed protocol, and discussed the purpose of the labs (Solano-Flores, 2010, Site). The team also reviewed how to use forms that the US organizing agency provided to note any components of their translations that students found confusing, unusual, or demanding unintended cognitive processes interfering with student performance (CAE, 2010, GS.37). The team members were to suggest changes to the PTs based on notes taken during the cognitive interview. The US organizing agency provided the team with electronic copies of all training materials.

The AHELO study's protocol for cognitive labs included several opportunities for interviewers to document progress. Country E's interviewers noted any peculiarities during the student's talk aloud on a form that the US organizing agency provided along with student explanations about them (CAE, 2010, GS.37). In addition, the interviewers used another form to note student responses to five general questions that interviewers were to ask each student (CAE, 2010, GS.37). Furthermore, the interviewers recorded the cognitive labs (CAE, 2010, GS.37). Finally, the Country E team provided the US organizing agency with notes from the cognitive labs (CAE, 2010, GS.47).

The study's timeline provided by the US organizing agency indicated that country teams were to conduct cognitive labs and make revisions to their translations based their results between August 15 and October 1, 2010 (CAE, 2010, GS.1). The US PI scheduled a call for September 2010 with the Country E team that included discussing cognitive lab progress on the agenda (Shavelson & Kurpius, 2010, End). During the call, Country E's team members explained that the team had not been able to begin work on cognitive labs (Shavelson & Kurpius, 2010, End). Furthermore, they explained that they would conduct the labs in November 2010 because of a national report that they would need to focus on that was due in October (Shavelson & Kurpius, 2010, End).

The Country E team had access to information regarding the cognitive labs. The US organizing agency provided a video modeling how to conduct the labs (CAE, 2010, GS.37). The Country E team also received an electronic copy of a script and data collection forms that they were to use (CAE, 2010, GS.37). Finally, the US agency provided a narrative of how to use the forms and script when conducting the interviews (CAE, 2010, GS.37). The team's national project manager explained that the materials were helpful and easy to use (Shavelson & Kurpius, 2010, End). However, one of the team members disagreed with modeling the think aloud portion of the lab for students (Solano-Flores, 2010, Visit).

To complete cognitive labs as a means to validate translations, Country E team members acquired support from people within Country E who were not part of the team. The team coordinated with higher education institutions (HEIs) to identify students who were eligible to participate in the process (Shavelson & Kurpius, 2010, End). In addition, Country E recruited students attending Country E HEIs who would agree to participate in the labs (Shavelson & Kurpius, 2010, End). Country E team members conducted and recorded all of the cognitive labs (Shavelson & Kurpius, 2010, End).

To complete cognitive labs as a means to validate translations, Country E team members acquired support from people outside of the country. Although some members of the Country E team had extensive expertise in cognitive labs, the team requested protocol to accommodate the specific goals of the AHELO study. The US organizing agency supported Country E team members in their efforts to conduct cognitive labs as a means to validate the translation results. Team members participated in training the US agency conducted that addressed how to conduct labs for the AHELO study, record data from the interviews, and use the information within the context of translation validation (CAE, 2010, GS.26; Solano-Flores, 2010, Visit; CAE, 2010, GS.37; Shavelson & Kurpius, 2010, End).

Task 13: Test assessment implementation process for target population usability.

Country A

The Country A team conducted cognitive labs with students using pencil and paper instead of with the online platform on which students would normally be assessed (CAE, 2010, GS.37). Country A team members were to review the language and the format of the performance tasks online (CAE, 2010, GS.48). However, country teams did not pilot the performance tasks as they would be implemented during the actual assessment with a representative sample of students. The AHELO study did not include an opportunity to test the internet platform with students from the participating countries (CAE, 2010, GS.1).

Country B

The AHELO study did not include an opportunity to test the internet platform with students from the participating countries (CAE, 2010, GS.1). The Country B team conducted cognitive labs with students using pencil and paper instead of with the online platform on which students would normally be assessed (CAE, 2010, GS.37). Country B team members were to review the language and the format of the performance tasks online (CAE, 2010, GS.48).

However, the Country B team did not pilot the performance tasks as they would be implemented during the actual assessment with a representative sample of students.

Country C

The AHELO study did not include an opportunity to test the internet platform with students from the participating countries (CAE, 2010, GS.1). The Country C team conducted cognitive labs with students using pencil and paper instead of with the online platform on which students would normally be assessed (CAE, 2010, GS.37). Country C team members were to review the language and the format of the performance tasks online (CAE, 2010, GS.48). However, the Country C team did not pilot the performance tasks as they would be implemented during the actual assessment with a representative sample of students.

Country D

The AHELO study did not include an opportunity to test the internet platform with students from the participating countries (CAE, 2010, GS.1). The Country D team conducted cognitive labs with students using pencil and paper instead of with the online platform on which students would normally be assessed (CAE, 2010, GS.37). Country D team members were to review the language and the format of the performance tasks online (CAE, 2010, GS.48). However, the Country D team did not pilot the performance tasks as they would be implemented during the actual assessment with a representative sample of students.

Country E

The AHELO study did not include an opportunity to test the internet platform with students from the participating countries (CAE, 2010, GS.1). The Country E team conducted cognitive labs with students using pencil and paper instead of with the online platform on which students would normally be assessed (CAE, 2010, GS.37). Country E team members were to review the language and the format of the performance tasks online (CAE, 2010, GS.48). However, the Country E team did not pilot the performance tasks as they would be implemented during the actual assessment with a representative sample of students.

Task 14: Provide students with an opportunity to become familiar with test format and expectations.

Country A

During the initial New York City meeting, country teams expressed concerns that their students would not be familiar with the performance tasks' format and response requirements (CAE, 2010, GS.26; CAE, 2010, Milestone 3). The US organizing agency responded quickly by creating a mini-performance task that Country A could provide to students so that they could practice with prior to taking the actual performance task test (CAE, 2010, GS.38). In January 2011 the US organizing agency sent country teams a timeline that included due dates for the mini performance task. By the third week in December 2011, the Country A team was to send the US organizing agency their translated copies of the mini-PT (CAE, 2010, GS.45; CAE, 2010, GS.38). Communication between Country A and the US organizing agency was timely.

The US organizing agency, with assistance from psychometricians with experience in performance task assessments created a mini-performance task that Country A (CAE, 2010, GS.38). Country A was to translate the mini-PT and make it available to students; students were able to become familiar with the assessment format and response requirements through the mini-PT practice test (CAE, 2010, GS.38). Country A's NPM, who had measurement experience and was familiar with the full performance tasks, reviewed the mini-PT (Shavelson & Kurpius, 2010, End; Ursin, personal communication, January 21, 2011).

Country A was responsible for translating the mini-PT that the US organizing agency created (CAE, 2010, GS.1; CAE, 2010, GS.38). The Country A NPM contributed to the mini-PT's translation process and submitted the final mini-PT (Ursin, personal communication, June xx, 2011). The Country A NPM did not have translation expertise (Ursin, 2010, CV). In addition, it is not clear who, if anyone, assisted the NPM with review of the mini-PT translation.

The Country A team was able to review and translate the mini-performance task (Ursin, personal communication, June 8, 2011). The Country A NPM was able to take on the additional work and collaborate with the US organizing agency and Country A team. The Country A team was able to make the mini-PT available prior to assessment administration (Keeley, personal communication, April 13, 2012).

The US organizing agency provided Country A with opportunities to review work completed with the mini-performance task. First, Country A was able to review the mini-PT for appropriateness (Ursin, personal communication, June 8, 2010). Second, translation of the mini-PT required that two translators create two independent translations and then reconcile them into one (CAE, 2010, GS.36). Third, the Country A team was able to review the mini-PT on the internet platform (Keeley, personal communication, June 6, 2011).

Country teams were to implement the dual translation process with the mini-performance task (CAE, 2010, GS.36). Country A team members participated in translation training that took place during the initial meeting in New York City (CAE, 2010, GS.26). The training addressed the dual translation process (CAE, 2010, GS.13). In addition, the US organizing agency provided documents addressing the translation process and mini-performance task (CAE, 2010, GS.36; CAE, 2010, GS.13; CAE, 2010, GS.38).

Following the guidelines that the US organizing agency provided, the Country A team documented their progress when working on the mini-performance task. Country A and the US organizing agency kept a copy of the original English version of the mini-PT (CAE, personal communication, May 11, 2010). Country A team members kept the two individual translations, notes from the reconciliation process, and the final reconciled mini-performance task (CAE, 2010, GS.4; CAE, 2010, GS.36; CAE, 2010, GS.13). In addition, the US organizing agency documented country teams' suggested modifications and final mini-PT changes (CAE, 2010, Milestone; CAE, 2010, Module).

The Country A team had to complete several activities while working on the mini-performance task. During the beginning of the summer of 2011, the Country A team had to

review the English version of the mini-performance task that the US organizing agency created and suggest modifications (CAE, personal communication, May 11, 2010). Country A was to translate the mini-performance task using the dual translation process by December 1, 2010 (CAE, 2010, GS.1, 2010). Finally, the Country A team was to review the online version of the mini-PT (Keeley, personal communication, June 6, 2010). Country A was able to review the original version of the mini-PT and translate it (Ursin, personal communication, June 7, 2010; Ursin, personal communication, January 21, 2011). Country A was then able to review the online platform (CAE, 2011, Milestone 3).

The US organizing agency provided the Country A team with several documents addressing the dual translation process (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). In addition, the US organizing agency created a mini-performance task that would require less resources and be easier to work with than a full performance task (CAE, 2010, GS.38) The Country A team did not indicate that they experienced challenges when working with the documents.

The US organizing agency required that Country A use the dual translation process with the mini-performance task (CAE, 2010, GS.38). The Country A team completed the work on the mini-PT (Keeley, personal communication, April 13, 2012). However, there is limited information on the staff involved in translating the mini-PT. The team also had to make the mini-PT available to students, which required assistance from higher education institutions.

The Country A team received support from outside of the country to provide students with an opportunity to become familiar with the test format and expectations. The US organizing agency created a mini-performance task that mirrored the full performance tasks students would take (CAE, 2010, GS.38). The US organizing agency also provided training and documentation addressing dual translation procedures that country teams would use with the mini-PT (CAE, 2010, GS.13; CAE, 2010, GS.4).

Country B

At the initial meeting in New York City, Country B team members expressed concern that students would not be familiar with the performance tasks' demands—specifically, the evaluation criteria (CAE, 2010, Module 3). In response the US organizing agency created a mini-performance task that Country B could provide to students in advance (CAE, 2010, Module 3). The mini-PT would serve as a means to share evaluation criteria and provide students with an opportunity to become familiar with the PT format (CAE, 2010, Module 3). The US agency also provided deadlines (CAE, 2010, GS.38). The Country B team was quick to communicate with the US agency about the mini-PT.

The US organizing agency, with the assistance of psychometricians specializing in performance assessments, created a mini-PT (CAE, 2010, Module 3; CAE, 2010, GS.38). Once they translated the PT, Country B could give students access to the mini-PT online prior to the actual test (CAE, 2010, Module 3). Students could then become familiar with the assessment format and response requirements (CAE, 2010, GS.38). The Country B team, which included a

measurement expert, reviewed the mini-PT and provided feedback (Young, personal communication, June 4, 2010).

The Country B team was able to review and translate the mini-PT the US organizing agency created (Young, personal communication, June 4, 2010). However, the team for Country B did not appear to have expertise in translation or assessment across diverse linguistic or cultural groups. (Solano-Flores, Visit, 2010). There is no information indicating who, if anyone, assisted the Country B team with work on the mini-PT.

The Country B NPM was able to coordinate with the US organizing agency and the Country B team to complete work on the mini-performance task. The team reviewed and translated the mini-PT the US organizing agency created (Young, personal communication, June 4, 2010).

Country B had two opportunities to review work completed with the mini-performance task. Once the US organizing agency create the mini-PT, the Country B team was able to review the mini-PT for appropriateness (Young, personal communication, June 4, 2010). In addition, translation of the mini-PT required that two translators create two independent translations and then reconcile them into one (CAE, 2010, GS.36). Finally, the Country B team reviewed the mini-PT on the internet platform (Keeley, personal communication, June 6, 2011).

Country teams were to implement the dual translation process with the mini-performance task (CAE, 2010, GS.36). During the initial meeting in New York City, the US organizing agency provided Country B team members with training on the dual translation process (CAE, 2010, GS.26; CAE, 2010, GS.13). In addition, the US organizing agency provided documents addressing the translation process (CAE, 2010, GS.36; CAE, 2010, GS.13; CAE, 2010, GS.38).

The US organizing agency provided Country B with guidelines for working with the mini-performance task that included several opportunities for the Country B team to document their progress. First, Country B and the US organizing agency kept a copy of the original English version of the mini-PT (CAE, personal communication, May 11, 2010). Second, Country B team members collected the two individual translations, notes from the reconciliation process, and the final reconciled mini-performance task (CAE, 2010, GS.4; CAE, 2010, GS.36; CAE, 2010, GS.13). In addition, the US organizing agency documented country teams' suggested modifications and final mini-PT changes (CAE, 2010, Milestone; CAE, 2010, Module).

To prepare the mini-performance task that the US organizing agency created for use with Country B students, the Country B team had to complete several steps. During the beginning of the summer of 2011, the team reviewed the English version of the mini-performance task that the US organizing agency created and suggested modifications (CAE, personal communication, May 11, 2010). In addition, the Country B team was to translate the mini-performance task by December 1, 2010 and then review the online version of the mini-PT (CAE, 2010, GS.1, 2010; Keeley, personal communication, June 6, 2010). There is no evidence that the team was able to complete the translation and online review on time.

The US organizing agency provided the Country B team with several documents addressing the dual translation process (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). In addition, the US organizing agency created a mini-performance task that would require less resources and be easier to work with than a full performance task (CAE, 2010, GS.38) The Country B team did not indicate that they experienced challenges when working with the documents.

The US organizing agency required that Country B apply the dual translation process with the mini-performance task (CAE, 2010, GS.38). The Country B team was not able to complete the steps involved in making the mini-PT available to students.

The Country B team received support from outside of the country to provide students with an opportunity to become familiar with the test format and expectations. The US organizing agency created a mini-performance task that mirrored the full performance tasks students would take (CAE, 2010, GS.38). The US organizing agency also provided training and documentation addressing dual translation procedures that country teams would use with the mini-PT (CAE, 2010, GS.13; CAE, 2010, GS.4).

Country C

During the initial New York City meeting, country teams shared that their students may not have been familiar with the performance tasks' format and response requirements (CAE, 2010, GS.26; CAE, 2010, Milestone 3). The US organizing agency responded quickly by creating a mini-performance task that Country C could provide to students so that they could practice with prior to taking the actual performance task test (CAE, 2010, GS.38). In January 2011 the US organizing agency sent country teams a timeline that included due dates for the mini performance task. By the third week in December 2011, the Country C team was to send the US organizing agency their translated copies of the mini-PT (CAE, 2010, GS.45; CAE, 2010, GS.38). Communication between Country C and the US organizing agency was timely.

The US organizing agency, with the assistance of psychometricians specializing in performance assessments, created a mini-PT (CAE, 2010, Module 3; CAE, 2010, GS.38). After translating the mini-PT, Country C could online access to the mini-PT prior to the actual test (CAE, 2010, Module 3). Students could become familiar with the assessment format and evaluation criteria (CAE, 2010, GS.38). There is limited information about the Country C staff who worked on the mini-PT translation process.

After the US organizing agency created a mini-PT to familiarize students with the assessment format and evaluation criteria, the Country C team was responsible for translation (CAE, 2010, GS.1). There is very limited information about Country C's process while working with the mini-PT. Therefore, it is not possible to determine if the team had staff with translation experience involved in the process.

The Country C team was able to review and translate the mini-performance task that the US organizing agency created (CAE, 2010, Milestone 3). The Country C team was able to make

the mini-PT available prior to assessment administration (Keeley, personal communication, April 13, 2012).

The US organizing agency provided Country C with guidelines on how to work on the mini-performance task. The guidelines included two opportunities to review work completed with the mini-performance task. First, Country C was able to review the mini-PT for appropriateness. Second, translation of the mini-PT required that two translators create two independent translations and then reconcile them into one (CAE, 2010, GS.36). Last, the Country C team had an opportunity to review the mini-PT on the internet platform (Keeley, personal communication, June 6, 2011).

The mini-performance task was to be translated using the dual translation process (CAE, 2010, GS.36). Country C team members participated in training address the dual translation process that took place during the initial meeting in New York City (CAE, 2010, GS.26; CAE, 2010, GS.13). In addition, the US organizing agency provided electronic copies of documents that addressed the translation process and mini-performance task (CAE, 2010, GS.36; CAE, 2010, GS.13; CAE, 2010, GS.38).

Country C and the US organizing agency documented Country C's progress while working on the mini-performance task. Once they received the original English version of the mini-PT, the Country C team collected the two initial individual translations and the reconciled translation (CAE, 2010, GS.4; CAE, 2010, GS.36). The US organizing agency documented Country C's progress in translation in several reports to the international organizing agency (CAE, 2010, Milestone; CAE, 2010, Module).

The Country C team had to complete several activities while working on the mini-performance task. During the beginning of the summer of 2011, the Country C team reviewed the English version of the mini-performance task that the US organizing agency created and suggest modifications (CAE, personal communication, May 11, 2010). Country C translated the mini-performance task using the dual translation process (CAE, 2010, Milestone 3). Finally, the Country C team was to review the online version of the mini-PT (Keeley, personal communication, June 6, 2010). Country C was then able to review the online platform (CAE, 2011, Milestone 3).

