MATHEMATICS ASSESSMENT IN THE RACE TO THE TOP ERA: AN EXPLORATORY STUDY OF THE SEMIOTIC RESOURCES IN LARGE-SCALE ASSESSMENT AND THEIR USE BY EMERGENT AND NON-EMERGENT BILINGUAL STUDENTS

by

CHELSEY SHADE

B.A., University of Colorado Boulder, 2007
M.A., University of Colorado Boulder, 2010

A dissertation submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
School of Education
Department of Educational Equity and Cultural Diversity
2017
The dissertation entitled:
Mathematics Assessment in the Race to the Top Era: An Exploratory Study of the
Semiotic Resources in Large-Scale Assessment and Their Use by Emergent and Non-
Emergent Bilingual Students
written by Chelsey Shade
has been approved for the School of Education,
Department of Educational Equity and Cultural Diversity
at the University of Colorado Boulder

Kathy Escamilla, Ph.D., School of Education
(Committee Chair)

Guillermo Solano-Flores, Ph.D., Graduate School of Education, Stanford University
(Committee Co-Chair)

Arturo Aldama, Ph.D., Department of Ethnic Studies

Susan Hopewell, Ph.D., School of Education

Edd Taylor, Ph.D., School of Education

Date _____________

The final copy of this dissertation 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.

IRB protocol # 14-0625
Current policy mandates that emergent bilingual (EB) students partake in standardized assessments before they are fully proficient in English. Additionally, standardized assessments are quickly converting to an online administration. The increase in design features inherent in computer-based assessments will most likely increase the number of construct-irrelevant factors and affect the accessibility of assessment items for EBs.

This study used frequency analyses, descriptive statistics, item semiotic complexity measures, Spearman’s rank-order correlation, one- and two-way analysis of variance, and Chi-square test of independence to examine the semiotic components and complexity of items included in a Smarter Balanced Assessment Consortium Grade 8 practice test, and how the semiotic features of the items affected non-EB and EB students as they solved four of the items.

Results indicate that the SBAC items have different ideational, interpersonal, and textual metafunctions as well as intersemiotic relations than items in curriculum resources. The average item semiotic complexity was approximately 69 components, indicating that students had to correctly interpret, as intended by the item writers, 69
different constituents to answer an item correctly. Regardless of language group, students’ reported actions and thinking involved very few semiotic components or intersemiotic relations belonging to the test items.

For all items combined, EBs had a higher total cognitive load than non-EBs. As item semiotic complexity increased, EBs’ scores decreased while non-EBs’ scores remained constant once a certain item semiotic complexity level was reached. Results indicate that as cognitive load decreased, total score increased for both language groups. Additionally, total cognitive load negatively affected both groups of students during problem-solving and non-EBs during interpretation.

This study provides the PARCC and SBAC assessment consortia with information relevant to designing test items using multiple semiotic resources used in computer-based assessments in ways that are sensitive to the characteristics of EB students. Results from this study contribute to the establishment of a research agenda on the relationship between the semiotic properties of test items and the achievement of EB students in large-scale tests. Additionally, the information gained from this study informs teachers on how they can use semiotic resources to support their students in the classroom.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Section summary .................................................. 39
How multimodal texts represent meaning .................. 40
Emergent bilingual students .................................. 40
Monolingual students .......................................... 42
Section summary .................................................. 45
How students use and construct multimodal texts ....... 45
Emergent bilingual students .................................. 45
Monolingual students .......................................... 48
Section summary .................................................. 50
Semiotics and the testing of linguistically diverse populations... 51
Section summary .................................................. 64
Cognitive Load Theory ........................................... 65
Working memory subsystems .................................. 65
Section summary .................................................. 73
Cognitive load: Intrinsic, extraneous, and germane cognitive loads ...................................................... 74
Section summary .................................................. 78
Summary of Literature Review .................................. 78

IV. METHODOLOGY .................................................. 81
Goals and Research Questions .................................. 81
General Information ............................................... 81
Context ................................................................. 81
Participants ............................................................ 83
Recruitment ............................................................ 87
Researcher’s role ...................................................... 88
Test materials .......................................................... 89
Study 1: Semiotic Resources Used in SBAC Test Items ....... 90
Coding ................................................................. 90
Analysis ................................................................. 95
Study 2: EB and Non-EB Students’ Sense Making of Semiotic Resources in SBAC Items

Instruments and procedures.................................................................97
Data collection......................................................................................98
Coding.................................................................................................99
Analysis...............................................................................................103

V. RESULTS............................................................................................106

Overview..............................................................................................106
Study 1.................................................................................................106
Study 2.................................................................................................106

Study 1: Semiotic Resources Used in SBAC Test Items.........................108
Semiotic composition.............................................................................108
Natural language.....................................................................................108
Mathematical symbols and expressions................................................110
Visual representations..........................................................................111
Semiotic resource complexities and item semiotic complexity.............114
Intersemiosis.........................................................................................117
Description of the items used in the cognitive interviews.....................119

Study 2: EB and Non-EB Students’ Making Sense of Semiotic Resources
in SBAC Items........................................................................................126
Overall performance and connection to item semiotic complexity...........126
Total cognitive load and association between item score and item
semiotic complexity...............................................................................129
Semiotic composition............................................................................132
Intersemiosis.........................................................................................140
Responding to items............................................................................141
Interpreting..........................................................................................142
Problem-solving...................................................................................143
Responding..........................................................................................144
Total cognitive load and responding to items............. 148

VI. DISCUSSION.......................................................................... 149
    Overview.................................................................................. 149
    Summary of Results................................................................. 150
        Study 1................................................................................. 150
        Study 2................................................................................. 151
    Findings as They Relate to the Conceptual Framework........... 154
    Implications............................................................................... 157
    Limitations................................................................................. 158
    Future Research....................................................................... 160
    Conclusion................................................................................ 161

REFERENCES.................................................................................. 162

APPENDIX

A. TEACHER RECRUITMENT LETTER........................................... 172
B. TEACHER CONSENT FORM...................................................... 174
C. STUDENT ASSENT FORM – ENGLISH...................................... 177
D. STUDENT ASSENT FORM – SPANISH....................................... 178
E. PARENT PERMISSION FORM – ENGLISH................................. 179
F. PARENT PERMISSION FORM – SPANISH................................... 182
G. ITEMS FROM THE 2015 SBAC GRADE 8 MATH PRACTICE
    TEST........................................................................................ 185
H. CODING FRAMEWORK FOR ITEM ANALYSIS............................ 200
I. EXAMPLES OF ITEM CODING.................................................... 208
J. PILOT STAGE OF THE DEVELOPMENT OF INSTRUMENTS........... 215
K. COGNITIVE INTERVIEW PROTOCOL: VERSION 1 – ENGLISH.... 226
L. COGNITIVE INTERVIEW PROTOCOL: VERSION 1 – SPANISH.... 235
M. SUMMARY OF SELECTED 2015 SBAC GRADE 8 MATH PRACTICE
    TEST ITEMS.............................................................................. 246
N. COGNITIVE INTERVIEW PROTOCOL: VERSION 2 – ENGLISH.... 249
O. COGNITIVE INTERVIEW PROTOCOL: VERSION 2 – SPANISH.... 256
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

P. CODING FRAMEWORK FOR INTERVIEW ANALYSIS..................264
Q. EXAMPLES OF INTERVIEW CODING........................................269
R. DESCRIPTIVE STATISTICS AND TWO-WAY ANALYSIS OF
   VARIANCE RESULTS FOR SEMIOTIC COMPOSITION.....................272
S. DESCRIPTIVE STATISTICS AND TWO-WAY ANALYSIS OF
   VARIANCE RESULTS FOR THREE-PART SEQUENCE OF
   RESPONDING TO AN ITEM..........................................................278
TABLES

Table

4.1. Frequencies and Percentages (Rounded) of Sociodemographic Variables for Participating High Schools.................................................................................................................. 82
4.2. Frequencies and Percentages (Rounded) of Sociodemographic Variables for Cognitive Interview Participants.......................................................................................... 85
4.3. Student Sociodemographic Variables for Classes Containing Cognitive Interview Participants............................................................................................................................................... 86
4.4. Sociodemographic Variables for Teachers............................................................................................................................................................................................................. 87
4.5. Item Coding System – Semiotic Components.............................................................................................................................................................................................................. 93
4.6. Functional Categories.............................................................................................................................................................................................................................. 95
4.7. Interview Coding System........................................................................................................................................................................................................................ 102
5.1. Frequencies and Percentages (Rounded) of the Semiotic Components of Natural Language........................................................................................................................................................................................................ 109
5.2. Frequencies and Percentages (Rounded) of the Semiotic Components of Mathematical Symbols and Expressions .............................................................................................................. 111
5.3. Frequencies and Percentages (Rounded) of the Semiotic Components of Visual Representations........................................................................................................................................................................................................ 113
5.4. Descriptive Statistics for Semiotic Resource Complexities and Item Semiotic Complexity........................................................................................................................................................................................................ 115
5.5. Descriptive Statistics for Semiotic Resource Complexities and Item Semiotic Complexity Disaggregated by Item Type and Knowledge Domain.............................................................................................................................................................................................................. 116
5.6. Semiotic Components of Item 1........................................................................................................................................................................................................................................... 121
5.7. Semiotic Components of Item 5........................................................................................................................................................................................................................................... 122
5.8. Semiotic Components of Item 9........................................................................................................................................................................................................................................... 123
5.9. Semiotic Components of Item 24........................................................................................................................................................................................................................................... 124
5.10. Cognitive Interview Items: Type, Knowledge Domain, Semiotic Resource Complexities, and Item Semiotic Complexity................................................................................................................................................................................................................... 125
5.11. Cognitive Interview Items: Intersemiotic Relations............................................................................................................................................................................................................................ 126
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

5.12. Cognitive Interview Items: Descriptive Statistics for Scores on Each Item and Total Score.................................................................................................................................126
5.13. Cognitive Interview Items: Descriptive Statistics for Total Cognitive Load... 129
5.15. Frequency, Proportion, and Chi-Square Test of Independence Results for the Metafunctional Components of the Semiotic Resources.................................................................133
5.16. Summary of Two-Way ANOVA Results for Semiotic Resources...............138
5.17. Frequency, Proportion, and Chi-Square Test of Independence Results for Intersemiotic Relations..........................................................................................................................141
5.18. Summary of Two-Way ANOVA Results for Responding to Items...............146
5.19. Spearman’s Rank-Order Correlation Between Total Cognitive Load and Variables Pertaining to Interpreting, Problem-Solving, and Responding........148
FIGURES

2.1. Example of intersemiosis.................................................................20
2.2. Potential semiotic resource complexities and item semiotic complexity of mathematics test item..........................................................24
2.3. Item 8 from the 2015 SBAC Grade 8 Math Practice Test..................25
2.4. Three-part sequence of responding to mathematics assessment items........33
4.1. Percent of Grade 9 and Grade 10 students meeting each proficiency level on the 2014 state-mandated mathematics assessment....................................83
4.2. Example of a Spanish pop-up glossary for the phrase, that makes in Item 20 from the 2017 SBAC Grade 8 Math Practice Test........................................90
5.1. Semiotic resource complexities and item semiotic complexity for each item..114
5.2. Relative frequency of each intersemiotic relation for each pair of semiotic resources..............................................................................117
5.3. Relative frequency of each intersemiotic relation for each pair of semiotic resources disaggregated by item type..................................................118
5.4. Relative frequency of each intersemiotic relation for each pair of semiotic resources disaggregated by knowledge domain.................................119
5.5. Average score on each item and total score disaggregated by language group .........................................................................................127
5.6. Relation between item semiotic complexity and average score (percent) disaggregated by language group.........................................................128
5.7. Average total cognitive load of each item and the items combined disaggregated by language group.................................................................130
5.8. Relation between item semiotic complexity and average total cognitive load disaggregated by language group.........................................................132
5.9. Estimated marginal means of components pertaining to the NL, MSE, and VR semiotic resources and the ideational, interpersonal, and textual metafunctions..........................................................139
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

5.10. Estimated marginal means of the dependent variables pertaining to the three-part sequence of responding to an item..........................................................147
SUMMARY

Since the enactment of the 2001 No Child Left Behind Act (NCLB), the number and importance of standardized assessments have been on the rise. Concurrently, emergent bilingual (EB) students are the fastest growing school-aged population (Flynn & Hill, 2005). Under NCLB, these students are required to participate in the assessments before they are fully proficient in the English language (Lohman, 2010). Thus, a great deal of measurement error is due to their English language proficiency (Solano-Flores & Li, 2013), and students, teachers, and schools are being penalized.

Scholars have only recently begun to examine and understand the effectiveness of certain accommodations for EB students on standardized assessments. Now that research is beginning to understand how to control for measurement error due to language proficiency, the format of the assessments is changing. Both the Smarter Balanced Assessment Consortium (SBAC) and the Partnership for Assessment of Readiness for College and Careers received grant monies to, in part, create a computer-based test administration (U.S. Department of Education, 2014). The increase in design features and tools inherent in computer-based assessments will most likely increase the number of construct-irrelevant factors and affect the accessibility of assessment items for EBs (Djonov, 2008).

Using research and theories from the fields of semiotics and cognitive science, the purposes of this dissertation was to examine the semiotic properties of SBAC mathematics assessments items, and how EB and non-EBs interact with these properties as they respond to the items. I conducted two studies to address these issues. The first study analyzed the semiotic components and complexity of the items on the 2015 SBAC
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Grade 8 Math Practice Test. The second study investigated how students interacted with the semiotic properties of the items by analyzing data from 60 cognitive interviews with eleventh-grade high school students (30 EBs and 30 non-EBs).

Results indicate that the SBAC items have different ideational, interpersonal, and textual metafunctions as well as intersemiotic relations than items in curriculum resources. The average item semiotic complexity was approximately 69 components, indicating that students had to correctly interpret these different constituents to answer an item correctly. Regardless of language group, students’ reported actions and thinking involved very few semiotic components or intersemiotic relations belonging to the test items.

For all item combined, EBs had a higher total cognitive load than non-EBs. As item semiotic complexity increased, EBs’ scored decreased while non-EBs’ scores remained constant once a certain item semiotic complexity level was reached. Results indicate that as cognitive load decreased, total score increased for both language groups. Additionally, total cognitive load negatively affected both groups of students during problem-solving and non-EBs during interpretation.

This study provides the PARCC and SBAC assessment consortia with information relevant to designing test items using multiple semiotic resources used in computer-based assessments in ways that are sensitive to the characteristics of EB students. Results from this study contribute to the establishment of a research agenda on the relationship between the semiotic properties of test items and the achievement of EB students in large-scale tests. Additionally, the information gained from this study informs teachers on how they can use semiotic resources to support their students in the classroom.
CHAPTER I

INTRODUCTION

Throughout history, countries have experienced shifts in population due to political, economic, and natural factors. These shifts impact education because they change the student demographics in the classroom. Educational practices and policies must keep pace with changing times to educate evolving student bodies.

Within the United States, emergent bilingual students (EBs) pose a complex set of challenges. Emergent bilinguals can be defined as students who will become bilingual through school and through acquiring English (García, Kleifgen, & Falchi, 2008)—While most policy documents and research use terms such as *English learners* or *English language learners* (ELLs), *second language learners*, and *students with limited English proficiency* (LEPs) to refer to these students, the term *emergent bilingual* is used in this dissertation in an attempt to eliminate deficit perspectives of students with linguistically diverse backgrounds and skills.

EB students constitute the fastest growing school-aged population. According to Flynn and Hill (2005), while the overall number of school-aged children increased by 19% between 1979 and 2003, the number of EBs grew by 124%. In the 2012-2013 school year, approximately 9.2% of the public school student population, an estimated 4.4 million students, were classified as *English learners* (Institute of Education Sciences, 2015).
Due to this rapid change in student demographics and the emphasis on high-stakes assessments, test developers need to be mindful of EBs throughout the process of assessment design. When EBs are not taken into consideration, high-stakes assessments incorrectly measure their content knowledge and confound it with language proficiency. These practices have harmful consequences. As an example, test results may be used incorrectly to place EB students into special education classes (Klingner, Artiles, & Barletta, 2006).

This study provides assessment designers with information relevant to designing test items using several semiotic resources in computer-based mathematics assessments in ways that are sensitive to the unique skills and characteristics of EB students. Furthermore, this study helps educators and educational researchers understand how students, especially EBs, use semiotic resources as they respond to assessment items and how they can be supported during classroom instruction and test preparation activities.

Overview and Definitions

This dissertation addresses aspects of social semiotics and cognitive perspectives in the assessment of EBs in mathematics. The critical concepts of social semiotics, working memory, cognitive load, and cognitive overload, are defined in this section, as they are key ideas being discussed throughout the study.

Social semiotics is the study of how signs are used in society to create meaning, and a semiotic resource is any system of meaning making (Halliday, 1978). A text is the material realization of a sign system that is created when different forms of social organization engage with the semiotic resource, such as novels, songs, and mathematical proofs (Halliday, 1978; Hodge & Kress, 1988). Multimodal social semiotics is the study
of how readers of multimodal texts (i.e., texts comprising more than one semiotic resource) use multiple semiotic resources separately and integratively—combining the meaning-making potentials of different semiotic resources (Royce as cited in Caple, 2008)—to produce one complete semantic unit (Caple, 2008). For this dissertation, I view all texts as multimodal, as there are infinite possible meanings that can be made from one text (see the Semiotics Perspectives section in Chapter II for a detailed description).

Along with the aforementioned concepts pertaining to social semiotics, working memory, cognitive load, and cognitive overload are also concepts critical to this study and will be defined. According to Baddeley (1986), working memory comprises a central executive system that coordinates two subsystems: the visuospatial sketchpad and the phonological loop (Baddeley as cited in Brünken, Plass, & Leutner, 2003). The visuospatial sketchpad stores visual and/or pictorial information, while the phonological loop is key to inner rehearsal of auditory material and aids in language comprehension (Verhoeven, Schnotz, & Paas, 2009).

The cognitive load of each subsystem can be defined as the sum of the intrinsic, the extraneous, and the germane cognitive loads. The intrinsic load is determined by the difficulty of the task, the number of elements that must be processed simultaneously, and the interactions between the complexity of the task and the ability level of the student. The extraneous load is determined by the amount of irrelevant information or number of stimuli in a task. The germane cognitive load is determined by the amount of cognitive resources students employ in the construction of schemata and automation.
overload is reached when the difference between the total cognitive load and the processing capacity of one of the subsystems approaches zero (Brünken et al., 2003).

Social semiotics, working memory, cognitive load, and cognitive overload are pertinent to my study as they explain how students, especially EBs, interpret and respond to test items and how their ability to do so may be affected by the semiotic composition and complexity of the items. The following section discusses the policy context under which standardized assessments are currently being developed and the need for this study.

Policy Context

This research took place in the context of the No Child Left Behind (NCLB) Act and the Race to the Top Fund—a grant program that rewards states and K-12 school districts for their educational reform movements (U.S. Department of Education, 2009). As they directly affect the assessment systems within the United States, they have particular relevance to this study.

No Child Left Behind Act. In 2001, the U.S. Congress reauthorized the Elementary and Secondary Education Act under NCLB, in which assessment was a key component (Lohman, 2010). NCLB was premised on the belief that higher standards and standardized assessments would improve America’s education system. Under NCLB, states must create measureable objectives and assessments to measure English language arts and mathematics skills for all students in grades 3-8 and tenth grade. Each school must assess at least 95% of its enrolled students, including 95% of the students in the following subpopulations: (a) economically disadvantaged students, (b) minority students, (c) students with disabilities, and (d) students with limited English proficiency.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

For schools to meet annual yearly progress, all students, including those in the four subpopulations, must meet or exceed the objectives. Failure to make annual yearly progress for two consecutive years can lead to sanctions, including school transfer options and/or corrective actions, such as replacing school staff; implementing new curriculum; and restructuring the internal organization of the school (Lohman, 2010).

**Race to the Top Fund.** In 2009, President Obama signed the American Recovery and Reinvestment Act into law. This act allocated 4.35 billion dollars to the Race to the Top Fund, which works concurrently with NCLB (U.S. Department of Education, 2009). To receive grant monies, states must implement reform plans in four separate areas:

1. Standards and Assessments—to prepare students to be successful in college and the workplace.
2. Data Systems—to measure student success and growth and inform instruction.
3. Teacher and Principal Development—to employ, develop, reward, and retain effective teachers and principals.
4. School Turn Around—to turn around low-achieving schools and substantially raise the achievement of students.

To fulfill the first criteria (Standards and Assessments), a state must participate in a consortium of states that cooperatively develops valid and reliable K-12 standards and assessments (U.S. Department of Education, 2009).

In 2010, the Race to the Top Assessment Program awarded approximately 180 million dollars to the Smarter Balanced Assessment Consortium (SBAC) and the Partnership for Assessment of Readiness for College and Careers (PARCC) to develop assessments that measure students’ college and career readiness and their achievement in
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

meeting these standards. Additionally, both SBAC and PARCC received a supplemental grant of 15.9 million dollars to analyze the gap between current and new standards; to support educators in interpreting assessment results and implementing new standards; and to enhance the technology used in the assessment systems, including computer-based test administration (U.S. Department of Education, 2014).

Assessment for Emergent Bilinguals Context

The enactment of NCLB highlighted the “achievement gap” between disadvantaged students (defined by NCLB as students from the four subpopulations mentioned above) and non-disadvantaged students because the latter usually outperform the former on standardized assessments. Many argue that the achievement gap between mainstream and emergent bilingual students is not due to academic factors, but rather due to other social factors such as opportunity to learn and socioeconomic status (Darling-Hammond, 2000; Nasir, Hand, & Taylor, 2008; Wright & Li, 2008). Many researchers in the fields of bilingual education and teaching English as a second language attribute the gap to differences in English language proficiency. Content tests rely heavily on students’ English language proficiency (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Abedi, Hofstetter, & Lord, 2010). Therefore, a great deal of measurement error in the testing of EB students is due to their proficiency in the language in which tests are administered (Solano-Flores & Li, 2013). Language, whether it be language proficiency of the student or the language of the test, is a source of construct-irrelevant variance that, in this case, systematically deflates EBs’ test scores due to systematic or uncontrolled measurement error (Haladyna & Downing, 2004). An indication that the importance of
this source of measurement error should not be underestimated is that research in related fields have investigated methods to control for the measurement error introduced by language in hopes of creating more valid assessments for EBs.

To comply with legislation that mandates EBs be included in assessment programs, efforts are made to minimize language as a construct-irrelevant factor. For example, EBs are provided with testing accommodations. These accommodations are modifications made to the test or testing procedures with the intent of minimizing measurement error due to language proficiency without changing the construct being assessed (Abedi, Hofstetter, & Lord, 2004).

There are a vast number of accommodations provided to EBs on standardized assessments, such as dictionaries, extended time, native language directions, or a separate testing environment. One of the major challenges in properly testing EBs is the lack of sound practices regarding the use of accommodations. A review made by Rivera and Collum (2004) revealed that there are 75 accommodations provided to EBs under states’ assessment policies. Many of those accommodations are borrowed from special education, are not necessarily designed to help minimize measurement error due to language proficiency, and are not appropriate for EBs. In fact, only 59% of the listed EB accommodations are relevant to the needs of these students, while the other 41% are only relevant to students with disabilities (Rivera & Collum, 2004).

Researchers have begun to examine the effectiveness of different accommodations provided to EBs. Abedi, Hofstetter, and Lord (2004) found that the most promising accommodations for EBs are modified English assessments and customized dictionaries. Using previous research, Pennock-Roman and Rivera's (2011) meta-analysis of
accommodations for EBs and non-EBs suggests that modified English accommodations are appropriate for EBs with intermediate levels of English language proficiency (ELP). Native language versions of tests are the most beneficial for EBs with low levels of ELP and literacy skills in their first language and who have received instruction in their native language. These findings suggest that the most promising accommodations for tests with time restrictions are pop-up English dictionaries and/or glossaries used in computer-administered assessments. The most promising accommodations for tests with extended time or no time restrictions are dual language assessments, bilingual glossaries, and traditional paper versions of English dictionaries and/or glossaries.

Abedi et al. (2004) and Pennock-Roman and Rivera (2011) stress that the use of accommodations for EB students should not be a blanket approach because EBs are a large, heterogeneous group with diverse linguistic skills and background characteristics. Rather, the selection of accommodations for EBs needs to be tailored to fit the students’ levels of English proficiency, native language literacy skills, and/or exposure to native language instruction.

While this work is promising in reducing measurement error caused by language factors, it is limited to traditional, paper-and-pencil assessments. The recommended list of accommodations might not be as effective in computer-based assessments. Computer-based assessments introduce a whole new array of interactive features and tools, such as hyperlinks and audio files (Norman, 1988). Yet it is unknown whether the current set of accommodations help EBs access these features, or even whether these features will act as accommodations in their own right. Furthermore, computer-based assessment may
introduce a whole new set of construct-irrelevant factors that have yet to be identified or investigated (see Djonov, 2008; Jewitt, 2005; Royce, 2002).

**Problems, Implications, and Concerns**

Computer-based test administration affords opportunities for computer-adapted testing, computer-based grading, interactive features, and embedded accessibility resources, such as native language glossaries and stacked translations (Solano-Flores, Shade, & Chrzanowski, 2014). However, complications arise due to the wide range of features and resources that are available in computer-based assessments. For example, it is known that a large number of tools increases the amount and complexity of the different computer-based features (Norman, 1988). In addition, the complexity of the environment used to administer tests, such as illustrations, cascade menus, audio clips, and hyperlinks are sources of considerable variability (Djonov, 2008). Thus, while these tools can potentially support EBs to gain access to the content of items, the variety and complexity of this wide range of accessibility resources may also hamper their performance.

The effect of computer-based features used in standardized-assessments on the performance of EB students has not been examined in detail. Related research has examined the complexity of multiple semiotic resources in paper-and-pencil tests and its relation to item difficulty for EBs. Martiniello (2009), for example, examined the association between linguistic complexity, the presence of non-linguistic representations, and item difficulty for EB and non-EB students. Results from her study indicate that the differential item functioning of items with greater linguistic complexity favored non-EBs
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

over EBs. However, the magnitude of the differential item functioning was reduced for EBs when the item contained non-linguistic representations (Martiniello, 2009).

While non-textual representations may help EBs to interpret items, their use is not as simple as Martiniello suggests, as an abundance of visual information may have an opposite effect than the intended. Solano-Flores and Wang (2011) examined the use of illustrations as a form of assessment accommodation for EBs. They found evidence that visual and textual information interacted in students making sense of science items.

Solano-Flores et al. (2014) also examined the use of illustrations as a form of accommodations on a science assessment. Their findings suggest that non-EB students use text more effectively than EB students to correctly interpret illustrations, apparently because EBs’ limited reading proficiency in English may limit the ability to integrate information provided textually and non-textually (Solano-Flores, Wang, Kachchaf, Soltero-Gonzalez, & Nguyen-Le, 2014). Clearly, the role that non-textual representations play in helping EBs to gain access to the content of assessment items is shaped by multiple factors.

While research has begun to examine the relation between the complexity of different semiotic resources and student performance, the development of projects by PARCC and SBAC has been more rapid than the ability of researchers to examine these issues and inform the consortia about their use and effectiveness. In regards to these issues, PARCC is silent about illustrations and other semiotic resources in its documents. SBAC does address non-linguistic representations, although its efforts are insufficient to provide proper guidance for test development. For example, SBAC defines *Simple Texts* as tasks that contain extensive illustrations that directly support and help students to interpret
written text. *Very Complex Texts* are defined as tasks that contain minimal illustrations that support and help students to interpret written text (Measured Progress/ETS Collaborative, 2012). SBAC asserts that test developers should specify the connections between graphics and text (i.e., graphics that mirror the text, graphics that support the text, or graphics that are not part of the text). It also specifies the different types of symbols and labels that are to be used in graphics (Smarter Balanced Assessment Consortium, 2012). Nonetheless, SBAC has yet to generate documents that discuss the complexity of different semiotic resources or how semiotic complexity may influence the performance of EBs on tests.