The US organizing agency provided the Country C team with several documents addressing the translation process (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). In addition, the US organizing agency created a mini-performance task that would require less resources and be easier to work with than a full performance task (CAE, 2010, GS.38) The Country C team did not indicate that they experienced challenges when working with the documents.

The US organizing agency required that Country B use the dual translation process with the mini-performance task (CAE, 2010, GS.38). The Country B team completed the work on the mini-PT (Keeley, personal communication, April 13, 2012). However, there is limited

information on the staff involved in translating the mini-PT. The team also had to make the mini-PT available to students, which required assistance from higher education institutions.

The Country C team received support from outside of the country to provide students with an opportunity to become familiar with the test format and expectations. The US organizing agency created a mini-performance task that mirrored the full performance tasks students would take (CAE, 2010, GS.38). The US organizing agency also provided training and documentation addressing dual translation procedures that country teams would use with the mini-PT (CAE, 2010, GS.13; CAE, 2010, GS.4).

Country D

Country teams expressed concerns that their students would not be familiar with the performance tasks' format and response requirements during the initial New York City meeting (CAE, 2010, GS.26; CAE, 2010, Milestone 3). The US organizing agency responded quickly by creating a mini-performance task that Country D could provide to students so that they could practice with prior to taking the actual performance task test (CAE, 2010, GS.38). In January 2011 the US organizing agency sent country teams a timeline that included due dates for the mini performance task. By the third week in December 2011, the Country D team was to send the US organizing agency their translated copies of the mini-PT (CAE, 2010, GS.45; CAE, 2010, GS.38). Communication between Country D and the US organizing agency was timely.

With the assistance of psychometricians specializing in performance assessments, the US organizing agency created a mini-PT (CAE, 2010, Module 3; CAE, 2010, GS.38). Once they translated the mini-PT, the Country D team could give students access to the mini-PT online prior to the actual test (CAE, 2010, Module 3). Students could then become familiar with the assessment format and response requirements (CAE, 2010, GS.38). The measurement expert on the Country D team reviewed the mini-PT and provided feedback (Urrea, personal communication, June 1, 2010).

The US organizing agency created and made available to the Country D team a mini-PT that would allow students to become familiar with the assessment format and evaluation criteria (CAE, 2010, GS.38). The Country D team reviewed and translated the mini-PT. Although the Country D team members had experience working in bilingual education and were bicultural (CAE, GS.11, 2010), the team did not possess translation expertise.

The Country D team reviewed and translated the mini-PT. One of the team members coordinated with the US organizing agency and the other team members (Urrea, June 01, 2011). The Country D team was able to make the mini-PT available prior to assessment administration (Keeley, personal communication, April 13, 2012).

Country D had two opportunities to review work completed with the mini-performance task. Once the US organizing agency created the mini-PT, Country D team members were able to review the mini-PT for local cultural responsiveness (Urrea, personal communication, June 1, 2010). Also, translation of the mini-PT required that two translators create two independent translations and then reconcile them into one (CAE, 2010, GS.36). In addition, the Country D

team had the opportunity to review the mini-PT on the internet platform (Keeley, personal communication, June 6, 2011).

Country teams were to implement the dual translation process with the mini-performance task (CAE, 2010, GS.36). The Country D team participated in translation training that took place during the initial meeting in New York City (CAE, 2010, GS.26). The training addressed the dual translation process (CAE, 2010, GS.13). In addition, the US organizing agency provided Country D with electronic copies of documents addressing the translation process and mini-performance task (CAE, 2010, GS.36; CAE, 2010, GS.13; CAE, 2010, GS.38).

Following the guidelines that the US organizing agency provided, the Country D team documented their progress while working on the mini-performance task. Country D and the US organizing agency kept a copy of the original English version of the mini-PT (CAE, personal communication, May 11, 2010). Also, Country D team members collected both individual translations, notes from the reconciliation process, and the final reconciled mini-PT (CAE, 2010, GS.4; CAE, 2010, GS.36; CAE, 2010, GS.13). In addition, the US organizing agency documented country teams' suggested modifications and final mini-PT changes (CAE, 2010, Milestone; CAE, 2010, Module).

The Country D team had to complete several activities to prepare the mini-performance task created by the US organizing agency for use by Country D students. During the beginning of the summer of 2011, the Country D team had to review the English version of the mini-performance task that the US organizing agency created and suggest modifications (CAE, personal communication, May 11, 2010). Country D also had to translate the mini-performance task using the dual translation process by December 1, 2010 (CAE, 2010, GS.1, 2010). Finally, the Country D team was to review the online version of the mini-PT (Keeley, personal communication, June 6, 2010). Country D was able to review the original version of the mini-PT and translate it (Rosas, personal communication, June 6, 2011). Also, Country D translated the mini-PT and reviewed the online platform (CAE, 2011, Milestone 3).

The US organizing agency provided the Country D team with several documents addressing the translation process (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). In addition, the US organizing agency created a mini-performance task that would require less resources and be easier to work with than a full performance task (CAE, 2010, GS.38) The Country D team did not indicate that they experienced challenges when working with the documents.

The US organizing agency required that Country D use the dual translation process with the mini-performance task (CAE, 2010, GS.38). The Country D team completed the work on the mini-PT (Keeley, personal communication, April 13, 2012). However, there is limited information on the staff involved in translating the mini-PT. The team also had to make the mini-PT available to students, which required assistance from higher education institutions.

The Country D team received support from outside of the country to provide students with an opportunity to become familiar with the test format and expectations. The US organizing

agency created a mini-performance task that mirrored the full performance tasks students would take (CAE, 2010, GS.38). The US organizing agency also provided training and documentation addressing dual translation procedures that country teams would use with the mini-PT (CAE, 2010, GS.13; CAE, 2010, GS.4).

Country E

During the initial New York City meeting, country teams expressed concerns that their students would not be familiar with the performance tasks' format and response requirements (CAE, 2010, GS.26; CAE, 2010, Milestone 3). The US organizing agency responded quickly by creating a mini-performance task that Country E could provide to students so that they could practice with prior to taking the actual performance task test (CAE, 2010, GS.38). In January 2011 the US organizing agency sent country teams a timeline that included due dates for the mini performance task. By the third week in December 2011, the Country E team was to send the US organizing agency their translated copies of the mini-PT (CAE, 2010, GS.45; CAE, 2010, GS.38). Communication between Country E and the US organizing agency was timely.

The US organizing agency, with the assistance of psychometricians specializing in performance assessments, created a mini-PT (CAE, 2010, Module 3; CAE, 2010, GS.38). Once they translated the PT, Country E could provide students access to the mini-PT online prior to the actual test (CAE, 2010, Module 3). Students could then become familiar with the assessment format and response requirements (CAE, 2010, GS.38). The Country E team, which included measurement expertise, reviewed the mini-PT and provided feedback (Turmo, personal communication, June 3, 2010).

The US organizing agency provided Country E with a mini-PT that would allow students to become familiar with the PT assessment format and evaluation criteria (CAE, 2010, GS.38). The Country E team was responsible for reviewing and translating the mini-PT (CAE, 2010, GS.1). The Country E team members had extensive experience with translation procedures used in international comparison studies (CAE, GS.11, 2010).

The Country E team was able to review and translate the mini-performance task (Turmo, personal communication, June 3, 2011). The Country E NPM was able to take on the additional work and collaborate with the US organizing agency and Country E team. The Country E team was able to make the mini-PT available prior to assessment administration (Keeley, personal communication, April 13, 2012).

After the US organizing agency created the mini-performance task, the agency provided Country E with guidelines for working with the mini-PT. Country D had opportunities to review work completed with the mini-performance task. First, Country D was able to review the mini-PT for appropriateness (Urrea, personal communication, June 1, 2010). Second, translation of the mini-PT required that two translators create two independent translations and then reconcile them into one (CAE, 2010, GS.36). Third, the Country E team was able to review the mini-PT on the internet platform (Keeley, personal communication, June 6, 2011).

Country E team members were to implement the dual translation process with the mini-performance task (CAE, 2010, GS.36). Country E team members participated in translation training that took place during the initial meeting in New York City (CAE, 2010, GS.26). The training addressed the dual translation process (CAE, 2010, GS.13). In addition, the US organizing agency provided documents addressing the translation process and mini-performance task (CAE, 2010, GS.36; CAE, 2010, GS.13; CAE, 2010, GS.38).

The US organizing agency and Country E documented Country E's progress with the mini-performance task translation following the guidelines that the US organizing agency provided. Country E and the US organizing agency kept a copy of the original English version of the mini-PT (CAE, personal communication, May 11, 2010). Also, Country E team members collected the two individual translations, notes from the reconciliation process, and the final reconciled mini-PT (CAE, 2010, GS.4; CAE, 2010, GS.36; CAE, 2010, GS.13). In addition, the US organizing agency documented country teams' suggested modifications and final mini-PT changes (CAE, 2010, Milestone; CAE, 2010, Module).

To prepare the mini-performance task that the US organizing agency created for use with Country E students, the Country E team had to complete several steps. During the beginning of the summer of 2011, the team reviewed the English version of the mini-performance task that the US organizing agency created and suggested modifications (CAE, personal communication, May 11, 2010). In addition, the Country E team translated the mini-performance task and reviewed the online version of the mini-PT (CAE, 2010, GS.1, 2010; Keeley, personal communication, June 6, 2010).

The US organizing agency provided the Country E team with several documents addressing the translation process (CAE, 2010, GS.13; CAE, 2010, GS.4; CAE, 2010, GS.14; CAE, 2010, GS.31; CAE, 2010, GS.36). In addition, the US organizing agency created a mini-performance task that would require less resources and be easier to work with than a full performance task (CAE, 2010, GS.38) The Country E team did not indicate that they experienced challenges when working with the documents.

The US organizing agency required that Country E use the dual translation process with the mini-performance task (CAE, 2010, GS.38). The Country E team completed the work on the mini-PT (Keeley, personal communication, April 13, 2012). However, there is limited information on the staff involved in translating the mini-PT. The team also had to make the mini-PT available to students, which required assistance from higher education institutions.

The Country E team received support from outside of the country to provide students with an opportunity to become familiar with the test format and expectations. The US organizing agency created a mini-performance task that mirrored the full performance tasks students would take (CAE, 2010, GS.38). The US organizing agency also provided training and documentation addressing dual translation procedures that country teams would use with the mini-PT (CAE, 2010, GS.13; CAE, 2010, GS.4).

Task 15: Hire scorers according to coordinating group's specifications.

Country A

Country teams asked for information about scorers during the initial meeting in New York City (CAE, 2010, GS.26). By March, the US organizing agency provided information about scorers. The agency recommended that each country hire five scorers per task, which would total ten scorers, from June 2011 through 2012 (CAE, 2010, GS.24). Each scorer would need two days of training and six hours for recalibration (CAE, 2010, GS.24). In March 2010, the US organizing agency also provided information on the lead scorer (CAE, 2010, GS.23). In October 2010 the US organizing agency sent out another document reminding countries about scorer recruitment and training (CAE, 2010, GS.45). Country teams gave countries over two months notice prior to lead scorer training scheduled to take place during the general meeting in March 2012 (Coates & XX, personal communication, December 23, 2011). There is no indication that either the Country A team or the US organizing agency had difficulty with untimely communication.

The US organizing agency recommended that Country A hire five scorers per task, which would total ten scorers, and recruit one lead scorer (CAE, 2010, GS.24). However, the agency did not recommend any specific qualifications for hiring the scorers (CAE, 2010, GS.24). Scorers would gain measurement knowledge through training. The scorers would train on the performance task expectations and scoring and on recalibration (CAE, 2010, GS.45).

The US organizing agency did not provide recommended qualifications that the Country A team was to use when hiring scorers (CAE, 2010, GS.24). There is no information about Country A's scorers; there is no indication that Country A scorers had expertise in translation.

Each country's national project managers had to coordinate with the US organizing agency, scorers, and other team members in order to hire and train scorers (CAE, 2010, GS.24; CAE, 2010, GS.23). Country A's NPM was able to coordinate all of the steps involved and complete the work associated with scorers (Keeley, personal communication, April 13, 2012).

There were no specific review opportunities while country teams hired scorers. However, scorers were to complete training that included reviewing their progress in their scoring abilities (CAE, 2010, GS.24; CAE, 2010, GS.23). Country A's scorers completed all required tasks (Keeley, personal communication, April 13, 2012).

There were training opportunities for Country A team members and scorers that they hired. Country A team members took part in initial scorer training that occurred during a general meeting in Paris the third week of November 2011 (Coates & XX, personal communication, December 23, 2011). Also, according to guidelines created by the US organizing agency, each scorer would participate in two days of scoring training and six hours of recalibration (CAE, 2010, GS.24). Country A scorers participated in additional scorer training that was scheduled to take place prior to test administration in 2012 (Keeley, personal communication, April 13, 2012).

The Country A team did not have opportunities to document the hiring of scorers for the AHELO study.

Country A team members were able to participate in scorer training that took place during the general Paris meeting the third week of November 2011. In addition, Country A hired scorers and they completed two days of training and six hours for recalibration that would take place through 2012 (Coates & XX, personal communication, December 23, 2011; Keeley, personal communication, April 13, 2012).

The US organizing agency provided two documents addressing hiring scorers. One document provided information on hiring scorers (CAE, 2010, GS.24). A second document included information regarding the lead scorer (CAE, 2010, GS.23). In addition, an international cooperating agency emailed countries information about scorer training (Coates & XX, personal communication, December 23, 2011). The Country A team did not give any indication that they found the documents difficult or challenging.

The Country A team needed to hire ten scorers and identify a lead scorer (CAE, 2010, GS.23; CAE, 2010, GS.24). However, there is no information about the scorers that the Country A team hired for the study.

The Country A team received support from outside of the country when hiring scorers. The US organizing agency provided information about the number of scorers that each country needed to hire (CAE, 2010, GS.24; CAE, 2010, GS.23). The US organizing agency also provided scorer training (Coates & XX, personal communication, December 23, 2011).

Country B

During the initial meeting in New York City Country B team members requested information about scorers (CAE, 2010, GS.26). A month after the meeting the US organizing agency provided information. The agency recommended that Country B hire a total of ten scorers—five scorers per task—from June 2011 through 2012 (CAE, 2010, GS.24). In addition to the actual scoring work, each scorer would need two days of training and six hours for recalibration (CAE, 2010, GS.24). In March 2010, the US organizing agency also provided information on the lead scorer (CAE, 2010, GS.23). In October 2010 the US organizing agency sent out another document reminding countries about scorer recruitment and training (CAE, 2010, GS.45). Countries had over two months notice prior to lead scorer training scheduled to take place during the general meeting in March 2012 (Coates & XX, personal communication, December 23, 2011). There is no indication that either the Country B team or the US organizing agency had difficulty with untimely communication.

The US organizing agency recommended that the Country B team hire five scorers per task, which would total ten scorers, as well as a lead scorer (CAE, 2010, GS.24). However, the US agency did not recommend qualifications for hiring the scorers (CAE, 2010, GS.24). Scorers would gain measurement knowledge through training. The scorers would train on the performance task expectations and scoring and on recalibration (CAE, 2010, GS.45).

The US organizing agency did not provide recommended qualifications that the Country B team was to use when hiring scorers (CAE, 2010, GS.24). There is no indication that Country B scorers had expertise in translation.

Each country's national project managers had to coordinate with the US organizing agency, scorers, and other team members in order to hire and train scorers (CAE, 2010, GS.24; CAE, 2010, GS.23). There is not evidence indicating that the Country B NPM was able to coordinate all of the steps involved and complete the work associated with scorers (Keeley, personal communication, April 13, 2012).

There were no specific review opportunities while country teams hired scorers. However, scorers were to complete training that included reviewing their progress in their scoring abilities (CAE, 2010, GS.24; CAE, 2010, GS.23). There is no evidence that Country B's scorers completed all required tasks (Keeley, personal communication, April 13, 2012).

Country B team members and scorers they hired for the AHELO study had opportunities to train on the scoring process. Country B team members took part in initial scorer training that occurred during a general meeting in Paris the third week of November 2011 (Coates & XX, personal communication, December 23, 2011). The US organizing agency asked Country B to allow each scorer to participate in two days of scoring training and six hours of recalibration (CAE, 2010, GS.24). However, there is no evidence that Country B scorers participated in additional scorer training that was scheduled to take place prior to test administration in 2012 (Keeley, personal communication, April 13, 2012).

The Country B team did not have opportunities to document the hiring of scorers for the AHELO study.

Country B team members were able to participate in scorer training that took place during the general Paris meeting the third week of November 2011. However, there is no indication that Country B team members hired scorers and had them complete two days of training and six hours for recalibration that would take place through 2012 (Coates & XX, personal communication, December 23, 2011; Keeley, personal communication, April 13, 2012).

The Country B team did not give any indication that they found the documents difficult or challenging. The US organizing agency provided two documents addressing hiring scorers. One document provided information on hiring scorers (CAE, 2010, GS.24). A second document included information regarding the lead scorer (CAE, 2010, GS.23). In addition, an international cooperating agency emailed countries information about scorer training (Coates & XX, personal communication, December 23, 2011).

The Country B team needed to hire ten scorers and identify a lead scorer (CAE, 2010, GS.23; CAE, 2010, GS.24). However, there is no information about the scorers that the Country B team hired for the study.

The Country B team received support from outside of the country when hiring scorers. The US organizing agency provided information about the number of scorers that each country needed to hire (CAE, 2010, GS.24; CAE, 2010, GS.23). The US organizing agency also provided scorer training (Coates & XX, personal communication, December 23, 2011).

Country C

During the initial February meeting in New York City teams requested information about scorers (CAE, 2010, GS.26). The following month the US organizing agency provided information. The agency suggested that—from June 2011 through 2012—Country C hire a total of ten scorers—five scorers per task (CAE, 2010, GS.24). In addition to the actual scoring work, Country C needed to pay for each scorer to attend two days of training and six hours for recalibration (CAE, 2010, GS.24). In March 2010, the US organizing agency also provided information on the lead scorer (CAE, 2010, GS.23). In October 2010 the US organizing agency sent out another document reminding countries about scorer recruitment and training (CAE, 2010, GS.45). Countries had over two months notice prior to lead scorer training scheduled to take place during the general meeting in March 2012 (Coates & XX, personal communication, December 23, 2011). There is no indication that either the Country C team or the US organizing agency had difficulty due to untimely communication.

According to the guidelines that the US organizing agency provided, the Country C team was to hire five scorers per task, for a total ten scorers, as well as a lead scorer (CAE, 2010, GS.24). However, the US agency did not recommend qualifications for hiring the scorers (CAE, 2010, GS.24). Scorers Country C hired would gain measurement knowledge through training. The scorers were to train on the performance task expectations and scoring and on recalibration (CAE, 2010, GS.45).

The US organizing agency did not provide recommended qualifications that the Country C team was to use when hiring scorers (CAE, 2010, GS.24). There is no information about Country C's scorers; there is no indication that Country C scorers had expertise in translation.

Each country's national project managers had to coordinate with the US organizing agency, scorers, and other team members in order to hire and train scorers (CAE, 2010, GS.24; CAE, 2010, GS.23). Country C's NPM was able to coordinate all of the steps involved and complete the work associated with scorers (Keeley, personal communication, April 13, 2012).

There were no specific review opportunities while country teams hired scorers. However, scorers were to complete training that included reviewing their progress in their scoring abilities (CAE, 2010, GS.24; CAE, 2010, GS.23). Country C's scorers completed all required tasks (Keeley, personal communication, April 13, 2012).

Country C team members had opportunities to train on the scoring process. In addition, the scorers they hired for the AHELO study also had training available. Country C team members took part in initial scorer training that occurred during a general meeting in Paris the third week of November 2011 (Coates & XX, personal communication, December 23, 2011). In addition, the US organizing agency asked that scorers participate in two days of scoring training and six hours of recalibration (CAE, 2010, GS.24). The Country C scorers participated in all training suggested (Keeley, personal communication, April 13, 2012).

The Country C team did not have opportunities to document the hiring of scorers for the AHELO study.

Country C team members participated in scorer training that took place during the general Paris meeting the third week of November 2011. In addition, the Country C team hired scorers and they completed two days of training and six hours for recalibration that would take place through 2012 (Coates & XX, personal communication, December 23, 2011; Keeley, personal communication, April 13, 2012).

The Country C team did not give any indication that they found the documents difficult or challenging. The US organizing agency provided two documents addressing hiring scorers. One document provided information on hiring scorers (CAE, 2010, GS.24). A second document included information regarding the lead scorer (CAE, 2010, GS.23). In addition, an international cooperating agency emailed countries information about scorer training (Coates & XX, personal communication, December 23, 2011).

The Country C team members needed to hire ten scorers and identify a lead scorer (CAE, 2010, GS.23; CAE, 2010, GS.24). However, there is no information about the scorers that the Country C team hired for the study.

The Country C team received support from outside of the country when hiring scorers. The US organizing agency provided information about the number of scorers that each country needed to hire (CAE, 2010, GS.24; CAE, 2010, GS.23). The US organizing agency also provided scorer training (Coates & XX, personal communication, December 23, 2011).

Country D

During the initial February meeting in New York City teams expressed a lack of clarity about scorer hiring and responsibility (CAE, 2010, GS.26). Within a month the US organizing agency provided information. From June 2011 through 2012, Country D needed to hire a total of ten scorers—five scorers per task (CAE, 2010, GS.24). The US agency also informed Country D that in addition to the actual scoring work, they needed to pay for each scorer to attend two days of training and six hours for recalibration (CAE, 2010, GS.24). In March 2010, the US organizing agency also provided information on the lead scorer (CAE, 2010, GS.23). In October 2010 the US organizing agency sent out another document reminding countries about scorer recruitment and training (CAE, 2010, GS.45). Country D had over two months notice prior to lead scorer training scheduled to take place during the general meeting in March 2012 (Coates & XX, personal communication, December 23, 2011). There is no indication that either the Country D team or the US organizing agency had difficulty due to untimely communication.

The US organizing agency provided guidelines for hiring scorers; the agency recommended that Country D hire five scorers per task, which would total ten scorers, as well as a lead scorer (CAE, 2010, GS.24). However, the US agency did not recommend qualifications for hiring the scorers (CAE, 2010, GS.24). Scorers would gain measurement knowledge through training. The scorers would train on the performance task expectations and scoring and on recalibration (CAE, 2010, GS.45).

The US organizing agency did not recommended translation qualifications that the Country D team was to use when hiring scorers (CAE, 2010, GS.24). There is no information

about Country D's scorers; there is no indication that Country A scorers had expertise in translation.

Each country's national project managers had to coordinate with the US organizing agency, scorers, and other team members in order to hire and train scorers (CAE, 2010, GS.24; CAE, 2010, GS.23). Country D's NPM was able to coordinate all of the steps involved and complete the work associated with scorers (Keeley, personal communication, April 13, 2012).

There were no specific review opportunities while country teams hired scorers. However, scorers were to complete training that included reviewing their progress in their scoring abilities (CAE, 2010, GS.24; CAE, 2010, GS.23). Country D's scorers completed all required tasks (Keeley, personal communication, April 13, 2012).

The Country D team, and scorers that they hired, had opportunities to train for the AHELO study's scoring process. Country D team members took part in initial scorer training that occurred during a general meeting in Paris the third week of November 2011 (Coates & XX, personal communication, December 23, 2011). Also, following the guidelines created by the US organizing agency, Country D scorers participated in additional scorer training that was scheduled to take place prior to test administration in 2012 (Keeley, personal communication, April 13, 2012).

The Country D team did not have opportunities to document the hiring of scorers for the AHELO study.

Country D team members were able to participate in scorer training that took place during the general Paris meeting the third week of November 2011. In addition, Country D hired scorers and they completed two days of training and six hours for recalibration that would take place through 2012 (Coates & XX, personal communication, December 23, 2011; Keeley, personal communication, April 13, 2012).

The Country D team did not give any indication that they found the documents difficult or challenging. The US organizing agency provided two documents addressing hiring scorers. One document provided information on hiring scorers (CAE, 2010, GS.24). A second document included information regarding the lead scorer (CAE, 2010, GS.23). In addition, an international cooperating agency emailed countries information about scorer training (Coates & XX, personal communication, December 23, 2011).

The Country D team needed to hire ten scorers and identify a lead scorer (CAE, 2010, GS.23; CAE, 2010, GS.24). However, there is no information about the scorers that the Country D team hired for the study.

The Country D team received support from outside of the country when hiring scorers. The US organizing agency provided information about the number of scorers that each country needed to hire (CAE, 2010, GS.24; CAE, 2010, GS.23). The US organizing agency also provided scorer training (Coates & XX, personal communication, December 23, 2011).

Country E

During the initial February meeting in New York City teams requested information about scorers (CAE, 2010, GS.26). The following month the US organizing agency provided information. The agency suggested that—from June 2011 through 2012—Country E hire a total of ten scorers—five scorers per task (CAE, 2010, GS.24). In addition to the actual scoring work, Country E needed to pay for each scorer to attend two days of training and six hours for recalibration (CAE, 2010, GS.24). In March 2010, the US organizing agency also provided information on the lead scorer (CAE, 2010, GS.23). In October 2010 the US organizing agency sent out another document reminding countries about scorer recruitment and training (CAE, 2010, GS.45). Countries had over two months notice prior to lead scorer training scheduled to take place during the general meeting in March 2012 (Coates & XX, personal communication, December 23, 2011). There is no indication that either the Country E team or the US organizing agency had difficulty due to untimely communication.

The US organizing agency recommended that the Country E team hire five scorers per task, which would total ten scorers, as well as a lead scorer (CAE, 2010, GS.24). However, the US agency did not recommend qualifications for hiring the scorers (CAE, 2010, GS.24). Scorers would gain measurement knowledge through training. The scorers would train on the performance task expectations and scoring and on recalibration (CAE, 2010, GS.45).

The US organizing agency did not provide recommended qualifications that the Country E team was to use when hiring scorers (CAE, 2010, GS.24). There is no information about Country E's scorers; there is no indication that Country E scorers had expertise in translation.

Each country's national project managers had to coordinate with the US organizing agency, scorers, and other team members in order to hire and train scorers (CAE, 2010, GS.24; CAE, 2010, GS.23). Country E's NPM was able to coordinate all of the steps involved and complete the work associated with scorers (Keeley, personal communication, April 13, 2012).

There were no specific review opportunities while country teams hired scorers. However, scorers were to complete training that included reviewing their progress in their scoring abilities (CAE, 2010, GS.24; CAE, 2010, GS.23). Country E's scorers completed all required tasks (Keeley, personal communication, April 13, 2012).

There were training opportunities for Country E team members and scorers that they hired. Country E team members took part in initial scorer training that occurred during a general meeting in Paris the third week of November 2011 (Coates & XX, personal communication, December 23, 2011). Also, according to guidelines created by the US organizing agency, each scorer would participate in two days of scoring training and six hours of recalibration (CAE, 2010, GS.24). Country E scorers participated in additional scorer training that was scheduled to take place prior to test administration in 2012 (Keeley, personal communication, April 13, 2012).

The Country E team did not have opportunities to document the hiring of scorers for the AHELO study.

Country E team members were able to participate in scorer training that took place during the general Paris meeting the third week of November 2011. In addition, Country E hired scorers

and they completed two days of training and six hours for recalibration that would take place through 2012 (Coates & XX, personal communication, December 23, 2011; Keeley, personal communication, April 13, 2012).

The Country E team did not give any indication that they found the documents difficult or challenging. The US organizing agency provided two documents addressing hiring scorers. One document provided information on hiring scorers (CAE, 2010, GS.24). A second document included information regarding the lead scorer (CAE, 2010, GS.23). In addition, an international cooperating agency emailed countries information about scorer training (Coates & XX, personal communication, December 23, 2011).

The Country E team needed to hire ten scorers and identify a lead scorer (CAE, 2010, GS.23; CAE, 2010, GS.24). However, there is no information about the scorers that the Country E team hired for the study.

The Country E team received support from outside of the country when hiring scorers. The US organizing agency provided information about the number of scorers that each country needed to hire (CAE, 2010, GS.24; CAE, 2010, GS.23). The US organizing agency also provided scorer training (Coates & XX, personal communication, December 23, 2011).

Task 16: Attend kick-off meeting as well as in-person and phone meeting to discuss progress.

Country A

Country A experienced timely communication with the US organizing agency and international organizing agency. With almost two weeks notice, Country A's national project manager participated in an initial conference call on January 14, 2010 (Shavelson, personal communication, January 11, 2010). Country A also participated in a conference call on September 27, 2010 and had over a week to plan for it (CAE, personal communication, September 22, 2010). Country A also participated in in-person meetings. Several team members attended the meeting in New York City that took place on February 15-18, 2010 after two months notice (Kurpius, personal communication, February 12, 2010). The team also had over a month to prepare for the AHELO Generic Strand Meeting in Paris on March 17 (Shavelson & Kurpius, personal communication, March 11, 2010). The national project manager also attended the AHELO NPM meeting in Paris on October 27-28, 2010 (Unknown, 2010, Participants; CAE, personal communication, September 30, 2010). The NPM also participated in a meeting that took place over November 23, 24, 25, 2011 (ACER, 2010, November). In addition, the Country A national project manager also attended the NPM meeting on March 29, 2011 (Tremblay, personal communication, February 18, 2011). Finally, one of the Country A translators participated in a conference call to address rubrics (Chia, 2011, Rubric).

Country A's national project manager had experience in higher education assessment practices (Ursin, 2010, CV). In addition, the Country A team included a measurement expert who specialized in statistics (CAE, 2010, GS.11). The NPM attended all but one meeting that specifically required the participation of one member of the translation team (Chia, 2010, Rubric). In addition, the measurement expert joined the NPM several meetings.

Translation expertise was not required for most meetings. One meeting, which dealt with translation of the rubric, required the input of at least one of the team's translators. One of Country A's translators, who possessed translation expertise, was able to participate in the meeting (Chia, 2010, Rubric).

Country teams were to participate in several meetings: in-person as well as conference calls. Country A's national project manager used project management expertise to prepare for and participate in every scheduled meeting throughout the study.

Country A did not have any review opportunities while attending in-person and telephone call meetings.

There was no need, or scheduled opportunities, to provide training for Country A team members to participate in telephone or in-person meetings.

There was documentation of several meetings. Information from Country A's initial conference call on January 14, 2010 was captured in an email (Shavelson, personal communication, January 11, 2010). The agenda for Country A's second conference call on September 27, 2010 was also documented in an email (CAE, personal communication, September 22, 2010). All information addressed in the meeting in the February 2010 New York City was captured in meeting minutes (CAE, 2010, GS.26). The agendas for the AHELO generic strand meetings in Paris were also captured in email (Shavelson & Kurpius, personal communication, March 11, 2010; CAE, personal communication, September 30, 2010). The international agency in charge of the NPM meeting that took place over November 23, 24, 25, 2011 captured information in a document (ACER, 2010, November). Finally, the interviewer took notes during the conference call to address rubrics (Chia, 2011, Rubric).

Country A team members did not indicate that they found the meeting dates challenging. Country A's national project manager participated in an initial conference call on January 14, 2010 and a conference call on September 27, 2010 (Shavelson, personal communication, January 11, 2010; CAE, personal communication, September 22, 2010). Country A also attended the initial kick-off meeting, a generic strand meeting in March, and the NPM meetings (Kurpius, personal communication, February 12, 2010; Shavelson & Kurpius, personal communication, March 11, 2010; CAE, personal communication, September 30, 2010; ACER, 2010, November; Tremblay, personal communication, February 18, 2011). Finally, one of the Country A translators participated in a conference call to address rubrics (Chia, 2011, Rubric).

The US organizing agency emailed Country A team members when planning meetings (Kurpius, personal communication, February 12, 2010; Shavelson & Kurpius, personal communication, March 11, 2010; CAE, personal communication, September 30, 2010; Tremblay, personal communication, February 18, 2011). The Country A team did not indicate that they found the emails challenging to follow.

During the official meetings scheduled by the organizing agencies, only the presence of team members was required. Country A did not need to acquire in-country support external to the team.

The US organizing agency coordinated the vast majority of meetings. The US agency provided Country A team members with the meeting agenda and logistical information for the initial meeting in New York (Shavelson, personal communication, January 11, 2010; Kurpius, personal communication, February 12, 2010). The agency also provided the Country A team with conference call lines and agendas for each telephone meeting (CAE, personal communication, September 22, 2010; Kurpius, personal communication February 12, 2010). The international coordinating agency also assisted Country A with meeting coordination and agendas (ACER, 2010, November; Tremblay, personal communication, February 18, 2011).

Country B

Country B participated in several in-person and telephone meetings during its participation in the AHELO study. Two members of the Country B team attended the initial meeting in New York City on February 15-18, 2010; they had over one month's notice (Kurpius, personal communication, February 12, 2010). Team members also participated in the AHELO Generic Strand Meeting at that took place in Paris on March 17 (Shavelson & Kurpius, personal communication, March 11, 2010). The team's national project manager also participated in the AHELO NPM meeting in Paris that took place on October 27-28, 2010 (Unknown, 2010, Participants; CAE, personal communication, September 30, 2010). The team's national project manager also attended the NPM meeting that occurred on November 23, 24, 25, 2011 (ACER, 2010, November). Team member also participated in conference calls on January 18, 2010 and September 28, 2010 (CAE, personal communication, September 22, 2010). Communication between the Country B team and other agencies seemed timely.

The Country B team included a measurement expert who specialized in assessment and statistics (CAE, 2010, GS.11). The expert participated in some of the meetings. The expert attended the initial meeting in New York City on February 15-18, 2010 (CAE, 2010, GS.26). The team member also participated in the AHELO Generic Strand Meeting that took place in Paris on March 17 (Shavelson & Kurpius, personal communication, March 11, 2010). Team member also participated in conference calls on January 18, 2010 and September 28, 2010 (CAE, personal communication, September 22, 2010). In addition, the team's measurement expert participated in the AHELO NPM meeting in Paris that took place on October 27-28, 2010 (Unknown, 2010, Participants; CAE, personal communication, September 30, 2010) and the NPM meeting that occurred on November 23, 24, 25, 2011 (ACER, 2010, November).

Translation expertise was not required for most meetings. One meeting, which dealt with translation of the rubric, required the input of at least one of the team's translators. The Country B team did not participate in the meeting.

Country teams were to participate in several meetings: in-person as well as conference calls. Country B's national project manager used project management expertise to prepare for and participate in most scheduled meetings throughout the study.

Country B did not have any review opportunities while attending in-person and telephone call meetings.