**Problem Statement: Testing and Semiotics**

As mentioned above, technology affords opportunities for computer-adapted testing, computer-based grading, and the inclusion of interactive features. However, it may also make test-taking more challenging for students with limited accessibility to and familiarity with technology, especially for EB students. While research in the fields of multiliteracies and multimodal social semiotics addresses the effect of multimodal resources on classroom instruction (see Djonov, 2008; Halliday, 1993; Halliday, 1978; Hodge & Kress, 1988; Jewitt, 2005; Knox, 2008; Lemke, 2003; O’Halloran, 2005; O’Halloran, 2000; Royce, 2002; The New London Group, 2000), additional research is needed that examines the interaction between computer-based standardized assessments, multimodal resources, and accessibility resources and their influence on assessment practices for EB and non-EB students. Also, while teachers may use information technology in their classrooms, the new computer-based test administration will change the ways in which they prepare students for standardized assessments. Teachers will
need to support students to make meaning of the problems and content of items using the sets of semiotic resources used in computer-based standardized assessments and classroom instructional materials. Yet, limited access to computers may limit the set of opportunities for EBs (many of whom live in poverty) to become familiar with the multiple semiotic resources involved in computer-administered testing. Additionally, teachers need to have a better understanding of how semiotic resources affect the teaching and learning processes to support their students.

**Purpose and Research Questions**

The use of social semiotics as a theoretical lens can help researchers identify how EBs use language and other meaning-making systems, separately and integratively, to construct knowledge, interpret items, and demonstrate content knowledge within the classroom and on standardized assessments. Because mathematics is multimodal by nature, it is necessary to examine how the semiotic resources are used in the assessment of EBs.

SBAC and PARCC administered their assessments for the first time during the spring of 2015. This was the first time in the history of testing in the U.S. in which state-mandated, large-scale assessments were administered online. Due to the novelty of the computer-administered assessments, the relation between the semiotic resources used in computer-administered items and the students’ response processes, especially those of EBs, has not yet been investigated. There is a gap in the research that this study addressed.

The main purpose of the study was to provide information on variables that are relevant to future investigations on the role of semiotics in computer-based mathematics
assessment for EB students and how this information can guide teachers’ practices. I identified the semiotic composition and complexity of mathematics items and the meaning-making processes of students as they responded to these items.

More specifically, I asked two research questions:

1. What semiotic resources are used in SBAC test items?
2. What commonalities and differences can be observed between EB and non-EB students in the ways in which they make sense of the semiotic resources used in SBAC items?

Due to the uncharted nature of online assessment evaluation, this study was necessarily exploratory.

**Significance**

This study provides the PARCC and SBAC assessment consortia with information relevant to designing test items using multiple semiotic resources used in computer-based assessments in ways that are sensitive to the characteristics of EB students. Results from this study contribute to the establishment of a research agenda on the relationship between the semiotic properties of test items and the achievement of EB students in large-scale tests. Additionally, the information gained from this study informs teachers on how they can use semiotic resources to support their students in the classroom.
CHAPTER II

CONCEPTUAL FRAMEWORK

The conceptual framework for this dissertation is based on two theories: multimodal social semiotics and cognitive load theory. The former was the primary approach used to analyze the semiotic composition and complexity of mathematics assessment items. The latter was the secondary approach used to examine how working memory and cognitive load influenced students’ interpretations of and responses to items.

Semiotics Perspectives

Semiotics is a very broad field. Stated simply, it is the science of signs in society (Hodge & Kress, 1988). A semiotic resource is any system of meaning-making, such as oral and written language, visuals, music, gestures, numbers, symbols, and many more. The following subsections will describe and apply the origins of multimodal social semiotics, the metafunctions of semiotic resources, and the process of intersemiosis to mathematics assessment items.

Semiotics. Semiotics emerged as its own field of study at the beginning of the 20th century through the writings of the Swiss linguist, Ferdinand de Saussure, the American philosopher, Charles Sanders Peirce, and others. de Saussure defined semiotics as the study of what signs are made of and what laws command them with specific, almost exclusionary, focus on linguistic systems (Mick, 1986). He defined a sign as an arbitrary
relationship between a signifier and the signified. A signifier is the sound that forms in somebody’s mind (sound-image), and the signified is the meaning that the sound-image produces (Silverman, 1983). For de Saussure, we do not learn the world, but rather learn the codes that mediate our experiences with it (Thayer, as cited in Mick, 1986).

Peirce differs from de Saussure in his attention to the relationship between signs and objects. According to Peirce, the process of meaning-making is a complicated interaction that involves a sign (the signifier), an interpretant (the signified), and an object (the referent). A sign represents an idea of an object to the participant. This idea creates an equivalent or a more developed sign within the mind of the participant (the interpretant), which is governed by culturally specific rules of thought (Peirce as cited in Silverman, 1983). For Peirce, signs are more than tools we use to mediate our experiences with the world; there is an existential relationship between signs and objects in which our experiences with reality affect our meaning-making systems, and the process of signification constructs our reality.

Social semiotics and semiotic metafunctions. Social semiotics was created in response to some of the critiques of structural semiotics (e.g., the work of de Saussure and Peirce), including the argument that structuralists focus on words and sentences and ignore social dimensions of semiotic systems (Hodge & Kress, 1988). Instead of focusing on static sign structures, social semioticians analyze the dynamic, socially situated sign processes of texts—the material realizations of sign systems that are created when different forms of social organization engage with them (Halliday, 1978; Hodge & Kress, 1988; Iedema, 2003).
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Halliday is one of the leading figures in social semiotics, as he developed a theoretical framework (i.e., systemic functional linguistics) to examine the social dimensions of language. One aspect of his theory contends that language comprises three meaning-making metafunctions: ideational, interpersonal, and textual (Halliday, 1978, 2004).

The purpose of the ideational metafunction is to interpret our external and internal experiences with the world and to describe the purpose and/or context for the use of language. One way to examine the ideational metafunction of language and other semiotic resources is to identify the functional categories that are present. There are four types of functional categories: identification, activity, circumstances, and attributes (Halliday, 2004). Identification describes who or what are represented by the semiotic resource. Activity describes the processes and/or actions that are taking place between the actor(s) and the recipient(s) or object(s) of the action within the semiotic resource. Circumstances are the elements that are concerned with the setting, the participants not involved in the actions, or the participants used by the actors. Attributes are the qualities and characteristics of the participants (Royce, 2002).

The interpersonal metafunction describes how language is used to socially position the people involved in a situation. This involves examining the relationships that are present, such as how the text addresses the reader. Lastly, the textual metafunction describes the structural rituals of a text, including the construction, sequence, organization, and continuity of a text.

**Multimodal social semiotics and intersemiosis.** This reconceptualization of semiotics, as the study of socially situated texts, highlighted the importance of other
semiotic resources in accordance with linguistic ones (Unsworth, 2008). Under this framework, visuals, dance, gestures, dress, etc., along with language, are all examples of texts. Thus, semioticians interested in non-linguistic semiotic resources adapted Halliday’s theories of social semiotics to multimodal texts (i.e., texts comprising more than one semiotic resource) and formed the subfield of multimodal social semiotics.

According to Caple (2008), users of multimodal texts employ two or more different semiotic resources in combination to produce one complete semantic unit. It is important to understand that within a multimodal text, various semiotic resources are intended to construct meaning in distinct ways. Even though individuals create distinct meanings from each resource, there is a synergistic relationship between the different resources that allows for intersemiosis—the construction of one integrative meaning (Royce as cited in Caple, 2008).

Take, for example, a multimodal text comprising natural language and an image. The reader may construct meaning through the words, the image, and/or the integration of the words and image. Figure 2.1 shows Miller’s (1942) famous poster which is a multimodal text containing an image of a woman and the words, *We Can Do It!* The image by itself may invoke one’s knowledge of Rosie the Riveter, World War II, or concepts of industry and equality. The words, *We Can Do It!* may conjure feelings of solidarity, strength, and hope. Integratively, the image and words may evoke feminist theories, especially those developed during World War II.
Several scholars have devoted their work to examining how meaning is created across semiotic resources (see Jewitt, 2005; Knox, 2008; Royce, 2002). One such way is to apply Halliday and Hasan’s (1976) theory of lexical cohesion to understand the ideational relationship between visuals and language in multimodal texts. Their theory claims that texts create meaning by using one or several linguistic relations: (a) Repetition, (b) Synonymy, (c) Antonymy, (d) Hyponymy, (e) Metonymy, (f) Collocation, and (g) Identity. Lexical Repetition occurs when a lexical item is repeated; for example, *dog* in:

> Logan bought a dog. The dog was cute.

Lexical Synonymy occurs when a lexical item is synonymous to the item before it; for example, *yell* and *scream*. Lexical Antonymy occurs when the lexical items have opposite meanings, such as *win* and *lose*. Lexical Hyponymy and Metonymy are difficult
to detangle at times. The former occurs when there is a class-subclass relationship between items; for instance, mammals and bears. The latter occurs when there is a part-whole relationship between lexical items, such as flower and stem. Lexical Collocation occurs when the presence of one item is expected with the presence of another; for example, fire and smoke. Lastly, lexical Identity refers to the identity of reference that occurs when the items are synonyms of the same or some higher level of generality (Halliday, 2004, p. 573). The words bluebirds and birds, for example, are synonymous. However, the word, birds is at a higher level of generality and can be used to identify bluebirds, while the word, bluebirds cannot be used to identify all birds.

**Multimodal social semiotics and mathematics.** Lemke (2003) applies the theories of multimodal social semiotics to describe the kind of meaning that users of mathematics construct and the evolution of the semiotic resources used to establish such meaning-making processes. According to Lemke, mathematics is a system of related social practices that cannot be simply identified by the systems of signs that are used, but rather by the kinds of meanings they can create (e.g., addition, division, and proportions). Lemke argues that natural language is ill-equipped to allow individuals to provide useful descriptions of natural phenomena. Thus, mathematics was created to fill this void and allow people to make sense of the world that varies categorically and/or continuously using three semiotic resources: natural language, mathematics (symbols and expressions [MSE]), and visual representations.

Using Halliday’s theories, O’Halloran (2005) identifies the metafunctions of the three semiotic resources of mathematics used in curriculum resources. The main ideational function of natural language is to construct relational processes such as be, have, and
Mathematical symbols are used to encode meaning unambiguously and efficiently. Ideationally, symbols are primarily concerned with relations between mathematical elements and variables. Meaning is achieved through the operative process, which is governed by the Rule of Order (e.g., parentheses, exponents, multiplication and division, and addition and subtraction [PEMDAS]). Interpersonally, symbols are employed to construct an authoritative and non-negotiable relationship with the reader. Textual meaning is achieved through highly conventionalized notation and spatial positioning.

The ideational metafunction of visual representations is to provide intuitive and perceptual depictions of the world. Visuals are used to create logical meaning through the employment of spatial and simultaneous ordering principles, unlike language and symbols that construct meaning temporally and sequentially (Hull & Nelson, 2005). Interpersonally, visuals hold a high-truth value and are used to direct the reader to the most important aspects of the visual. Textually, visuals are used to focus the attention of
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

the viewer to the experiential aspect of the mathematical content of the image (de Oliveira & Cheng, 2011).

This dissertation expands on the aforementioned theories by introducing the ideas of semiotic resource complexity and item semiotic complexity. I define semiotic resource complexity as the summation of the metafunctional components of a semiotic resource. For the purposes of this dissertation, I am only examining the semiotic resource complexities of NL, MSE, and VR, as modeled by equations 2.1, 2.2, and 2.3:

\[
NL_{C} = \sum NL \text{ Ideational components} + \sum NL \text{ Interpersonal components} + \sum NL \text{ Textual components} \quad [\text{Eq.2.1}]
\]

\[
MSE_{C} = \sum MSE \text{ Ideational components} + \sum MSE \text{ Interpersonal components} + \sum MSE \text{ Textual components} \quad [\text{Eq.2.2}]
\]

\[
VR_{C} = \sum VR \text{ Ideational components} + \sum VR \text{ Interpersonal components} + \sum VR \text{ Textual components} \quad [\text{Eq.2.3}]
\]

where \( NL_{C} \) represents the semiotic resource complexity of NL, \( MSE_{C} \) represents the semiotic resource complexity of MSE, and \( VR_{C} \) represents the semiotic resource complexity of VR.

I define item semiotic complexity as the summation of the three semiotic resource complexities. An item’s semiotic complexity (\( I_{C} \)) is given by the following equation:

\[
I_{C} = NL_{C} + MSE_{C} + VR_{C} \quad [\text{Eq.2.4}]
\]

Figure 2.2 models the potential semiotic resource complexities and the item semiotic complexity of a mathematics test item.
Figure 2.2. Potential semiotic resource complexities and item semiotic complexity of a mathematics test item.
Application. All texts are multimodal (Hodge & Kress, 1988), including assessment items. Figure 2.3 shows a multimodal test item from the 2015 SBAC Grade 8 Math Practice Test. This item contains natural language (For each number, indicate whether it is rational or irrational, Rational, and Irrational), MSE ($\frac{4}{7}$, $\sqrt{30}$, $\frac{21}{\sqrt{4}}$, $\pi$, $-27$), and visual representations (the categorical table and the boxes for students to check).

![Image of test item]


Due to the complexity of the conceptual framework, I will only expand upon some of the metafunctional features of each semiotic resource. One of the ideational purposes of the natural language is to tell students what to do (1), thus the interpersonal metafunction is directive (2). Textually, it contains one sentence (3), uses bold words (4), and is placed above (5) and inside the visual representation (6). Thus, the natural language semiotic
resource complexity \((NL_C)\) equals 6 (one ideational component, one interpersonal component, and three textual components, as labeled above).

Ideationally, the MSE represents rational (1) and irrational numbers (2). It has a non-negotiable interpersonal relationship with students (3). Textually, it contains Arabic numerals (4), operations (5), and symbols (6). The MSE semiotic resource complexity \((MSE_C)\) for this item is also 6.

One of the ideational purposes of the VR is to represent categories (1). Interpersonally, the VR is indicative (2); it indicates to students how they are supposed to answer the question. Textually, it is a schematic representation (3), it is located below the text (4) and students must use it to select their answers (5). The VR semiotic resource complexity \((VR_C)\) is 5.

Thus, the item semiotic complexity \((I_C)\), albeit only partial, for this problem is 17 as modeled in the equations below:

\[
I_C = NL_C + MSE_C + VR_C = 6 + 6 + 5 = 17
\]

[Eq.2.5]

(Note that these are not all the possible metafunctional components, simply a few of them.)

Now we will examine the process of intersemiosis. In this problem, there is Intersemiotic Collocation between NL and VR because the NL tells students to indicate whether each number is rational or irrational, thus it is expected that there will be a visual space for students to select their answers. Additionally, there is Intersemiotic Identity
between the NL and MSE because the text says, *each number* and the MSE is used to identify the specific numbers.

**Cognitive Perspectives**

Semiotic resources are used to interpret and represent our experiences with the world; they allow us to develop socially and cognitively as we learn in our semiotic environment. Cognition is the set of all mental processes related to knowledge, such as attention, memory, evaluation, computation, and problem solving (Von Eckardt, 1995). Thus, the cognitive processes of learning are just as critical to knowledge construction as are the tools that individuals use to construct meaning. The following subsections will describe and apply cognitive perspectives of memory, working memory more specifically, and cognitive load to the process of responding to mathematics test items.

**Memory.** Central to cognition is memory—the process in which information is encoded (received, processed, and combined with other received information), stored, and retrieved for the use in a process or activity (Tyler, Hertel, McCallum, & Ellis, 1979). Information is stored in three different types of memory: sensory memory, working memory, and long-term memory. Several experimental studies have examined the types of information encoded in each type of memory, as well as the number of pieces of information each type can attend to and/or the duration of each. According to these investigations, sensory information (e.g., sight, taste, and sound) is retained in the sensory memory for less than one second after the information has been encoded (Walsh, 1978). Working memory encodes information acoustically and, to a lesser extent, visually. According to Miller (as cited in Sweller, 2005), working memory can only hold approximately seven pieces of information and can only integrate between two and four
Working memory is also constrained by its duration; it can only attend to one piece of information for approximately 20 seconds without rehearsal (Peterson & Peterson as cited in Sweller, 2005). Long-term memory encodes information semantically and can store much larger quantities of information for an undetermined amount of time, potentially unlimited (Baddeley & Warrington, 1970).

According to Sweller (2005), knowledge is made possible through working memory and long-term memory. In fact, learning only occurs if there has been a change in one’s long-term memory. This change can be defined as the acquisition of schemas that organize multiple pieces of information into one element. An individual can process a schema automatically or consciously depending on his or her knowledge of it and the frequency with which it is used and practiced. For example, the multiplication of two single-digit numbers (e.g., 2x4) is automatic for many students. This schema allows students to solve more complex computations, such as the multiplication of multi-digit numbers.

Working memory comprises a central executive system and two subsystems. The central executive system is composed of the schemas from the long-term memory. These schemas tell the working memory what to do, how to do it, and when to do it. If pre-existing schemas cannot be applied to new information, the working memory must solve the problem at hand and evaluate the proposed solution. This information will then be organized and integrated with prior knowledge (Oviatt, 2006; Sweller, 2005; Verhoeven et al., 2009).

The two subsystems of the working memory are the phonological loop and the visuospatial sketchpad (Baddeley as cited in Brünken, Plass, & Leutner, 2003). The
phonological loop allows for inner rehearsal of auditory material and aids in language comprehension (Verhoeven et al., 2009). The visuospatial sketchpad retains the visual and/or pictorial information gained from one’s visual perception of reality and images. Also, according to Mayer's (2001) dual-coding theory, verbal and pictorial information are processed and represented in two separate but interrelated systems within the visuospatial sketchpad. These systems have limited capacity and are independent.

**Cognitive load.** Cognitive load is the total amount of mental effort being used in the working memory and there are three types: intrinsic, extraneous, and germane (Sweller, 2005). The intrinsic load is determined by the nature of the learning task, the number of elements that must be processed simultaneously, and the interaction between the complexity of the task and the learner’s level of expertise. The intrinsic load of an item from an eighth-grade mathematics test should be less for a junior in high school than it is for a student in eighth grade because the older student should have more expertise. The extraneous cognitive load is determined by the presentation of the task and extraneous features. For example, non-essential pictures or background music may distract some of the student’s cognitive resources away from attending to the essential pieces of information. The germane cognitive load is determined by the amount of resources that learners invest in the construction of schemata and automation. For instance, students can support the construction of schemata by summarizing what they have read or building concept maps of the material (Brünken et al., 2003; van Merriënboer & Ayres, 2005). DeLeeuw and Mayer (2008) contend that the intrinsic cognitive load can be measured by having students rate their mental effort after each item; the germane cognitive load can be measured by having students measure the level of difficulty of the
lesson at the end of the instructional materials; and the extraneous cognitive load can be measured by adding a secondary task to a lesson and measuring students’ response times to this task.

For each subsystem, the total cognitive load can be defined as the sum of the intrinsic, the extraneous, and the germane cognitive loads. Cognitive overload is reached when the difference between the total cognitive load and the processing capacity of one of the subsystems approaches zero (Brünken et al., 2003). Cognitive load theorists contend that the three types of cognitive load are additive; if one type increases, then another must decrease to avoid overload.

**Application.** Using the aforementioned theories, I measured the intrinsic cognitive load by having students rate their mental effort after each item on a scale of one to 10. I measured the germane cognitive load by having the students rate the level of difficulty of all the items combined on a scale of one to 10. I measured the extraneous cognitive load by having students identify extraneous features in each item because I did not have the capability to add a secondary task to the items. Using the theory that cognitive load is additive, I define the total cognitive load ($CL_T$) of an item as the summation of the intrinsic rating, the germane rating, and the number of extraneous features present in an item, as given by the following equation:

$$CL_T = \text{Intrinsic Item Rating} + \text{Germane Rating} + \sum \text{Extraneous features} \quad [\text{Eq.2.6}]$$
where Intrinsic Item Rating and Germane Rating are between one and 10 and the summation of extraneous features does not have a limit. Cognitive overload can be examined with the total cognitive load negatively affects student performance.

A multimodal text, as is the case of items on a mathematics test, is defined by the semiotic resources that compose it. Thus, the total cognitive load of an item is determined by the presence or absence of semiotic resources and their components. Semiotic components related to the content of the item affect the intrinsic cognitive load. For instance, the item, \(2 + 2 = ?\) has a lower intrinsic cognitive load than the item, *Solve the following equation for* \(x\): \(2x + 5 = 12\) because there are more semiotic components. The extraneous cognitive load of an item is affected by the non-content related semiotic resources and/or components, such as the presence of a visual representation of a car in a problem concerning gas mileage. Lastly, the germane cognitive load is determined by the presence or absence of semiotic resources related to schema construction.

As semiotic systems are cultural tools developed and modified to meet certain functions in a society (Halliday, 1978), some people may struggle to make meaning from semiotic resources developed in cultures or languages that are different from their own. In the context of this dissertation, emergent bilingual (EB) students who are being tested in their second language may experience cognitive overload before their native-English speaking peers. Take, for example, an item that contains NL and MSE and that already has high intrinsic and extraneous cognitive loads. A native-English speaking student may not reach cognitive overload because he only needs to attend to elements in the pictorial representation of the MSE since the English language is an automatic schema in his long-term memory. An EB student may reach cognitive overload because not only does she
have to attend to the visual elements of the MSE, but she also must attend to the verbal components of the NL, thus creating an overload in the processing capacity of the visuospatial sketchpad.

If the semiotic components of an item cause cognitive overload for students, then their working memory is negatively affected. This could result in students being unable to recall the appropriate schema; or if they can recall the correct schema, then they may not be able to incorporate the new pieces of information correctly. If the students do not have the correct schema in their long-term memory, then cognitive overload may prohibit them from using other schemas to problem solve and come up with the correct answer.

I view the act of responding to an item as a three-part sequence that involves interpreting, problem solving or recalling the necessary information, and responding (Figure 2.4). In this model, students must figure out what they need to do, how they are to do it, and how they need to represent their answer. It is important to note that this sequence does not have to be linear. For example, students may read the question, problem solve, re-read the question, revise their problem-solving method, write their answer, re-read the question, and then modify the format of their final response.
Semiotic composition and complexity affect total cognitive load, which, in turn, affects working memory, and working memory affects students’ ability to respond to an item. Thus, I will examine the association between total cognitive load and student performance. Because cognitive overload could negatively affect students at any time during the test-taking process, I will examine the effects of total cognitive overload on each part of the sequence (interpret, solve, and respond).

**Final Comments**

The foundations of the conceptual framework and the theories used to address the research questions originate from the areas of multimodal social semiotics and cognitive science. The conceptual framework has focused on the processes through which multimodal texts are used to construct meaning and their complexity. Cognitive perspectives have been used to analyze the relationships between semiotic composition
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

and complexity, working memory, total cognitive load, and cognitive overload as they relate to the test-taking process.
CHAPTER III

REVIEW OF THE LITERATURE

The first section of this review of the literature provides an overview of multimodal social semiotics within education. More specifically, it discusses studies that examine the multimodal nature of mathematics, how multimodal texts represent meaning to both emergent bilingual (EB) and monolingual students, how EBs and monolingual students construct their own multimodal texts, and the use of non-linguistic semiotic resources in the assessment of EBs. The second section outlines key studies pertaining to cognitive load theory with an emphasis on the subsystems of the working memory and cognitive load.

Multimodal Social Semiotics and Education

Multimodality of mathematics. Mathematics is used to continuously construct meaning through the use and integration of natural language, symbols, and visuals (O’Halloran, 2000). Due to its multimodal nature, a considerable quantity of research exists on semiotics and mathematics. The following section outlines how natural language, symbols, and visuals are used to create ideational, interpersonal, and textual meaning within multimodal mathematics texts.

Natural language. de Oliveira and Cheng (2011) provided representative samples of their findings from two qualitative studies that investigated how the multisemiotic nature of mathematics could potentially create challenges for EB students. For the purposes of
this investigation, they examined problems from two mathematics textbooks for grades 1-5 and a standardized state assessment.

The mathematics problems were analyzed to examine how natural language used within mathematics discourse, and nominalization, more specifically, positioned EB students as they interacted with the texts. Nominal groups are nouns derived from verbs describing processes and adjectives denoting qualities, such as the product of two numbers (de Oliveira & Cheng, 2011). The authors state that the ideational function for nominal groups within these texts was to specify the requirements to be met to solve the problem. The following is an example of a question containing a nominal group from the Harcourt Math curriculum, Grade 2: “Name a group of nickels, dimes and quarters that has the same value as the 1 half dollar” (Harcourt Math cited in de Oliveira & Cheng, 2011, p. 260). de Oliveira and Cheng claim that EB students could potentially have misinterpreted the problems due to the complex structure of the nominal groups. They recommend textbook and assessment designers use less complex nominal structures and align activity sequences with the linguistic order of the problems to facilitate item comprehension.

The large sample of textbook questions used in these two studies contributes to the generalizability of the findings regarding the structure and use of nominal groups. Yet, to further support their findings, de Oliveira and Cheng could have conducted cognitive interviews with students to see whether they may have misinterpreted the questions due to the complexity of the nominal groups.

**Symbols.** Kirshner (1989) conducted a study examining the role that the visual syntax of mathematical symbols and expressions plays in the meaning-making process.
Kirshner argues that visual structures of symbols are highly correlated with the semantic meaning underlying the syntax. For example, while the signs in the expression $3x^2$ are arbitrarily placed, the location of the symbols indexes the underlying operations. The sign $x^2$ indicates one must square the value of $x$, and the positions of the 3 and $x$ indicate the multiplication of 3 and the squared value of $x$.

To test his theory, Kirshner created an instrument containing three 10-item subtests. The first subtest used standard notation, the second subtest used simple, non-standard notation, and the third subtest used complex, non-standard notation. The non-standard notation used different spacing and symbols. For instance, "$3x + 4$" (standard) was written as “$3MxA4$” (simple) or “$3M \times A4$” (complex). The sample of students comprised 381 students from grades 9, 11, and college freshmen in calculus.

The analysis revealed that almost all the students could calculate the expressions using standard notation. Only a few students could evaluate the expressions using non-standard notations. While there is no historical evidence that mathematical symbols and expressions were written to cue the underlying semantic meanings, Kirshner asserts the visual syntax of algebra has become unconsciously assimilated and used to facilitate the syntactic analysis of algebraic expressions. In other words, meaning is created through the textual organization of symbols. Kirshner’s findings suggest textbook designers and teachers should spend more time teaching algebraic syntax to facilitate students’ mathematical meaning-making processes.

One limitation of Kirshner’s study is that the items were scored as either correct or incorrect, regardless of process. The students could have miscalculated or made other

---

1 The term, *sign* is used here in its semiotic, not mathematical sense.
errors not related to syntax. Kirshner could have conducted cognitive interviews to examine how students arrived at each answer to support his claim that the visual syntax supported students’ ability to solve items correctly.

**Visuals.** Hegarty and Kozhevnikov (1999) used the work of Presmeg (1992) to analyze the effectiveness of different types of mathematical visuals during problem solving. Presmeg identified five different kinds of visuals used by students to solve mathematics problems: (a) images of concrete objects, (b) pattern imagery (relationships depicted using a visual-spatial scheme), (c) kinesthetic imagery, (d) dynamic transformations of geometric figures, and (e) visualization of memorized formulas. Presmeg argues that the use of images of concrete objects has a negative effect on students’ ability to correctly solve mathematics problems because they distract from relevant information. Conversely, the use of visual-spatial schematics improves students’ mathematical performance because they can depict pure relationships between objects and variables.

Hegarty and Kozhevnikov expand Presmeg’s categorizations by dividing pattern imagery into pictorial representations—images representing concrete objects or persons—or schematic representations—images representing spatial relationships between objects and spatial transformations. The researchers coded fifteen items from a sixth-grade mathematics test comprising pictorial representations, schematic representations, and non-pictorial-schematic representations. Hegarty and Kozhevnikov also conducted cognitive interviews with 33 sixth grade boys to discover how they solved the problems and administered tests of verbal reasoning, non-verbal reasoning, and spatial ability.
Results from the study indicated that students used pictorial-schematic representations to solve 55% of the items. Of these items, students were more likely to use pictorial representations (41%) than schematic representations (14%). Even though students favored pictorial representations, there was a negative correlation between students’ use of pictorial representations and their problem-solving ability. In contrast, there was a positive correlation between the use of schematic representations and problem solving ability. Furthermore, spatial ability positively correlated with the use of schematic representations and negatively correlated with the use of pictorial representations.