There was no need, or scheduled opportunities, to provide training for Country B team members to participate in telephone or in-person meetings.

Staff from the US organizing agency took notes during the entire meeting that took place in New York City on February 15-18, 2010 (CAE, 2010, GS.26). Information about the AHELO generic strand meeting at that took place in Paris on March 17 was captured in an email (Shavelson & Kurpius, personal communication, March 11, 2010). Information about the AHELO NPM meeting in Paris that took place on October 27-28, 2010 was also captured via email (CAE, personal communication, September 30, 2010). The international coordinating agency created a document with information regarding the NPM meeting that took place on November 23, 24, 25, 2011 (ACER, 2010, November). Information about conference calls that occurred on January 18, 2010 and September 28, 2010 was captured in email (CAE, personal communication, September 22, 2010).

Country B did not find the meeting dates challenging. Team members attended the initial meeting in New York City (CAE, 2010, GS.26). Team members also participated in the AHELO generic strand meeting on March 17, the AHELO NPM meeting that took place on October 27-28, 2010, and the NPM meeting that occurred on November 23, 24, 25, 2011 (Shavelson & Kurpius, personal communication, March 11, 2010; CAE, personal communication, September 30, 2010; ACER, 2010, November). Team member also participated in conference calls on January 18, 2010 and September 28, 2010 (CAE, personal communication, September 22, 2010).

The US organizing agency emailed Country B team members when planning meetings (Kurpius, personal communication, February 12, 2010; Shavelson & Kurpius, personal communication, March 11, 2010; CAE, personal communication, September 30, 2010; Tremblay, personal communication, February 18, 2011). The Country B team did not indicate that they found the emails challenging to follow.

During the official meetings scheduled by the organizing agencies, only the presence of team members was required. Country B did not need to acquire in-country support external to the team.

The US organizing agency coordinated the vast majority of meetings. The US agency provided Country B team members with the meeting agenda and logistical information for the initial meeting in New York (Shavelson, personal communication, January 11, 2010; Kurpius, personal communication, February 12, 2010). The agency also provided the Country B team with conference call lines and agendas for each telephone meeting (CAE, personal communication, September 22, 2010; Kurpius, personal communication February 12, 2010). The international coordinating agency also assisted Country B with meeting coordination and agendas (ACER, 2010, November; Tremblay, personal communication, February 18, 2011).

Country C

The Country C team participated in several meetings that took place in-person and over the telephone. Two team members participated in the initial New York City meeting that took place on February 15-18, 2010 (Kurpius, personal communication, February 12, 2010). The team

also attended the AHELO generic strand meeting the occurred in Paris on March 17 (Shavelson & Kurpius, personal communication, March 11, 2010). The team's national project manager attended the AHELO NPM meeting in Paris—October 27-28, 2010 (Unknown, 2010, Participants; CAE, personal communication, September 30, 2010). Finally, the national project manager attended the NPM meeting on November 23, 24, 25, 2011 (ACER, 2010, November). The national project manager also participated in a conference call on January 15, 2010. There is no indication that there was a challenge with timely communication.

The Country C team participated in several meetings over the course of the study. However, there is no evidence that a measurement expert was present at any of them.

Translation expertise was not required for most meetings. One meeting, which dealt with translation of the rubric, required the input of at least one of the team's translators. Although the Country C NPM participated, he did not possess translation expertise (Chia, 2010, Rubric). Country teams were to participate in several meetings: in-person as well as conference calls. Country B's national project manager used project management expertise to prepare for and participate in most scheduled meetings throughout the study.

Country teams were to participate in several meetings: in-person as well as conference calls. Country B's national project manager used project management expertise to prepare for and participate in most scheduled meetings throughout the study.

Country C did not have any review opportunities while attending in-person and telephone call meetings.

There was no need, or scheduled opportunities, to provide training for Country C team members to participate in telephone or in-person meetings.

Staff from the US organizing agency took notes during the entire meeting that took place in New York City on February 15-18, 2010 (CAE, 2010, GS.26). Information regarding the AHELO generic strand meeting that occurred in Paris on March 17 was captured via email (Shavelson & Kurpius, personal communication, March 11, 2010). Information about the AHELO NPM meeting in Paris—October 27-28, 2010 was addressed in an email by the US organizing agency (CAE, personal communication, September 30, 2010). The international coordinating agency captured information about the NPM meeting on November 23, 24, 25, 2011 (ACER, 2010, November).

Country C team members did not find the meeting dates challenging. Team members attended the initial meeting in New York City (CAE, 2010, GS.26). Team members also participated in the AHELO generic strand meeting on March 17, the AHELO NPM meeting that took place on October 27-28, 2010, and the NPM meeting that occurred on November 23, 24, 25, 2011 (Shavelson & Kurpius, personal communication, March 11, 2010; CAE, personal communication, September 30, 2010; ACER, 2010, November). The NPM also participated in a conference call on January 15, 2010.

The US organizing agency emailed Country C team members when planning meetings (Kurpius, personal communication, February 12, 2010; Shavelson & Kurpius, personal

communication, March 11, 2010; CAE, personal communication, September 30, 2010; Tremblay, personal communication, February 18, 2011). The Country C team did not indicate that they found the emails challenging to follow.

During the official meetings scheduled by the organizing agencies, only the presence of team members was required. Country C did not need to acquire in-country support external to the team.

The US organizing agency coordinated the vast majority of meetings. The US agency provided Country C team members with the meeting agenda and logistical information for the initial meeting in New York (Shavelson, personal communication, January 11, 2010; Kurpius, personal communication, February 12, 2010). The agency also provided the Country C team with conference call lines and agendas for each telephone meeting (CAE, personal communication, September 22, 2010; Kurpius, personal communication February 12, 2010). The international coordinating agency also assisted Country C with meeting coordination and agendas (ACER, 2010, November; Tremblay, personal communication, February 18, 2011).

Country D

There was constant timely communication between the Country D team and organizing agencies. Several team members attended the initial New York City meeting on February 15-18, 2010 (Kurpius, personal communication, February 12, 2010). Team members also participated in the AHELO generic strand meeting that took place in Paris on March 17 (Shavelson & Kurpius, personal communication, March 11, 2010). The national project manager also attended the AHELO NPM meeting in Paris on October 27-28, 2010 (Unknown, 2010, Participants; CAE, personal communication, September 30, 2010). Finally, the national project manager participated in the NPM meeting that occurred November 23, 24, 25, 2011 (ACER, 2010, November). Country D also participated in conference calls. The team communicated with the US organizing agency over the telephone on January 18, 2010, July 27, 2010, and September 29, 2010 (CAE, personal communication, September 22, 2010; Shavelson, personal communication, July 26, 2010). Finally, one of the Country D translators participated in a conference call to address rubrics (Chia, 2011, Rubric).

Country D included a measurement expert as part of their core team The measurement expert participated in several meetings. The expert attended the initial New York City meeting on February 15-18, 2010 (Kurpius, personal communication, February 12, 2010). In addition, the measurement expert attended all but one telephone meetings, which only required a translation experts. The measurement expert communicated with the US organizing agency over the telephone on January 18, 2010, July 27, 2010, and September 29, 2010 (CAE, personal communication, September 22, 2010; Shavelson, personal communication, July 26, 2010).

Translation expertise was not required for most meetings. One meeting, which dealt with translation of the rubric, required the input of at least one of the team's translators. One of Country D's translators was able to participate in the meeting; however, the person who helped

the team with the translation did not have professional translation expertise, certification, or academic background (Chia, 2010, Rubric).

Country teams were to participate in several meetings: in-person as well as conference calls. Country D's national project manager used project management expertise to prepare for and participate in every scheduled meeting throughout the study.

Country D did not have any review opportunities while attending in-person and telephone call meetings.

There was no need, or scheduled opportunities, to provide training for Country D team members to participate in telephone or in-person meetings.

Staff from the US organizing agency took notes during the entire meeting that took place in New York City on February 15-18, 2010 (CAE, 2010, GS.26). Information regarding the AHELO generic strand meeting that occurred in Paris on March 17 was captured via email (Shavelson & Kurpius, personal communication, March 11, 2010). Information about the AHELO NPM meeting in Paris—October 27-28, 2010 was addressed in an email by the US organizing agency (CAE, personal communication, September 30, 2010). The international coordinating agency captured information about the NPM meeting on November 23, 24, 25, 2011 (ACER, 2010, November). The US organizing agency captured information about telephone conference calls that took place on January 18, 2010, July 27, 2010, and September 29, 2010 via email (CAE, personal communication, September 22, 2010; Shavelson, personal communication, July 26, 2010). Finally, the US organizing agency's staff took meeting minutes during a conference call in the fall of 2011 (Chia, 2011, Rubric).

Country D team members did not find the meeting dates challenging. Team members attended the initial meeting in New York City (CAE, 2010, GS.26). Team members also participated in the AHELO generic strand meeting on March 17, the AHELO NPM meeting that took place on October 27-28, 2010, and the NPM meeting that occurred on November 23, 24, 25, 2011 (Shavelson & Kurpius, personal communication, March 11, 2010; CAE, personal communication, September 30, 2010; ACER, 2010, November). In addition, the team communicated with the US organizing agency over the telephone on January 18, 2010, July 27, 2010, and September 29, 2010 (CAE, personal communication, September 22, 2010; Shavelson, personal communication, July 26, 2010). Also, one of the Country D translators participated in a conference call to address rubrics (Chia, 2011, Rubric).

The US organizing agency emailed Country D team members when planning meetings (Kurpius, personal communication, February 12, 2010; Shavelson & Kurpius, personal communication, March 11, 2010; CAE, personal communication, September 30, 2010; Tremblay, personal communication, February 18, 2011). The Country D team did not indicate that they found the emails challenging to follow.

During the official meetings scheduled by the organizing agencies, only the presence of team members was required. Country D did not need to acquire in-country support external to the team.

The US organizing agency coordinated the vast majority of meetings. The US agency provided Country D team members with the meeting agenda and logistical information for the initial meeting in New York (Shavelson, personal communication, January 11, 2010; Kurpius, personal communication, February 12, 2010). The agency also provided the Country D team with conference call lines and agendas for each telephone meeting (CAE, personal communication, September 22, 2010; Kurpius, personal communication February 12, 2010). The international coordinating agency also assisted Country D with meeting coordination and agendas (ACER, 2010, November; Tremblay, personal communication, February 18, 2011).

Country E

The Country E team and organizing agencies exchanged timely communication. Country E team members participated in the initial New York City meeting on February 15-18, 2010 (Kurpius, personal communication, February 12, 2010). The team also participated in the AHELO generic strand meeting that occurred in Paris on March 17 (Shavelson & Kurpius, personal communication, March 11, 2010). Country E's national project manager also attended the AHELO NPM meeting in Paris scheduled for October 27-28, 2010 (Unknown, 2010, Participants; CAE, personal communication, September 30, 2010). Finally, the country's national project manager also attended the NPM meeting that took place on November 23, 24, 25, 2011 (ACER, 2010, November). The Country E team also part in telephone conference calls. Country E team members spoke with US representative in January 2010 and on September 27, 2010 (CAE, personal communication, September 22, 2010).

The national project manager for Country E had some experience in large-scale assessment studies. Also, the Country E team included a measurement expert as part of the core group involved in the study. The Country E team measurement expert participated in the initial New York City meeting on February 15-18, 2010 (Kurpius, personal communication, February 12, 2010). The team member also participated in the AHELO generic strand meeting that occurred in Paris on March 17 (Shavelson & Kurpius, personal communication, March 11, 2010). Country E's national project manager also attended the AHELO NPM meeting in Paris scheduled for October 27-28, 2010 (Unknown, 2010, Participants; CAE, personal communication, September 30, 2010). Finally, the country's national project manager also attended the NPM meeting that took place on November 23, 24, 25, 2011 (ACER, 2010, November). The Country E team also part in telephone conference calls. Country E team members spoke with US representative in January 2010 and on September 27, 2010 (CAE, personal communication, September 22, 2010).

Translation expertise was not required for most meetings. One meeting, which dealt with translation of the rubric, required the input of at least one of the team's translators. Country E had contracted translators from an external translation company; therefore, Country E was not able to participate in the meeting.

Country teams were to participate in several meetings: in-person as well as conference calls. Country E's national project manager used project management expertise to prepare for and participate in most scheduled meetings throughout the study.

Country E did not have any review opportunities while attending in-person and telephone call meetings.

There was no need, or scheduled opportunities, to provide training for Country E team members to participate in telephone or in-person meetings.

Staff from the US organizing agency took notes during the entire meeting that took place in New York City on February 15-18, 2010 (CAE, 2010, GS.26). Information regarding the AHELO generic strand meeting that occurred in Paris on March 17 was captured via email (Shavelson & Kurpius, personal communication, March 11, 2010). Information about the AHELO NPM meeting in Paris—October 27-28, 2010 was addressed in an email by the US organizing agency (CAE, personal communication, September 30, 2010). The international coordinating agency captured information about the NPM meeting on November 23, 24, 25, 2011 (ACER, 2010, November). The US organizing agency communicated information about conference call via email (CAE, personal communication, September 22, 2010).

Country E team members did not find the meeting dates challenging. Team members attended the initial meeting in New York City (CAE, 2010, GS.26). Team members also participated in the AHELO generic strand meeting on March 17, the AHELO NPM meeting that took place on October 27-28, 2010, and the NPM meeting that occurred on November 23, 24, 25, 2011 (Shavelson & Kurpius, personal communication, March 11, 2010; CAE, personal communication, September 30, 2010; ACER, 2010, November). The Country E team also part in telephone conference calls on January 2010 and on September 27, 2010 (CAE, personal communication, September 22, 2010).

The US organizing agency emailed Country E team members when planning meetings (Kurpius, personal communication, February 12, 2010; Shavelson & Kurpius, personal communication, March 11, 2010; CAE, personal communication, September 30, 2010; Tremblay, personal communication, February 18, 2011). The Country E team did not indicate that they found the emails challenging to follow.

During the official meetings scheduled by the organizing agencies, only the presence of team members was required. Country E did not need to acquire in-country support external to the team.

The US organizing agency coordinated the vast majority of meetings. The US agency provided Country E team members with the meeting agenda and logistical information for the initial meeting in New York (Shavelson, personal communication, January 11, 2010; Kurpius, personal communication, February 12, 2010). The agency also provided the Country E team with conference call lines and agendas for each telephone meeting (CAE, personal communication, September 22, 2010; Kurpius, personal communication February 12, 2010). The international

coordinating agency also assisted Country E with meeting coordination and agendas (ACER, 2010, November; Tremblay, personal communication, February 18, 2011).

Task 17: Submit feedback on process and technical reports regarding progress.

Country A

Country A's team members had multiple opportunities to provide feedback. Communication about feedback between the organizing agencies and the Country A team members was usually timely. The US agency gave Country A opportunities to provide feedback on the study's workplan and specific activities throughout the four-day meeting that took place in New York (CAE, 2010, GS.26). The US agency also asked Country A for feedback on the entire translation and adaptation process during the last conference call of 2010 (Shavelson, 2010, End). Country teams also had an opportunity to provide feedback on translation review and validation procedures throughout the two-day site visits (Solano-Flores, 2010, Visit). In addition, during an interview conducted during the site visit, Country A's team and translation team were asked for feedback regarding team selection and translation and adaptation completed (Solano-Flores & Chia, 2010, Interview). Also, an international agency working on the project requested country feedback. The international agency created an online exchange where they posted all documents created for the project; Country A's national project manager was given access to the exchange and asked to provide feedback regarding documents used in 2010 (Coates, personal communication, December 20, 2010). For 2011, the international agency created a wiki/blog through which Country A's team members could provide feedback on the study's progress and exchange information (Coates, personal communication, December 20, 2011). The international coordinating agency had also provided a format for NPM diaries that Country A was to use to provide feedback about progress in 2010 (CAE, personal communication, February 11, 2011). However, Country A did not find out about this until one year after the project had begun (CAE, personal communication, February 11, 2011). Country A experienced similar challenge with timing of communication during an interview. During the fall of 2011, a US agency's staff member interviewed a member of Country A's translation team asking for feedback specific to the scoring rubric (Chia, 2010, Rubric). Although almost one year had passed since the translation work was completed, the interviewee was able to answer questions in great detail based on his notes from the process (Chia, 2010, Rubric).

Country A's team national project manager had experience with assessment practices in higher education (Ursin, 2010, CV). In addition, one of Country A's team members had extensive training and experience in measurement (CAE, 2010, GS.11). The Country A team's measurement expert provided feedback during the initial meeting in New York and the general team meetings hosted by the international coordinating agency; the NPM provided feedback during every scheduled meeting (CAE, 2010, GS.26; Shavelson, 2010, End; Solano-Flores, 2010, Visit). Although the international agency created an online exchange for Country A's national project manager to provide feedback regarding documents used in 2010 there is no information about Country A's contribution (Coates, personal communication, December 20,

2010). Similarly, there is no information about Country A's feedback regarding 2011 progress on the wiki/blog that the international agency created (Coates, personal communication, December 20, 2011). Since it was not required, there was no measurement expertise provided during the rubric interview (Chia, 2010, Rubric).