These findings suggest that schematic representations contribute to meaning construction through the representation of spatial relationships and spatial transformations. Teachers should encourage students to use schematic representations and discourage them from using pictorial representations during problem solving activities. Future studies should examine how different dimensions or types of schematic representations affect problem solving.

Section summary. Each of the studies discussed in this section examined one of the three main, coarse-grained semiotic resources of mathematics—natural language, symbols and expressions, and visual representations. de Oliveira and Cheng (2011) examined the ideational and interpersonal metafunctions of natural language within mathematics; Kirshner (1989) analyzed the textual metafunction of algebraic symbols and expressions; and Hegarty and Kozhevnikov (1999) investigated the ideational meaning of two different types of visual representations. These studies add to the growing body of research that examines how mathematical meaning is constructed using
several semiotic resources. The following subsection outlines studies pertaining to how multimodal texts represent meaning and how these texts affect or may potentially affect students (EB and monolingual) in the classroom.

**How multimodal texts represent meaning.** Scholars within fields such as multiliteracies and multimodal social semiotics are primarily concerned with how multimodal texts construct meaning. One of the milestones within these fields has been the success of applying Halliday’s (1978) theory of textual metafunctions (i.e., ideational, interpersonal, and textual) and Halliday and Hasan’s (1976) theories of lexical cohesion (i.e., Repetition, Synonymy, Antonymy, Hyponymy, Metonymy, and Collocation) to multimodal texts to explain how meaning is constructed.

**Emergent bilingual students.** In his 2002 work, Royce asserts that teachers of English as a second language should be able to analyze multimodal texts and teach their EB students how to read such texts to develop multimodal communicative competence. He claims that this skill is vital in the 21st century due to the rise of computer technology that combines text, audio, video, and images to create intersemiotic meaning.

To test the claim that multimodal texts create meaning through the integration of several semiotic resources, he conducted an intersemiotic ideational analysis of the text and image of the water cycle from an introductory environmental science textbook. Royce created descriptive glosses for each image that described the participants, activity, circumstances, and attributes of each visual. The descriptive glosses were compared to the semantic meaning of the text. Royce used these comparisons to determine the type of intersemiotic relationships that existed between the visual and text.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Royce’s analysis revealed that most the intersemiotic relations between image and text belonged to the intersemiotic Synonymy category—the pictures and text had a similar semantic meaning—or intersemiotic Metonymy—the pictures exemplified the text (e.g., a picture of a lake exemplifying the textual phrase, *bodies of water*). One of the most critical findings from this investigation is that visual and verbal semiotic resources complement each other and are used to create intersemiotically coherent multimodal texts (Royce, 2002). Royce’s development of intersemiotic ideational analysis is critical to the field of multimodal social semiotics because it is a systematic method that can be used to analyze how visuals and language create meaning.

Royce asserts that educators should teach Halliday’s metafunctions to emergent bilinguals as a way to immerse themselves into and interpret multimodal texts. Additionally, he recommends that teachers should teach EBs to interpret images before reading to ease them into the reading and discuss the confirmation or disconfirmation of their expectations after reading. While Royce’s recommendations are reasonable, they should be taken with caution because their effects were never empirically tested.

Knox (2008) conducted a multimodal discourse analysis of three different versions of a newspaper article about obesity (a traditional hardcopy newspaper, an online newsbite, and an online newsbit) to model how the authors of these sources used multimodal texts to represent meaning. Knox first examined how text and images were used integratively in the newsbite—stories with one paragraph, a headline, a lead, and a hyperlink—to create meaning. The newsbite contained a picture of the backsides of two obese women plus the title, *Obesity Crisis*. Below the picture and title was the following lead, *The average weight of young women is ballooning, a survey says.*
The multimodal discourse analysis revealed that intersemiotic Repetition (same meaning) and Collocation (expectancy relations) were used to construct meaning through images and text. For example, the lexical term, *obesity*, was repeated using the image of two obese women (intersemiotic repetition). The image was also placed against a sky-blue backdrop at an angle that placed the figures above the viewers. This technique was used to create an association between the image and the term, *ballooning* (intersemiotic Collocation). Lastly, the term, *crisis*, was associated with obese people. Thus, a negative judgment was being made about obese people (interpersonal meaning).

Overall, Knox discovered that intersemiotic Synonymy was used frequently to co-articulate the meaning of the different versions of the article. Knox also discovered that interpersonal relations between image and text were used to evaluate the aesthetic beauty of the object and judge the people portrayed in the image. Knox argues that EBs need to learn to become critical receivers of texts and learn how multimodal texts are positioning them as readers and viewers (Knox, 2008).

Research is needed to examine whether these findings can be replicated across several different online newspapers. Additionally, research is needed to examine whether readers, especially EBs, are correctly interpreting the intended meaning of multimodal texts.

*Monolingual students.* In her 2005 article on multimodal reading and writing practices, Jewitt used illustrative examples from her research on multimodality, learning, and technology in education. The examples were taken from a series of observations of a Year 7 science classroom and a Year 9 English classroom from a school in London. The examples that Jewitt highlights demonstrate how screen-based technologies affect
students’ reading and writing practices in these classrooms. Results from Jewitt’s multimodal analysis of the science classroom indicate that the designers of the computer curriculum used text and images to construct ideational meaning; text was used to label images of specialized science vocabulary, and images were used to represent scientific observations. For example, text was used to label the process of water changing from the solid to the liquid state and images were used to illustrate this process (intersemiotic Repetition).

Royce (2002) has shown that intersemiotic Repetition may support student learning because students can make meaning of a concept using one or both semiotic resources. While multiple resources may be available to students to access a text, many students favor one resource over another, and thus do not integrate the information from both. The students in Jewitt’s study favored the visual representation of water molecules and did not correctly interpret the image (even though it was labeled) because there was a large conceptual gap between the image and their everyday experiences with water molecules. Conceptual gaps occur when the semiotic resource comes in conflict with students’ experiences with the concrete objects they are meant to represent. In this situation, the students misinterpreted the image of water molecules as particles in a glass of water because they were not yet familiar with the concept of molecule, they misread the content of the image, and they did not integrate the information provided by the text and the image. Thus, while intersemiotic Repetition may support student learning, it cannot be used in isolation to teach new concepts; teachers still play a critical role in delivering content instruction and teaching students how to read multimodal texts so that they can use intersemiotic Repetition to support their learning.
Jewitt also conducted a multimodal analysis of a computer version of the novel *Of Mice and Men* in an English language arts classroom. She discovered that the designers of the computer-based text used different semiotic resources to construct different interpersonal meanings. Text font, color, and layout were used to construct characters’ identities. For example, a different font was used to position two characters as outsiders. Handwriting font was used to indicate that the piece was personal and/or fictional, and Courier font was used to indicate that a piece of information was factual. The different font types gave students clues about the interpersonal metafunction of the piece; students could prepare to engage imaginatively when they saw font that looked like handwriting and prepare to receive factual, non-imaginative information when they saw Courier font.

Overall, the reading and writing practices of both classrooms were affected by the multimodal computer-based curricula. Jewitt claims that multimodal texts allow for multiple interpretations and provide different modes of entry into texts for different students. Compared to traditional, paper-based texts, the semiotic resources within computer-based curricula are used more frequently and with more specific intentions to create interpersonal and ideational meaning. Additionally, Jewitt found that students tended to misinterpret multimodal texts if there was a large conceptual gap between different semiotic resources.

These findings should be interpreted with caution because the sample of observations was small (five observations of two classrooms for a total of 10 observations). Further research is needed to examine whether Jewitt’s findings are generalizable across several online science and English language arts curricula and across a representative sample of students.
Section summary. All the researchers in this section conducted intersemiotic ideational analyses of various texts to identify how text and image were used to construct meaning separately and integratively. The most common intersemiotic relationships among these studies were intersemiotic Repetition and Synonymy. Unlike Royce and Knox, Jewitt examined how the intersemiotic relationships affected or could potentially affect students’ interpretations of texts. Much of the research indicates that intersemiotic relationships assist students in interpreting texts. This research suggests that in order to ensure or maximize their effectiveness, educators should teach students how to construct meaning from multimodal texts.

How students use and construct multimodal texts. This section discusses the research that examines how students use semiotic resources to construct their own ideational and interpersonal meanings and adapt multimodal texts to fit their own sociocultural historical contexts. A critical finding from this work is that students use semiotic resources separately and integratively to construct meaning that could not be made using unimodal texts.

Emergent bilingual students. Using multimodal and critical pedagogy approaches, Ajayi (2008) examined how 33 high school EB students in an advanced English as a second language (ESL) class constructed vocabulary definitions through the use of multimodal representations and socially relevant instructional material. It is important to note that of the 33 students, 26 were immigrants from different countries in Latin America. After completing pre-reading activities and reading an article, the students were divided into groups and completed two of three activities related to the lesson vocabulary. Students who chose the first activity read the article again while highlighting
vocabulary words. They then used contextual clues, brainstorming, and group discussions to negotiate and write definitions for the vocabulary words. Students who chose the second activity created a campaign advertisement and a slogan. The product had to represent the students’ interpretation of the article using the vocabulary. Students who chose the third activity created a cartoon strip with captions to recreate the article. Again, the students were required to use the vocabulary in the cartoon.

Ajayi chose and analyzed one group’s product from each activity. Ajayi argues that the group that created an advertisement and the group that created a cartoon strip used placement to construct meaning. The former put the given information—information provided by the text—on the left and new information—information the students deemed important—on the right. The latter put ideal images (e.g., the promise of America) on the top of the page and more realistic images (e.g., immigrants in jail for not having a driver’s license) on the bottom of the page. Both groups used different viewpoints (e.g., first person or bird’s eye view visuals) to make the reader feel as though they were part of the situation and to make them empathize with the immigrants. The group that wrote definitions for the vocabulary words created socially constructed definitions that were connected to their lives, as opposed to the original, more abstract definitions found in the dictionary.

Ajayi argues that the multimodal text allowed for wider interpretations, interpretations that were connected to the students’ social realities. The multimodal text also allowed students to demonstrate a deeper, more critical understanding of the text than what they could have conveyed through writing due to their limited English skills.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

For example, students used illustrations to show their concern for the current discourse about immigration and critically examined their social conditions as immigrants.

Ajayi recommends that students should learn to identify and analyze the multimodal properties of different types of text across diverse social and cultural contexts so that they become critical users and creators of such texts. Ajayi’s sample only comprised students in an advanced ESL class. Thus, this study should be replicated with a larger sample of EB students with varying levels of English proficiency.

In a study of slightly younger students, Ajayi (2009) took a multimodal and critical pedagogy approach to investigating seventh grade EB students’ use of multimodal texts. Ajayi taught a three-week lesson about advertisements to 18 EB middle school students in an advanced ESL class. After learning how to deconstruct meaning from an advertisement for a cellular phone, the students made their own advertisements, explained them in writing, and presented them to the class.

Ajayi found that the students interpreted and recreated the cellular phone advertisement so that the portrayed meaning matched their own life experiences. Also, students used visuals to create a new identity and challenge views about their world. In congruence with the findings from his earlier article, Ajayi found that multimodal texts provided EB students with multiple modes of entry into the text, and that the students used multimodal texts to convey their understanding and interpretation of the text.

Ultimately, reading and constructing multimodal texts encouraged EBs to move away from normative interpretations and, instead, deconstruct texts based on their own knowledge structures (Ajayi, 2009).
As with his 2008 study, Ajayi’s 2009 findings need to be replicated with a larger sample of students from a wider range of English abilities. Moreover, Ajayi treated his instruction to the class as the treatment condition. While this does not discredit his findings, it does limit their generalizability to other classrooms because the typical classroom teacher most likely does not have the same skill set or knowledge base as Ajayi, who is an expert in multimodal instruction for ESL classrooms. To generalize across advanced ESL classrooms, in which he is not the teacher, and improve multimodal instruction in these classes, Ajayi should formalize his lesson plan(s) and strategies so that teachers may provide the same kind of instruction to their students.

**Monolingual students.** In their 2005 article, Hull and Nelson illustrated how multimodal texts generate meanings that are more than and distinct from the collective contribution of each semiotic resource. They performed a fine-grained multimodal analysis of a digital story created by a young man, Randy, in the Digital Underground Storytelling for You(th) Community Technology Center. The Center is a university and community partnership that was designed as a tool for “making powerful forms of signification (tools for and practices of digital multimodal composing) available to children and adults who did not otherwise have such access at home or at school” (Hull & Nelson, 2005, p. 230). Randy’s digital story was a compilation of music, spoken word, text, and images. The researchers focused on the images, text, and spoken word for this investigation.

The analysis revealed that Randy used text and images to orient viewers to the physical and cultural time and space of the digital story. More specifically, words and images were used to portray the universal struggles and hopes of African American
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

males. For example, Randy used images of Malcolm X, Tupac Shakur, Marcus Garvey, and Biggie Smalls—four men whose images carry significant political, social, and cultural meaning that cannot be recreated through text alone. He also used synonymous relationships to index the social meaning portrayed through the text by using images. For instance, the image of a handshake appeared directly after the word, truth, appeared on the screen. The images used in the final segment of the story were pictures of Randy and his neighborhood. Randy used these images to locate himself in the universal struggles and hopes of the African American male (Hull & Nelson, 2005).

Overall, Hull and Nelson’s findings suggest that Randy used images, text, and spoken word to construct meaning that could not have been achieved through a unimodal text. Hull and Nelson’s findings would have been strengthened had they interviewed Randy to discover his intentions and compare them to their results as a form of triangulation.

Jewitt, Kress, Ogborn, and Tsatsarelis (2001) adopted a multimodal semiotic perspective to investigate how seventh grade students reconstructed symbols provided to them by their teacher to facilitate their own learning. Jewitt et al. observed a science lesson focused on the characteristics of onion cells. The teacher began the lesson by giving instructions; modeling how to use the microscope and prepare the slide; and providing a picture of what the students should see under their microscopes. The teacher told the students that onion cells were like building blocks that looked like honeycombs. The students’ final products included written descriptions of their procedures and illustrations of what they saw under the microscopes.

The researchers analyzed four different student products for this investigation. Results from the multimodal analysis revealed that all the students recreated the teacher’s
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

analogies so that they aligned with their own background knowledge and experiences. For example, while working together, Students A and B stated that their onion cells looked like a brick wall, which was common in the students’ neighborhoods. Even though these two students were in a group together, their images were rather different. Student A drew uniform bricks with clear boundaries, while Student B emphasized the irregularities of the boundaries between bricks. Student C also used a brick wall analogy, but emphasized the irregularities between the bricks even more than Student B in his drawing. Even though Student D wrote in her procedure that she was looking for a honeycomb, she declared that onion cells looked like “wavy weaves” and drew a wave-like image.

At the beginning of the lesson, the teacher compared the image of onion cells to building blocks and honeycombs. However, none of the students used these analogies in their descriptions. Rather, they recreated the signs provided to them by their teacher to construct their own meaning. Jewitt et al. argue that students recreate signs and symbols given to them during instruction to construct knowledge that is meaningful to their own backgrounds and experiences.

These findings should be interpreted with caution due to the interpretive nature of the findings and the small sample size of participants. Had Jewitt et al.’s study been informed by cognitive science research they may have been able to analyze how the students integrated new information utilizing their existing schemas. They could have also used cognitive interviews to verify their findings with the students’ intentions.

Section summary. The authors in this section discussed how students constructed their own multimodal texts within several different disciplines. A commonality of these
studies is the finding that both EB and monolingual students recreated multimodal texts to align with their own background experiences and knowledge (Ajayi, 2008, 2009; Hull & Nelson, 2005; Jewitt et al., 2001). While both groups of students recreated multimodal texts, EB students relied on multimodal texts more than non-EBs to convey a deeper, more critical understanding of texts due to their limited English skills (Ajayi, 2008, 2009). Overall, the students used multimodal texts to construct meaning that could not be created using only one semiotic resource. The studies in this section either outlined how multimodal texts construct meaning, how students make meaning from multimodal texts, or how students create their own multimodal texts within different subject areas and contexts. The following section takes a more narrowed approach as it only examines the effect of multimodal texts on EBs as they take content assessments.

Semiotics and the testing of linguistically diverse populations. While there is a substantial amount of research regarding how multimodal texts are used to construct meaning and how teachers and students interpret them during classroom activities, research pertaining to how assessment items affect students’ ability to interpret and solve these items is limited. The literature reviewed below examines the role that illustrations play in assessment.

Martiniello (2009) examined the relationship between items’ measures of non-mathematical linguistic complexity, non-linguistic forms of representation, and information on item parameters based on item response theory—a psychometric of scaling. More specifically, she performed a differential item functioning (DIF) analysis—a technique that has been employed to compare the performance of two groups (for example, defined by race, gender, or linguistic status) on a given specific item after
controlling for the performance of the two groups on the overall test. DIF has been used to examine the extent to which an item is bias against or in favor of one group, the focal group (in this case, EBs) with respect to a reference group (in this case, non-EB students) (see Clauser & Mazor, 1998).

In the first stage of this study, items from the English version of a state-mandated Grade 4 mathematics test were rated for linguistic complexity; the presence and type of visual representations (i.e., text only, primarily pictorial, or primarily schematic) were coded; and the measures of DIF were calculated for EBs and non-EBs.

Martiniello’s findings indicated that, on average, items with greater linguistic complexity favored non-EBs, while items with schematic representations favored EBs. Linguistic complexity and its interaction with schematic representation were significant and accounted for 66% of the variation in DIF. This significant interaction effect highlights that the overall impact of linguistic complexity on DIF was reduced for EB students with the presence of schematic representations (Martiniello, 2009). Martiniello recommends that mathematics assessment items be written at levels appropriate to EBs’ English language proficiencies and be devoid of complex grammatical structures and low-frequency, non-mathematical vocabulary. Moreover, DIF measures should be used to evaluate standardized mathematics assessments to detect bias.

Martiniello had access to a large number of EBs \( N = 3,179 \), which is rare among studies involving EBs due to the small number of EBs usually found in individual schools. Even though the sample could have been potentially large enough to allow disaggregation by English language proficiency levels, Martiniello treated the group as though it was homogeneous. Martiniello could have strengthened her results by
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Describing the makeup of the EB group in more detail and/or disaggregating by proficiency level. Additionally, only 24 EBs were included in cognitive interviews. A key limitation of this study (as seen through the lens of my theoretical framework) is that Martiniello included both visual-spatial patterns and algebraic expressions in the schematic representations category. Because these two semiotic resources could alternatively be thought as being separate meaning-making systems, research is needed to examine how each resource affects the impact of linguistic complexity on DIF between EB and non-EB students. Solano-Flores, Barnett-Clarke, and Kachchaf (2013) also examined how the complexity of mathematics items affected EB students’ performance. However, they had a much more detailed definition of semiotic complexity, which will be discussed in further detail in the following paragraphs.

Solano-Flores et al. (2013) analyzed and measured the semiotic complexity (i.e., load) of mathematics items from an assessment used to evaluate the effectiveness of a mathematics curriculum. The authors coded 132 four-option, multiple-choice questions from the content knowledge (CK) and academic language (AL) assessments for fourth and fifth grade students. The presence of 32 features, which were divided into five different modalities, was coded dichotomously. Solano-Flores et al. also analyzed the pre- and post-tests results of 1,343 students by test, grade, and language status (EBs and non-EBs).

The authors found that AL items had a statistically significant higher semiotic load than CK items. AL items had a higher proportion of items that belonged to the mathematics register, natural/mathematics language, and testing register semiotic modalities, while CK items had a higher proportion of notation and visual semiotic
modalities. Additionally, the score gap between EBs and non-EBs increased from the pre- and post-tests for both assessments. Although students scored higher on AL items, these items were more challenging and the score gains were smaller for EBs.

Findings from this study support the notion that EB students should not be included in large-scale assessments before they have developed English language skills at a proficiency level that allow them to understand all aspects of testing. The authors contend that, EB students “appeared not to have developed a consolidated meaning-making system necessary to meet the different sets of interpretive demands of the two types of items with similar effectiveness” (p. 154). Thus, teachers should support EB students as they develop a consolidated meaning-making system.

One of the critical implications of this study is that test developers can use information about the frequency of semiotic features and modalities to create assessment items more systematically. One of the limitations of this study is that the authors treated the group of EBs as homogeneous and did not explicate the composition of the group. The following study discusses just how test developers should systematically design the format of and images used in assessment items.

Kopriva (2008) dedicates a chapter of her book on the testing of emergent bilinguals to access-based item development. Using a review of the literature, her own work within the field, and released items from a state’s fifth grade science assessment, Kopriva created a framework for analyzing whether items support EBs’ access to the content of items. The two dimensions of the framework that are most relevant to this study are item format and illustration features.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Under a multimodal social semiotic framework, texts create meaning through the textual metafunction. For example, the use of italicized or bold words constructs a certain kind of meaning, such as identifying key information. To provide access to the content of items for EBs, Kopriva recommends test developers separate key ideas; clearly indicate the question; use relevant titles for longer items; use a set of symbols to indicate certain test requirements (such as a stop sign to tell students to stop working); provide examples; highlight key words and/or phrases; and use boxes and/or lines to indicate where students should respond. While this is not an exhaustive list, the critical points are that (a) item format should be standardized across all items in an assessment, and (b) the meaning-making potential of format features should be used to increase item accessibility.

Kopriva also devotes a section of her chapter explicating the features of illustrations that increase EB students’ abilities to access items. Kopriva’s framework for illustrations includes, but is not limited to, the following recommendations: (a) illustrations should provide contextual support (i.e., use visuals that are relevant to the task and/or clarify the text), (b) item designers should illustrate content words (e.g., verbs, nouns, adverbs, and adjectives) and not function words (e.g., articles and prepositions), (c) illustrations should mirror or replace text, (d) pictures should be used to facilitate problem solving and response options, (e) designers should use visuals that are culturally relevant to the respondents, (f) illustrations should be simple (black and white, line drawings, no supplemental information), (g) designers should draw first person visuals with speech bubbles, (h) visual frames should be used to sequence events and arrange speech over time, and (i) key information should be made clear (Kopriva, 2008).
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Overall, item format features and illustrations can be used to allow EBs to access the content and meet the requirements of items which might not have been ascertained through linguistic resources alone due to their limited English skills. Although Kopriva indicates her recommendations were created using a review of the literature and findings from her own work, it is not always clear when the recommendations are based on empirical evidence and when they are principle-based. Additionally, the interaction effects of the features in Kopriva’s frameworks have not been empirically investigated in combination or across different populations of EB students. Solano-Flores and Wang (2011) also provide recommendations for systematically developing illustrations. Unlike Kopriva, however, the researchers used their framework to code and analyze illustrations from national and international assessments.

Solano-Flores and Wang (2011) provide a conceptual framework for designing and analyzing vignette illustrations used in science assessments. Vignette illustrations are illustrations added to test items originally designed without an illustration (Solano-Flores, 2011). The conceptual framework was designed to address a gap in the research—the lack of a systematic method to create and analyze illustrations used in science assessments with the intent to increase item accessibility for EBs. The researchers provide empirical evidence from two studies to support their conceptual framework.

Solano-Flores and Wang’s conceptual framework used illustration dichotomous variables to describe the presence or absence of different features. Overall, there were 85 variables (e.g., 2D [the visual contains two dimensional objects that depict length and width] and schematic [the visual is void of any unnecessary elements]) divided among 21 categories (e.g., projection [the possible dimensions of the objects] and image
concreteness [the possible types of images]), which were separated into five dimensions (i.e., representation of objects and background, metaphorical visual language, text in illustration, representation of variables, constants, and functions, and illustration-text interaction).

The purpose of their first study was to investigate how illustrations could be used to make science items more accessible to EBs. The framework was used to standardize the creation of the vignette illustrations used in the study. All of the vignettes had the following features: (a) represented only one to two constituents that might have been difficult for EBs to comprehend, (b) contained line drawings, (c) were simplified and realistic, (d) represented concrete objects and first person experiences, (e) were free of text, (f) did not represent any sequence of actions or stages, (g) minimized graphical features or metaphorical visual language, (h) were in black and white, and (i) were placed to the right of the questions (Solano-Flores & Wang, 2011). Creating the vignette illustrations was an iterative process that involved a multidisciplinary team.

The authors also examined the characteristics of illustrations used in state, national (U.S. and China), and international (Trends in International Mathematics and Science Study) science assessments. Using their conceptual framework, Solano-Flores and Wang coded a sample of 800 released multiple-choice and constructed-response eighth and ninth grade science items. Their findings suggested that China tended to have more and more varied illustration features than images from the U.S. This indicates that there are important cultural differences in the creation and complexity of science assessment illustrations, which should be examined in further detail.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

One of the limitations of this study is that the categories for language, symbols, and illustrations were confounded. Therefore, the findings cannot be attributed to the separate meaning-making potential of illustrations, but to the separate and integrated meaning-making systems of all three coarse-grained semiotic resources (language, symbols, and illustrations).

Wang (2012) elaborated on the coding system she and Solano-Flores used in their 2011 study. She analyzed a smaller sample of test illustrations ($N = 416$) from international (TIMSS) and national (U.S. and China) Grade 8/9 Earth Science, Life Science, and Physical Science assessments. The purpose of Wang’s study was to examine the differences in illustration complexity across item type (e.g., multiple-choice), assessment system type (e.g., middle school exit exam), content area, and national origin. Wang also investigated the correlation between illustration complexity and student performance on the TIMSS assessment.

Results from descriptive statistics and frequency analyses revealed that, on average, each illustration had 22 features. Additionally, there was not a statistically significant difference between the complexity of illustrations across item types, assessment systems, content, or country of origin of the items. Although the difference was not statistically significant, the Chinese illustrations tended to have a wider range of features than U.S. and TIMSS illustrations.

There was, however, a statistically significant difference in illustration complexity for the Context in Illustration dimension, which is the dimension that Wang added to the original framework. Overall, Chinese science assessment illustrations made more connections to the country’s national, cultural, political, and historical contexts than
illustrations from the U.S. and TIMSS. Results from correlational analyses indicate that there was a statistically significant positive correlation between illustration complexity for the *Context in Illustration* dimension and item difficulty (i.e., $p$-value—the proportion of students who answer a given item correctly) on the TIMSS assessment. For all students, regardless of region, their performance improved as the number of contextual features in the illustration increased.

While Wang analyzed a large number of test illustrations, her results are limited to illustrations found in Earth Science, Life Science, and Physical science assessments within the U.S., China, and TIMSS. Another limitation of her study is the fact that the coding system is generic and not content specific. It is plausible that illustration features differ greatly across content areas. Future research should examine this limitation. Lastly, Wang had limited access to data on student performance—she was limited to item $p$-values for each region that participated in TIMSS.

A more detailed discussion of the first study used to support Solano-Flores and Wang’s (2011) conceptual framework can be found in Solano-Flores, Wang, Kachchaf, Soltero-Gonzalez, and Nguyen-Le's (2014) investigation. In this study, Solano-Flores et al. systematically developed and investigated the use of vignette illustrations as testing accommodations for EBs on a multiple-choice science assessment. A multidisciplinary team developed the vignettes using a set of accommodation specifications and functions that were developed by the research team and a team of science and bilingual teachers (Solano-Flores, 2011; Solano-Flores & Wang, 2011). Each vignette was a concrete representation of one to two constituents that were likely to be difficult for EB students to comprehend.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

After several iterations of vignette development, the researchers investigated the effectiveness of the vignettes as a form of accommodation for native Spanish-speaking EB students. Seven hundred and twenty-eight eighth grade students (506 non-EBs and 222 EBs) took two different versions of a 27-multiple-choice science test in two formats, illustrated and non-illustrated. Half of the items in each format contained vignettes and the other half did not. Both language groups performed better on items with vignettes. The non-EB group outperformed the EB group on items with and without vignettes. Results from a two-way analysis of variance (ANOVA) revealed that there was a statistically significant difference and a large effect size due to language group, and a statistically significant difference but a negligible effect size due to format.