The US agency asked Country A for feedback on the entire translation and adaptation process during the last conference call of 2010 (Shavelson, 2010, End). However, there was no translation expert included in the call. Country A's translation team had an opportunity to provide feedback on translation review and validation procedures throughout the two-day site visits (Solano-Flores, 2010, Visit). In addition, the translation team participated in the interview conducted during the site visit that asked for feedback regarding team selection and translation and adaptation completed (Solano-Flores & Chia, 2010, Interview). Finally, a US agency staff member got feedback specific to the scoring rubric from a member of Country A's translation team (Chia, 2010, Rubric).

Country A used project management expertise to provide feedback regarding the study's process and material. The NPM coordinated travel so that team members could attend the four-day meeting that took place in New York and provide feedback on the study's work plan and specific activities (CAE, 2010, GS.26). The NPM also coordinated with the US agency to provide feedback on the entire translation and adaptation process during the last conference call of 2010 (Shavelson, 2010, End). Country A's NPM also coordinated the two-day site visit so that the team and translators provided feedback on translation review and validation procedures (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010, Interview). In addition, the Country A NPM helped coordinate a translator's participation in an interview that addressed the scoring rubric (Chia, 2010, Rubric).

Country A did not have any review opportunities while submitting feedback.

Country A did not have any training opportunities while submitting feedback.

In several reports to the international coordinating agency, the US organizing agency documented Country A's feedback. The US agency wrote a progress report that included Country A's questions and concerns regarding the item sample size and intended constructs as well as the country team's progress in selecting performance tasks for the study, understanding the translation and adaptation framework, and student sampling (CAE, 2010, GS.30). The agency also documented Country A's feedback regarding translation, adaptation, translation team qualification, and finalizing translated performance tasks based on cognitive lab results (CAE, 2010, Module A). The US agency also document Country A's feedback regarding finalizing the translation team, the translation review team, materials create for translation and translation review, site visits, and meeting schedule (CAE, 2010, Progress). In another report the US organizing agency documented the Country A team's feedback regarding telecommunication meetings, scoring, recruiting staff to assist country teams, recruitment of translation team, site visits, sampling, test administration, and material created throughout the study (CAE, 2011, Module A).

The Country A team was able to provide the US organizing agency feedback about the study's procedures, materials, and required resources (CAE, 2010, Module A; CAE, 2011, Module A). There is no evidence that Country A's team members had difficulty with the deadlines provided for submitting feedback.

The Country A team was able to provide the US organizing agency feedback about the study's procedures, materials, and required resources through meetings, conference calls, and emails (CAE, 2010, Module A; CAE, 2011, Module A; CAE, 2010, GS.26; Shavelson & Kurpius, 2010, End). There were no materials for providing feedback.

At times, Country A required in-country support external to the team when providing feedback about the study. In addition to Country A's team members, the Country A translation team shared feedback during an interview conducted during the site visit and throughout the on site training (Solano-Flores & Chia, 2010, Interview; Solano-Flores, 2010, Visit). In addition, one of the translation team members provided feedback regarding scoring rubric during an interview (Chia, 2010, Rubric).

Country A's team members did not require support external to the team and outside of the country to provide feedback about the study.

Country B

Country B's team members had multiple opportunities to provide feedback. Communication about feedback between the organizing agencies and the Country B team members was usually timely. The US agency gave Country B opportunities to provide feedback on the study's workplan and specific activities throughout the four-day meeting that took place in New York (CAE, 2010, GS.26). The US agency also asked Country B for feedback on the entire translation and adaptation process during the last conference call of 2010 (Shavelson, 2010, End). Country teams also had an opportunity to provide feedback on translation review and validation procedures throughout the two-day site visits (Solano-Flores, 2010, Visit). In addition, during an interview conducted during the site visit, Country B's country team and translation team were asked for feedback regarding team selection and translation and adaptation completed (Solano-Flores & Chia, 2010, Interview). Also, an international agency working on the project requested country feedback. The international agency created an online exchange where they posted all documents created for the project; Country B's national project manager was given access to the exchange and asked to provide feedback regarding documents used in 2010 (Coates, personal communication, December 20, 2010). For 2011, the international agency created a wiki/blog through which Country B's team members could provide feedback on the study's progress and exchange information (Coates, personal communication, December 20, 2011). The international coordinating agency had also provided a format for NPM diaries that Country B was to use to provide feedback about progress in 2010 (CAE, personal communication, February 11, 2011). However, Country B did not find out about this until one year after the project had begun (CAE, personal communication, February 11, 2011). In addition,

although multiple attempts were made to include Country B in an interview regarding the scoring rubric, the team did not respond and was not able to participate (Chia, 2010, Rubric).

Country B had a measurement expert as part of the country team. Country B's measurement expert provided feedback throughout the four-day meeting that took place in New York (CAE, 2010, GS.26). The measurement expert also provided feedback on translation review and validation procedures throughout the two-day site visits (Solano-Flores, 2010, Visit). In addition, during an interview conducted during the site visit, Country B's measurement expert gave feedback regarding team selection and the translation and adaptation completed (Solano-Flores & Chia, 2010, Interview). There is no evidence that the Country B team's measurement expert had access to, or provided feedback on, the online exchange where the international agency posted all documents created for the project in 2010 (Coates, personal communication, December 20, 2010). Likewise, there is no indication that the measurement expert had access to, or participated in, the international agency's wiki/blog or NPM diaries through which Country B could provide feedback on the study's progress (Coates, personal communication, December 20, 2011). In addition, Country B did not participate in an interview regarding the scoring rubric (Chia, 2010, Rubric).

Country B did not obtain assistance from professional translators. Therefore, the team was not able to provide feedback that included translation expertise.

Country B demonstrated project management expertise when providing feedback regarding the study's process and material. The NPM coordinated travel so that team members could attend the four-day meeting that took place in New York and provide feedback on the study's work plan and specific activities (CAE, 2010, GS.26). The NPM also coordinated with the US agency to provide feedback on the entire translation and adaptation process during the last conference call of 2010 (Shavelson, 2010, End). Country B's NPM also coordinated the two-day site visit so that the team and translators provided feedback on translation review and validation procedures (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010, Interview). However, although multiple attempts were made to include Country B in an interview regarding the scoring rubric, the team did not respond and was not able to participate (Chia, 2010, Rubric).

Country B did not have any review opportunities while submitting feedback.

Country B did not have any training opportunities while submitting feedback.

In several reports to the international coordinating agency, the US organizing agency documented Country B's feedback. The US agency wrote a progress report that included Country B's questions and concerns regarding the item sample size and intended constructs as well as the country team's progress in selecting performance tasks for the study, understanding the translation and adaptation framework, and student sampling (CAE, 2010, GS.30). The agency also documented Country B's feedback regarding translation, adaptation, translation team qualification, and finalizing translated performance tasks based on cognitive lab results (CAE, 2010, Module A). The US agency also document Country B's feedback regarding finalizing the translation team, the translation review team, materials create for translation and translation

review, site visits, and meeting schedule (CAE, 2010, Progress). In another report the US organizing agency documented the Country B team's feedback regarding telecommunication meetings, scoring, recruiting staff to assist country teams, recruitment of translation team, site visits, sampling, test administration, and material created throughout the study (CAE, 2011, Module A).

The Country B team was able to provide the US organizing agency feedback about the study's procedures, materials, and required resources (CAE, 2010, Module A; CAE, 2011, Module A). There is no evidence that Country B's team members had difficulty with the deadlines provided for submitting feedback.

The Country B team was able to provide the US organizing agency feedback about the study's procedures, materials, and required resources through meetings, conference calls, and emails (CAE, 2010, Module A; CAE, 2011, Module A; CAE, 2010, GS.26; Shavelson & Kurpius, 2010, End). There were no materials for providing feedback.

At times, Country B required in-country support external to the team when providing feedback about the study. In addition to Country B's team members, the staff that the team hired to complete translations shared feedback during an interview conducted during the site visit and throughout the on site training (Solano-Flores & Chia, 2010, Interview; Solano-Flores, 2010, Visit).

Country B's team members did not require support external to the team and outside of the country to provide feedback about the study.

Country C

Country C's team members had multiple opportunities to provide feedback. Communication about feedback between the organizing agencies and the Country C team members was usually timely. The US agency gave Country C opportunities to provide feedback on the study's workplan and specific activities throughout the four-day meeting that took place in New York (CAE, 2010, GS.26). The US agency also asked Country C for feedback on the entire translation and adaptation process during the last conference call of 2010 (Shavelson, 2010, End). Country teams also had an opportunity to provide feedback on translation review and validation procedures throughout the two-day site visits (Solano-Flores, 2010, Visit). In addition, during an interview conducted during the site visit, Country C's team and translation team were asked for feedback regarding team selection and translation and adaptation completed (Solano-Flores & Chia, 2010, Interview). Also, an international agency working on the project requested country feedback. The international agency created an online exchange where they posted all documents created for the project; Country C's national project manager was given access to the exchange and asked to provide feedback regarding documents used in 2010 (Coates, personal communication, December 20, 2010). For 2011, the international agency created a wiki/blog through which Country C's team members could provide feedback on the study's progress and exchange information (Coates, personal communication, December 20, 2011). The international coordinating agency had also provided a format for NPM diaries that Country C was to use to

provide feedback about progress in 2010 (CAE, personal communication, February 11, 2011). However, Country C did not find out about this until one year after the project had begun (CAE, personal communication, February 11, 2011). During the fall of 2011, a US agency's staff member interviewed a member of Country C's country team asking for feedback specific to the scoring rubric (Chia, 2010, Rubric). Due to the amount of time that had passed, members of the translation team were not available and the new NPM did not possess knowledge about the rubric's translation process (Chia, 2010, Rubric).

There is no indication that the Country C team included a measurement expert. Therefore, there is no evidence that the Country C team provided feedback from a measurement expert.

The US agency asked Country C for feedback on the entire translation and adaptation process during the last conference call of 2010 (Shavelson, 2010, End). However, there was no feedback from translation experts during the call (Shavelson, 2010, End). Translation team members provided feedback on translation review and validation procedures throughout the two-day site visits (Solano-Flores, 2010, Visit). In addition, during an interview conducted during the site visit, Country C's translation team contributed feedback regarding team selection and translation and adaptation completed (Solano-Flores & Chia, 2010, Interview). Members of the translation team were not available to participate in the interview about the rubric translation process and the new NPM did not possess translation expertise (Chia, 2010, Rubric).

Country C used project management expertise to provide feedback regarding the study's process and material. The NPM coordinated travel so that team members could attend the four-day meeting that took place in New York and provide feedback on the study's work plan and specific activities (CAE, 2010, GS.26). The NPM also coordinated with the US agency to provide feedback on the entire translation and adaptation process during the last conference call of 2010 (Shavelson, 2010, End). Country C's NPM also coordinated the two-day site visit so that the team and translators provided feedback on translation review and validation procedures (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010, Interview). In addition, the Country C NPM helped coordinate a translator's participation in an interview that addressed the scoring rubric (Chia, 2010, Rubric). Although Country C did not find out about the NPM diaries until one year after the project had begun the NPM attempted to access the diaries and asked for technical assistance (CAE, personal communication, February 11, 2011; Al-Rashed, personal communication, February 13, 2011). In addition, the new NPM did not possess knowledge about the rubric's translation process but tried to provide information about translation of the scoring rubric during an interview with a staff member from the US organizing agency (Chia, 2010, Rubric).

Country C did not have any review opportunities while submitting feedback.

Country C did not have any training opportunities while submitting feedback.

In several reports to the international coordinating agency, the US organizing agency documented Country C's feedback. The US agency wrote a progress report that included

Country C's questions and concerns regarding the item sample size and intended constructs as well as the country team's progress in selecting performance tasks for the study, understanding the translation and adaptation framework, and student sampling (CAE, 2010, GS.30). The agency also documented Country C's feedback regarding translation, adaptation, translation team qualification, and finalizing translated performance tasks based on cognitive lab results (CAE, 2010, Module A). The US agency also document Country C's feedback regarding finalizing the translation team, the translation review team, materials create for translation and translation review, site visits, and meeting schedule (CAE, 2010, Progress). In another report the US organizing agency documented the Country C team's feedback regarding telecommunication meetings, scoring, recruiting staff to assist country teams, recruitment of translation team, site visits, sampling, test administration, and material created throughout the study (CAE, 2011, Module A).

The Country C team was able to provide the US organizing agency feedback about the study's procedures, materials, and required resources (CAE, 2010, Module A; CAE, 2011, Module A). There is no evidence that Country C's team members had difficulty with the deadlines provided for submitting feedback.

The Country C team was able to provide the US organizing agency feedback about the study's procedures, materials, and required resources through meetings, conference calls, and emails (CAE, 2010, Module A; CAE, 2011, Module A; CAE, 2010, GS.26; Shavelson & Kurpius, 2010, End). There were no materials for providing feedback.

At times, Country C required in-country support external to the team when providing feedback about the study. In addition to Country C's team members, the Country C translation team shared feedback during an interview conducted during the site visit and throughout the on site training (Solano-Flores & Chia, 2010, Interview; Solano-Flores, 2010, Visit).

Country C's team members did not require support external to the team and outside of the country to provide feedback about the study.

Country D

Country D's team members had multiple opportunities to provide feedback. Communication about feedback between the organizing agencies and the Country D team members was usually timely. The US agency gave Country D opportunities to provide feedback on the study's workplan and specific activities throughout the four-day meeting that took place in New York (CAE, 2010, GS.26). The US agency also asked Country D for feedback on the entire translation and adaptation process during the last conference call of 2010 (Shavelson, 2010, End). Country teams also had an opportunity to provide feedback on translation review and validation procedures throughout the two-day site visits (Solano-Flores, 2010, Visit). In addition, during an interview conducted during the site visit, Country D's country team and translation team were asked for feedback regarding team selection and translation and adaptation completed (Solano-Flores & Chia, 2010, Interview). Also, an international agency working on the project requested country feedback. The international agency created an online exchange where they

posted all documents created for the project; Country D's national project manager was given access to the exchange and asked to provide feedback regarding documents used in 2010 (Coates, personal communication, December 20, 2010). For 2011, the international agency created a wiki/blog through which Country D's team members could provide feedback on the study's progress and exchange information (Coates, personal communication, December 20, 2011). The international coordinating agency had also provided a format for NPM diaries that Country D was to use to provide feedback about progress in 2010 (CAE, personal communication, February 11, 2011). However, Country D did not find out about this until one year after the project had begun (CAE, personal communication, February 11, 2011). During the fall of 2011, a US agency's staff member interviewed a member of Country D's translation team asking for feedback specific to the scoring rubric (Chia, 2010, Rubric). Due to the amount of time that had passed, the interviewee needed to review her notes (Chia, 2010, Rubric). In addition, lack of experience in translation made it difficult for the translator to provide specific information about the use of rubrics in Country D (Chia, 2010, Rubric).

The Country D team included a measurement expert who was able to provide information at times. The measurement expert provided feedback throughout the four-day meeting that took place in New York (CAE, 2010, GS.26). The measurement expert also provided feedback about the entire translation and adaptation process during the last conference call of 2010 with the US agency's PI (Shavelson, 2010, End). The measurement expert also gave feedback on translation review and validation procedures throughout the two-day site visits (Solano-Flores, 2010, Visit). Although the international agency created an online exchange where they posted all documents created for the project in 2010 as well as a wiki/blog and NPM diary for the entire study's procedures in 2011 there is no evidence Country D's measurement expert was given access to them or provided feedback (Coates, personal communication, December 20, 2010; Coates, personal communication, December 20, 2011; CAE, personal communication, February 11, 2011). There was also no measurement expertise included in the feedback provided during the rubric interview (Chia, 2010, Rubric).

Country D did not obtain assistance from professional translators. Therefore, the team was not able to provide feedback that included translation expertise.

Country D used project management expertise while providing feedback regarding the study's process and material. The NPM coordinated travel so that team members could attend the four-day meeting that took place in New York and provide feedback on the study's work plan and specific activities (CAE, 2010, GS.26). The NPM also coordinated with the US agency to provide feedback on the entire translation and adaptation process during the last conference call of 2010 (Shavelson, 2010, End). Country D's NPM also coordinated the two-day site visit so that the team and translators provided feedback on translation review and validation procedures (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010, Interview). In addition, the Country D NPM helped coordinate a translator's participation in an interview that addressed the scoring rubric (Chia, 2010, Rubric).

Country D did not have any review opportunities while submitting feedback.

Country D did not have any training opportunities while submitting feedback.

In several reports to the international coordinating agency, the US organizing agency documented Country D's feedback. The US agency wrote a progress report that included Country D's questions and concerns regarding the item sample size and intended constructs as well as the country team's progress in selecting performance tasks for the study, understanding the translation and adaptation framework, and student sampling (CAE, 2010, GS.30). The agency also documented Country D's feedback regarding translation, adaptation, translation team qualification, and finalizing translated performance tasks based on cognitive lab results (CAE, 2010, Module A). The US agency also document Country D's feedback regarding finalizing the translation team, the translation review team, materials create for translation and translation review, site visits, and meeting schedule (CAE, 2010, Progress). In another report the US organizing agency documented the Country D team's feedback regarding telecommunication meetings, scoring, recruiting staff to assist country teams, recruitment of translation team, site visits, sampling, test administration, and material created throughout the study (CAE, 2011, Module A).

The Country D team was able to provide the US organizing agency feedback about the study's procedures, materials, and required resources (CAE, 2010, Module A; CAE, 2011, Module A). There is no evidence that Country D's team members had difficulty with the deadlines provided for submitting feedback.

The Country D team was able to provide the US organizing agency feedback about the study's procedures, materials, and required resources through meetings, conference calls, and emails (CAE, 2010, Module A; CAE, 2011, Module A; CAE, 2010, GS.26; Shavelson & Kurpius, 2010, End). There were no materials for providing feedback.

At times, Country D required in-country support external to the team when providing feedback about the study. In addition to Country D's team members, the staff that the team hired to complete the translation process shared feedback during an interview conducted during the site visit and throughout the on site training (Solano-Flores & Chia, 2010, Interview; Solano-Flores, 2010, Visit). In addition, one of the translation team members provided feedback regarding scoring rubric during an interview (Chia, 2010, Rubric).

Country D's team members did not require support external to the team and outside of the country to provide feedback about the study.

Country E

Country E's team members had multiple opportunities to provide feedback. Communication about feedback between the organizing agencies and the Country E team members was usually timely. The US agency gave Country E opportunities to provide feedback on the study's workplan and specific activities throughout the four-day meeting that took place in New York (CAE, 2010, GS.26). The US agency also asked Country E for feedback on the entire translation and adaptation process during the last conference call of 2010 (Shavelson, 2010,

End). Country teams also had an opportunity to provide feedback on translation review and validation procedures throughout the two-day site visits (Solano-Flores, 2010, Visit). In addition, during an interview conducted during the site visit, Country E's team and translation team were asked for feedback regarding team selection and translation and adaptation completed (Solano-Flores & Chia, 2010, Interview). Also, an international agency working on the project requested country feedback. The international agency created an online exchange where they posted all documents created for the project; Country E's national project manager was given access to the exchange and asked to provide feedback regarding documents used in 2010 (Coates, personal communication, December 20, 2010). For 2011, the international agency created a wiki/blog through which Country E's team members could provide feedback on the study's progress and exchange information (Coates, personal communication, December 20, 2011). The international coordinating agency had also provided a format for NPM diaries that Country E was to use to provide feedback about progress in 2010 (CAE, personal communication, February 11, 2011). However, Country E did not find out about this until one year after the project had begun (CAE, personal communication, February 11, 2011). During the fall of 2011, a US agency's staff member tried to interview a member of Country E's translation team to ask for feedback specific to the scoring rubric (Chia, 2010, Rubric). Due to the amount of time that had passed, members of the translation team were not available and the new NPM did not possess knowledge about the rubric's translation process (Chia, 2010, Rubric).

Several of Country E's team members had measurement expertise and a background in international test comparisons. In addition, the team included a measurement expert (CAE, 2010, GS.11). Country E included measurement expertise in their feedback on the study's workplan and specific activities throughout the four-day meeting that took place in New York (CAE, 2010, GS.26). Country E's feedback on the entire translation and adaptation process during the last conference call of 2010 also included measurement expertise (Shavelson, 2010, End). Country E also included measurement expertise in their feedback on translation review and validation procedures throughout the two-day site visits (Solano-Flores, 2010, Visit). There is no information about Country E's feedback on online exchange that included all documents created for the project in 2010, the wiki/blog for 2011 progress, or the NPM diaries (Coates, personal communication, December 20, 2010; Coates, personal communication, December 20, 2011; CAE, personal communication, February 11, 2011). Also, Country E was not able to contribute feedback during the interview addressing the rubric's translation process (Chia, 2010, Rubric).

Country E provided feedback that included translation expertise on the entire translation and adaptation process during the last conference call of 2010 (Shavelson, 2010, End). Country E's translation team also had an opportunity to provide feedback on translation review and validation procedures throughout the two-day site visits (Solano-Flores, 2010, Visit). In addition, during an interview conducted during the site visit, Country E's translation team gave feedback regarding team selection and translation and adaptation completed (Solano-Flores & Chia, 2010,

Interview). Country E's translation team members were not available and the new NPM did not possess knowledge about the rubric's translation process (Chia, 2010, Rubric).

Country E used project management expertise to provide feedback regarding the study's process and material. The NPM coordinated travel so that team members could attend the four-day meeting that took place in New York and provide feedback on the study's work plan and specific activities (CAE, 2010, GS.26). The NPM also coordinated with the US agency to provide feedback on the entire translation and adaptation process during the last conference call of 2010 (Shavelson, 2010, End). Country E's NPM also coordinated the two-day site visit so that the team and translators provided feedback on translation review and validation procedures (Solano-Flores, 2010, Visit; Solano-Flores & Chia, 2010, Interview). In addition, the Country E NPM tried to coordinate a translator's participation in an interview that addressed the scoring rubric but was unsuccessful (Chia, 2010, Rubric).

Country E did not have any review opportunities while submitting feedback.

Country E did not have any training opportunities while submitting feedback.

In several reports to the international coordinating agency, the US organizing agency documented Country E's feedback. The US agency wrote a progress report that included Country E's questions and concerns regarding the item sample size and intended constructs as well as the country team's progress in selecting performance tasks for the study, understanding the translation and adaptation framework, and student sampling (CAE, 2010, GS.30). The agency also documented Country E's feedback regarding translation, adaptation, translation team qualification, and finalizing translated performance tasks based on cognitive lab results (CAE, 2010, Module A). The US agency also document Country E's feedback regarding finalizing the translation team, the translation review team, materials create for translation and translation review, site visits, and meeting schedule (CAE, 2010, Progress). In another report the US organizing agency documented the Country E team's feedback regarding telecommunication meetings, scoring, recruiting staff to assist country teams, recruitment of translation team, site visits, sampling, test administration, and material created throughout the study (CAE, 2011, Module A).

The Country E team was able to provide the US organizing agency feedback about the study's procedures, materials, and required resources (CAE, 2010, Module A; CAE, 2011, Module A). There is no evidence that Country E's team members had difficulty with the deadlines provided for submitting feedback.

The Country E team was able to provide the US organizing agency feedback about the study's procedures, materials, and required resources through meetings, conference calls, and emails (CAE, 2010, Module A; CAE, 2011, Module A; CAE, 2010, GS.26; Shavelson & Kurpius, 2010, End). There were no materials for providing feedback.

At times, Country E required in-country support external to the team when providing feedback about the study. In addition to Country E's team members, the Country E translation

team shared feedback during an interview conducted during the site visit and throughout the on site training (Solano-Flores & Chia, 2010, Interview; Solano-Flores, 2010, Visit).

Country E's team members did not require support external to the team and outside of the country to provide feedback about the study.

Task 18: Recruit institutions and students to participate in the assessment.

Country A

Conversations between the Country A team and the organizing agencies about recruiting higher education institutions (HEIs) and students started at the initial meeting in New York City (CAE, 2010, GS.26). In 2010, Country A received information about recruitment. Country A needed a sample of ten institutions that reflected the diversity of Country A's higher education system to implement the generic strand assessment (ACER, 2010, Engaging). Country A's NPM was to engage leaders and decision-makers involved in institutions as well as faculty and students (ACER, 2010, Engaging). In addition, Country A's NPM was to secure executive sign-off from each institution participating in the study (ACER, 2010, Engaging). In 2011, the international coordinating agency provided Country A with more specific information about student sampling (ACER, 2011, Sampling). Country A was also to select a probabilistic sample of 200 students from within each participating HEI with the expectation that HEIs would see a 75 percent response rate (ACER, 2011, Sampling; CAE, 2010, GS.26). Students were to be at the end of a three or four-year undergraduate degree in a participating HEI (ACER, 2010, Sampling).

There is no information about the amount of input that Country A's measurement expert had during the identification and recruitment of higher education institutions and students.

There is no evidence that Country A required translation expertise while recruiting institutions and students.

There is no information about the amount of input that Country A's national project manager had during the identification and recruitment of higher education institutions and students.

There is no information about review opportunities available to Country A's national project manager during the identification and recruitment of higher education institutions and students.

Country A had access to training on sampling for the study. An international cooperating agency provided Country A with documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). In addition, Country A had an opportunity to attend training on sampling that was provided in a meeting that took place in November 2011 (Keeley, personal communication, October 3, 2011).

The international coordinating agency documented Country A's progress while sampling HEIs and students (Coates & Richardson, personal communication, December 21, 2011). The Country A team was to confirm the number and names of participating HEIs two months prior to test implementation (Coates & Richardson, personal communication, December 21, 2011).

It is unclear if Country A found the deadline for providing sampling results to the international coordinating agency challenging (Coates & Richardson, personal communication, December 21, 2011).

An international cooperating agency provided Country A with documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). There is no evidence that country teams found the material difficult to use.

Country teams needed in-country support external to the team. The Country A team had to recruit higher education institutions (HEIs) and students to participate in the assessment. The Country A team needed a sample of ten institutions that reflected the diversity of the higher education system to implement the generic strand assessment (ACER, 2010, Engaging). The Country A team was to engage leaders and decision-makers involved in institutions as well as faculty and students (ACER, 2010, Engaging). Country A team members were also to recruit 200 students from within each participating HEI (ACER, 2011, Sampling; CAE, 2010, GS.26).

An international coordinating agency provided Country A team members with guidance for recruiting and sampling HEIs and students for the study. The international cooperating agency provided documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). In addition, training on sampling was provided in November's meeting (Keeley, personal communication, October 3, 2011).

Country B

Conversations between the Country B team and the organizing agencies about recruiting higher education institutions (HEIs) and students started at the initial meeting in New York City (CAE, 2010, GS.26). In 2010, Country B received information about recruitment. Country B needed a sample of ten institutions that reflected the diversity of Country B's higher education system to implement the generic strand assessment (ACER, 2010, Engaging). Country B's NPM was to engage leaders and decision-makers involved in institutions as well as faculty and students (ACER, 2010, Engaging). In addition, Country B's NPM was to secure executive sign-off from each institution participating in the study (ACER, 2010, Engaging). In 2011, the international coordinating agency provided Country B with more specific information about student sampling (ACER, 2011, Sampling). Country B was also to select a probabilistic sample of 200 students from within each participating HEI with the expectation that HEIs would see a 75 percent response rate (ACER, 2011, Sampling; CAE, 2010, GS.26). Students were to be at the end of a three or four-year undergraduate degree in a participating HEI (ACER, 2010, Sampling).

There is no information about the amount of input that Country B's measurement expert had during the identification and recruitment of higher education institutions and students.

There is no evidence that Country B required translation expertise while recruiting institutions and students.

There is no information about the amount of input that Country B's national project manager had during the identification and recruitment of higher education institutions and students.

There is no information about review opportunities available to Country B's national project manager during the identification and recruitment of higher education institutions and students.

Country B had access to training on sampling for the study. An international cooperating agency provided Country B with documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). In addition, Country B had an opportunity to attend training on sampling that was provided in a meeting that took place in November 2011 (Keeley, personal communication, October 3, 2011).

The international coordinating agency documented Country B's progress while sampling HEIs and students (Coates & Richardson, personal communication, December 21, 2011). The Country B team was to confirm the number and names of participating HEIs two months prior to test implementation (Coates & Richardson, personal communication, December 21, 2011).

It is unclear if Country B found the deadline for providing sampling results to the international coordinating agency challenging (Coates & Richardson, personal communication, December 21, 2011).

An international cooperating agency provided Country B with documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). There is no evidence that country teams found the material difficult to use.

Country teams needed in-country support external to the team. The Country B team had to recruit higher education institutions (HEIs) and students to participate in the assessment. The Country B team needed a sample of ten institutions that reflected the diversity of the higher education system to implement the generic strand assessment (ACER, 2010, Engaging). The Country B team was to engage leaders and decision-makers involved in institutions as well as faculty and students (ACER, 2010, Engaging). Country B team members were also to recruit 200 students from within each participating HEI (ACER, 2011, Sampling; CAE, 2010, GS.26).

An international coordinating agency provided Country B team members with guidance for recruiting and sampling HEIs and students for the study. The international cooperating agency provided documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). In addition, training on sampling was provided in November's meeting (Keeley, personal communication, October 3, 2011).

Country C

Conversations between the Country C team and the organizing agencies about recruiting higher education institutions (HEIs) and students started at the initial meeting in New York City (CAE, 2010, GS.26). In 2010, Country C received information about recruitment. Country C needed a sample of ten institutions that reflected the diversity of Country C's higher education system to implement the generic strand assessment (ACER, 2010, Engaging). Country C's NPM

was to engage leaders and decision-makers involved in institutions as well as faculty and students (ACER, 2010, Engaging). In addition, Country C's NPM was to secure executive sign-off from each institution participating in the study (ACER, 2010, Engaging). In 2011, the international coordinating agency provided Country C with more specific information about student sampling (ACER, 2011, Sampling). Country C was also to select a probabilistic sample of 200 students from within each participating HEI with the expectation that HEIs would see a 75 percent response rate (ACER, 2011, Sampling; CAE, 2010, GS.26). Students were to be at the end of a three or four-year undergraduate degree in a participating HEI (ACER, 2010, Sampling).

There is no evidence that Country C had a measurement expert participating in the study.

There is no evidence that Country C required translation expertise while recruiting institutions and students.

There is no information about the amount of input that Country C's national project manager had during the identification and recruitment of higher education institutions and students.

There is no information about review opportunities available to Country C's national project manager during the identification and recruitment of higher education institutions and students.

Country C had access to training on sampling for the study. An international cooperating agency provided Country C with documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). In addition, Country C had an opportunity to attend training on sampling that was provided in a meeting that took place in November 2011 (Keeley, personal communication, October 3, 2011).

The international coordinating agency documented Country C's progress while sampling HEIs and students (Coates & Richardson, personal communication, December 21, 2011). The Country C team was to confirm the number and names of participating HEIs two months prior to test implementation (Coates & Richardson, personal communication, December 21, 2011).

It is unclear if Country C found the deadline for providing sampling results to the international coordinating agency challenging (Coates & Richardson, personal communication, December 21, 2011).

An international cooperating agency provided Country C with documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). There is no evidence that country teams found the material difficult to use.

Country teams needed in-country support external to the team. The Country C team had to recruit higher education institutions (HEIs) and students to participate in the assessment. The Country C team needed a sample of ten institutions that reflected the diversity of the higher education system to implement the generic strand assessment (ACER, 2010, Engaging). The Country C team was to engage leaders and decision-makers involved in institutions as well as

faculty and students (ACER, 2010, Engaging). Country C team members were also to recruit 200 students from within each participating HEI (ACER, 2011, Sampling; CAE, 2010, GS.26).

An international coordinating agency provided Country C team members with guidance for recruiting and sampling HEIs and students for the study. The international cooperating agency provided documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). In addition, training on sampling was provided in November's meeting (Keeley, personal communication, October 3, 2011).

Country D

Conversations between the Country D team and the organizing agencies about recruiting higher education institutions (HEIs) and students started at the initial meeting in New York City (CAE, 2010, GS.26). In 2010, Country D received information about recruitment. Country D needed a sample of ten institutions that reflected the diversity of Country D's higher education system to implement the generic strand assessment (ACER, 2010, Engaging). Country D's NPM was to engage leaders and decision-makers involved in institutions as well as faculty and students (ACER, 2010, Engaging). In addition, Country D's NPM was to secure executive sign-off from each institution participating in the study (ACER, 2010, Engaging). In 2011, the international coordinating agency provided Country D with more specific information about student sampling (ACER, 2011, Sampling). Country D was also to select a probabilistic sample of 200 students from within each participating HEI with the expectation that HEIs would see a 75 percent response rate (ACER, 2011, Sampling; CAE, 2010, GS.26). Students were to be at the end of a three or four-year undergraduate degree in a participating HEI (ACER, 2010, Sampling).

There is no information about the amount of input that Country D's measurement expert had during the identification and recruitment of higher education institutions and students.

There is no evidence that Country D required translation expertise while recruiting institutions and students.

There is no information about the amount of input that Country D's national project manager had during the identification and recruitment of higher education institutions and students.

There is no information about review opportunities available to Country D's national project manager during the identification and recruitment of higher education institutions and students.

Country D had access to training on sampling for the study. An international cooperating agency provided Country D with documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). In addition, Country D had an opportunity to attend training on sampling that was provided in a meeting that took place in November 2011 (Keeley, personal communication, October 3, 2011).

The international coordinating agency documented Country D's progress while sampling HEIs and students (Coates & Richardson, personal communication, December 21, 2011). The

Country D team was to confirm the number and names of participating HEIs two months prior to test implementation (Coates & Richardson, personal communication, December 21, 2011).

It is unclear if Country D found the deadline for providing sampling results to the international coordinating agency challenging (Coates & Richardson, personal communication, December 21, 2011).

An international cooperating agency provided Country D with documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). There is no evidence that country teams found the material difficult to use.

Country teams needed in-country support external to the team. The Country D team had to recruit higher education institutions (HEIs) and students to participate in the assessment. The Country D team needed a sample of ten institutions that reflected the diversity of the higher education system to implement the generic strand assessment (ACER, 2010, Engaging). The Country D team was to engage leaders and decision-makers involved in institutions as well as faculty and students (ACER, 2010, Engaging). Country D team members were also to recruit 200 students from within each participating HEI (ACER, 2011, Sampling; CAE, 2010, GS.26).

An international coordinating agency provided Country D team members with guidance for recruiting and sampling HEIs and students for the study. The international cooperating agency provided documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). In addition, training on sampling was provided in November's meeting (Keeley, personal communication, October 3, 2011).