In the second study, the authors selected a subset of 16 of the original 27 items and created two different formats, vignettes with text and vignettes without text. A sample of the original population of students (40 EBs and 40 non-EBs) received each format and was asked to describe the vignettes in writing. Four independent raters scored the students’ descriptions according to their alignment with the intended content of the vignette. Both language groups had better alignment with the vignettes that contained text. The non-EB group had better alignment with both the texted and non-texted vignettes than the EB group. Solano-Flores et al. found statistically significant differences and large effect sizes due to linguistic group and format but not the interaction effect. Results from a generalizability study revealed a larger score variation due to the interaction of student-format-item for non-EBs (62%) than EBs (42%).

Solano-Flores et al. conclude that both EB and non-EB students used text to make sense of vignettes. EB students were most likely not as successful in correctly
interpreting illustrations because of their limited reading skills in English. Solano-Flores et al. recommend the connection between text constituents and their illustrations be made more explicit since students used both resources to interpret items. This study has the same limitation as Solano-Flores and Wang’s 2011 study—the categories for language and symbols were confounded with the categories for illustrations. Additionally, the EB sample was treated as a homogeneous group of students.

Based on findings from their previous studies involving the evaluation and creation of illustrations used in science items, Solano-Flores and Wang (2015) examined the complexity of illustrations used in the Programme for International Student Assessment (PISA)-2009 science assessment. They also examined the correlation between these features and the performance of students from Shanghai-China, the United States, and Mexico. These jurisdictions were chosen for two reasons: (a) they have very different languages and cultures; and (b) they represented very different points in the ranking of the 65 participating jurisdictions; Shanghai, the U.S., and Mexico ranked 1st, 23rd, and 50th, respectively, in the PISA-2009 science assessment (OECD, 2010).

Solano-Flores and Wang used data on the items’ p-values provided by the Organisation for Economic Co-operation and Development (OECD, 2010). Illustration complexity was measured by coding the presence or absence of 111 different features, which were grouped into 23 categories and five dimensions. To investigate the relationship between student performance and illustration complexity, the authors examined the Pearson correlation coefficient between each item’s p-value and the total number of features present. They also modeled the correlation between item p-value and the number of different illustration dimensions.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Findings indicate the average number of illustration features was 19.2. The $p$-values were consistently higher (i.e., more students answered the items correctly) for Shanghai, followed by the U.S., and then Mexico. The correlation between the total number of features and item $p$-value was practically zero for Shanghai ($r = 0.001$) and negative for the U.S. ($r = -0.194$) and Mexico ($r = -0.246$). Additional analyses revealed that the mean correlation between the number of features and the magnitude of the $p$-value was practically zero regardless of the number of dimensions modeled for Shanghai. Conversely, for the U.S. and Mexico, the higher the number of illustration dimensions, the higher the negative correlation between illustration complexity and item $p$-value with the magnitude being higher for Mexico than the U.S.

Overall, the items with more complex illustrations tended to be more difficult for students from the U.S. and Mexico. Additional analyses revealed that correct responses were less likely to occur for items in which the objects and background dimension was more complex, especially for Mexican students. This supports research claiming that unwarranted visual information tends to impede, rather than support, meaning-making (Mayer, Heiser, & Lonn, 2001). Correct responses were more likely to occur when the illustrations provided more textual information, which suggests that textual information may support students’ interpretations of non-linguistic forms of information.

Solano-Flores and Wang assert that illustration comprehension plays a critical role in student achievement in science assessments. They thus recommend that illustration features be examined as potential sources of meaning-making that shape item comprehension and student performance. Solano-Flores and Wang conclude by stating that illustration features are potentially critical in systematically examining and designing
illustrations for science assessment items. These findings should be replicated with a larger sample of illustrations and jurisdictions and with a sample of students whose linguistic background and proficiency levels are known.

Solano-Flores, Wang, and Shade (2015) addressed one of the limitations of the aforementioned study by replicating the investigation using all 65 participating PISA-2009 jurisdictions. The authors used a 2x2 between-subjects factorial design examining the differences between the magnitude and direction of the Pearson correlation between item difficulty and illustration complexity for jurisdictions from high- and low-ranking regions (based on overall PISA-2009 science scores) and Western and non-Western regions. They also examined the magnitude and direction of the correlation as a function of the number of illustration dimensions present in each illustration. Thirty items with their own illustrations were used for this investigation. Two of the five original dimensions (Metaphorical Visual Language and Form and Function) were removed for the purpose of this study due to the low frequency of observations and restricted variation of frequency of features, respectively.

Solano-Flores, Wang, and Shade observed a higher mean correlation for higher-ranking than lower-ranking jurisdictions. Mean values were also higher for non-Western jurisdictions in comparison to Western jurisdictions. A two-way ANOVA revealed statistically significant differences due to Ranking and Region, but not their interaction effect. Additionally, the correlation tended to increase positively with the number of dimensions for the majority of non-Western jurisdictions, and this correlation was noticeable among higher-ranking jurisdictions. For the majority of Western jurisdictions, the correlation increased negatively or not at all with the number of dimensions. Results
from a Ranking x Region x Dimensions repeated ANOVA revealed statistically significant differences in the magnitude of the correlation due to the interaction of Dimension and Ranking and Dimension and Region.

The authors contend that illustration complexity tends to be more influential on student performance for higher-ranking than lower-ranking jurisdictions. More specifically, more complex illustrations support students from higher-ranking jurisdictions in making sense of items, but hinder students from lower-ranking jurisdictions. The authors claim that the differences due to Region (Western and non-Western) point to the need for research examining how cultural factors contribute to how students make meaning from multimodal science items. This study highlights the need for science teachers, science standards, and science assessments and assessment frameworks to treat the ability to process multimodal information as an aspect of science proficiency rather than an intrinsic component.

The main limitation of this study is how the authors defined culture—as Western and non-Western world regions. However, because the authors had limited information pertaining to cultural factors, they did not make generalizations based on region. Furthermore, they acknowledged that future research is needed to address this limitation.

**Section summary.** Martiniello’s (2009) study was one of the first studies examining whether and how the presence of illustrations differentially affected the achievement of EBs and non-EBs on mathematics assessments. From this study, we learned that the presence of schematic representations could reduce the overall impact of linguistic complexity on DIF for EBs. Solano-Flores et al. (2013) improved upon these methods by examining how semiotic features present in two different types of mathematics items
affected EB students. Overall, EB students did not have a consolidated meaning-making system that allowed them to interpret the two types of items with similar efficacy. Kopriva (2008) was one of the first scholars to develop systematic guidelines for the creation of illustrations and format features as tools that allow EBs to access the content of items. Results from Solano-Flores and Wang’s (2011, 2015), Wang’s (2012) and Solano-Flores, Wang, and Shade’s (2015) studies point at cultural differences in the creation, complexity, and comprehension of science assessment illustrations. While the aforementioned studies indicate that illustrations help EBs interpret text, findings from Solano-Flores et al. (2014) indicate they may not be as beneficial for EBs as was once expected. In their study, both groups of students used text to interpret vignettes. However, EBs were most likely not as successful in correctly interpreting illustrations because of their limited English skills.

Cognitive Load Theory

The following section discusses various empirical studies that provide evidence for the existence of two partially independent information-processing subsystems of the working memory: the visual/pictorial channel (also known as the visuospatial sketchpad or visual information-processing channel) and the auditory/verbal channel (also known as the phonological loop or auditory information-processing channel). This section also discusses empirical studies pertaining to the three types of cognitive load—extraneous, intrinsic, and germane—and the construct of cognitive overload.

Working memory subsystems. Mayer and Anderson (1992) provide evidence from two studies in support of the dual-channel assumption and the contiguity principle. As discussed in the conceptual framework, the dual-channel assumption asserts that visual
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

and auditory information are processed and mentally represented in separate but interconnected subsystems (Baddeley as cited in Brünken et al., 2003; Mayer, 2001). The contiguity principle states that learning is facilitated through multimedia instruction—instruction using words and pictures (Mayer and Moreno, 2003)—when spoken text and pictures are used simultaneously because students can integrate information from both subsystems.

The purpose of both experiments was to compare the retention/recall and problem solving/transfer performance of students who received different combinations of animated (A) and narrated (N) multimedia instructional materials. One group received concurrent animated and narrated instruction, four groups received different versions of successive animated and narrated instruction (ANANAN, NANANA, AAANNN, NNANAAA), one group received animated instruction only, one group received narrated instruction only, and one group received no instruction (control group).

The participants of the first experiment included 136 college students who were not familiar with household repairs. The second experiment had 144 college students who were not familiar with auto mechanics. It must be noted that while the sample sizes were relatively large, there were only 17 students and 18 students in each of the eight groups for experiments one and two, respectively. The first set of participants watched a multimedia instructional video about bicycle tire pumps, and the second set of participants learned about the braking systems of cars. The students took a questionnaire prior to the video; watched the video; and completed a retention/recall test and a problem solving/transfer test after the video.
In both experiments, all the treatment groups significantly outperformed the control groups on the recall test. None of the treatment groups differed significantly from each other. Additionally, the concurrent group significantly outperformed all other groups on the transfer tests in both experiments. Again, none of these groups significantly differed from each other, even the control groups. The authors conclude that effective multimedia instruction should include concurrent animation and narration. Their findings support the dual-channel assumption and the contiguity principle because students in the concurrent group could build connections between their audio and visual processing systems.

Emerging from the dual-channel theory is the dual-coding effect, which states that the use of both verbal and non-verbal representations of essential information in multimedia instruction facilitates learning (Moreno & Valdez, 2005). In their 2005 work, Moreno and Valdez investigated the dual-coding effect using multimedia instructional programs about the causal effect of lightning.

To test the dual-coding effect, Moreno and Valdez created three different versions of instructional materials about lightning. The first version contained different frames of pictures (P) explaining the causal steps that lead to lightning; the second version contained different frames of words (W); and the third version contained both pictures and words (WP). Importantly, the words and pictures were integrated and non-redundant (the text and narration were not the same nor did they occur simultaneously). Ninety-eight college students were randomly assigned to one of six different groups. Each participant completed a questionnaire pertaining to their background knowledge of lightning; received the treatment; and completed assessments of retention and transfer. The subjects also rated how difficult it was for them to learn about lightning.
A multivariate analysis of variance was conducted treating code (P, W, and WP) and interactivity (interactive and non-interactive) as between-subjects factors. The scores from the recall, transfer, and difficulty measures were used as the dependent variables. There was a significantly large main effect for code when recall was used as the dependent measure. Group WP outperformed Group W, and Group W outperformed Group P. When the transfer test was used as the dependent measure, there was a moderate and significant main effect for code. Again, Group WP scored better than Group W, which scored better than Group P. The results were in reverse (difficulty was higher for Group P, followed by Group W, and then Group WP) when difficulty was used as the dependent measure.

Overall, Moreno and Valdez’s findings provide evidence in support of the dual-coding effect because the WP group outperformed the other two groups. They recommend using non-redundant, integrated words and pictures in multimedia instruction about causal systems to support student learning because they are more beneficial than using just words or pictures. Again, generalizability of these findings is limited by the small sample sizes (approximately 16 students) within each cell in the design.

In their 2001 article, Mayer, Heiser, and Lonn compared older ideas of cognitive learning to newer ones. Older theories of learning claim adding on-screen text to a narrated animation facilitates learning because the material is represented in multiple ways, allowing students to focus their cognitive resources on the mode they favor. The more recent split-attention hypothesis states that the combination of on-screen text and animation impedes learning because students must split their visual attention between the two modes. Thus, students are unable to attend to all the essential pieces of information.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

This is referred to as the redundancy effect; the presence of two visual representations of information causes an overload in the visual information-processing channel (see dual-channel theory; Mayer, 2001).

Another, older cognitive theory hypothesis, the emotional interest hypothesis, asserts that adding interesting, non-essential (i.e., seductive) information to instructional material will also facilitate learning because it increases students’ interest. Contrarily, the seductive hypothesis claims that the presence of seductive information impedes learning because it disrupts the flow of information, which is known as the coherence effect, and leads to the faulty construction of knowledge schemas.

Using four different experiments, Mayer, Heiser, and Lonn investigated these different hypotheses using multimedia instructional materials about lightning. In their first experiment, the authors tested whether the presence of on-screen text and seductive material facilitated or impeded learning. Seventy-eight college students were randomly assigned to four treatments: no-text/no-seductive details group (control), text/no-seductive details group, no-text/seductive details group, and the text/seductive details group. All students completed a questionnaire prior to the treatment; watched the videos; and completed retention and transfer post-tests.

Results from a two-way ANOVA revealed that students who received on-screen text summaries scored significantly lower on the tests of recall and transfer than students who received no on-screen text. These findings indicate the addition of on-screen text impedes student learning, which supports the redundancy effect and split-attention hypothesis. Results from the second two-way ANOVA indicated students who received
the seductive details performed significantly worse than students who did not receive them. These findings support the seductive hypothesis and the coherence effect.

The second experiment sought to examine the findings from the first two-way ANOVA in more detail. For example, on the one hand, students in the group with the added text could have performed poorly because the on-screen text information was competing with the narrated information. On the other hand, students could have been reconciling information from the on-screen text with that from the narrated text.

The participants for the second experiment included 109 college students who were randomly assigned to one of three groups: no text (control), summary text, and a full text transcription of the narration. Using a one-way ANOVA, Mayer et al. found that students who received no text performed significantly better than the summary text and full text groups on the recall and transfer assessments. There was no statistical difference between students in the summary text and full text groups. These findings support the split-attention hypothesis because the control group outperformed both treatment groups, indicating the redundancy effect was caused by an overload of the visual channel. Results from this experiment support the findings of Mayer and Moreno in their 1998 work involving multimedia instructional units about lightning and car braking systems (see Mayer & Moreno, 1998).

The third and fourth experiments were created to test the emotional interest hypothesis. To investigate this hypothesis, the researchers inserted short videos depicting lightning storms into the instructional presentations used in the first two experiments. Thirty-eight college students were involved in the third study. These students were randomly placed in either the no video group or the seductive video group. The
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

procedure was the same in the third and fourth experiments as it was in the first two experiments. Results from the first analysis revealed no statistical difference between the two groups on the recall test. However, the no video group scored significantly better than the video group on the transfer test. These findings support the seduction hypothesis and provide evidence against the emotional interest hypothesis.

The fourth experiment examined another aspect of the emotional interest and seductive hypotheses. It explored whether presenting interesting material before the lesson will increase interest and attentiveness and whether presenting it afterwards will not increase interest. The seductive hypothesis claims students should perform worse when interesting material is presented before instruction because these videos facilitate the construction of inappropriate schemas.

Thirty-two college students participated in the fourth experiment. Half of the students received the interesting videos before the instructional material and half of the students received them afterwards. Results from the recall assessment showed no statistical difference between the two groups even though the video-after group performed better than the video-before group. However, the students in the video-after group did perform significantly better than the video-before group on the transfer assessment, which supports the seductive hypothesis.

Overall, results from these studies support the redundancy effect for multimedia materials; adding redundant text to a narrated instructional lesson impeded learning. The addition of redundant information caused an overload in the visual processing system, and students were not able to successfully attend to all the essential pieces of information. Thus, multimedia materials should not use on-screen text with narrated animation. The
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

presence of interesting, construct-irrelevant material also inhibited learning. If interesting material is used at all, it should be presented after students receive the essential pieces of information and have already developed appropriate schemas. It should be noted that only one of the four studies contained more than 30 students per group. These studies should be replicated with larger sample sizes to validate their findings. Thus far, the studies discussed in this section (and in the field in general) have only included monolingual students in their samples. The subsequent study, however, addresses this gap in the literature and investigates multimedia instruction with students learning a foreign language.

Guichon and McLornan (2008) also examined how different combinations of audio, visuals, and on-screen text affected students’ comprehension. Their study, however, focused on students learning a foreign language and the use of native language (L1) and second language (L2) subtitles as on-screen text. Guichon and McLornan used four different versions of a news report (audio-only, audio and video, audio with video and L2 subtitles, and audio with video and L1 subtitles). Forty native French-speaking students with intermediate English skills were evenly placed into the four treatment groups. Students wrote a summary of the report for twenty minutes immediately after viewing it.

The summaries were scored dichotomously for the presence of 35 semantic units that the authors believed were central to understanding the report. The findings revealed that students who were exposed to the audio-only treatment discussed the smallest percentage of semantic units in their summaries (19.7%), followed by audio and video (25.1%), audio-video-L1 subtitles (29.7%), and audio-video-L2 subtitles (30.2%). However, there was no significant difference between the subtitle groups. A more detailed analysis
revealed that students could only provide a precise identification of the characters and settings if they had access to the video; comprehension was negatively affected if the audio and visual information were not directly connected; and L1 subtitles appeared to have negatively affected comprehension because students focused on them rather than the L2 oral message (Guichon & McLornan, 2008).

Guichon and McLornan contend that associating images in L2 and text in L2 is an effective way to increase students’ comprehension of multimodal texts. However, this finding was not significant. Subtitles distracted viewers from information provided by the video and audio and were dependent on students’ reading skills. In contrast to Moreno and Valdez’ findings (2005), Guichon and McLornan found that cognitive overload can occur when audio and visual information do not concur on the kind of information they represent. The researchers encourage multimedia designers to utilize several different but redundant semiotic resources to improve foreign language learners’ comprehension. These findings must be replicated with larger sample sizes and should be replicated with EB students.

Section summary. The aforementioned studies provide evidence in support of the dual-channel assumption, the contiguity principle, the dual-coding effect, the redundancy effect, and the seductive hypothesis. Findings support the notion that multimedia instructional tools should contain concurrent animation and narration and non-redundant, integrated words and pictures. These tools should not contain on-screen text and narration and/or interesting but irrelevant information. It should be noted that L2 subtitles and redundant information might be beneficial for students learning a second language and EBs (Echevarria, Vogt, & Short, 2008; Guichon & McLornan, 2008).
Cognitive load: Intrinsic, extraneous, and germane cognitive loads. Cognitive load is a theoretical construct that cannot be measured directly because it cannot be directly observed. Naturally, researchers do not agree on how to measure it. DeLeeuw and Mayer (2008) used three different measures of cognitive load to test whether it is a unitary construct or a construct comprising three parts: the extraneous cognitive load, the intrinsic cognitive load, and the germane cognitive load (Sweller, 1999).

The authors created two versions of a multimedia lesson about how an electric motor works to manipulate the extraneous cognitive load. The first version was non-redundant and contained animation and narration. The second version was redundant and contained animation, narration, and text. This version was redundant because the text and narration were the same and occurred simultaneously. The redundant lesson should have increased the extraneous cognitive load because the students would have had to waste cognitive resources on interpreting and holding on to information that was repeated in two different modes. The authors manipulated the intrinsic cognitive load by using four simple sentences and four complex sentences in each version. The complex sentences should have required more intrinsic cognitive load because the students would have had to interpret more information to understand the content. The germane cognitive load was measured by comparing low-performing to high-performing students on tests of information transfer.

DeLeeuw and Mayer used three methods to measure cognitive load to test whether it was a unitary or triarchic construct. To measure the extraneous cognitive load, the authors added a secondary task to the lesson and measured students’ response times to this task. For this task, students pressed the space bar once they noticed the color of the
screen beginning to change. The intrinsic cognitive load was measured by having students rate their mental effort, on a scale of one to nine, after each simple and complex sentence. The germane cognitive load was measured by having students measure the level of difficulty of the lesson, on a scale of one to nine, at the end of the instructional materials.

The authors found that response times were longer for students in the redundant condition; students reported using high mental effort after complex sentences; and students with low transfer abilities rated the lesson as more difficult than students with high transfer abilities. Additionally, none of the measures of cognitive load correlated highly in either experiment. Their findings indicate that cognitive load is not a unitary construct, but rather a triarchic construct comprising extraneous, intrinsic, and germane cognitive loads. The different loads can be measured using response times to a secondary task, mental effort ratings, and difficulty ratings, respectively.

Oviatt (2006) also used the triarchic model of cognitive load to compare the cognitive loads of different user interfaces. More specifically, she used theories of human-centered design and cognitive load theory to analyze the effectiveness of four different types of interfaces as students solved geometry problems. According to human-centered design principles, users adjust to the limitations of their working memories by using multimodal forms of communication. Thus, interfaces should be multimodal and flexible to facilitate usability. Interfaces should also be designed around the experiences, knowledge, behaviors, and work experiences of its users and support the representation systems of the task. For example, interfaces should use and allow its users to use algebraic symbols and expressions, language, and images for mathematics tasks. According to cognitive load
theory, user interfaces should reduce the extraneous cognitive load relative to the amount of germane cognitive load to improve student performance.

Oviatt compared several different measures of cognitive load for 20 high school students using a traditional paper-and-pencil (PP) format and three different computer-based interfaces: (a) a digital stylus and paper interface (DP); (b) a pen tablet interface with stylus input (PT); and (c) a graphical tablet interface with keyboard, mouse, and stylus input (GT). The results revealed that students performed better with the DP interface than with the PT interface and performed better using the PT interface than with the GT interface. Moreover, the cognitive load of low-performing students increased when they used the PT and GT interfaces. Students completed the problems faster and were more attentive to their mathematics when using the PP and DP interfaces in comparison to the GT interface.

Using cognitive load theory, Oviatt contends that the DP interface improved student performance because it was most similar to the traditional paper-and-pencil format. The DP interface also improved student performance because it did not contain as much extraneous material as the other two interfaces, such as different font colors or answer selection tools. Interpreted through a human-centered design theory lens, students performed better using the DP interface because it was easier for them to communicate multimodally. For example, students could draw diagrams or pictures to support their thinking and reasoning, which was not possible when using the other two interfaces.

Oviatt argues for the importance of designing interfaces that reduce cognitive load so that new technologies do not create a digital divide between low- and high-performing
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

students. Oviatt’s sample size was relatively small \((N = 20)\); thus, she could have conducted cognitive interviews with the students to support her findings.

Many of the methods used to measure cognitive load are subjective and/or indirect. To address this gap in research methodologies, Ikehara and Crosby (2005) conducted an investigation that used electric sensors to measure physiological responses as proxy measures of cognitive load (task difficulty, or intrinsic cognitive load, in this case) during a computer-based problem solving task involving fractions. These sensors measured eye movement, skin conductivity, relative body temperature, heart rate, and the pressures applied to the computer mouse at various times.

The sample comprised 13 male volunteers from the United States Air Force Academy. The instrument was a computer-based task that presented the subjects with several fractions inside ovals. The subjects had to select the fractions that were less than \(1/3\) before they disappeared. To measure cognitive load, the subjects completed three easy tasks, three hard tasks, three easy tasks, and three hard tasks.

Results indicated that eye movement, skin conductivity, and the pressure applied to the computer mouse were all predictors of task difficulty. Eye movement decreased as task difficulty increased because the subjects were more focused on the task. In opposition to other findings, the authors discovered that skin conductivity decreased as task difficulty increased. Lastly, the pressure applied to the mouse before selecting the answer increased for more difficult tasks.

Ikehara and Crosby conclude that the physiological measures mentioned above can be more sensitive than traditional measures of cognitive load. Thus, they could be used as alternative or supplemental measurements of task difficulty. Findings from their study
should not be generalized due to the small sample \((N = 13)\). Further research should examine whether eye movement, skin conductivity, and computer mouse pressure predict task difficulty for individuals who are not in the armed forces, as members of the Air Force could have systematic differences from a random sample of the general population. Additionally, future studies should investigate whether and how these physiological measures predict extrinsic and germane cognitive loads.

**Section summary.** The findings, recommendations, and theories of the authors discussed in this subsection all address issues of cognitive load. Findings from DeLeeuw and Mayer (2008) and Oviatt’s (2006) studies all support the hypothesis that cognitive load is a triarchic construct comprising extraneous, intrinsic, and germane cognitive loads. Each type of cognitive load can be measured by adding a construct-irrelevant task, asking students to rate their mental effort after each question, and asking students to rate the difficulty level of the material, respectively (DeLeeuw & Mayer, 2008). The design of computer interfaces affects cognitive load; to reduce cognitive load, interfaces should mimic the design around the experiences of the users and the representation systems of the task. They should also have a limited amount of extraneous material and allow users to communicate multimodally (Oviatt, 2006). Ikehara and Crosby (2005) argue that physiological measures of intrinsic cognitive load should be used instead of or in addition to traditional measures that are currently being used.

**Summary of Literature Review**

Semiotic resources represent meaning using ideational, interpersonal, and textual metafunctions. Within multimodal texts, semiotic resources are used to construct meaning separately and integratively to produce a complete semantic unit that cannot be
reproduced using unimodal texts. Mathematics is a multisemiotic subject that represents meaning using three primary, coarse-grained resources—language, visuals, and symbols.

One of the common findings in the studies pertaining to the production of multimodal texts was that students, EBs and non-EBs, use semiotic resources separately and integratively to construct meaning that cannot be made with unimodal texts. The construction of multimodal texts is critical for EB students because it allows them to communicate what they know without relying on their knowledge of the English language.

Findings from research regarding the use of illustrations as tools that provide EBs access to the content of items and/or as a form of assessment accommodations have mixed results. Several scholars argue that schematic representations support EBs, while others contend that visuals should comprise concrete, pictorial representations. Additionally, some fields of research suggest multimodal texts are beneficial for EBs, while others claim multimodal texts create cognitive overload.

The aforementioned studies have made great contributions to the fields of multimodal social semiotics, education, the assessment of EBs, and cognitive load theory. However, research is needed that integrates knowledge from all four domains. There is a lack of research on multiliteracies as they pertain to computer-based assessment. In addition, research on semiotic resources of mathematics and their effects on EBs is sparse. Scholars have begun to examine how images affect EBs on mathematics and science assessments, but these studies are limited to images and traditional, paper-and-pencil assessments. Furthermore, a great deal of the work in these three fields does not discuss the demands on working memory caused by multiple semiotic resources, including
multiple languages. Conversely, cognitive load theory research is limited to cause-and-effect, multimedia instructional materials with monolingual English-speaking students.

This dissertation fills these gaps by examining how the semiotic properties of multimodal items on a computer-administered mathematics test affect how EB and non-EB students interpret and respond to the items and how these resources affect cognitive load. Unlike other studies that only examine the three main types of semiotic resources used in mathematics (i.e., language, symbols, and visuals), my dissertation views other coarse- and fine-grained semiotic resources, such as computer features and text location, respectively, as their own meaning-making systems. This dissertation adds to the field of assessment by examining how items on a computer-administered mathematics practice test create ideational, interpersonal, and textual meaning through the use and integration of several semiotic resources. Computer-administered standardized tests are relatively new assessment tools. This dissertation examined whether students can create multimodal texts using this platform; whether the platform limits students’ abilities to demonstrate their knowledge; whether the limited capacities of the platform are detrimental to the performance of EBs; and whether and how the platform causes cognitive overload.
CHAPTER IV

METHODOLOGY

Goals and Research Questions

The purpose of this exploratory study was to examine the semiotic correspondence across patterns in the design of tests and students’ test-taking strategies in the context of large-scale mathematics assessment, as they concern emergent bilingual students (EBs).

Two studies were conducted, one to answer each of the following research questions:

1. What semiotic resources are used in SBAC test items?
2. What commonalities and differences can be observed between EB and non-EB students in the ways in which they make sense of the semiotic resources used in SBAC items?

This chapter discusses first the aspects common to both studies (context, recruitment of participants, researcher’s role, test materials), then the methods used in each study.

General Information

Context. Table 4.1 shows the sociodemographic information for each of the participating high schools. This study took place in three high schools within a large, suburban district in a mid-Western state that is a member of the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium. The district enrolls approximately 77,000 students. There are 52 schools in operation within the district, including eight high schools.
Table 4.1

*Frequencies and Percentages (Rounded) of Sociodemographic Variables for Participating High Schools*

<table>
<thead>
<tr>
<th>Sociodemographic Variables</th>
<th>School 1</th>
<th>School 2</th>
<th>School 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
</tr>
<tr>
<td><strong>Student Enrollment</strong></td>
<td>1,962</td>
<td>100</td>
<td>1,784</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1,242</td>
<td>63</td>
<td>589</td>
</tr>
<tr>
<td>Black</td>
<td>43</td>
<td>2</td>
<td>59</td>
</tr>
<tr>
<td>Asian</td>
<td>88</td>
<td>5</td>
<td>77</td>
</tr>
<tr>
<td>Native Hawaiian/Pacific Islander</td>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Hispanic</td>
<td>545</td>
<td>28</td>
<td>1,032</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>18</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Two or More Ethnicities</td>
<td>22</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Free and Reduced Lunch</td>
<td>639</td>
<td>32</td>
<td>774</td>
</tr>
<tr>
<td>English Language Learner</td>
<td>86</td>
<td>4</td>
<td>356</td>
</tr>
</tbody>
</table>

*Note.* Data obtained from 2013 October count. All variable and category names are those used by the district.

Figure 4.1 shows the publicly available data concerning the percent of 9th and 10th grade students who scored at each proficiency level (unsatisfactory, partially proficient, proficient, and advanced) in the 2014 state standardized mathematics assessment.

Amongst the three schools, School 2 had the largest percentage of students at the unsatisfactory level for both grade levels. Contrarily, School 3 had the highest percentage of students at the proficient and advanced levels for both grade levels.
Figure 4.1. Percent of Grade 9 and Grade 10 students meeting each proficiency level on the 2014 state-mandated mathematics assessment. Disaggregated by school.

The data were also disaggregated by English language proficiency level—Non-English Proficient (NEP), Limited English Proficient (LEP), and Fluent English Proficient (FEP) students—categories used by the district. None of the schools reported scores for NEP students because there were fewer than 16 at each school. Achievement on the 2014 state assessment varied between schools for LEPs. Overall, School 2 had the highest percentage of FEP students scoring at the unsatisfactory or partially proficient levels. School 3 had the highest percentage of FEP students scoring at the proficient and advanced levels.

Participants. Tables 4.2, 4.3, and 4.4 show, respectively, the sociodemographic information on the students who participated in the study, their classes, and their teachers. Sixty 11th grade students, 30 EBs and 30 non-EBs, from 13 different classrooms participated in the cognitive interviews. These numbers ensured a minimum level of
statistical power for analyses. Grade 11 was chosen to minimize variation due to grade level. In an ideal design, three EBs and three non-EBs from each of the 13 classes would have been interviewed. However, differences in the number of EBs in each class, the attendance rates of these students, and other practical limitations prevented me from interviewing the same number of EBs and non-EBs across classes (which would have controlled for differences attributable to class or teacher). However, I made the attempt to interview the same number of EBs and non-EBs within each class.

All the teachers were White (one teacher chose not to share her race), spoke English as their first language, and held teaching licenses in mathematics (Teacher 7 held teaching licenses in mathematics, science, and computer science).
### Table 4.2

**Frequencies and Percentages (Rounded) of Sociodemographic Variables for Cognitive Interview Participants**

<table>
<thead>
<tr>
<th></th>
<th>Emergent Bilinguals (N=30)</th>
<th>Non-Emergent Bilinguals (N=30)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td><strong>Race/ Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hispanic</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>Native American</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Asian</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Biracial</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>14</td>
<td>47</td>
</tr>
<tr>
<td>Male</td>
<td>16</td>
<td>53</td>
</tr>
<tr>
<td><strong>English Language Proficiency</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not English Proficient</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Limited English Proficient</td>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>Fluent English Proficient</td>
<td>16</td>
<td>53</td>
</tr>
<tr>
<td><strong>School</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School 1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>School 2</td>
<td>26</td>
<td>87</td>
</tr>
<tr>
<td>School 3</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td><strong>Teacher and Section</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>T3</td>
<td>10</td>
<td>33</td>
</tr>
<tr>
<td>T5.1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>T5.2</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>T6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T7.1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>T7.2</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>T8.1</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>T8.2</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>T8.3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>T9</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>T10</td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>

*Note.* Race/ethnicity categories were those used by students. English language proficiency categories were those used by the district.
Table 4.3

Sociodemographic Variables for Classes Containing Cognitive Interview Participants

<table>
<thead>
<tr>
<th>School</th>
<th>Teacher &amp; Section</th>
<th>Class</th>
<th>Number of Students</th>
<th>ELBs (Rounded Percentages)</th>
<th>Grade Level (Rounded Percentages)</th>
<th>Sex (Rounded Percentages)</th>
<th>Race/Ethnicity (Rounded Percentages)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Grade 10</td>
<td>Grade 11</td>
<td>Grade 12</td>
</tr>
<tr>
<td>1</td>
<td>T1</td>
<td>CMIC 3</td>
<td>26</td>
<td>15</td>
<td>0</td>
<td>85</td>
<td>15</td>
</tr>
<tr>
<td>1</td>
<td>T2</td>
<td>CMIC 3</td>
<td>19</td>
<td>21</td>
<td>6</td>
<td>68</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>T3</td>
<td>Math</td>
<td>27</td>
<td>52</td>
<td>0</td>
<td>96</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Anal/Trig</td>
<td></td>
<td></td>
<td>Grade 10</td>
<td>Grade 11</td>
<td>Grade 12</td>
</tr>
<tr>
<td>2</td>
<td>T5.1</td>
<td>CMIC 2</td>
<td>26</td>
<td>58</td>
<td>73</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>T5.2</td>
<td>CMIC 2</td>
<td>25</td>
<td>40</td>
<td>80</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>T6</td>
<td>CMIC 4</td>
<td>20</td>
<td>20</td>
<td>60</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Grade 10</td>
<td>Grade 11</td>
<td>Grade 12</td>
</tr>
<tr>
<td>2</td>
<td>T7.1</td>
<td>CMIC 3</td>
<td>22</td>
<td>27</td>
<td>86</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>T7.2</td>
<td>CMIC 3</td>
<td>27</td>
<td>48</td>
<td>59</td>
<td>41</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>T8.1</td>
<td>CMIC 3</td>
<td>31</td>
<td>39</td>
<td>65</td>
<td>36</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>T8.2</td>
<td>CMIC 3</td>
<td>26</td>
<td>42</td>
<td>27</td>
<td>54</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Grade 10</td>
<td>Grade 11</td>
<td>Grade 12</td>
</tr>
<tr>
<td>2</td>
<td>T8.3</td>
<td>CMIC 3</td>
<td>25</td>
<td>48</td>
<td>92</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>T9</td>
<td>CMIC 2</td>
<td>30</td>
<td>17</td>
<td>67</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>T10</td>
<td>Algebra II</td>
<td>24</td>
<td>17</td>
<td>0</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

Note. CMIC is the Contemporary Mathematics in Context Curriculum used within the district. Math Anal/Trig is the Math Analysis and Trigonometry class that is taught in some schools and corresponds to CMIC IV. Race/Ethnicity categories were those used by students.
Recruitment. The district approved my application to conduct research in its classrooms in December of 2014. After receiving the approval letter, I made initial contact with the principals of the participating high schools and informed them about my study. I gave the principals a recruitment letter (Appendix A) to send to the mathematics teachers. The recruitment letter notified the teachers about the purpose of the study, the characteristics of the students that were being recruited (i.e., 11th grade EBs and non-EBs), and my contact information. Each participating teacher received 20 dollars as an incentive and token of appreciation for their time. Each participating teacher signed a consent form (Appendix B) to partake in the study.

After recruiting 10 teachers and randomly selecting classrooms for the observations and cognitive interviews, I informed the students about my study and provided them with an assent form to be signed by the students and a permission form to be signed by their parents. Both forms were provided in English and Spanish (see Appendices C-F). If a
student and/or their parents did not consent to participating in the study, then they were not included in the recordings, no demographic information was collected on them, and they were not included in cognitive interviews.

**Researcher’s role.** I was a major influence in the coding of the data in this study because I collected the data. Thus, my role as a researcher and my identity must be discussed as they interacted with and affected the participants, data collection, and data analysis.

As a result of high-stakes assessments and corrective action for schools that do not meet annual yearly progress, schools are currently very tense work environments. To avoid increasing teachers’ stress levels, I assured them that I was not evaluating them or their teaching practices in any shape or form. I also attempted to foster professional and supportive relationships with the teachers so that they were not anxious about my presence.

While I limited my interactions with students during observations so as not to influence the data, I believe that I formed some sort of relationship with them outside of classroom observations (e.g., before and after class). This relationship, I feel, eased students’ anxiety levels during classroom observations and cognitive interviews so that they felt free to talk to me about their thinking. Additionally, the students could answer the interview questions in English, Spanish, or both although all the students chose to respond in English.

My identity as a white, middle-class female also had an impact on the data. While my identity did not reflect most of the EB students’ identities, I have conversational fluency in Spanish and have worked with similar populations of students as a
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

mathematics teacher. Conversely, my identity did reflect most of the teachers’ identities. The similarities between my identity and those of the majority of the teaching population most likely only reduced the teachers’ anxiety levels during classroom observations.

**Test materials.** The Smarter Balanced Assessment Consortium (SBAC) practice test was chosen for this study, even though the state belongs to the PARCC consortium, because SBAC has a larger variety of semiotic resources and accessibility resources available on its online practice tests. “Ultimately, these accessibility resources are intended to increase the validity of interpretations of Smarter Balanced assessment scores by reducing the amount of measurement error attributable to factors that are irrelevant to the constructs test items are intended to assess” (Solano-Flores et al., 2014, p. 1). In addition, by using SBAC items, students were less likely to be familiar with the items used for this investigation.

SBAC has several accessibility resources built into their system that administrators, teachers, and/or students can select or turn on before test administration. Stacked, full translations are offered in Spanish only. These tests are in Spanish with the English translation underneath the Spanish text. In addition, pop-up glossary translations are available in English only or English and an additional language (Arabic, Cantonese, Spanish, Korean, Mandarin, Punjabi, Russian, Filipino, Ukrainian, and Vietnamese). Figure 4.2 shows Item 20 from the current Grade 8 SBAC Math Practice Test before and after activating the Spanish pop-up glossary for the phrase, *that makes.*
As of May 2015, SBAC offered three different types of practice tests for Grades 3 through 12: (a) Math Training Test ($n = 7$), (b) Math Performance Task ($n = 7$), and (c) Math Practice Test ($n = 29$). Even though this study focused on 11th grade students, the 2015 Grade 8 Math Practice Test was chosen for several reasons. Since 11th grade students are commonly placed in one of three mathematics classes (Contemporary Mathematics in Context [CMIC] II, CMIC III, or CMIC IV), not all participating students were taking the same mathematics class. Since the purpose of this study was not to measure the students’ mathematics skills, per se, using a test of a lower grade allowed elimination of error due to mathematics proficiency level.

**Study 1: Semiotic Resources Used in SBAC Test Items**

**Coding.** Screenshots of the items on the practice test were coded on two different occasions: to examine the semiotic components and complexity of each item and to examine intersemiosis. Appendix G shows all the items from the practice test.

---

**Figure 4.2.** Example of a Spanish pop-up glossary for the phrase, *that makes* in Item 20 from the 2017 SBAC Grade 8 Math Practice Test. The image on the left is the item before activating the glossary, and the image on the right is the item after activating the glossary. Source: SBAC Grade 8 Math Practice Test. Taken from Smarter Balanced Assessment Consortium. (2017). Retrieved from [http://sbac.portal.airast.org/practice-test/](http://sbac.portal.airast.org/practice-test/).
The semiotic components of the items were coded deductively and inductively (Forman & Damschroder, 2008; Miles & Huberman, 1994; Zhang & Wildemuth, 2009). I used deductive coding first to understand my data in relation to my conceptual framework. Then I used inductive coding to create new codes, which were grounded in the data, and to refine or eliminate a priori codes.

The items on the practice test were coded deductively using five main theoretical frameworks: (a) Halliday’s (1978) systemic functional linguistics, (b) Halliday and Hasan’s (1976) theory of lexical cohesion, (c) Lemke’s (2003) notion of the multimodality of mathematics, (d) O’Halloran’s (2005) identification of the metafunctions of the three semiotic resources of mathematics, and (e) Solano-Flores’ and Wang’s (2011) conceptual framework on illustration features.

Inductive codes were created using open coding with constant comparison (see Strauss & Corbin, 1994). After comparing a given item’s semiotic component to the deductive codes, I either coded it using one of the a priori codes or created a new code. Subsequent item semiotic components were compared to both the a priori codes and the new codes. If the properties had similar meanings to the other properties, I coded them using the existing codes; if they had different meanings, then I created new codes. I continued this process iteratively, moving between items and codes, until all the semiotic components were accounted for in the coding system.

Two coding systems were applied to the item features. The first focused on the semiotic components of items. Unlike previous approaches to examining semiotic components of items, this coding system did not focus on one metafunction or one semiotic resource. Rather, the three metafunctions of each semiotic resource were
analyzed and compared. More specifically, the natural language (NL), mathematical symbols and expressions (MSE), and visual representations (VR) of each item were examined in terms of multiple item dichotomous variables (IDVs) grouped into different dimensions belonging to each metafunction (ideational, interpersonal, and textual). IDVs are defined as features whose combined presence or absence defines the semiotic components of a given item (Solano-Flores & Wang, 2011; Wang, 2012). Table 4.5 shows the three metafunctions, their dimensions, and corresponding IDVs for each semiotic resource. Appendix H shows a more detailed codebook containing definitions for each metafunction, dimension, and IDV. Each IDV was coded dichotomously as present (1) or absent (0).
Table 4.5

Item Coding System – Semiotic Components

<table>
<thead>
<tr>
<th>NL.1. Interpersonal metafunction</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL.1.1. Interpersonal: 1. directive, 2. factual, 3. inquisitive, 4. rational, 5. false</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NL.2. Ideational metafunction</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL.2.2. Text functional categories: 1. participants, 2. activity, 3. attributes, 4. circumstances</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NL.3. Textual metafunction</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL.3.1. Text unit: 1. symbol, 2. letter, 3. abbreviation, 4. word, 5. phrase, 6. sentence, 7. multiple sentences, 8. bullet points</td>
</tr>
<tr>
<td>NL.3.2. Text emphasis: 1. bolding</td>
</tr>
<tr>
<td>NL.3.3. Text location: 1. left of response space (RS), 2. above RS, 3. separated by visual, 4. separated by MSE, 5. embedded in schematic visual, 6. embedded in concrete visual, 7. embedded in RS, 8. embedded in response tool</td>
</tr>
<tr>
<td>NL.3.4. Text complexity: 1. multi-word noun phrase, 2. multi-word verb construction, 3. present tense, 4. past tense</td>
</tr>
<tr>
<td>NL.3.5. Text constituents: 1. simple subject, 2. compound subject, 3. simple predicate, 4. compound predicate, 5. pronoun, 6. adjective, 7. adverb, 8. direct object, 9. indirect object</td>
</tr>
<tr>
<td>NL.3.6. Text register: 1. natural language, 2. testing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MSE.1. Interpersonal metafunction</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE.1.1. Interpersonal: 1. factual, 2. neutral, 3. false, 4. non-negotiable, 5. helpful</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MSE.2. Ideational metafunction</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE.2.1. MSE purpose: 1. discrete numbers not on graph, 2. discrete numbers on graph, 3. measurement, 4. equation, 5. transformation, 6. coordinate point, 7. function, 8. linear system, 9. symbol, 10. operation, 11. variable</td>
</tr>
<tr>
<td>MSE.2.2. MSE functional categories: 1. participants, 2. activity, 3. attributes, 4. circumstances</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MSE.3. Textual metafunction</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE.3.1. MSE unit: 1. math symbol, 2. operations, 3. Arabic numeral, 4. variable, 5. scientific notation</td>
</tr>
<tr>
<td>MSE.3.2. MSE emphasis: 1. bolding, 2. shading</td>
</tr>
<tr>
<td>MSE.3.3. MSE location: 1. embedded in NL, 2. between stem/prompt, 3. left of RS, 4. above RS, 5. embedded in schematic visual, 6. embedded in concrete visual, 7. embedded in RS, 8. embedded in response tool</td>
</tr>
<tr>
<td>MSE.3.4. Reference to MSE: 1. explicit, 2. not stated</td>
</tr>
<tr>
<td>MSE.3.5. Stated actions to perform w/ MSE: 1. drag, 2. indicate/select</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VR.1. Interpersonal metafunction</th>
</tr>
</thead>
<tbody>
<tr>
<td>VR.1.1. Interpersonal: 1. indicative, 2. neutral, 3. factual, 4. helpful</td>
</tr>
</tbody>
</table>
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

VR.2. Ideational metafunction
   VR.2.1. VR purpose: 1. response, 2. identity, 3. measurement, 4. discrete number, 5. category, 6. infinite, 7. distance, 8. time, 9. function, 10. concrete image, 11. shape, 12. map
   VR.2.2. VR functional categories: 1. participants, 2. activity, 3. attributes, 4. circumstances

VR.3. Textual metafunction
   VR.3.1. VR type: 1. concrete, 2. schematic, 3. symbolic attribute, 4. symbol 5. RS
   VR.3.2. Composition: 1. single image, 2. compound image
   VR.3.3. VR location: 1. below text, 2. embedded in stem/prompt, 3. embedded in RS, 4. embedded in response tool, 5. right of text
   VR.3.4. Reference to VR: 1. explicit, 2. not stated
   VR.3.5. Stated actions to perform w/ VR: 1. drag, 2. enter, 3. observe, 4. select, 5. graph/draw
   VR.3.6. Image concreteness: 1. realistic, 2. unrealistic, 3. scheme, 4. NA
   VR.3.7. Background: 1. with background, 2. without background
   VR.3.8. View: 1. side/external, 2. internal, 3. NA
   VR.3.9. Dimension: 1. 2D, 2. 3D
   VR.3.10. Relative position of objects: 1. relevant, 2. irrelevant
   VR.3.11. Relative scale of objects: 1. proportionate, 2. disproportionate
   VR.3.12. Color: 1. black and white, 2. gray scale, 3. color
   VR.3.13. Schematic type: 1. number line, 2. categorical table, 3. function, 4. coordinate graph, 5. points, 6. line/segment, 7. 2D shape, 8. 3D shape
   VR.3.15. Intervals: 1. equal between axes, 2. unequal between axes, 3. NA
   VR.3.16. Coordinate plane: 1. grid, 2. 1 quadrant, 3. 2 quadrants, 4. 3 quadrants, 5. 4 quadrants, 5. no coordinate plane

Note. NL is natural language, MSE is mathematical symbols and expressions, and VR is visual representations. Metafunctions are in bold letters, dimensions are underlined, and variables are in italics.

The second coding system used Halliday’s (1978, 2004) notions of functional categories and textual correspondence and Royce’s (2002) process of intersemiotic ideational analysis to examine intersemiosis. Unlike the first coding system, which focused on the semiotic components of items, this coding system did not create dichotomous codes for the presence or absence of features. Rather, it identified the components in each semiotic resource that belonged in each functional category. Table 4.6 shows the functional categories and their definitions.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Table 4.6

Functional Categories

<table>
<thead>
<tr>
<th>Functional category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>The represented constituents.</td>
</tr>
<tr>
<td>Activity</td>
<td>The action that is occurring between the actors, recipients, and/or objects of the action.</td>
</tr>
<tr>
<td>Attributes</td>
<td>The participants’ characteristics.</td>
</tr>
<tr>
<td>Circumstances</td>
<td>The elements concerned with the setting.</td>
</tr>
</tbody>
</table>

The functional category constituents were then compared across each pair of semiotic resource (i.e., NL and VR, NL and MSE, and VR and MSE) and scored dichotomously (presence = 1; absence = 0) according to the seven intersemiotic relations: (a) Repetition – repetition of the same semantic meaning, (b) Synonymy – similar relations, (c) Antonymy – opposite relations or inverse relations, (d) Hyponymy – class-subclass relations, (e) Metonymy – part-whole relations, (f) Collocation – expectancy relations, and (g) Identity – relational processes. I added inverse relations to Intersemiotic Antonymy, as natural language was used to describe the inverse properties of a number in one of the items. This process mirrored Royce’s (2002) intersemiotic ideational analysis between NL and VR, but it also expanded upon the process as it was also applied to the relations between NL and MSE and between MSE and VR.

Appendix I shows four examples of the application of the coding systems.

Analysis. After coding each item using both coding systems, I examined the absolute and relative frequencies of IDVs coded 1. The frequency and percentages of IDVs were disaggregated by semiotic resource and metafunction. Frequency analyses were also conducted to examine which types of intersemiotic relations were most common and for which pairs of semiotic resources.
The item semiotic complexity \((I_C)\) of each item was calculated by adding together the semiotic resource complexities, as shown in equations 4.1 through 4.4:

\[ I_C = NL_C + MSE_C + VR_C \]  
\[ NL_C = \sum NL \text{ Ideational components} + \sum NL \text{ Interpersonal components} + \sum NL \text{ Textual components} \]  
\[ MSE_C = \sum MSE \text{ Ideational components} + \sum MSE \text{ Interpersonal components} + \sum MSE \text{ Textual components} \]  
\[ VR_C = \sum VR \text{ Ideational components} + \sum VR \text{ Interpersonal components} + \sum VR \text{ Textual components} \]

where \(c\) is complexity, \(NL\) is natural language, \(MSE\) is mathematical symbols and expressions, and \(VR\) is visual representation. A line graph was constructed to model and compare the semiotic resource complexities and item semiotic complexity of each item \((NL_C, MSE_C, \text{ and } VR_C, \text{ and } I_C)\). Descriptive statistics were also calculated to compare complexities across items.

The semiotic resource complexities and item semiotic complexity of each item were disaggregated and compared according to item type (i.e., multiple choice, graphing, and constructed response) and knowledge domain (i.e., number system, expressions and equations, geometry, functions, ratios and proportional relations, and statistics and probability).

The semiotic composition and complexity, as well as the intersemiotic relations, of the four interview items were then elaborated upon to facilitate the findings and discussion of Study 2.
Study 2: EB and Non-EB students’ Sense Making of Semiotic Resources in SBAC Items

**Instruments and procedures.** I developed an interview protocol to examine how EB and non-EB students interacted with the semiotic composition and complexity of four items from the 2015 SBAC Grade 8 Math Practice Test as they responded to them. Refer to Appendix J, *Pilot Stage of the Development of Instruments*, for a detailed description of the development of the interview protocol. The original protocol was modified using information gathered from six pilot interviews and 20 classroom observations. The original protocol was written in English (Appendix K) and translated into Spanish (Appendix L).

The final interview protocol comprised four phases: Introduction, Concurrent Reporting, Retrospective Reporting, and Follow-Up Questions. The Introduction Phase allowed me to introduce myself and the goal of the activity. I explained the purpose for and procedures of the cognitive interview. I also asked the students questions concerning demographic information (e.g., age, ethnicity/race, and spoken languages).

The Concurrent Reporting Phase allowed me to indirectly observe students’ cognitive processes as they interacted with and solved the four items. During this phase, students had to read the questions aloud and describe what they were thinking and doing and why. As recommended by Ericsson and Simon (1993), I minimized my influence on students’ thinking by limiting my interactions with them. For example, I limited my speech to “What are you thinking about?” or “Why did you do that?” when students were quiet, so as not to interfere. Items 1, 5, 9, and 24 were chosen for this phase because (a) students are required to master the content of each of these items in eighth grade; (b) they are
exposed to the material several times in eighth grade and again in at least one, if not all, high school math classes; (c) the vocabulary is not as specialized as that used in other items; and (d) they only contain a few steps. Appendix M contains descriptions of each selected item (1, 5, 9, and 24) and a table of their English and/or Spanish glossaries.

The Retrospective Reporting Phase allowed me to ask specific questions pertaining to the semiotic resources of each item and students’ cognitive processes. In this phase, I asked students scripted and spontaneous probes. Examples of the scripted probes include asking students what the questions were asking them to do, and asking them to rate the mental effort they used to solve each problem. Spontaneous probes emerged naturally during each interview and captured important information that was not addressed in the scripted probes (Beatty & Willis, 2007; Willis, 1999). Items 1, 5, 9, and 24 were also used for this phase.

The Follow-Up Phase allowed me to ask questions that were relevant to the task, but not specific to the semiotic components of the items or the problem-solving process. I asked these questions after the students completed the Concurrent and Retrospective Phases for each of the four items. These questions addressed the germane cognitive load of the four items; their preparation for, perceived achievement on, and thoughts about the PARCC assessment; and issues pertaining to classroom instruction and general test-taking. While the items I used were SBAC items, the follow-up questions were equally applicable for SBAC and PARCC. Appendices N and O contain the final interview protocol in English and Spanish, respectively.

**Data collection.** I conducted cognitive interviews with 30 Grade 11 EB and 30 Grade 11 non-EB students using the interview protocol described above. The interviews
were audio recorded and screenshots were taken to capture the students’ online responses and scratch work. The think-aloud portion of the interviews were transcribed to ensure that I captured all the important information from the think-aloud process.

Students were interviewed individually and used a laptop to complete the activity. While they were told that they were free to respond in English, Spanish, or both, depending on their preferences, all of them chose to speak only in English. Students also had the option to use the stacked translation accessibility resource, which gives the question in English and Spanish, but none of them chose to use it. Students were given no time restriction to answer the items using the online interface.

**Coding.** The audio recordings, transcriptions, and screenshots of students’ responses and work were coded deductively and inductively to examine how semiotic complexity, and total cognitive load influenced students’ interpretations of and responses to items, along with their problem-solving processes.

The coding process began by listening to the audiorecordings of each interview. Semantic units—a sentence, part of a sentence, or two or more sentences that represented a given idea (Gebre & Polman, 2016)—were transcribed into an interview coding form. The coding form contained categories that were created from the conceptual framework (i.e., semiotic metafunctions, intersemiosis, interpreting, problem-solving, responding, intrinsic cognitive load, germane cognitive load, and extraneous cognitive load).

The semantic units that were transcribed into the interview coding form were then coded deductively and inductively using open coding with constant comparison (as described in the item analysis data coding section). The interviews were coded deductively using the item coding system (see Appendix H), Sweller’s (2005) work on
knowledge and memory, Brünken et al.’s (2003) theories of cognitive load and overload, and the three-part sequence of responding to items that I developed in my conceptual framework. The transcriptions from the think-aloud portion of the interview and the screen shots of the students’ answer and their scratch work were coded using the same process as the audiorecordings.

Table 4.7 shows the categories, their dimensions, and the corresponding variables for the interview coding system. Appendix P shows a more detailed interview codebook containing definitions for each category, dimension, and variable. Appendix Q shows an example of the application of the interview coding system. The final interview coding system contains two continuous variables (item rating and total item rating) and 119 dichotomous variables.

The first part of the coding system examined how students’ reported actions and thinking involved the ideational, interpersonal, and textual metafunctions of the NL, MSE, and VR and the process of intersemiosis (Categories 1-9). For example, when asked to describe the picture from Item 9 in as much detail as possible, a student said, “It’s a picture of a tree and a ladder leaning up against it. It shows that the ladder is 13 feet and the distance from the tree to the ladder is five feet” (personal communication, May 4, 2015). The student identified several textual components, such as the use of concrete visuals and Arabic numerals. When asked to describe how the text and graph were related in Item 24, a student said, “The numbers actually show the linear relationship that is mentioned in the text” (personal communication, April 29, 2015). The student, indirectly, identified the intersemiotic relation of Identity between NL and MSE.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

The second part of the coding system identified student response processes pertaining to cognitive load (Categories 10-12). Note that the intrinsic cognitive load (item rating) and the germane cognitive load (total item rating) variables are not dichotomous but rather continuous, as students rated the difficulty of each item and the difficulty of the items overall on a scale of 1 to 10, as recommended by DeLeeuw and Mayer (2008), respectively.

The third part of the coding system addressed the three-part sequence of responding to mathematics assessment items that I developed in my conceptual framework (Categories 13-15). For example, when a student said, “I just guessed,” (personal communication, April 21, 2015) I coded this as Guess. Or, when a student said, “It was pretty easy to move the fractions [in Item 1]” (personal communication, April 10, 2015), I coded this as Response Tool Easy.