Country E

Conversations between the Country E team and the organizing agencies about recruiting higher education institutions (HEIs) and students started at the initial meeting in New York City (CAE, 2010, GS.26). In 2010, Country E received information about recruitment. Country E needed a sample of ten institutions that reflected the diversity of Country E's higher education system to implement the generic strand assessment (ACER, 2010, Engaging). Country E's NPM was to engage leaders and decision-makers involved in institutions as well as faculty and students (ACER, 2010, Engaging). In addition, Country E's NPM was to secure executive sign-off from each institution participating in the study (ACER, 2010, Engaging). In 2011, the international coordinating agency provided Country E with more specific information about student sampling (ACER, 2011, Sampling). Country E was also to select a probabilistic sample of 200 students from within each participating HEI with the expectation that HEIs would see a 75 percent response rate (ACER, 2011, Sampling; CAE, 2010, GS.26). Students were to be at the end of a three or four-year undergraduate degree in a participating HEI (ACER, 2010, Sampling).

There is no information about the amount of input that Country E's measurement expert had during the identification and recruitment of higher education institutions and students.

There is no evidence that Country E required translation expertise while recruiting institutions and students.

There is no information about the amount of input that Country E's national project manager had during the identification and recruitment of higher education institutions and students.

There is no information about review opportunities available to Country E's national project manager during the identification and recruitment of higher education institutions and students.

Country E had access to training on sampling for the study. An international cooperating agency provided Country E with documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). In addition, Country E had an opportunity to attend training on sampling that was provided in a meeting that took place in November 2011 (Keeley, personal communication, October 3, 2011).

The international coordinating agency documented Country E's progress while sampling HEIs and students (Coates & Richardson, personal communication, December 21, 2011). The Country E team was to confirm the number and names of participating HEIs two months prior to test implementation (Coates & Richardson, personal communication, December 21, 2011).

It is unclear if Country E found the deadline for providing sampling results to the international coordinating agency challenging (Coates & Richardson, personal communication, December 21, 2011).

An international cooperating agency provided Country E with documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). There is no evidence that country teams found the material difficult to use.

Country teams needed in-country support external to the team. The Country E team had to recruit higher education institutions (HEIs) and students to participate in the assessment. The Country E team needed a sample of ten institutions that reflected the diversity of the higher education system to implement the generic strand assessment (ACER, 2010, Engaging). The Country E team was to engage leaders and decision-makers involved in institutions as well as faculty and students (ACER, 2010, Engaging). Country E team members were also to recruit 200 students from within each participating HEI (ACER, 2011, Sampling; CAE, 2010, GS.26).

An international coordinating agency provided Country E team members with guidance for recruiting and sampling HEIs and students for the study. The international cooperating agency provided documents addressing HEI and student sampling that provided guidance for recruitment (ACER, 2010, Engaging; ACER, 2011, Sampling). In addition, training on sampling was provided in November's meeting (Keeley, personal communication, October 3, 2011).

Appendix J: Three matrices for each country, one for each evidence type: confirming evidence bit (CEB), disconfirming evidence bit (DEB), and no evidence bit (NEB).

Total cells=198
c_n=type of criterion n; t_n: task
FI: Country A; KO:Country B; KU: Country C; ME: Country D; NO:Country E

CEBs

FI_1
#=144

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	
t_1	1	1	1	1	1	1	1	1	1	1	1	11
t_2	1	1	1	1	1		1	1	1	1	1	10
t_3	1	1		1	1	1	1	1	1		1	9
t_4	1	1	1	1	1	1	1	1	1	1		10
t_5	1	1			1	1	1	1		1	1	8
t_6	1	1	1	1		1	1		1	1	1	9
t_7	1		1	1		1	1		1	1	1	8
t_8	1	1	1	1	1	1	1		1	1	1	10
t_9	1		1	1		1	1		1	1	1	8
t_{10}	1	1			1	1	1	1	1		1	8
t_{11}	1	1		1					1		1	5
t_{12}	1	1		1	1	1	1	1	1	1	1	10
t_{13}												0
t_{14}	1	1		1	1	1	1	1	1	1	1	10
t_{15}	1	1		1	1	1		1	1		1	8
t_{16}	1	1	1	1			1	1	1		1	8
t_{17}		1	1	1			1	1		1		6
t_{18}	1					1	1		1	1	1	6
	16	14	9	14	10	13	15	11	15	12	15	

KO_1
#=111

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	
t_1		1			1	1	1		1	1	1	7
t_2	1	1		1	1		1	1	1	1	1	9
t_3	1	1		1	1	1	1	1	1		1	9
t_4	1	1	1	1	1	1	1	1	1	1		10
t_5	1	1			1	1	1	1			1	7
t_6	1			1		1	1		1	1	1	7
t_7							1	1	1	1	1	5
t_8	1	1			1	1	1		1	1	1	8
t_9						1	1		1	1	1	5
t_{10}	1				1	1	1	1	1		1	7
t_{11}		1					1				1	3
t_{12}						1	1			1	1	4
t_{13}												0
t_{14}	1	1		1	1	1	1		1		1	8
t_{15}	1	1							1		1	4
t_{16}	1	1		1			1	1	1		1	7
t_{17}		1		1			1	1		1		5
t_{18}	1					1	1		1	1	1	6
	11	11	1	7	8	11	16	8	13	10	15	

KU_1
#=117

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	
t_1			1	1	1		1		1	1	1	7
t_2	1		1	1	1		1	1	1	1	1	9
t_3	1			1	1	1	1	1	1		1	8
t_4	1		1	1	1	1	1	1	1	1		9
t_5					1	1	1	1			1	5
t_6	1		1	1		1	1		1	1	1	8
t_7			1	1		1	1	1	1	1	1	8
t_8	1		1		1	1	1		1	1	1	8
t_9			1			1	1		1	1	1	6
t_{10}	1				1	1	1	1	1		1	7
t_{11}									1		1	2
t_{12}				1		1	1		1	1	1	6
t_{13}												0
t_{14}	1			1	1	1	1	1	1	1	1	9
t_{15}	1	1		1	1	1		1	1		1	8
t_{16}	1			1			1	1	1		1	6
t_{17}			1	1			1	1		1		5
t_{18}	1					1	1		1	1	1	6
	10	1	8	11	9	12	15	10	15	11	15	

ME_1
#=131

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	
t_1	1	1	1	1	1	1	1	1	1	1	1	11
t_2	1	1		1	1		1	1	1	1	1	9
t_3	1	1		1	1	1	1	1	1		1	9
t_4	1	1	1	1	1	1	1	1	1	1		10
t_5	1	1			1	1	1	1			1	7
t_6	1	1				1	1	1	1	1	1	8
t_7	1			1		1	1		1	1	1	7
t_8	1	1			1	1	1		1		1	7
t_9	1			1		1	1		1	1	1	7
t_{10}	1	1			1	1	1	1	1		1	8
t_{11}	1	1		1			1				1	5
t_{12}	1	1		1		1	1			1	1	7
t_{13}												
t_{14}	1	1		1	1	1	1	1	1	1	1	10
t_{15}	1	1		1	1	1		1	1		1	8
t_{16}	1	1		1			1	1	1		1	7
t_{17}		1		1			1	1		1		5
t_{18}	1					1	1		1	1	1	6
	16	14	2	12	9	13	16	11	13	10	15	

NO_1
#=151

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	
t_1	1	1	1	1	1	1	1	1	1	1	1	11
t_2	1	1	1	1	1		1	1	1	1	1	10
t_3	1	1	1	1	1	1	1	1	1		1	10
t_4	1	1	1	1	1	1	1	1	1	1		10
t_5	1	1	1		1	1	1	1		1	1	9
t_6	1	1	1	1		1	1	1	1	1	1	10
t_7		1	1	1	1	1	1	1	1	1	1	10
t_8	1	1	1		1	1	1		1	1	1	9
t_9	1	1	1	1	1	1	1		1	1	1	10
t_{10}	1	1	1		1	1	1	1	1		1	9
t_{11}	1	1	1	1				1			1	6
t_{12}	1	1	1	1		1	1		1	1	1	9
t_{13}												0
t_{14}	1	1	1	1	1	1	1	1	1	1	1	11
t_{15}	1	1		1	1	1		1	1		1	8
t_{16}	1	1		1			1	1	1		1	7
t_{17}		1	1	1			1	1		1		6
t_{18}	1					1	1		1	1	1	6
	15	16	14	13	11	13	15	13	14	12	15	

DEBs

FI_{-1} #=42

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	
t_1												
t_2						-1						-1
t_3			-1							-1		-2
t_4											-1	-1
t_5			-1	-1					-1			-3
t_6					-1			-1				-2
t_7		-1			-1			-1				-2
t_8								-1				-1
t_9		-1			-1			-1				-3
t_{10}												
t_{11}						-1	-1	-1				-3
t_{12}			-1									-1
t_{13}	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-11
t_{14}												
t_{15}			-1				-1					-2
t_{16}					-1	-1				-1		-3
t_{17}	-1				-1	-1			-1		-1	-5
t_{18}			-1									-1
	-2	-3	-6	-2	-6	-5	-3	-6	-3	-3	-3	

KO_{-1} #=68

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	
t_1	-1		-1	-1				-1				-4
t_2			-1			-1						-2
t_3			-1							-1		-2
t_4											-1	-1
t_5			-1	-1					-1	-1		-4
t_6		-1	-1		-1			-1				-4
t_7	-1	-1	-1	-1	-1	-1						-6
t_8								-1				-1
t_9	-1	-1		-1	-1			-1				-5
t_{10}												
t_{11}	-1			-1		-1		-1	-1			-5
t_{12}	-1		-1		-1			-1	-1			-5
t_{13}	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-11
t_{14}								-1				-1
t_{15}			-1	-1	-1	-1	-1	-1				-6
t_{16}			-1		-1	-1				-1		-4
t_{17}	-1		-1		-1	-1			-1		-1	-6
t_{18}			-1									-1
	-7	-4	-12	-7	-8	-7	-2	-9	-5	-4	-3	

KU_{-1}
#=62

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	
t_1	-1	-1				-1		-1				-4
t_2		-1				-1						-2
t_3		-1	-1							-1		-3
t_4		-1									-1	-2
t_5	-1	-1	-1	-1					-1	-1		-6
t_6		-1			-1			-1				-3
t_7	-1	-1			-1							-3
t_8		-1						-1				-2
t_9	-1	-1		-1	-1			-1				-5
t_{10}												
t_{11}						-1	-1	-1				-3
t_{12}	-1				-1			-1				-3
t_{13}	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-11
t_{14}												
t_{15}			-1				-1					-2
t_{16}		-1	-1		-1	-1				-1		-5
t_{17}	-1	-1			-1	-1			-1		-1	-6
t_{18}		-1	-1									-2
	-7	-13	-6	-3	-7	-6	-3	-7	-3	-4	-3	

ME_{-1}
#=55

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	
t_1												
t_2			-1			-1						-2
t_3			-1							-1		-2
t_4											-1	-1
t_5			-1	-1					-1	-1		-4
t_6			-1	-1	-1							-3
t_7		-1	-1		-1			-1				-4
t_8			-1					-1		-1		-3
t_9		-1	-1		-1			-1				-4
t_{10}												
t_{11}						-1		-1	-1			-3
t_{12}			-1		-1			-1	-1			-4
t_{13}	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-11
t_{14}			-1									-1
t_{15}			-1				-1					-2
t_{16}			-1		-1	-1				-1		-4
t_{17}	-1		-1		-1	-1			-1		-1	-6

t_{18}			-1									-1
	-2	-3	-14	-3	-7	-5	-2	-6	-5	-5	-3	

NO.-1
#=37

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	
t_1												
t_2						-1						-1
t_3										-1		-1
t_4											-1	-1
t_5				-1					-1			-2
t_6					-1							-1
t_7	-1											-1
t_8								-1				-1
t_9								-1				-1
t_{10}												
t_{11}						-1	-1		-1			-3
t_{12}					-1			-1				-2
t_{13}	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-11
t_{14}												
t_{15}			-1				-1					-2
t_{16}			-1		-1	-1				-1		-4
t_{17}	-1				-1	-1			-1		-1	-5
t_{18}			-1									-1
	-3	-1	-4	-2	-5	-5	-3	-4	-4	-3	-3	

C Matrices

FI_1	t_1	1	1	1	1	1	1	1	1	1	1	1	11
KO_1	t_1		1			1	1	1		1	1	1	7
KU_1	t_1			1	1	1		1		1	1	1	7
ME_1	t_1	1	1	1	1	1	1	1	1	1	1	1	11
NO_1	t_1	1	1	1	1	1	1	1	1	1	1	1	11
		3	4	4	4	5	4	5	3	5	5	5	47

Figure T1. CEB matrix with totals for each evidence type for Task 1 by country.

FI_1	t_2	1	1	1	1	1	0	1	1	1	1	1	10
KO_1	t_2	1	1	0	1	1	0	1	1	1	1	1	9
KU_1	t_2	1	0	1	1	1	0	1	1	1	1	1	9
ME_1	t_2	1	1	0	1	1	0	1	1	1	1	1	9
NO_1	t_2	1	1	1	1	1	0	1	1	1	1	1	10
		5	4	3	5	5	0	5	5	5	5	5	47

Figure T2. CEB matrix for each evidence type across countries and their sums for Task 2.

FI_1	t_3	1	1	0	1	1	1	1	1	1	0	1	9
KO_1	t_3	1	1	0	1	1	1	1	1	1	0	1	9
KU_1	t_3	1	0	0	1	1	1	1	1	1	0	1	8
ME_1	t_3	1	1	0	1	1	1	1	1	1	0	1	9
NO_1	t_3	1	1	1	1	1	1	1	1	1	0	1	10
		5	4	1	5	5	5	5	5	5	0	5	45

Figure T3. CEB matrix with totals for each evidence type for Task 3 by country.

FI_1	t_4	1	1	1	1	1	1	1	1	1	1	0	10
KO_1	t_4	1	1	1	1	1	1	1	1	1	1	0	10
KU_1	t_4	1	0	1	1	1	1	1	1	1	1	0	9
ME_1	t_4	1	1	1	1	1	1	1	1	1	1	0	10
NO_1	t_4	1	1	1	1	1	1	1	1	1	1	0	10
		5	4	5	5	5	5	5	5	5	5	0	49

Figure T4. CEB matrix with totals for each evidence type for Task 4 by country.

FI_1	t_5	1	1	0	0	1	1	1	1	0	1	1	8
KO_1	t_5	1	1	0	0	1	1	1	1	0	0	1	7
KU_1	t_5	0	0	0	0	1	1	1	1	0	0	1	5
ME_1	t_5	1	1	0	0	1	1	1	1	0	0	1	7
NO_1	t_5	1	1	1	0	1	1	1	1	0	1	1	9
		4	4	1	0	5	5	5	5	0	2	5	36

Figure T5. CEB matrix with totals for each evidence type for Task 5 by country.

FI_1	t_6	1	1	1	1	0	1	1	0	1	1	1	9
KO_1	t_6	1	0	0	1	0	1	1	0	1	1	1	7
KU_1	t_6	1	0	1	1	0	1	1	0	1	1	1	8
ME_1	t_6	1	1	0	0	0	1	1	1	1	1	1	8
NO_1	t_6	1	1	1	1	0	1	1	1	1	1	1	10
		5	3	3	4	0	5	5	2	5	5	5	42

Figure T6. CEB matrix with totals for each evidence type for Task 6 by country.

FI_1	t_7	1	0	1	1	0	1	1	0	1	1	1	8
KO_1	t_7	0	0	0	0	0	0	1	1	1	1	1	5
KU_1	t_7	0	0	1	1	0	1	1	1	1	1	1	8
ME_1	t_7	1	0	0	1	0	1	1	0	1	1	1	7
NO_1	t_7	0	1	1	1	1	1	1	1	1	1	1	10
		2	1	3	4	1	4	5	3	5	5	5	38

Figure T7. CEB matrix with totals for each evidence type for Task 7 by country.

FI_1	t_8	1	1	1	1	1	1	1	0	1	1	1	10
KO_1	t_8	1	1	0	0	1	1	1	0	1	1	1	8
KU_1	t_8	1	0	1	0	1	1	1	0	1	1	1	8
ME_1	t_8	1	1	0	0	1	1	1	0	1	0	1	7
NO_1	t_8	1	1	1	0	1	1	1	0	1	1	1	9
		5	4	3	1	5	5	5	0	5	4	5	42

Figure T8. CEB matrix for each evidence type across countries and their sums for Task 8.

FI_1	t_8	1	0	1	1	0	1	1	0	1	1	1	8
KO_1	t_8	0	0	0	0	0	1	1	0	1	1	1	5
KU_1	t_8	0	0	1	0	0	1	1	0	1	1	1	6
ME_1	t_8	1	0	0	1	0	1	1	0	1	1	1	7
NO_1	t_8	1	1	1	1	1	1	1	0	1	1	1	10
		3	1	3	3	1	5	5	0	5	5	5	36

Figure T9. CEB matrix for each evidence type across countries and their sums for Task 9.

FI_1	t_{10}	1	1	0	0	1	1	1	1	1	0	1	8
KO_1	t_8	1	0	0	0	1	1	1	1	1	0	1	7
KU_1	t_8	1	0	0	0	1	1	1	1	1	0	1	7
ME_1	t_8	1	1	0	0	1	1	1	1	1	0	1	8
NO_1	t_8	1	1	1	0	1	1	1	1	1	0	1	9
		5	3	1	0	5	5	5	5	5	0	5	39

Figure T10. CEB matrix for each evidence type across countries and their sums for Task 10.