The last part of the coding system addressed the item changes proposed by students and their thoughts pertaining to the calculator tool and pop-up glossaries. For example, when a student said, “You should add more words [to Item 1] to describe how to solve it” (personal communication, April 2, 2015), I coded this as Add NL.
### Table 4.7

**Interview Coding System**

1. **NL – Ideational metafunction**
   1.1. Text purpose: 1. directions, 2. clarity, 3. visualize, 4. think more
   1.2. Text cohesion: 1. Repetition
2. **NL – Interpersonal**
   2.1. Interpersonal: 1. directive, 2. factual, 3. helpful
3. **NL – Textual metafunction**
   3.1. Text unit: 1. abbreviation, 2. word, 3. phrase
   3.2. Text emphasis: 1. bolding
4. **MSE – Ideational metafunction**
   4.1. MSE purpose: 1. discrete numbers on graph, 2. identity
   4.2. Textual cohesion: 1. Identity, 2. Repetition, 3. Synonymy
5. **MSE – Textual metafunction**
   5.1. MSE unit: 1. Arabic numerals/values
6. **VR – Ideational metafunction**
   6.1. VR purpose: 1. infinity, 2. identity, 3. discrete numbers, 4. direction, 5. inequality, 6. measurement, 7. function, 8. categories, 9. decorative, 10. visualize, 11. easy to look at
7. **VR – Interpersonal metafunction**
   7.1. Interpersonal: 1. Directive, 2. helpful, 3. factual
8. **VR – Textual metafunction**
   8.1. Schematic type: 1. number line, 2. categorical table, 3. line/segment
   8.2. Axes: 1. x-axis, 2. y-axis
   8.3. Intervals: 1. equal between axes, 2. unequal between axes
   8.4. Context: 1. unfamiliar
9. **Intersemiosis (for each pair of semiotic resource)**
10. **Intrinsic cognitive load**
    10.1. Item rating: 1. item rating
11. **Germane cognitive load**
    11.1. Rating: 1. 4-item rating
    11.2. Key semiotic resources: 1. NL, 2. MSE, 3. VR
    11.3. Emphasis: 1. bold, 2. underline, 3. color, 4. italicize, 5. large font size, 6. not specified
    11.4. Tools: 1. list of equations
12. **Extraneous cognitive load**
    12.1. Extraneous features: 1. MSE answer box, 2. real-life example, 3. decorative pieces
13. **Interpretation**
    13.1. Interpretation: 1. correct, 2. correct with help, 3. incorrect
    13.2. Process: 1. struggles with word(s), 2. asks clarifying questions
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

13.3. **Student statements:** 1. confusing, 2. NL clear, 3. NL unnecessary, 4. VR unnecessary

14. **Problem-Solving**

14.1. **Schema:** 1. correct, 2. incorrect

14.2. **Process:** 1. logical estimation, 2. math error, 3. does not know, 4. math question, 5. hint/question, 6. guess, 7. difficulty in English, 8. checks answer

14.3. **Semiotic Resources:** 1. NL verbal, 2. NL paper, 3. MSE paper, 4. schematic paper, 5. concrete paper, 6. hands

15. **Responding**

15.1. **Representation:** 1. does, 2. does not

15.2. **Response tool:** 1. easy, 2. hard, 3. unfamiliar with parts, 4. asks questions, 5. help, 6. calculator

15.3. **Suggestions:** 1. add NL, 2. add MSE, 3. remove MSE

16. **Item changes**

16.1. **Suggested changes:** 1. add NL, 2. NL clearer, 3. NL location, 4. replace NL with MSE, 5. add schematic VR, 6. add concrete VR, 7. remove concrete VR, 8. add color, 9. change response tool

17. **Extra tools**

17.1. **Calculator:** 1. likes, 2. hard, 3. modify, 4. prefers TI, 5. asked questions, 6. help

17.2. **Glossary:** 1. used, 2. made sense, 3. helpful, 4. more detail, 5. dialect, 6. add non-essential gloss, 7. add essential gloss, 8. add VR, 9. did not help

---

**Note.** NL is natural language, MSE is mathematical symbols and expressions, and VR is visual representations. The categories are in bold letters, dimensions are underlined, and variables are in italics. The intrinsic item rating (10.1) and germane rating (11.1) are not dichotomous, rather continuous on a scale from 1 to 10.

---

**Analysis.** The students’ responses to each of the four items were scored using the *SBAC: Practice Test Scoring Guide Grade 8 Mathematics* (SBAC, 2014). A total score was computed for each student by adding his/her score on Items 1, 5, 9, and 24.

Percentages were calculated for item score and total score. The total cognitive load of each item was computed by adding each student’s intrinsic item rating, his/her germane rating, and his/her mention of extraneous components, as modeled in the following equation:

\[
CL_T = \text{Intrinsic Item Rating} + \text{Germane Rating} + \sum \text{Extraneous features} \quad [\text{Eq.4.2}]
\]
Descriptive statistics were calculated for the score and the total cognitive load of each item and all items combined. Spearman’s rank-order correlations were calculated to examine the correlation between the score and the total cognitive load of each item and all items combined. Scatterplots were created to examine the relation between item semiotic complexity and total cognitive load, as well as the relation between item semiotic complexity and score. These relations were examined for the sample as a whole and disaggregated by language group. A series of one-way analysis of variance (ANOVA) was conducted to compare the effect of language group on score and total cognitive load of each item.

Frequency analyses were conducted to examine the frequency with which each language group’s reported actions and thinking involved the metafunctional components identified in Study 1. Proportions were calculated using the frequency results, and Chi-square tests of independence were calculated to examine whether the proportion of non-EBs and EBs whose actions and thinking involved the metafunctional components was significantly different. A series of two-way ANOVAs was conducted to compare the main effects of language group (non-EB and EB) and item (Items 1, 5, 9, and 24) and the interaction effect between language group and item on students’ reported actions and thinking involving each semiotic resource (NL, MSE, and VR) and each metafunction (ideational, interpersonal, and textual).

Frequency analyses were conducted to examine the frequency with which each language group’s reported actions and thinking involved the intersemiotic relations identified in Study 1. Proportions were calculated using the frequency results, and Chi-square tests of independence were calculated to examine whether the proportion of non-EBs and EBs whose actions and thinking involved the intersemiotic relations was significantly different. Other data, not included in this dissertation, were collected.
square tests of independence were calculated to examine whether the proportion of non-EBs and EBs whose actions and thinking involved intersemiotic relations was significantly different.

Lastly, a series of two-way ANOVAs was conducted to compare the main effects of language group (non-EB and EB) and item (Items 1, 5, 9, and 24) and the interaction effect between language group and item on the following variables: Student Struggled During Item Interpretation, Glossary Helped, Student Suggested NL Change, Student Did Not Try, Student Struggled During Problem-Solving, Student Solved Using Something Other than the Computer, Student Suggested an MSE Change, Response Does Not Represent Student’s Thinking, and Student Says Response Tool Was Hard. Spearman’s rank-order correlations were also computed to compare the correlation between the aforementioned variables and total cognitive load.
CHAPTER V

RESULTS

Overview

Study 1. I conducted a series of frequency analyses to examine the metafunctional components of each semiotic resource in each item. Frequency analyses were also conducted to examine which types of intersemiotic relations were most common and for which pairs of semiotic resources. I computed the semiotic resource complexities and the item semiotic complexity of each item. Descriptive statistics were calculated to compare semiotic resource complexities and item semiotic complexity across items, item type, and knowledge domain. These analyses were repeated using only the four items included in the cognitive interviews (Items 1, 5, 9, and 24. These results were used to help understand these items better and guide analyses of the cognitive interview data.

Study 2. Each student received a score on Items 1, 5, 9, and 24. I calculated a total score for each student by adding his/her scores on all four items. The scores were computed into percentages. The total cognitive load of each item was calculated by adding each student’s intrinsic item rating, germane item rating, and mention of extraneous components.

Descriptive statistics were computed for the score and the total cognitive load of each item and all items combined. Spearman’s rank-order correlations were computed to examine the correlation between the score and total cognitive load of each item and all
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

items combined. Scatterplots were created to examine the association between item semiotic complexity and total cognitive load and between item semiotic complexity and score for the entire sample and disaggregated by language group. A one-way analysis of variance (ANOVA) was conducted to compare the effect of language group (emergent bilingual [EB] and non-EB) on score and total cognitive load of each item.

Frequency analyses were conducted to examine the frequency with which each language group’s reported actions and thinking involved the metafunctional components and intersemiotic relations identified in Study 1. Proportions were calculated from the frequency results and Chi-square tests of independence were computed to examine whether the proportion of non-EBs and EBs whose actions and thinking involved the metafunction components and intersemiotic relations were significantly different.

Descriptive statistics and two-way ANOVAs were conducted to compare the main effects of language group (non-EB and EB) and item (Items 1, 5, 9, and 24) and the interaction effect between language group and item on students’ reported actions and thinking involving each semiotic resource (NL, MSE, and VR) and each metafunction (ideational, interpersonal, and textual), as well as the dependent variables related to interpreting, problem-solving, and responding (i.e., the three-part sequence of responding to an item). Spearman’s rank-order correlations were also computed to compare the correlation between the dependent variables related to the three-part sequence and total cognitive load.
Study 1: Semiotic Resources Used in SBAC Test Items

  **Semiotic composition.**

  *Natural language.* Table 5.1 shows the absolute and relative frequencies of the item dichotomous variables (IDVs) coded 1 corresponding to the ideational, interpersonal, and textual metafunctions of the natural language (NL) used in the items. The table shows only the IDVs observed in at least 50% of the items IDVs to allow for an examination of the most frequent variables. These findings indicate that the NL used in these mathematics test items had different metafunctions than those used in the curriculum materials found by O’Halloran (2005). NL was not primarily used to construct relational processes or nominal groups, but rather to give students directions on what they needed to do to solve the items. Naturally, NL was not primarily rational or objective, but rather directive. The textual metafunction was not used to carry mathematical arguments forward, but rather to give directions using features such as present tense, simple predicates, and Spanish glosses.
Table 5.1

Frequencies and Percentages (Rounded) of the Semiotic Components of Natural Language

<table>
<thead>
<tr>
<th>Item Dichotomous Variables</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ideational Metafunction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Directions</td>
<td>24</td>
<td>83</td>
</tr>
<tr>
<td>Relational processes</td>
<td>17</td>
<td>59</td>
</tr>
<tr>
<td>Nominalization</td>
<td>22</td>
<td>76</td>
</tr>
<tr>
<td><strong>Interpersonal Metafunction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Directive</td>
<td>25</td>
<td>86</td>
</tr>
<tr>
<td>Factual</td>
<td>21</td>
<td>72</td>
</tr>
<tr>
<td><strong>Textual Metafunction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Word</em></td>
<td>15</td>
<td>52</td>
</tr>
<tr>
<td><em>Multiple sentences</em></td>
<td>24</td>
<td>83</td>
</tr>
<tr>
<td>Emphasis</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Bolding</em></td>
<td>17</td>
<td>59</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Above response space</em></td>
<td>17</td>
<td>59</td>
</tr>
<tr>
<td><em>Embedded in response space</em></td>
<td>15</td>
<td>52</td>
</tr>
<tr>
<td>Complexity</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Has multi-word noun phrases</em></td>
<td>22</td>
<td>76</td>
</tr>
<tr>
<td><em>Present tense</em></td>
<td>29</td>
<td>100</td>
</tr>
<tr>
<td>Constituents</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Simple subject</em></td>
<td>20</td>
<td>69</td>
</tr>
<tr>
<td><em>Simple predicate</em></td>
<td>29</td>
<td>100</td>
</tr>
<tr>
<td><em>Adjective</em></td>
<td>24</td>
<td>83</td>
</tr>
<tr>
<td><em>Direct object</em></td>
<td>28</td>
<td>97</td>
</tr>
<tr>
<td>Register</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Testing</em></td>
<td>28</td>
<td>97</td>
</tr>
<tr>
<td><strong>Glossary Components</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>English</em></td>
<td>15</td>
<td>52</td>
</tr>
<tr>
<td><em>Spanish</em></td>
<td>29</td>
<td>100</td>
</tr>
<tr>
<td><em>Natural language register</em></td>
<td>21</td>
<td>72</td>
</tr>
<tr>
<td><em>Mathematics register</em></td>
<td>15</td>
<td>52</td>
</tr>
<tr>
<td><em>Testing register</em></td>
<td>28</td>
<td>97</td>
</tr>
</tbody>
</table>

*Note.* Total number of items = 29.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Mathematical symbols and expressions. Table 5.2 shows the absolute and relative frequencies of the IDVs coded 1 corresponding to the ideational, interpersonal, and textual metafunctions of the mathematical symbols and expressions (MSE) used in the items. The table shows only the IDVs observed in at least 50% of the items IDVs to allow for an examination of the most frequent variables. These findings indicate that the MSE used in these mathematics test items had a similar but more specific ideational metafunction, different interpersonal metafunctions, and similar but more detailed textual metafunctions than those used in curriculum materials. The main ideational function of MSE was to represent relations between mathematical elements and variables, but, in this practice test, the relation was that of a numerical value. Interpersonally, MSE was factual and/or neutral as it provided students with the values they needed to solve the items or provided them with MSE to enter their answers, respectively. This is contrary to the MSE used in curriculum materials that is authoritative and non-negotiable, as the MSE in these items was not used to tell students how to solve, but rather to allow them to demonstrate their knowledge. The textual metafunction of the MSE used highly conventionalized notation, such as Arabic numerals, and spatial positioning, but the findings also discovered other structural rituals, such as MSE location and reference.
Table 5.2

**Frequencies and Percentages (Rounded) of the Semiotic Components of Mathematical Symbols and Expressions**

<table>
<thead>
<tr>
<th>Item Dichotomous Variables</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideational Metafunction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Represent discrete numbers/values (not on graph)</td>
<td>16</td>
<td>55</td>
</tr>
<tr>
<td>Interpersonal Metafunction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factual</td>
<td>20</td>
<td>69</td>
</tr>
<tr>
<td>Neutral</td>
<td>20</td>
<td>69</td>
</tr>
<tr>
<td>Textual Metafunction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arabic numeral/value</td>
<td>29</td>
<td>100</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Embedded in natural language</td>
<td>19</td>
<td>66</td>
</tr>
<tr>
<td>Above response space</td>
<td>17</td>
<td>59</td>
</tr>
<tr>
<td>Embedded in schematic visual</td>
<td>20</td>
<td>69</td>
</tr>
<tr>
<td>Embedded in response space</td>
<td>19</td>
<td>66</td>
</tr>
<tr>
<td>Reference to mathematical symbols and expressions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit</td>
<td>18</td>
<td>62</td>
</tr>
</tbody>
</table>

*Note.* Total number of items = 29.

**Visual representations.** Table 5.3 shows the absolute and relative frequencies of the IDVs coded 1 corresponding to the ideational, interpersonal, and textual metafunctions of the visual representations (VR) used in the items. The table shows only the IDVs observed in at least 50% of the items IDVs to allow for an examination of the most frequent variables. These findings indicate that the VR used in these mathematics test items had a different ideational metafunction, a similar interpersonal metafunction, and similar but more detailed textual metafunctions than those used in curriculum materials. Unlike curriculum materials that use VR to represent perceptual depictions of the world, the main ideational function of VR used in these items was to simply create a space for
students to respond. Interpersonally, the visuals still held a high-truth value as they often presented students with facts. Textually, the VR in these items was used to focus the attention of the viewer to the experiential aspects of the mathematical content using features such as visuals without a background and black and white visuals.
Table 5.3

Frequencies and Percentages (Rounded) of the Semiotic Components of Visual Representations

<table>
<thead>
<tr>
<th>Item Dichotomous Variables</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ideational Metafunction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response</td>
<td>29</td>
<td>100</td>
</tr>
<tr>
<td><strong>Interpersonal Metafunction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>17</td>
<td>59</td>
</tr>
<tr>
<td>Factual</td>
<td>18</td>
<td>62</td>
</tr>
<tr>
<td><strong>Textual Metafunction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illustration type</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Schematic</em></td>
<td>20</td>
<td>69</td>
</tr>
<tr>
<td><em>Symbolic attribute</em></td>
<td>16</td>
<td>55</td>
</tr>
<tr>
<td><em>Symbol</em></td>
<td>15</td>
<td>52</td>
</tr>
<tr>
<td>Composition</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Compound image</em></td>
<td>25</td>
<td>86</td>
</tr>
<tr>
<td>Reference to illustration</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Explicit</em></td>
<td>22</td>
<td>76</td>
</tr>
<tr>
<td>Background</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Without</em></td>
<td>28</td>
<td>97</td>
</tr>
<tr>
<td>View</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>From the side/external</em></td>
<td>18</td>
<td>62</td>
</tr>
<tr>
<td>Dimension</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>2D</em></td>
<td>27</td>
<td>93</td>
</tr>
<tr>
<td>Relative position of objects</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Relevant</em></td>
<td>21</td>
<td>72</td>
</tr>
<tr>
<td><em>Irrelevant</em></td>
<td>23</td>
<td>79</td>
</tr>
<tr>
<td>Relative scale of objects</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Proportionate</em></td>
<td>25</td>
<td>86</td>
</tr>
<tr>
<td>Color</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Black and white</em></td>
<td>28</td>
<td>97</td>
</tr>
</tbody>
</table>

*Note.* Total number of items = 29.
Overall, the NL, MSE, and NL used in these test items had different metafunctional components than the curriculum materials examined by O’Halloran (2005). The items on the SBAC practice test made meaning differently than curriculum materials. Teachers should inform students how test items make meaning and how it differs from class materials to improve student achievement on assessments.

**Semiotic resource complexities and item semiotic complexity.** Figure 5.1 shows the semiotic resource complexities and the item semiotic complexity for each item. MSE was less complex than NL and VR in every item, except for Item 1 in which MSE was greater than NL but less than VR. Natural language and VR had similar ranges across items, but the number of IDVs usually differed between the two in an item. Overall, the semiotic resource and item complexities was inconsistent across items, which could potentially affect student achievement.

![Figure 5.1. Semiotic resource complexities and item semiotic complexity for each item.](image-url)
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Table 5.4 shows descriptive statistics for the semiotic resource complexities \((NL_C, MSE_C, VR_C)\) and item semiotic complexity \((I_C)\) for each item. MSE had the smallest average \((\bar{x} = 13.90)\) while VR had the largest average \((\bar{x} = 28.31)\). The average \(I_C\) was 69.62; on average, students had to make meaning, as intended by the item writers, from approximately 69 different components. These results should be taken with caution due to the small sample of items.

Table 5.4

*Descriptive Statistics for Semiotic Resource Complexities and Item Semiotic Complexity*

<table>
<thead>
<tr>
<th>Type of Complexity</th>
<th>Average</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural Language Semiotic Resource</td>
<td>27.41</td>
<td>5.81</td>
<td>14</td>
<td>38</td>
</tr>
<tr>
<td>Complexity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematical Symbols and Expressions</td>
<td>13.90</td>
<td>2.91</td>
<td>7</td>
<td>20</td>
</tr>
<tr>
<td>Semiotic Resource Complexity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual Representations Semiotic</td>
<td>28.31</td>
<td>6.69</td>
<td>19</td>
<td>37</td>
</tr>
<tr>
<td>Resource Complexity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item Semiotic Complexity</td>
<td>69.62</td>
<td>8.77</td>
<td>52</td>
<td>87</td>
</tr>
</tbody>
</table>

*Note.* Total number of items = 29 and total number of IDVs coded 1 = 2,019.

Table 5.5 shows descriptive statistics for the semiotic resource complexities \((NL_C, MSE_C, VR_C)\) and item semiotic complexity \((I_C)\) disaggregated by item type and knowledge domain. Of the different item formats, multiple choice items had the smallest average item semiotic complexity \((I_C = 66.91)\) and graphing items had the largest \((I_C = 74.83)\). Of the six knowledge domains, Statistics and Probability had the smallest average item semiotic complexity \((I_C = 67.00)\) and Ratios and Proportional Relations had the largest \((I_C = 80.00)\). Again, the inconsistency between semiotic resource
complexities and item semiotic complexity could potentially affect student achievement and introduce construct-irrelevant variance. These results should be taken cautiously given the limited number of items representing each item type and knowledge domain.

Table 5.5

Descriptive Statistics for Semiotic Resource Complexities and Item Semiotic Complexity

Disaggregated by Item Type and Knowledge Domain

<table>
<thead>
<tr>
<th>Categorical Variable</th>
<th>Frequency</th>
<th>NL Mean (SD)</th>
<th>MSE Mean (SD)</th>
<th>VR Mean (SD)</th>
<th>Item Semiotic Complexity Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Item Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constructed response</td>
<td>12</td>
<td>25.83 (6.44)</td>
<td>13.67 (3.50)</td>
<td>30.00 (5.86)</td>
<td>69.50 (9.06)</td>
</tr>
<tr>
<td>Graphing</td>
<td>6</td>
<td>28.33 (6.09)</td>
<td>12.67 (0.52)</td>
<td>33.83 (2.14)</td>
<td>74.83 (6.71)</td>
</tr>
<tr>
<td>Multiple choice</td>
<td>11</td>
<td>28.64 (5.03)</td>
<td>14.82 (2.86)</td>
<td>23.45 (6.14)</td>
<td>66.91 (8.85)</td>
</tr>
<tr>
<td><strong>Knowledge Domain</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number system</td>
<td>3</td>
<td>22.67 (5.03)</td>
<td>15.33 (3.51)</td>
<td>29.33 (6.11)</td>
<td>67.33 (7.77)</td>
</tr>
<tr>
<td>Expressions and equations</td>
<td>11</td>
<td>26.36 (7.23)</td>
<td>13.91 (1.58)</td>
<td>25.91 (7.74)</td>
<td>66.18 (10.15)</td>
</tr>
<tr>
<td>Geometry</td>
<td>8</td>
<td>27.50 (4.41)</td>
<td>14.63 (3.58)</td>
<td>29.00 (7.07)</td>
<td>71.13 (7.43)</td>
</tr>
<tr>
<td>Functions</td>
<td>5</td>
<td>30.20 (3.63)</td>
<td>13.00 (2.92)</td>
<td>31.40 (4.62)</td>
<td>74.60 (7.34)</td>
</tr>
<tr>
<td>Ratios and proportional relations</td>
<td>1</td>
<td>32.00 (NA)</td>
<td>15.00 (NA)</td>
<td>33.00 (NA)</td>
<td>80.00 (NA)</td>
</tr>
<tr>
<td>Statistics and probability</td>
<td>1</td>
<td>34.00 (NA)</td>
<td>7.00 (NA)</td>
<td>26.00 (NA)</td>
<td>67.00 (NA)</td>
</tr>
</tbody>
</table>

*Note.* NL is natural language, MSE is mathematical symbols and expressions, and VR is visual representations. Total number of items = 29 and total number of IDVs coded 1 = 2,019.
Intersemiosis. Figure 5.2 shows the relative frequencies with which intersemiotic relations were observed for each pair of semiotic resources. Of the 104 IDVs coded 1, the most common intersemiotic relation was Collocation between NL and VR (28%). The most common intersemiotic relation between NL and MSE was Identity (20%), and the most common intersemiotic relation between VR and MSE was Repetition (12%). Thus, SBAC items create meaning by using NL to indicate the presence of a VR; using NL to identify MSE at a higher level of generality; and using VR to repeat the meaning of MSE. Furthermore, 58% of the relations were between natural language and visual representations, 27% of the relations were between natural language and mathematical symbols and expressions, and 15% of the relations were between visual representations and mathematical symbols and expressions. Thus, SBAC items create meaning mostly through the integration of NL and VR.
Figure 5.3 shows the relative frequencies with which intersemiotic relations were observed for each pair of semiotic resources disaggregated by item type. Of the 104 IDVs coded 1, thirty-nine percent of the intersemiotic relations belonged to constructed response items, 36% belonged to multiple choice items, and 25% belonged to graphing items. The most common intersemiotic relation for constructed response items was Repetition between NL and VR (13%). The most common intersemiotic relation for multiple choice items was Identity between NL and MSE (13%). The most common intersemiotic relations for graphing items were Repetition and Collocation between NL and VR and Repetition between MSE and VR (6% each).

Figure 5.4 shows the relative frequencies with which intersemiotic relations were observed for each pair of semiotic resources disaggregated by knowledge domain.
104 IDVs coded 1, thirty-seven percent of the intersemiotic relations belonged to Expressions and Equations items and only 2% belonged to Statistics and Probability. The analyses that produced the results in Figures 5.3 and 5.4 should be replicated with a larger sample of items that have an equal distribution of item type and knowledge domain in order to make any general claims about the difference in intersemiosis relation types and frequencies.

*Figure 5.4. Relative frequency of each intersemiotic relation for each pair of semiotic resources disaggregated by knowledge domain.*

**Description of the items used in the cognitive interviews.** Tables 5.6, 5.7, 5.8, and 5.9 show the metafunctional components, or IDVs, of each semiotic resource for the items included in the cognitive interviews: Items 1, 5, 9, and 24, respectively. While the metafunction components varied across the four items, they had some similarities. Ideationally, NL was used to give students directions; MSE was to represent discrete numbers/values that were not on a graph or represent a measurement; and VR was used to create a response space for students. Interpersonally, NL was directive; MSE was
neutral; and VR was factual. Textually, NL used the present tense, simple predicates, and Spanish glosses; MSE used Arabic numerals; and VR was embedded within the response space.
### Table 5.6

**Semiotic Components of Item 1**

<table>
<thead>
<tr>
<th>Metafunction</th>
<th>Natural Language</th>
<th>Mathematical Symbols and Expressions</th>
<th>Visual Representations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideational</td>
<td>Give directions; functional categories: participants, activity, circumstances</td>
<td>Represent discrete numbers/values not on graph; label discrete points on graph/number line; represent a symbol; represent an operation; functional categories: participants, activity, attributes</td>
<td>Response; identity; discrete numbers; infinite; functional categories: participants, activity, attributes</td>
</tr>
<tr>
<td>Interpersonal</td>
<td>Directive</td>
<td>Neutral; non-negotiable; helpful</td>
<td>Indicative; factual</td>
</tr>
<tr>
<td>Textual</td>
<td>Sentence; left of response space; multi-word noun phrase; present tense; simple predicate; adjective; direct object; indirect object; testing resister; glossary: English, Spanish, NL register, testing register</td>
<td>Math/scientific symbol; operations; Arabic numeral/value; shading; embedded in schematic visual; embedded in RS; embedded in response tool; explicit reference; stated action to drag</td>
<td>Schematic; symbolic attribute (lines/arrows); response space; single image; embedded in response space; embedded in response tool; right of text; explicit reference; without background; from the side; 2D; position of objects relevant; scale of objects proportionate; grayscale; number line; x-axis</td>
</tr>
</tbody>
</table>
### Table 5.7

**Semiotic Components of Item 5**

<table>
<thead>
<tr>
<th>Metafunction</th>
<th>Natural Language</th>
<th>Mathematical Symbols and Expressions</th>
<th>Visual Representations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideational</td>
<td>Give directions; relational processes; non-mathematical context/background; nominalization; label schematic visual components; label response tools; functional categories: participants, activity</td>
<td>Label discrete points on graph/number line; represent a measurement; functional categories: attributes, circumstances</td>
<td>Response; discrete number; infinite; distance; time; functional categories: participants, activity, attributes</td>
</tr>
<tr>
<td>Interpersonal</td>
<td>Directive; factual</td>
<td>Factual; neutral</td>
<td>Indicative; factual; helpful</td>
</tr>
<tr>
<td>Textual</td>
<td>Abbreviation; word; phrase; multiple sentences; bullet points/lists; bolding; left of response space; embedded in schematic visual; embedded in response space; embedded in response tool; multi-word verb constructions; present tense; past tense; simple subject; simple predicate; compound predicate; pronoun; adjective; adverb; direct object; NL register; testing register; glossary: English, Spanish, NL register, testing register</td>
<td>Arabic numeral/value; embedded in NL; left of response space; embedded in schematic visual; embedded in response space; reference not stated</td>
<td>Schematic; symbolic attribute; symbol; response space; compound image; embedded in response space; embedded in response tool; right of text; reference to pic explicit; actions – graph/draw/make/complete schematic visual; without background; from the side; 2D; relevant and irrelevant position of objects; proportionate scale of objects; black and white; coordinate graph; line/segment; x-axis; y-axis; unequal between axes; grid; 1 quadrant</td>
</tr>
</tbody>
</table>
Table 5.8

**Semiotic Components of Item 9**

<table>
<thead>
<tr>
<th>Metafunction</th>
<th>Natural Language</th>
<th>Mathematical Symbols and Expressions</th>
<th>Visual Representations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideational</td>
<td>Give directions; relational processes; nominalization; label concrete visual components; functional categories: participants, activity, circumstances</td>
<td>Represent discrete numbers/values not on graph; represent a measurement; represent a symbol; functional categories: attributes, circumstances</td>
<td>Response; measurement; distance; concrete image; functional categories: participants, activity, circumstances</td>
</tr>
<tr>
<td>Interpersonal</td>
<td>Directive; factual</td>
<td>Factual; neutral; helpful</td>
<td>Indicative; neutral; factual; helpful</td>
</tr>
<tr>
<td>Textual</td>
<td>Symbol; abbreviation; multiple sentences; bolding; above response space; separated by visual; embedded in concrete visual; multi-word noun phrases; multi-word verb constructions; present tense; simple subject; compound subject; simple predicate; adjective; direct object; NL register; testing register; glossary: English, Spanish, NL register, testing register</td>
<td>Math/scientific symbol; operations; Arabic numeral/value; bolding; embedded in NL; between stem/prompt; above response space; embedded in concrete visual; embedded in response tool; reference not stated</td>
<td>Concrete; symbolic attribute; symbol; response space; compound image; embedded in stem/prompt; embedded in response space; embedded in response tool; reference explicit; action – enter; realistic; with background; from the side; 3D; position relevant and irrelevant; scale of objects disproportionate; black and white</td>
</tr>
</tbody>
</table>
### Semiotic Components of Item 24

<table>
<thead>
<tr>
<th>Metafunction</th>
<th>Natural Language</th>
<th>Mathematical Symbols and Expressions</th>
<th>Visual Representations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideational</td>
<td>Give directions; relational processes; nominalization; label schematic visual</td>
<td>Represent discrete numbers/values not on graph; represent measurement;</td>
<td>Response; time; function; functional categories: participants, activity</td>
</tr>
<tr>
<td></td>
<td>components; functional categories: participants, activity, attributes</td>
<td>represent function; functional categories: attributes</td>
<td></td>
</tr>
<tr>
<td>Interpersonal</td>
<td>Directive; factual</td>
<td>Factual; neutral</td>
<td>Neutral; factual</td>
</tr>
<tr>
<td>Textual</td>
<td>Abbreviation; word; multiple sentences; bolding; above response space; separated</td>
<td>Math/scientific symbol; operations; Arabic numeral/value; between stem/prompt;</td>
<td>Schematic; symbol; response space; compound image; embedded in stem/prompt;</td>
</tr>
<tr>
<td></td>
<td>by visual; embedded in schematic visual; multi-word noun phrase; present tense;</td>
<td>above response space; embedded in schematic; embedded in response tool; reference</td>
<td>embedded in response space; embedded in response tool; reference explicit; actions –</td>
</tr>
<tr>
<td></td>
<td>simple subject; simple predicate; adjective; direct object; NL register; testing</td>
<td>not stated</td>
<td>enter; without background; from the side; 2D; position of objects relevant and</td>
</tr>
<tr>
<td></td>
<td>register; glossary: English, Spanish, NL register, testing register, mathematics</td>
<td></td>
<td>irrelevant; scale proportionate; black and white; function</td>
</tr>
<tr>
<td></td>
<td>register</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Table 5.10 shows the item type, knowledge domain, semiotic resource complexity, and item semiotic complexity of each of the four items included in the cognitive interviews. Item 5 had the most complex NL, VR, and overall item semiotic complexity; and Item 1 had the most complex MSE.

Table 5.10

*Cognitive Interview Items: Type, Knowledge Domain, Semiotic Resource Complexities, and Item Semiotic Complexity*

<table>
<thead>
<tr>
<th>Item</th>
<th>Item Type</th>
<th>Knowledge Domain</th>
<th>NL</th>
<th>MSE</th>
<th>VR</th>
<th>Item Semiotic Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Constructed Response</td>
<td>Number System</td>
<td>18</td>
<td>19</td>
<td>28</td>
<td>65</td>
</tr>
<tr>
<td>5</td>
<td>Graph</td>
<td>Functions</td>
<td>36</td>
<td>12</td>
<td>36</td>
<td>84</td>
</tr>
<tr>
<td>9</td>
<td>Constructed Response</td>
<td>Geometry</td>
<td>30</td>
<td>18</td>
<td>31</td>
<td>79</td>
</tr>
<tr>
<td>24</td>
<td>Constructed Response</td>
<td>Functions</td>
<td>29</td>
<td>14</td>
<td>27</td>
<td>70</td>
</tr>
</tbody>
</table>

*Note.* NL is natural language, MSE is mathematical symbols and expressions, and VR is visual representations.

Table 5.11 shows the intersemiotic relations for the four items included in the cognitive interviews. Repetition was the most common intersemiotic relation of the four items. Identity was the most common relation between NL and MSE; Collocation (expectancy) and Repetition were the most common relations between NL and VR; and Repetition was the most relation between MSE and VR. Overall, the items created meaning by either repeating the visual semantic unit verbally and/or mathematically. All the items used NL to indicate the presence of the VR (i.e., Collocation).
Study 2: EB and Non-EB Students’ Making Sense of Semiotic Resources in SBAC Items

Overall performance and connection to item semiotic complexity. Table 5.12 shows descriptive statistics for the students’ scores on each item used during the cognitive interviews and their total score on all four items. The minimum total score was 0, the maximum total score was 5, and the average total score was 1.87 (37.4%). Overall, students performed the best on Item 1 and the worst on Item 24.

Table 5.12

Cognitive Interview Items: Descriptive Statistics for Scores on Each Item and Total Score

<table>
<thead>
<tr>
<th>Item</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0.32</td>
<td>0.47</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>1</td>
<td>0.45</td>
<td>0.50</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>1</td>
<td>0.18</td>
<td>0.39</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>5</td>
<td>1.87</td>
<td>1.70</td>
</tr>
</tbody>
</table>
Figure 5.5 shows the average score on each item and total score disaggregated by language group. EBs had a higher score than non-EBs on every item, except Item 5, and the total score. A series of one-way ANOVA was conducted to compare the effect of language group on average score. Results indicate that the effect of language group on average score for Item 1 was not significant, $F(1, 58) = 0.18, p = 0.67, \eta^2_p = 0.00$;

average score for Item 5 was not significant, $F(1, 58) = 0.08, p = 0.79, \eta^2_p = 0.00$;

average score for Item 9 was not significant, $F(1, 58) = 3.38, p = 0.07, \eta^2_p = 0.06$;

average score for Item 24 was not significant, $F(1, 58) = 0.99, p = 0.33, \eta^2_p = 0.02$; and

the average total score was not significant, $F(1, 58) = 0.83, p = 0.37, \eta^2_p = 0.01$. While the difference between groups was not significant for any variable, the difference in total score might have been significant if students had taken all 29 practice test items.

Figure 5.5. Average score on each item and total score disaggregated by language group.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Figure 5.6 shows the relation between item semiotic complexity and average score, as a percentage, for each item and the items overall disaggregated by language group. Again, EBs tended to score higher than non-EB students. This is surprising given that EB students usually underperform their non-EB peers on standardized assessments. The figure also shows that EBs’ scores started higher and had a larger negative slope than non-EBs’ scores as item semiotic complexity increased. Non-EBs’ scores started lower than EBs’ scores and had a lower negative slope as item semiotic complexity increased. Thus, as item semiotic complexity increased, EBs’ scores decreased while non-EBs’ scores remained constant once a certain item semiotic complexity level was reached. This finding reveals that for these EB students, item semiotic complexity had a greater negative impact on their scores than non-EB students. Generalizations should be made with caution as only four items were included in this analysis. Similar analyses should be conducted with all items to prove whether this claim is true.

*Figure 5.6*. Relation between item semiotic complexity and average score (percent) disaggregated by language group.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Total cognitive load and association between item score and item semiotic complexity. Table 5.13 shows the descriptive statistics for the total cognitive load of each item and the items combined. The average total cognitive load was 10.09. Item 1 had the highest total cognitive load and Item 24 had the lowest.

Table 5.13
Cognitive Interview Items: Descriptive Statistics for Total Cognitive Load

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.84</td>
<td>3.40</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>10.33</td>
<td>3.63</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>10.38</td>
<td>3.84</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>24</td>
<td>8.83</td>
<td>3.77</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>10.09</td>
<td>3.07</td>
<td>2.75</td>
<td>16.75</td>
</tr>
</tbody>
</table>

Figure 5.7 shows the average total cognitive load of each item and the items combined disaggregated by language group. For each item and the items combined, EB students had a higher total cognitite load than non-EBs. A series of one-way ANOVAs was conducted to compare the effect of language group on average total cognitive load for each item. Results indicate that the effect of language group on average total cognitive load for Item 1 was not significant, $F(1,58) = 2.04, p = 0.16, \eta^2_p = 0.03$; the effect of language group on average total cognitive load for Item 5 was not significant, $F(1,58) = 3.09, p = 0.08, \eta^2_p = 0.05$; the effect of language group on average total cognitive load for Item 9 was not significant, $F(1,58) = 3.30, p = 0.08, \eta^2_p = 0.05$; the effect of language group on average total cognitive load for Item 24 was significant, $F(1,58) = 4.90, p = 0.03, \eta^2_p = 0.08$; and the effect of language group on average total
cognitive load for the items combined was significant, $F(1,58) = 4.78, p = 0.03, \eta_p^2 = 0.08$. Even though EB students outperformed their non-EB peers, they had a significantly higher total cognitive load than non-EBs for Item 24 and all items combined. These findings could indicate that the extraneous features made it more difficult for EBs than non-EBs to access the content of the item or were designed in such a way that it was more difficult for them to recall the appropriate schemas.

![Figure 5.7](image)

*Figure 5.7. Average total cognitive load of each item and all items combined disaggregated by language group.*

Table 5.14 shows the Spearman-rank-order correlation between score and total cognitive load of each item and all items combined. All the statistically significant correlations were negatively associated; for these items and their corresponding language groups, as total cognitive load decreased, score increased. These findings indicate that
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

students’ scores are negatively affected as they approach or reach cognitive overload. Furthermore, since EB students approached cognitive overload sooner than non-EBs, their scores were affected more than non-EBs’ scores.

Table 5.14

*Cognitive Interview Items: Spearman’s Rank-Order Correlation Between Score and Total Cognitive Load*

<table>
<thead>
<tr>
<th>Item</th>
<th>Both Language Groups</th>
<th>Non-EBs</th>
<th>EBs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.502*</td>
<td>-0.413*</td>
<td>-0.597*</td>
</tr>
<tr>
<td>5</td>
<td>-0.232</td>
<td>-0.053</td>
<td>-0.401*</td>
</tr>
<tr>
<td>9</td>
<td>-0.171</td>
<td>-0.341</td>
<td>-0.121</td>
</tr>
<tr>
<td>24</td>
<td>-0.121</td>
<td>-0.427*</td>
<td>0.073</td>
</tr>
<tr>
<td>Total</td>
<td>-0.339*</td>
<td>-0.409*</td>
<td>-0.357</td>
</tr>
</tbody>
</table>

*Note.* *Correlation is significant at the 0.05 level.*

Figure 5.8 shows the relation between item semiotic complexity and average total cognitive load disaggregated by language group. The average total cognitive load for EBs was significantly higher than non-EBs for Item 24 and the items combined. It is hard to distinguish a distinct association between the two variables with only four items. However, the overall trend seems to indicate that as item semiotic complexity increased, total cognitive load increased slightly. Additionally, the total cognitive load of EBs increased at a greater rate than that of non-EBs. This findings indicates that item semiotic complexity negatively affects EB students more than non-EBs. These results should be taken with caution and the analyses duplicated with a larger sample of items to validate whether these claims are true.
Semiotic composition. Table 5.15 shows the absolute frequencies and relative proportions of the variables coded 1 related to how students’ reported actions and thinking involved the ideational, interpersonal, and textual metafunctions of the NL, MSE, and VR used in Items 1, 5, 9, and 24 that were also identified in Study 1.3 Students’ reported actions and thinking only corresponded to 11.8% of the NL IDVs, 3.1% of the MSE IDVs, and 9.8% of the VR IDVs listed in Table 5.6 through 5.9. These findings indicate that students, regardless of language group, could potentially benefit from learning about the different metafunctional components of assessment items.

Table 5.15 also shows the results of the Chi-square tests of independence that were conducted to examine whether the proportion of non-EBs whose reported actions and thinking involved the metafunctional components was significantly different from the

---

3 Other data, not included in this dissertation, were collected.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

proportion of EBs. The bold NL used in Item 24 was significant, $X^2 (1, N = 60) = 4.29, p = 0.04$, indicating that a higher proportion (1.00) of non-EBs’ reported actions and thinking involved this component compared to EBs (0.87). Overall, the reported actions and thinking involving metafunctional components did not differ between groups.

Table 5.15

Frequency, Proportion, and Chi-Square Test of Independence Results for the Metafunctional Components of the Semiotic Resources

<table>
<thead>
<tr>
<th>Interview Variables that Corresponded to IDVs</th>
<th>Non-Emergent Bilingual (n = 30)</th>
<th>Emergent Bilingual (n = 30)</th>
<th>Chi-Square Test of Independence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Proportion</td>
<td>Frequency</td>
</tr>
<tr>
<td><strong>Item 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE Ideational</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Represent discrete numbers on graph</em></td>
<td>29</td>
<td>0.967</td>
<td>29</td>
</tr>
<tr>
<td>VR Ideational</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>To show infinity</em></td>
<td>25</td>
<td>0.833</td>
<td>27</td>
</tr>
<tr>
<td><em>To show identity</em></td>
<td>29</td>
<td>0.967</td>
<td>29</td>
</tr>
<tr>
<td><em>To show discrete numbers</em></td>
<td>16</td>
<td>0.533</td>
<td>12</td>
</tr>
<tr>
<td>VR Textual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Schematic visual</em></td>
<td>27</td>
<td>0.900</td>
<td>29</td>
</tr>
<tr>
<td><strong>Item 5</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL Ideational</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>To give directions</em></td>
<td>3</td>
<td>0.100</td>
<td>6</td>
</tr>
<tr>
<td>NL Interpersonal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Directive</em></td>
<td>3</td>
<td>0.100</td>
<td>6</td>
</tr>
<tr>
<td><em>Factual</em></td>
<td>29</td>
<td>0.967</td>
<td>28</td>
</tr>
<tr>
<td>NL Textual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Bold</em></td>
<td>28</td>
<td>0.933</td>
<td>25</td>
</tr>
<tr>
<td><em>Abbreviated</em></td>
<td>21</td>
<td>0.700</td>
<td>19</td>
</tr>
<tr>
<td>VR Interpersonal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Helpful</em></td>
<td>22</td>
<td>0.733</td>
<td>27</td>
</tr>
</tbody>
</table>
Two different sets of two-way ANOVAs were conducted to compare the main effects of language group (EB and non-EB) and item (Items 1, 5, 9, and 24) and the interaction effect between language group and item on students’ reported actions and thinking involving components pertaining to the NL, MSE, and VR semiotic resources and to the ideational, interpersonal, and textual metafunctions. Appendix R shows the descriptive statistics and two-way ANOVA results for all six dependent variables.
To create the dependent variables for the first set, all interview variables belonging to NL were summed together, all interview variables belonging to MSE were summed together, and all interview variables belonging to VR were summed together.

When using NL Components as the dependent variable, the interaction between the effects of language group and item \((p = 0.55)\) was not significant. The main effect for item was significant, \(F(1,232) = 179.44, p = 0.00, \eta^2_p = 0.070\). A post hoc Tukey test showed that there was a statistical difference between all items; students’ reported actions and thinking involved more components pertaining to the NL of Item 5, followed by Item 24, Item 9, and then Item 1.

When using MSE Components as the dependent variable, the interaction between the effects of language group and item \((p = 0.95)\) was not significant. The main effect for item was significant, \(F(1,232) = 492.13, p = 0.00, \eta^2_p = 0.86\). A post hoc Tukey test showed that there was not a statistical difference between Items 1 and 24 or Items 5 and 9; students’ reported actions and thinking involved more components from the MSE of Items 1 and 24 than Items 5 and 9.

When using VR Components as the dependent variable, the interaction between the effects of language group and item \((p = 0.61)\) was not significant. The main effect for item was significant, \(F(1,232) = 582.27, p = 0.00, \eta^2_p = 0.88\). A post hoc Tukey test showed that there was a statistical difference between all items; students’ reported actions and thinking involved more components from the VR of Item 5, followed by Item 1, Item 24, and Item 9.

To create the dependent variables for the second set, all interview variables belonging to the ideational metafunction were summed together, all interview variables belonging to
the interpersonal metafunction were summed together, and all interview variables belonging to the textual metafunction were summed together.

When using *Ideational Components* as the dependent variable, the interaction between the effects of language group and item \((p = 0.55)\) was not significant. The main effect for item was significant, \(F(1,232) = 394.92, p = 0.00, \eta^2_p = 0.84\). A post hoc Tukey test showed that there was a statistical difference between all items; students’ reported actions and thinking involved more ideational components from Item 1, followed by Item 9, Item 24, and then Item 5.

When using *Interpersonal Components* as the dependent variable, the interaction between the effects of language group and item \((p = .20)\) was not significant. The main effect for item was significant, \(F(1,232) = 163.26, p = 0.00, \eta^2_p = 0.68\). A post hoc Tukey test showed that there was a statistical difference between all items; students’ reported actions and thinking involved more interpersonal components from Item 5, followed by Item 9, Item 24, and then Item 1.

When using *Textual Components* as the dependent variable, the interaction between the effects of language group and item \((p = 0.11)\) was not significant. The main effect for item was significant, \(F(1,232) = 877.54, p = 0.00, \eta^2_p = 0.92\). A post hoc Tukey test showed that there was a statistical difference between all items; students’ reported actions and thinking involved more textual components from Item 5, followed by Item 24, Item 1, and then Item 9.

Table 5.16 summarizes the two-way ANOVA results described above. Figure 5.9 shows the estimated marginal means results from the two-way ANOVAs. Overall, students’ reported actions and thinking involving semiotic resource components and
mathematical components differed between items, but not between language groups or the interaction between item and language group. This finding is not surprising given that the items have different mathematical components.
Table 5.16

Summary of Two-Way ANOVA Results for Semiotic Resources

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Item</th>
<th>Item F</th>
<th>Item p</th>
<th>Mean (SD)</th>
<th>Language group</th>
<th>Language group</th>
<th>Language group</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\eta^2_p$</td>
<td></td>
<td></td>
<td>Item 1</td>
<td>Item 5</td>
<td>Item 9</td>
<td>Item 24</td>
<td>$\eta^2_p$</td>
</tr>
<tr>
<td>NL Components</td>
<td>0.70</td>
<td>179.4*</td>
<td>0.00</td>
<td>0.00</td>
<td>2.80</td>
<td>1.80</td>
<td>2.15</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.00)</td>
<td>(0.90)</td>
<td>(0.61)</td>
<td>(0.86)</td>
<td></td>
</tr>
<tr>
<td>MSE Components</td>
<td>0.86</td>
<td>492.1*</td>
<td>0.00</td>
<td>0.97</td>
<td>0.00</td>
<td>0.00</td>
<td>0.88</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.18)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.32)</td>
<td></td>
</tr>
<tr>
<td>VR Components</td>
<td>0.88</td>
<td>582.3*</td>
<td>0.00</td>
<td>3.23</td>
<td>4.20</td>
<td>0.23</td>
<td>0.63</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.65)</td>
<td>(0.68)</td>
<td>(0.43)</td>
<td>(0.69)</td>
<td></td>
</tr>
<tr>
<td>Ideational</td>
<td>0.84</td>
<td>394.9*</td>
<td>0.00</td>
<td>3.27</td>
<td>0.15</td>
<td>0.90</td>
<td>0.63</td>
<td>0.00</td>
</tr>
<tr>
<td>Components</td>
<td></td>
<td></td>
<td></td>
<td>(0.63)</td>
<td>(0.36)</td>
<td>(0.30)</td>
<td>(0.69)</td>
<td></td>
</tr>
<tr>
<td>Interpersonal</td>
<td>0.68</td>
<td>163.3*</td>
<td>0.00</td>
<td>0.00</td>
<td>1.92</td>
<td>1.13</td>
<td>0.57</td>
<td>0.00</td>
</tr>
<tr>
<td>Components</td>
<td></td>
<td></td>
<td></td>
<td>(0.00)</td>
<td>(0.56)</td>
<td>(0.47)</td>
<td>(0.67)</td>
<td></td>
</tr>
<tr>
<td>Textual</td>
<td>0.92</td>
<td>877.5*</td>
<td>0.00</td>
<td>0.93</td>
<td>4.93</td>
<td>0.00</td>
<td>2.47</td>
<td>0.01</td>
</tr>
<tr>
<td>Components</td>
<td></td>
<td></td>
<td></td>
<td>(0.25)</td>
<td>(0.84)</td>
<td>(0.00)</td>
<td>(0.72)</td>
<td></td>
</tr>
</tbody>
</table>

Note. * $p < .05$. 

MATHEMATICS ASSESSMENT: SEMITOCXS AND EMERGENT BILINGUALS
Figure 5.9. Estimated marginal means of components pertaining to the NL, MSE, and VR semiotic resources and the ideational, interpersonal, and textual metafunctions. All main effects for item are significant.
Intersemiosis. Results from Study 1 indicate that there were 16 intersemiotic relations in the four items included in the cognitive interviews. During cognitive interviews, students’ reported actions and thinking only involved half of the intersemiotic relations identified in Study 1. None of the students’ reported actions or thinking involved the three intersemiotic relations that were present between NL and MSE; students’ reported actions and thinking involved five of the 10 intersemiotic relations that were present between NL and VR; and students’ reported actions and thinking involved all three of the intersemiotic relations that were present between MSE and VR. Furthermore, students’ reported actions and thinking did not involve any of the three Identity relations. They did, however, involve one of the four Collocation relations, five of the seven Repetition relations, and both Metonymy relations.

Table 5.17 shows the absolute frequencies, proportions, and Chi-square test of independence results for which students’ reported actions and thinking involved the eight intersemiotic relations described above. A series of Chi-square test of independence was calculated to examine whether the proportion of non-EBs was significantly different from EBs. None of the interactions were significant; thus, the reported actions and thinking of neither language group was more likely than the other to involve the intersemiotic relations listed in Table 5.17. These findings indicate that students could potentially benefit from learning how to identify and use intersemiotic relations to facilitate problem solving, especially Intersemiotic Identity and intersemiotic relations between NL and MSE.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Table 5.17

*Frequency, Proportion, and Chi-Square Test of Independence Results for Intersemiotic Relations*

<table>
<thead>
<tr>
<th>Interview Variables that Corresponded to Study 1 IDVs</th>
<th>Non-Emergent Bilingual (n = 30)</th>
<th>Emergent Bilingual (n = 30)</th>
<th>Chi-Square Test of Independence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Proportion</td>
<td>Frequency</td>
</tr>
<tr>
<td>Item 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE and VR Metonymy</td>
<td>10</td>
<td>0.333</td>
<td>15</td>
</tr>
<tr>
<td>MSE and VR Repetition</td>
<td>0</td>
<td>0.000</td>
<td>2</td>
</tr>
<tr>
<td>Item 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL and VR Collocation</td>
<td>1</td>
<td>0.033</td>
<td>0</td>
</tr>
<tr>
<td>NL and VR Metonymy</td>
<td>3</td>
<td>0.100</td>
<td>2</td>
</tr>
<tr>
<td>NL and VR Repetition</td>
<td>27</td>
<td>0.900</td>
<td>22</td>
</tr>
<tr>
<td>MSE and VR Repetition</td>
<td>26</td>
<td>0.867</td>
<td>23</td>
</tr>
<tr>
<td>Item 9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL and VR Repetition</td>
<td>23</td>
<td>0.767</td>
<td>23</td>
</tr>
<tr>
<td>Item 24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL and VR Repetition</td>
<td>3</td>
<td>0.100</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* NL is natural language. MSE is mathematical symbols and expressions, and VR is visual representations. Chi-square test of independence were not conducted for the intersemiotic relations with fewer than five cases in each cell.

**Responding to items.** A series of two-way ANOVA was conducted to compare the main effects of language group (EB and non-EB) and item (Items 1, 5, 9, and 24) and the interaction effect between language group and item on the dependent variables that were created by adding similar interview variables in Table 4.7 related to interpreting, problem-solving, and responding. Appendix S shows the descriptive statistics and two-way ANOVA results.
**Interpreting.** Interpreting had three categories: *Student Struggled During Item Interpretation* (interview variables 13.1.2, 13.2.1, 13.2.2, 13.3.1), *Glossary Helped* (interview variables 17.2.1, 17.2.2, 17.2.3), and *Student Suggested NL Change* (interview variables 15.3.1, 16.1.1, 16.1.2, and 16.1.3).

When using *Student Struggled During Item Interpretation* as the dependent variable, the Language Group X Item interaction effect was significant, $F(1, 232) = 2.54, p = 0.05, \eta^2_p = 0.03$. This interaction appeared to reflect the fact that EBs struggled more than non-EBs interpreting Items 5 and 24, with Item 5 being the hardest to interpret. Interestingly, Item 5 had the most complex NL semiotic resource complexity. Frequency analyses show that 20.8% of the EBs stated that a question was confusing, and 30% of the EBs asked clarifying questions regarding what a question was asking.

When using *Glossary Helped* as the dependent variable, the interaction between the effects of language group and item ($p = .272$) was not significant. The main effect for language group was significant, $F(1, 232) = 7.08, p = 0.01, \eta^2_p = 0.03$), indicating that the glossary feature was more helpful for EBs than non-EBs during item interpretation. Frequency analyses show that while 7.5% of the EBs stated that the glossaries made sense, only 1.7% used this feature to help them interpret the items.

When using *Student Suggested NL Change* as the dependent variable, the interaction between the effects of language group and item ($p = .084$) was not significant. However, both the language group, $F(1, 232) = 4.82, p = 0.03, \eta^2_p = 0.02$, and item, $F(1, 232) = 5.10, p = 0.00, \eta^2_p = 0.06$, main effects were significant. These results indicate that more EB students suggested that the NL should be changed while they were
interpreting the items. Frequency analyses show that 20.8% of EBs suggested to make
the NL clearer. A post hoc Tukey test showed that there was a significant difference
between Items 1 and 24 and 9 and 24; fewer students suggested to change the NL used in
Items 1 and 9 when compared to Item 24.

Problem-solving. Problem-solving had four categories: Student Did Not Try
(interview variables 14.2.3 and 14.2.6), Student Struggled During Problem-Solving
(interview variables 14.2.4, 14.2.5, and 14.2.7), Student Solved Using Something Other
than the Computer (interview variables 14.3.2, 14.3.3, 14.3.4, 14.3.5, and 14.3.6), and
Student Suggested an MSE Change (interview variables 15.3.2, 15.3.3, and 16.1.4).

When using Student Did Not Try as the dependent variable, the interaction between
the effects of language group and item ($p = 0.94$) was not significant. The main effect
for item was significant, $F(1,232) = 3.12, p = 0.03, \eta^2_p = 0.04$. A post hoc Tukey test
showed that there was a significant difference between Items 1 and 24, with more
students not trying Item 1 than Item 24. Frequency analyses show that 17.5% of the
students said that they guessed when solving Item 1. Interestingly, Item 1 had the most
complex MSE semiotic resource complexity, indicating that the intrinsic cognitive load
of this item might have been higher than the other items.

When using Student Struggled During Problem-Solving as the dependent variable, the
interaction between the effects of language group and item ($p = 0.49$) was not
significant. The main effect for item was significant, $F(1,232) = 6.24, p = 0.00, \eta^2_p =
0.08$. A post hoc Tukey test showed that there was a significant difference between Item
24 and the other three items, with fewer students struggling to solve Item 24. Frequency
analyses show that 40% of the students received a hint from me and 16.7% of the
students asked a math question while they were solving. These findings could indicate that Item 24 had the lowest intrinsic cognitive load.