FI_1	t_{11}	1	1	0	1	0	0	0	0	1	0	1	5
KO_1	t_8	0	1	0	0	0	0	1	0	0	0	1	3
KU_1	t_8	0	0	0	0	0	0	0	0	1	0	1	2
ME_1	t_8	1	1	0	1	0	0	1	0	0	0	1	5
NO_1	t_8	1	1	1	1	0	0	0	1	0	0	1	6
		3	4	1	3	0	0	2	1	2	0	5	21

Figure T11. CEB matrix for each evidence type across countries and their sums for Task 11.

FI_1	t_{12}	1	1	0	1	1	1	1	1	1	1	1	10
KO_1	t_8	0	0	0	0	0	1	1	0	0	1	1	4
KU_1	t_8	0	0	0	1	0	1	1	0	1	1	1	6
ME_1	t_8	1	1	0	1	0	1	1	0	0	1	1	7
NO_1	t_8	1	1	1	1	0	1	1	0	1	1	1	9
		3	3	1	4	1	5	5	1	3	5	5	36

Figure T12. CEB matrix for each evidence type across countries and their sums for Task 12.

FI_1	t_{13}	0	0	0	0	0	0	0	0	0	0	0	0
KO_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
KU_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
ME_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
NO_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
		0	0	0	0	0	0	0	0	0	0	0	0

Figure T13. CEB matrix for each evidence type across countries and their sums for Task 13.

FI_1	t_{14}	1	1	0	1	1	1	1	1	1	1	1	10
KO_1	t_8	1	1	0	1	1	1	1	0	1	0	1	8
KU_1	t_8	1	0	0	1	1	1	1	1	1	1	1	9
ME_1	t_8	1	1	0	1	1	1	1	1	1	1	1	10
NO_1	t_8	1	1	1	1	1	1	1	1	1	1	1	11
		5	4	1	5	5	5	5	4	5	4	5	48

Figure T14. CEB matrix for each evidence type across countries and their sums for Task 14.

FI_1	t_8	1	1	0	1	1	1	0	1	1	0	1	8
KO_1	t_8	1	1	0	0	0	0	0	0	1	0	1	4
KU_1	t_8	1	1	0	1	1	1	0	1	1	0	1	8
ME_1	t_8	1	1	0	1	1	1	0	1	1	0	1	8
NO_1	t_8	1	1	0	1	1	1	0	1	1	0	1	8
		5	5	0	4	4	4	0	4	5	0	5	36

Figure T15. CEB matrix for each evidence type across countries and their sums for Task 15.

FI_1	t_8	1	1	1	1	0	0	1	1	1	0	1	8
KO_1	t_8	1	1	0	1	0	0	1	1	1	0	1	7
KU_1	t_8	1	0	0	1	0	0	1	1	1	0	1	6
ME_1	t_8	1	1	0	1	0	0	1	1	1	0	1	7
NO_1	t_8	1	1	0	1	0	0	1	1	1	0	1	7
		5	4	1	5	0	0	5	5	5	0	5	35

Figure T16. CEB matrix for each evidence type across countries and their sums for Task 16.

FI_1	t_8	0	1	1	1	0	0	1	1	0	1	0	6
KO_1	t_8	0	1	0	1	0	0	1	1	0	1	0	5
KU_1	t_8	0	0	1	1	0	0	1	1	0	1	0	5
ME_1	t_8	0	1	0	1	0	0	1	1	0	1	0	5
NO_1	t_8	0	1	1	1	0	0	1	1	0	1	0	6
		0	4	3	5	0	0	5	5	0	5	0	27

Figure T17. CEB matrix for each evidence type across countries and their sums for Task 17.

FI_1	t_8	1	0	0	0	0	1	1	0	1	1	1	6
KO_1	t_8	1	0	0	0	0	1	1	0	1	1	1	6
KU_1	t_8	1	0	0	0	0	1	1	0	1	1	1	6
ME_1	t_8	1	0	0	0	0	1	1	0	1	1	1	6
NO_1	t_8	1	0	0	0	0	1	1	0	1	1	1	6
		5	0	0	0	0	5	5	0	5	5	5	30

Figure T18. CEB matrix for each evidence type across countries and their sums for Task 18.

D Matrices

FI_{-1}	t_1												0
KO_{-1}	t_1	-1		-1	-1				-1				-4
KU_{-1}	t_1	-1	-1				-1		-1				-4
ME_{-1}	t_1												0
NO_{-1}	t_1												0
		-2	-1	-1	-1	0	-1	0	-2	0	0	0	-8

Figure T1. DEB matrix with totals for each evidence type for Task 1 by country.

FI_{-1}	t_2	0	0	0	0	0	-1	0	0	0	0	0	-1
KO_{-1}	t_2	0	0	-1	0	0	-1	0	0	0	0	0	-2
KU_{-1}	t_2	0	-1	0	0	0	-1	0	0	0	0	0	-2
ME_{-1}	t_2	0	0	-1	0	0	-1	0	0	0	0	0	-2
NO_{-1}	t_2	0	0	0	0	0	-1	0	0	0	0	0	-1
		0	-1	-2	0	0	-5	0	0	0	0	0	-8

Figure T2. DEB matrix for each evidence type across countries and their sums for Task 2.

FI_{-1}	t_3	0	0	-1	0	0	0	0	0	0	-1	0	-2
KO_{-1}	t_3	0	0	-1	0	0	0	0	0	0	-1	0	-2
KU_{-1}	t_3	0	-1	-1	0	0	0	0	0	0	-1	0	-3
ME_{-1}	t_3	0	0	-1	0	0	0	0	0	0	-1	0	-2
NO_{-1}	t_3	0	0	0	0	0	0	0	0	0	-1	0	-1
		0	-1	-4	0	0	0	0	0	0	-5	0	-10

Figure T3. DEB matrix with totals for each evidence type for Task 3 by country.

FI_{-1}	t_4	0	0	0	0	0	0	0	0	0	0	-1	-1
KO_{-1}	t_4	0	0	0	0	0	0	0	0	0	0	-1	-1
KU_{-1}	t_4	0	-1	0	0	0	0	0	0	0	0	-1	-2
ME_{-1}	t_4	0	0	0	0	0	0	0	0	0	0	-1	-1
NO_{-1}	t_4	0	0	0	0	0	0	0	0	0	0	-1	-1
		0	-1	0	0	0	0	0	0	0	0	-5	-6

Figure T4. DEB matrix with totals for each evidence type for Task 4 by country.

FI_{-1}	t_5	0	0	-1	-1	0	0	0	0	-1	0	0	-3
KO_{-1}	t_5	0	0	-1	-1	0	0	0	0	-1	-1	0	-4
KU_{-1}	t_5	-1	-1	-1	-1	0	0	0	0	-1	-1	0	-6
ME_{-1}	t_5	0	0	-1	-1	0	0	0	0	-1	-1	0	-4
NO_{-1}	t_5	0	0	0	-1	0	0	0	0	-1	0	0	-2
		-1	-1	-4	-5	0	0	0	0	-5	-3	0	-19

Figure T5. DEB matrix with totals for each evidence type for Task 5 by country.

FI_{-1}	t_6	0	0	0	0	-1	0	0	-1	0	0	0	-2
KO_{-1}	t_6	0	-1	-1	0	-1	0	0	-1	0	0	0	-4
KU_{-1}	t_6	0	-1	0	0	-1	0	0	-1	0	0	0	-3
ME_{-1}	t_6	0	0	-1	-1	-1	0	0	0	0	0	0	-3
NO_{-1}	t_6	0	0	0	0	-1	0	0	0	0	0	0	-1
		0	-2	-2	-1	-5	0	0	-3	0	0	0	-13

Figure T6. DEB matrix with totals for each evidence type for Task 6 by country.

FI_{-1}	t_7	0	-1	0	0	-1	0	0	-1	0	0	0	-3
KO_{-1}	t_7	-1	-1	-1	-1	-1	-1	0	0	0	0	0	-6
KU_{-1}	t_7	-1	-1	0	0	-1	0	0	0	0	0	0	-3
ME_{-1}	t_7	0	-1	-1	0	-1	0	0	-1	0	0	0	-4
NO_{-1}	t_7	-1	0	0	0	0	0	0	0	0	0	0	-1
		-3	-4	-2	-1	-4	-1	0	-2	0	0	0	-17

Figure T7. DEB matrix with totals for each evidence type for Task 7 by country.

FI_{-1}	t_8	0	0	0	0	0	0	0	-1	0	0	0	-1
KO_{-1}	t_8	0	0	0	0	0	0	0	-1	0	0	0	-1
KU_{-1}	t_8	0	-1	0	0	0	0	0	-1	0	0	0	-2
ME_{-1}	t_8	0	0	-1	0	0	0	0	-1	0	-1	0	-3
NO_{-1}	t_8	0	0	0	0	0	0	0	-1	0	0	0	-1
		0	-1	-1	0	0	0	0	-5	0	-1	0	-8

Figure T8. DEB matrix for each evidence type across countries and their sums for Task 8.

FI_1	t_8	0	-1	0	0	-1	0	0	-1	0	0	0	-3
KO_1	t_8	-1	-1	0	-1	-1	0	0	-1	0	0	0	-5
KU_1	t_8	-1	-1	0	-1	-1	0	0	-1	0	0	0	-5
ME_1	t_8	0	-1	-1	0	-1	0	0	-1	0	0	0	-4
NO_1	t_8	0	0	0	0	0	0	0	-1	0	0	0	-1
		-2	-4	-1	-2	-4	0	0	-5	0	0	0	-18

Figure T9. DEB matrix for each evidence type across countries and their sums for Task 9.

FI_1	t_8	0	0	0	0	0	-1	-1	-1	0	0	0	-3
KO_1	t_8	-1	0	0	-1	0	-1	0	-1	-1	0	0	-5
KU_1	t_8	0	0	0	0	0	-1	-1	-1	0	0	0	-3
ME_1	t_8	0	0	0	0	0	-1	0	-1	-1	0	0	-3
NO_1	t_8	0	0	0	0	0	-1	-1	0	-1	0	0	-3
		-1	0	0	-1	0	-5	-3	-4	-3	0	0	-17

Figure T11. DEB matrix for each evidence type across countries and their sums for Task 11.

FI_1	t_8	0	0	-1	0	0	0	0	0	0	0	0	-1
KO_1	t_8	-1	0	-1	0	-1	0	0	-1	-1	0	0	-5
KU_1	t_8	-1	0	0	0	-1	0	0	-1	0	0	0	-3
ME_1	t_8	0	0	-1	0	-1	0	0	-1	-1	0	0	-4
NO_1	t_8	0	0	0	0	-1	0	0	-1	0	0	0	-2
		-2	0	-3	0	-4	0	0	-4	-2	0	0	-15

Figure T12. DEB matrix for each evidence type across countries and their sums for Task 12.

FI_1	t_8	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-11
KO_1	t_8	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-11
KU_1	t_8	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-11
ME_1	t_8	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-11
NO_1	t_8	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-11
		-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-55

Figure T13. DEB matrix for each evidence type across countries and their sums for Task 13.

FI_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
KO_1	t_8	0	0	0	0	0	0	0	-1	0	0	0	-1
KU_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
ME_1	t_8	0	0	-1	0	0	0	0	0	0	0	0	-1
NO_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
		0	0	-1	0	0	0	0	-1	0	0	0	-2

Figure T14. DEB matrix for each evidence type across countries and their sums for Task 14.

FI_1	t_8	0	0	-1	0	0	0	-1	0	0	0	0	-2
KO_1	t_8	0	0	-1	-1	-1	-1	-1	-1	0	0	0	-6
KU_1	t_8	0	0	-1	0	0	0	-1	0	0	0	0	-2
ME_1	t_8	0	0	-1	0	0	0	-1	0	0	0	0	-2
NO_1	t_8	0	0	-1	0	0	0	-1	0	0	0	0	-2
		0	0	-5	-1	-1	-1	-5	-1	0	0	0	-14

Figure T15. DEB matrix for each evidence type across countries and their sums for Task 15.

FI_1	t_8	0	0	0	0	-1	-1	0	0	0	-1	0	-3
KO_1	t_8	0	0	-1	0	-1	-1	0	0	0	-1	0	-4
KU_1	t_8	0	-1	-1	0	-1	-1	0	0	0	-1	0	-5
ME_1	t_8	0	0	-1	0	-1	-1	0	0	0	-1	0	-4
NO_1	t_8	0	0	-1	0	-1	-1	0	0	0	-1	0	-4
		0	-1	-4	0	-5	-5	0	0	0	-5	0	-20

Figure T16. DEB matrix for each evidence type across countries and their sums for Task 16.

FI_1	t_8	-1	0	0	0	-1	-1	0	0	-1	0	-1	-5
KO_1	t_8	-1	0	-1	0	-1	-1	0	0	-1	0	-1	-6
KU_1	t_8	-1	-1	0	0	-1	-1	0	0	-1	0	-1	-6
ME_1	t_8	-1	0	-1	0	-1	-1	0	0	-1	0	-1	-6
NO_1	t_8	-1	0	0	0	-1	-1	0	0	-1	0	-1	-5
		5	-1	-2	0	-5	-5	0	0	-5	0	-5	-28

Figure T17. DEB matrix for each evidence type across countries and their sums for Task 17.

FI_1	t_8	0	0	-1	0	0	0	0	0	0	0	0	-1
KO_1	t_8	0	0	-1	0	0	0	0	0	0	0	0	-1
KU_1	t_8	0	-1	-1	0	0	0	0	0	0	0	0	-2
ME_1	t_8	0	0	-1	0	0	0	0	0	0	0	0	-1
NO_1	t_8	0	0	-1	0	0	0	0	0	0	0	0	-1
		0	-1	-5	0	0	0	0	0	0	0	0	-6

Figure T18. DEB matrix for each evidence type across countries and their sums for Task 18.

N Matrices

FI_{-1}	t_8	0	0	0	0	0	0	0	0	0	0	0	0
KO_{-1}	t_8	0	0	1	1	0	0	0	0	0	0	0	2
KU_{-1}	t_8	0	0	0	1	0	0	0	0	0	0	0	1
ME_{-1}	t_8	0	0	0	1	0	0	0	0	0	0	0	1
NO_{-1}	t_8	0	0	0	1	0	0	0	0	0	0	0	1
		0	0	1	4	0	0	0	0	0	0	0	5

Figure T8. NEBs for each evidence type across countries and their sums for Task 8.

FI_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
KO_1	t_8	0	0	1	0	0	0	0	0	0	0	0	1
KU_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
ME_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
NO_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
		0	0	1	0	0	0	0	0	0	0	0	1

Figure T9. NEBs for each evidence type across countries and their sums for Task 9.

FI_1	t_8	0	0	1	1	0	0	0	0	0	1	0	3
KO_1	t_8	0	1	1	1	0	0	0	0	0	1	0	4
KU_1	t_8	0	1	1	1	0	0	0	0	0	1	0	4
ME_1	t_8	0	0	1	1	0	0	0	0	0	1	0	3
NO_1	t_8	0	0	0	1	0	0	0	0	0	1	0	2
		0	2	4	5	0	0	0	0	0	5	0	16

Figure T10. NEBs for each evidence type across countries and their sums for Task 10.

FI_1	t_8	0	0	1	0	1	0	0	0	0	1	0	3
KO_1	t_8	0	0	1	0	1	0	0	0	0	1	0	3
KU_1	t_8	1	1	1	1	1	0	0	0	0	1	0	6
ME_1	t_8	0	0	1	0	1	0	0	0	0	1	0	3
NO_1	t_8	0	0	0	0	1	0	0	0	0	1	0	2
		1	1	4	1	5	0	0	0	0	5	0	17

Figure T11. NEBs for each evidence type across countries and their sums for Task 11.

FI_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
KO_1	t_8	0	1	0	1	0	0	0	0	0	0	0	2
KU_1	t_8	0	1	1	0	0	0	0	0	0	0	0	2
ME_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
NO_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
		0	2	1	1	0	0	0	0	0	0	0	4

Figure T12. NEBs for each evidence type across countries and their sums for Task 12.

FI_1	t_8	0	0	1	0	0	0	0	0	0	0	0	1
KO_1	t_8	0	0	1	0	0	0	0	0	0	1	0	2
KU_1	t_8	0	1	1	0	0	0	0	0	0	0	0	2
ME_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
NO_1	t_8	0	0	0	0	0	0	0	0	0	0	0	0
		0	1	3	0	0	0	0	0	0	1	0	5

Figure T14. NEB matrix for each evidence type across countries and their sums for Task 14.

FI_1	t_8	0	0	0	0	0	0	0	0	0	1	0	1
KO_1	t_8	0	0	0	0	0	0	0	0	0	1	0	1
KU_1	t_8	0	0	0	0	0	0	0	0	0	1	0	1
ME_1	t_8	0	0	0	0	0	0	0	0	0	1	0	1
NO_1	t_8	0	0	0	0	0	0	0	0	0	1	0	1
		0	0	0	0	0	0	0	0	0	5	0	5

Figure T15. NEB matrix for each evidence type across countries and their sums for Task 15.

FI_1	t_8	0	1	0	1	1	0	0	1	0	0	0	4
KO_1	t_8	0	1	0	1	1	0	0	1	0	0	0	4
KU_1	t_8	0	0	0	1	1	0	0	1	0	0	0	3
ME_1	t_8	0	1	0	1	1	0	0	1	0	0	0	4
NO_1	t_8	0	1	0	1	1	0	0	1	0	0	0	4
		0	4	0	5	5	0	0	5	0	0	0	19

Figure T18. NEB matrix for each evidence type across countries and their sums for Task 18.