When using Student Solved Using Something Other than the Computer as the dependent variable, the interaction between the effects of language group and item \((p = 0.34)\) was not significant. The main effect for item was significant, \(F(1, 232) = 32.65, p = 0.00, \eta^2_p = 0.30\). A post hoc Tukey test showed that there was a significant difference between Item 9 and the other three items, with more students using something other than the computer to solve the item. Frequency results show that 36.7\% of the students used MSE on paper to solve Item 9.

When using Student Suggested an MSE Change (i.e., adding a list of equations to the problem/test) as the dependent variable, the interaction between the effects of language group and item \((p = 1.00)\) was not significant. The main effect for item was significant, \(F(1, 232) = 100.18, p = 0.00, \eta^2_p = 0.56\). A post hoc Tukey test showed that there was a significant difference between Item 9 and the other three items, as it was the only item students suggested adding a list of equations to.

**Responding.** Responding had two categories: Response Does Not Represent Student’s Thinking (interview variables 15.1.2) and Student Says Response Tool was Hard to Use (interview variables 15.2.2, 15.2.3, 15.2.4, 15.3.5, 15.2.6).

When using Response Does Not Represent Student’s Thinking as the dependent variable, the interaction between the effects of language group and item \((p = 0.29)\) was not significant. The main effect for item was significant, \(F(1, 232) = 30.85, p = 0.00, \eta^2_p = 0.29\). A post hoc Tukey test showed that there was not a significant difference between Item 1 and Item 24 and Item 5 and Item 9; Items 1 and 24 had a
statistically significant higher average of students whose responses did not match their thinking compared to Items 5 and 9. This finding is vital because students, regardless of language group, answered Items 1 and 24 correctly but did not receive credit because of the format of the response tools.

When using *Student Says Response Tool was Hard to Use* as the dependent variable, the interaction between the effects of language group and item ($p = 0.60$) was not significant. The main effect for item was significant, $F(1, 232) = 60.09, p = 0.00, \eta^2_p = 0.44$. A post hoc Tukey test showed that there was a statistically significant difference between Items 1 and 5, 1 and 9, 5 and 9, and 5 and 24. Item 1’s response tool was easier than that of 5 and 9; and Item 5’s response tool was harder than that of 1, 9, and 24.

Frequency results show that 19.6% of the students were unfamiliar with parts of the response tools, and 17.1% of the students received directions from me on how to use the response tool. This finding indicates that SBAC needs to improve the usability of their response tools.

Table 5.18 summarizes the two-way ANOVA results described above. Figure 5.10 shows the estimated marginal means results from the two-way ANOVAs.
### Table 5.18

**Summary of Two-Way ANOVA Results for Responding to Items**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Item $\eta^2_p$</th>
<th>Item F</th>
<th>Item $p$</th>
<th>Mean (SD)</th>
<th>Language group $\eta^2_p$</th>
<th>Language group F</th>
<th>Language Group $p$</th>
<th>Mean (SD)</th>
<th>Non-EBs</th>
<th>EB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Struggled During Item Interpretation</td>
<td>0.25</td>
<td>25.7*</td>
<td>0.00</td>
<td>Item 1</td>
<td>0.30 (0.50)</td>
<td>0.03 (0.96)</td>
<td>0.01 (0.50)</td>
<td>0.50 (0.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glossary Helped</td>
<td>0.03</td>
<td>2.29</td>
<td>0.08</td>
<td>Item 5</td>
<td>0.13 (0.34)</td>
<td>0.03 (0.26)</td>
<td>0.01 (0.09)</td>
<td>0.11 (0.41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student Suggested NL Change</td>
<td>0.06</td>
<td>5.1*</td>
<td>0.00</td>
<td>Item 9</td>
<td>0.15 (0.40)</td>
<td>0.02 (0.52)</td>
<td>0.03 (0.42)</td>
<td>0.29 (0.49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student Did Not Try</td>
<td>0.04</td>
<td>3.1*</td>
<td>0.03</td>
<td>Item 24</td>
<td>0.38 (0.56)</td>
<td>0.00 (0.48)</td>
<td>0.48 (0.42)</td>
<td>0.22 (0.49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student Solved Using Something Other Than Computer</td>
<td>0.30</td>
<td>32.7*</td>
<td>0.00</td>
<td></td>
<td>1.02 (1.10)</td>
<td>0.01 (0.43)</td>
<td>0.12 (0.43)</td>
<td>0.89 (1.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student Struggled During Problem Solving</td>
<td>0.08</td>
<td>6.2*</td>
<td>0.00</td>
<td></td>
<td>1.07 (1.01)</td>
<td>0.24 (0.62)</td>
<td>0.16 (0.91)</td>
<td>1.05 (1.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student Suggested MSE Change</td>
<td>0.56</td>
<td>100.2*</td>
<td>0.00</td>
<td></td>
<td>0.63 (0.49)</td>
<td>0.25 (0.40)</td>
<td>0.63 (0.49)</td>
<td>0.36 (0.50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response Does Not Represent Student’s Thinking</td>
<td>0.29</td>
<td>30.9</td>
<td>0.00</td>
<td></td>
<td>0.18 (0.49)</td>
<td>0.33 (0.42)</td>
<td>0.36 (0.50)</td>
<td>0.48 (0.50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student Says Response Tool Was Hard to Use</td>
<td>0.44</td>
<td>60.1*</td>
<td>0.00</td>
<td></td>
<td>0.62 (0.89)</td>
<td>0.33 (0.48)</td>
<td>1.00 (0.84)</td>
<td>0.63 (0.78)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. * $p < .05$
Figure 5.10. Estimated marginal means of the dependent variables pertaining to the three-part sequence of responding to an item. All effects for Students Struggled During Item Interpretation are significant; the main effect for language group is significant for Glossary Helped; both main effects are significant for Student Suggested NL Change; the main effect for item was significant for all other dependent variables.
Total cognitive load and responding to items. Table 5.19 shows the Spearman’s rank-order correlation between total cognitive load and the variables pertaining to interpreting, problem-solving, and responding that were identified above. There was a significant positive correlation between total cognitive load and students not trying on items and students struggling during problem-solving for both language groups. Thus, as total cognitive load increased, students did not try more and they struggled more during problem solving. This result indicates that total cognitive load negatively impacts students’ abilities to recall appropriate schemas and/or incorporate new information into the correct schemas during problem solving. Interestingly, there was a significant positive correlation between total cognitive load and struggled during interpretation for non-EBs. Thus, as total cognitive load increased for non-EBs, they struggled more to interpret the item.

Table 5.19
Spearman’s Rank-Order Correlation Between Total Cognitive Load and Variables Pertaining to Interpreting, Problem-Solving, and Responding

<table>
<thead>
<tr>
<th>Variable</th>
<th>Both Language Groups</th>
<th>Non-EBs</th>
<th>EBs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Struggled During Interpretation</td>
<td>0.135*</td>
<td>0.207*</td>
<td>0.041</td>
</tr>
<tr>
<td>Gloss Helped</td>
<td>0.098</td>
<td>-0.156</td>
<td>0.127</td>
</tr>
<tr>
<td>Suggested NL Change</td>
<td>0.064</td>
<td>-0.032</td>
<td>0.085</td>
</tr>
<tr>
<td>Did Not Try</td>
<td>0.330*</td>
<td>0.361*</td>
<td>0.301*</td>
</tr>
<tr>
<td>Struggled During Problem-Solving</td>
<td>0.264*</td>
<td>0.209*</td>
<td>0.280*</td>
</tr>
<tr>
<td>Used Something Other Than the Computer</td>
<td>0.034</td>
<td>0.083</td>
<td>-0.077</td>
</tr>
<tr>
<td>Suggested MSE Change</td>
<td>0.144*</td>
<td>0.156</td>
<td>0.138</td>
</tr>
<tr>
<td>Response Does Not Represent Thinking</td>
<td>0.049</td>
<td>0.072</td>
<td>0.058</td>
</tr>
<tr>
<td>Says Response Tool is Hard</td>
<td>0.058</td>
<td>0.106</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Note. * Correlation is significant at the 0.05 level.
CHAPTER VI

DISCUSSION

Overview

Current policy mandates the inclusion of emergent bilingual (EB) students in state standardized assessments before they are fully proficient in the language of the test, English. The format of previous assessments was paper and pencil. Research was just starting to understand how to make the content of these assessments accessible to EBs, such as using modified assessments and/or glossaries.

New standardized assessments, including those developed by Smarter Balanced Assessment Consortium (SBAC) and the Partnership for Assessment of Readiness for College and Careers, have a different format than previous tests; they are administered using an online platform. Research pertaining to the accessibility of online assessments for EBs needs to examine whether and how the new platform differentially affects EBs before students and schools are penalized.

This dissertation applied theories from the fields of semiotics and cognitive science to examine the nature of items on the 2015 SBAC Grade 8 Math Practice Test and the commonalities and differences between EB and non-EB students in the ways in which they solved selected items from the test. The purpose of this dissertation was to provide information on variables that are relevant to future investigations on the role of semiotics
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

in computer-based mathematics assessments for EB students and how this information can guide teachers’ practices.

Summary of Results

I conducted two main sets of analyses; the first set examined the semiotic composition and complexity of the items on the 2015 SBAC Grade 8 Math Practice Test, and the second examined the commonalities and differences between non-EB and EB students as they interacted with four of the items.

Study 1. Frequency analyses revealed the ideational, interpersonal, and textual metafunctional components of the natural language (NL), mathematical symbols and expressions (MSE), and visual representations (VR) used in the items. Ideationally, NL was used to give students directions; MSE was used to represent discrete numbers/values that were not on a graph; and VR was used to create a response space for students. Interpersonally, NL was directive; MSE was factual and/or neutral; and VR was factual. Textually, NL used the present tense, simple predicates, and Spanish glosses; MSE used Arabic numerals; and VR used black and white visuals without a background.

Descriptive statistics revealed that MSE was the least complex metafunction, which can most likely be contributed to its highly-conventionalized notations and spatial positioning. NL and VR had very similar complexity levels. The average item semiotic complexity was 69.62 components, indicating that students had to correctly interpret, as intended by the item writers, 69 different components to answer an item correctly. Of the three item types, multiple choice items were the least complex and graphing items were the most. Of the six knowledge domains, Statistics and Probability items were the least complex and Ratios and Proportional Relations items were the most.
Meaning was created between the different semiotic resources in the items using six intersemiotic relations: Anyonymy, Collocation, Hyponymy, Identity, Metonymy, and Repetition. The frequency with which each relation was used differed across each pair of semiotic resources. The most common relation between NL and VR was Collocation; the most common relation between NL and MSE was Identity; and the most common relation between MSE and VR was Repetition. The majority of intersemiotic relations were between NL and VR, while the least were between MSE and VR. Additionally, constructed response items had the most intersemiotic relations amongst item types and Expressions and Equations had the most amongst knowledge domains. These findings should be interpreted with caution due to the limited statistical power.

**Study 2.** EB students scored higher than non-EB students on every item, except Item 5, and the items combined. While these differences were not statistically significant, the total score might have been significant if all 29 practice test items had been used. Findings indicate that as item semiotic complexity increased, average score decreased. Moreover, EBs’ scores decreased more rapidly than non-EBs’ scores which remained constant once the item semiotic complexity reached a particular value. Similar analyses should be conducted with a larger sample of items to confirm the association between item semiotic complexity and score.

Even though EBs had higher scores than non-EBs, they thought the items were harder than their peers, as the average total cognitive load for Item 24 and the items combined were significantly higher. When examining the correlation between score and total cognitive load, approximately 47% of the correlations had a statistically significant inverse relation; as total cognitive load decreased, score increased. Results also indicate
that as item semiotic complexity increased, total cognitive load increased slightly. Again, this increase was more rapid for EBs than non-EBs. Analyses need to be duplicated with more items to prove significance before any generalizations should be made.

Results indicate that, regardless of language group, students’ reported actions and thinking involved very few NL, MSE, and VR metafunctional components of the mathematics test items. The only instance in which the proportion differed between the two language groups was in Item 24, as a higher proportion of non-EBs’ reported actions and thinking involved the use of bold NL for emphasis. The main effect for item was significant when examining the mean difference of students’ reported actions and thinking involving components belonging to NL, MSE, and VR and components belonging to the ideational, interpersonal, and textual metafunctions. This result was expected because each item creates its own meaning using different combinations of semiotic resources and metafunctions.

Frequency results indicate that students’ reported actions and thinking only involved half of the intersemiotic relations that were present in the four items. None of the students’ reported actions or thinking involved the intersemiotic relations between NL and MSE, while their reported actions and thinking involved half of the intersemiotic relations between NL and VR and all of the intersemiotic relations between MSE and VR. Additionally, students’ reported actions and thinking did not involve any of the three Identity relations, but they did involve all of the Metonymy relations. Results show that the reported actions and thinking of neither language group was more likely than the other to involve these intersemiotic relations.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Analyses were also conducted to examine the influence of language group and item number on the variables pertaining to the three-part sequence of responding to an item: interpreting, problem-solving, and responding. During interpretation, results show that EBs struggled to interpret the items more than non-EBs; Item 5 was the most difficult to interpret, followed by Item 24, and Items 1 and 9; EBs struggled more than non-EBs to interpret Items 24 and 5. EB students claimed that the glossary feature helped them to interpret the items more than non-EBs. More EBs suggested that the NL needed to be modified during interpretation than did non-EBs; fewer students suggested to change the NL used in Items 1 and 9 when compared to Item 24.

During problem-solving, significantly more students did not try to solve Item 1 in comparison to Item 24. Similarly, significantly fewer students struggled to solve Item 24 than the other three items. Interestingly, more students used something other than the computer (e.g., drawing a schematic visual or using algebra to solve the problem on paper) while they were solving Item 9 in comparison to the other three items. Item 9 was also the only item in which students suggested changing the MSE (i.e., adding a list of equations to the problem/test).

During responding, Items 1 and 24 had a significantly higher average of students whose responses did not match their thinking compared to Items 5 and 9. Item 1’s response tool was easier than that of 5 and 9; and Item 5’s response tool was harder than that of 1, 9, and 24. While the mean differences were not significant between language groups, the findings that students’ responses did not match their thinking, in a negative way, for half of the items, and that the items’ response tools had different difficulty levels are significant and should be addressed by item developers.
When examining the correlation between total cognitive load and the three-part sequence of responding to an item, results indicate that as total cognitive load increased, all students, regardless of language group, did not try more and struggled more to solve the problem. Interestingly, as cognitive load increased, non-EBs struggled to interpret the items more.

Findings as They Relate to the Conceptual Framework

The findings from Study 1 support O’Halloran’s (2005) findings pertaining to the ideational and interpersonal metafunctions of the natural language used in mathematics. More specifically, results indicate that the ideational metafunction of natural language is to construct relational processes and nominal groups, and the interpersonal metafunction is objective. Unlike her study, however, results show that the main ideational purpose of NL is to give students directions, and the interpersonal metafunction is also directive. This study elaborates on O’Halloran’s claim that NL is organized to carry mathematical arguments forward by describing the components of the textual metafunction that allows for this process to occur.

O’Halloran claims that, ideationally, symbols represent relations between mathematical elements and variables; interpersonally, symbols are authoritative and non-negotiable; and textual meaning is created through conventionalized notation and spatial positioning. Results from this study show that the main ideational purpose of the MSE used on the practice test was to represent discrete numbers or values that were not on a graph; and interpersonally, the MSE was factual and/or neutral. This study also found additional textual components of MSE, such as MSE location, that can be added to conventionalized notation and spatial positioning.
O’Halloran also states that the ideational purpose of VR is to provide intuitive depictions of the world. In this study, however, the main ideational purpose of visuals was to create a response space for students. The findings from this dissertation support her claim that, interpersonally, visuals hold a high-truth value as they were found to be factual as well as neutral. The findings from this dissertation add to de Oliveira and Cheng's (2011) findings that images guide readers to the mathematical content of an image by explaining the textual features of the VR, such as location, view, and dimension.

Results from this study show that the least complex semiotic resource used in the mathematics practice test was MSE, which is most likely attributed to its highly-conventionalized notation and spatial positioning, and the most complex was VR. The average item semiotic complexity was 69.62 components. To increase their chances of responding to an item on the practice test correctly, students must have correctly interpreted, as intended by the item writer, an average of 69 different pieces of information from across three different semiotic resources. Moreover, item semiotic complexity varied across item type and knowledge domain, which could have potentially affected students’ ability to interpret and respond to the different item types and knowledge domain with the same level of success.

Intersemiosis studies have only been conducted using images and natural language. In these studies (see Knox, 2008; Royce, 2002), Synonymy and Repetition were the most frequent types of intersemiotic relations between images and text. Both authors suggest that these relations may help emergent bilingual students interpret multimodal texts. Contrary to Knox and Royce, this study found that the most common intersemiotic
relation between NL and VR was Collocation (expectancy relations) followed by Repetition.

This study adds to the field of intersemiosis as it also examined the types of intersemiotic relations between NL and MSE and MSE and VR. The most common relation between NL and MSE was Identity; NL was commonly used as a synonym of an MSE constituent at a higher level of generality. The most common relation between MSE and VR was Repetition, as images often visually repeated the mathematical concept. Furthermore, findings revealed that the majority of intersemiotic relations were between NL and VR.

In this study, students’ total cognitive load on each item was determined by adding together their item rating (intrinsic cognitive load), their total item rating (germane cognitive load), and their identification of extraneous material. This study found that EBs had a higher total cognitive load than non-EBs on Item 24 and the items combined. Additionally, this study found that as total cognitive load decreased, score increased. Thus, had EB students’ total cognitive load been decreased, they would have scored higher on the items, and, most likely, significantly outperformed their non-EB peers.

Results from this study seem to indicate that as item semiotic complexity increased, total cognitive load increased slightly. This rate was more rapid for EBs, indicating that item semiotic complexity had a larger negative impact on total cognitive load and score for EBs. These results need to be replicated with a larger sample of items before any generalizations should be made, but this could have a significant impact on the field.

Cognitive overload is reached when the difference between the total cognitive load and the processing capacity of one of the subsystems of the working memory (i.e., the
phonological loop and the visuospatial sketchpad) approaches zero (Brünken et al., 2003). This study hypothesized that cognitive overload could negatively affect working memory, which, in turn, could negatively affect a student from recalling the appropriate schema or incorporating new information into the correct schema. Naturally, cognitive overload would affect a student’s ability to respond to an item. Results indicate that this hypothesis is true, but analyses need to be replicated with a larger set of items and students before generalizations should be made.

This study viewed the act of responding to an item as a three-part sequence involving interpreting, problem-solving, and responding. Results from this study found that an increase in total cognitive load negatively affected problem-solving for EBs and non-EBs. Findings indicate that as cognitive load increased, students tried to solve the item less and struggled more during problem-solving. Thus, in order to improve the validity of the SBAC Grade 8 Math Practice Test for EBs, SBAC should modify the items to decrease total cognitive load. This can be achieved by eliminating any extraneous features, or, more importantly since total cognitive load was found to negatively impact problem-solving, adding elements that facilitate schema recall, such as a list of equations. Additionally, SBAC needs to improve the usability of their response tools so that students’ responses reflect their thinking and they receive credit for their content knowledge.

Implications

Results indicate that, on average, students must make meaning from approximately 69 different components. However, students’ reported actions and thinking, regardless of language group, involved very few metafunctional components of the items. Students
could potentially benefit from learning how to identify the different metafunctions of the NL, MSE, and VR used in mathematics test items. For example, students could perform better if they could identify the purpose of the NL used in the item and the relationship it is building with them as the reader. Students could also benefit from learning how to identify the different types of intersemiotic relations that are present within a multimodal test item, as students only identified half of the relations in the practice test. Item comprehension could be facilitated if students knew how meaning was being created between NL and MSE, NL and VR, and MSE and VR.

Differential item functioning (DIF) is often used to determine whether test items systematically advantage one group over another (Hambleton & Rodgers, 1995). While DIF measures may indicate to item writers that an item is differentially affecting a certain population, it does not tell item writers why. The frameworks for calculating item semiotic complexity and students’ total cognitive load presented in this dissertation address this methodological gap. Item writers could use these frameworks on a larger scale to determine which items are differentially affecting language groups and why. Instead of looking at DIF measures, item writers could examine the item semiotic complexity and total cognitive load of items and use these measures to guide item development and modification.

Limitations

One of the limitations of this study, one that most likely cannot be avoided in high school settings, is the demographics of the sample of students who participated in cognitive interviews. First, the students were in one of five different mathematics classes (Contemporary Mathematics 1-4, Trigonometry and Math Analysis, and Algebra II), and
thus had varying levels of mathematics proficiency. Second, 60% of the sample came from the classes of two different teachers (Teachers 3 and 8). Third, I was not able to interview the same number of EBs and non-EBs from each classroom due to practical limitations. The last two limitations could add possible bias due to overrepresentation of students from a particular teacher. While these limitations were shaped by practical issues, such as student attendance, consent and assent return rates, the structure of high school mathematics classes, and the general trend to place EBs in only a few classrooms, they must be recognized as they limit the generalizations of my study.

Another limitation of my study is that I did not have access to documents concerning the students’ schooling history, such as whether they had studied mathematics in their first language. Their exposure to mathematics classes in their first language could have affected how they responded to the item and used the accessibility resources, especially the glossary feature. However, the impact of this limitation is most likely miniscule as only two of the EB students were labeled “Not English Proficient.” Additionally, very few, if any, bilingual secondary schools exist in the state in which this study was conducted.

A major limitation of this study is that I did not have a second, independent scorer code a sample of the items or a sample of the students’ responses. The lack of a second scorer inhibits the ability to gather inter-rater reliability information. Thus, it is difficult to consider the external validity of this dissertation, hence the reason why this dissertation in exploratory. One reason for not having a second, independent scorer was because most of the coding was inductive, thus making it extremely difficult to calculate or even expect any sort of inter-rater reliability. Future studies should use the item and interview
coding schemes deductively to collect inter-rater reliability information and allow for external validity.

The last limitations of this study concerns the practice test items. First, SBAC did not have item measures, such as the $p$-value, for this set of practice test items, thus I could not use these measures to validate my findings. Second, I did not have access to SBAC’s bank of items for the Grade 8 Math Test. Thus, while this study could have been improved by increasing the number of items included in the item analysis to examine the significance of disaggregating by item type and knowledge domain, I was only able to analyze the 29 items on the practice test. Third, due to practical limitations such as time constraints, only four items were included in the cognitive interviews. This prohibited me from making any significant findings pertaining to the relation between item semiotic complexity and how they affected students as they responded to the items. Fourth, I identified a wide variety of variables for a relatively small number of items in Study 1, which limits the generalizations that can be made from this study.

**Future Research**

This study used deductive and inductive techniques to create two frameworks: one that identified the semiotic components of mathematics test items, and another that identified how the semiotic resources of the items affected students as they interacted with the items. Future research should use these frameworks to examine the relationship between the variables that were identified in this study and the achievement of EBs on large-scale tests. For example, which aspects of item semiotic complexity, ideational, interpersonal, or textual facilitates item comprehensions? Which aspects hinder item comprehension? Future research should also examine the relation between item semiotic
complexity, total cognitive load, and student performance with a larger set of items and students. Additionally, future research should use more fine-tuned measures of intrinsic, germane, and extraneous cognitive loads to examine how total cognitive load affects EBs as they respond to mathematics test items.

**Conclusion**

Under current policy, bilingual students are mandated to partake in standardized assessments before they are fully proficient in the language of the test, English. Since standardized assessments will not be disappearing any time soon, test designers should create assessments with bilingual students in mind throughout the entire process, instead of after the fact. Examining how EB students make meaning from the semiotic components of test items, how the semiotic components and complexity of test items affect students’ total cognitive load and problem-solving abilities, and how total cognitive load affects student performance sheds light onto how we can make more valid and reliable mathematics assessments for bilingual students.

One picture is worth a thousand words. Yes, but only if you look at the picture and say or think the thousand words.

– William Saroyan
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

References


MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

New York: Continuum International Publishing Group.


http://doi.org/10.1080/09500693.2016.1258129


http://doi.org/10.1016/j.system.2007.11.005


MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS


MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS


MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

http://doi.org/10.1016/j.learninstruc.2009.02.002


Appendix A
Teacher Recruitment Letter

TEACHERS WANTED!
FOR RESEARCH STUDY

STUDY: Mathematics Assessment in the Race to the Top Era: An Exploratory Study on the Role of Semiotic Resources in the Teaching, Learning, and Assessment of Emergent Bilingual Students

PRINCIPAL INVESTIGATOR: Chelsey Shade

PURPOSE: The purpose of this study is to examine the semiotic practices used in classroom settings and large-scale assessments and how they influence emergent bilingual (EB; commonly referred to as ELLs) and non-EB students’ reasoning and demonstration of mathematics.

CRITERIA FOR ELIGIBILITY:

- Currently teaching a high school math class comprised of mostly juniors
- Have at least 3 native Spanish-speaking EBs in your “junior level” mathematics class

PARTICIPATION BENEFITS:

- Increased awareness about the semiotic resources you use in your own teaching
- Increased awareness about how your students’ (EBs and non-EBs) reasoning is influenced by the presence of different semiotic resources
- Increased awareness about the semiotic resources used in the Smarter Balanced Assessment Consortium (SBAC) mathematics practice assessment
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

- Monetary compensation – You will receive a $20 gift card for your time

COMMITMENTS:

- Your selected class(es) will be observed two times between January and March.
- Your classroom observations will be video and/or audio recorded.
- Six students per class (3 EBs and 3 non-EBs) will be interviewed one time during class (if convenient), before school, after school, or during the student’s off period.
- You will provide basic background information about yourself (e.g., sex, years teaching, degrees/certifications) and your students (e.g., sex, grade, age, language background and proficiency level).

LOCATION:

- Your classroom!

CONTACT INFORMATION

Principal Investigator: Chelsey Shade
Email Address: Chelsey.shade@colorado.edu
Phone Number: 303-551-5106
Address: 249 UCB, Boulder, CO 80309
Appendix B

Teacher Consent Form

Title of research study: Mathematics Assessment in the Race to the Top Era: An Exploratory Study on the Role of Semiotic Resources in the Teaching, Learning, and Assessment of Emergent Bilingual Students

Investigator: Chelsey Shade

Why am I being invited to take part in a research study?
We invite you to take part in a research study because you are currently teaching high school level mathematics to high school emergent bilingual students (EBs; commonly referred to as “ELLs”) and non-EBs.

What should I know about a research study?
- Someone will explain this research study to you.
- Whether or not you take part is up to you.
- You can choose not to take part.
- You can agree to take part and later change your mind.
- Your decision will not be held against you.
- You can ask all the questions you want before you decide.

Who can I talk to?
If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team: Chelsey Shade; Chelsey.shade@colorado.edu; 303-551-5106

This research has been approved by an Institutional Review Board (“IRB”). You may talk to them at (303) 735-3702 or irbadmin@colorado.edu if:
- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You have questions about your rights as a research subject.
- You want to get information or provide input about this research.

Why is this research being done?
The purpose of the research is to examine the semiotic practices used in classroom settings and large-scale assessments and how they influence emergent bilingual and non-EB bilingual students’ reasoning and demonstration of mathematics. While SBAC and PARCC have begun to develop accessibility resources that enable EBs to access the content of the items, little attention has been paid to examining how the presence and complexity of multimodal resources, such as illustrations, graphs, and drag-and-drop
features, affect the performance of EB students. Ultimately, this investigation will inform both assessment development and mathematics teaching practices for the benefit of EB students.

**How long will the research last?**
We expect that you will be in this research study from the beginning of January until the end of March (approximately three months).

**How many people will be studied?**
We expect about 10 teachers and approximately 300 students will be in this research study.

**What happens if I say yes, I want to be in this research?**
If you say yes, you can expect the following:

- Your selected class(es) will be observed two times between January and March.
- Your classroom observations will be video and/or audio recorded.
- Six students per class (3 EBs and 3 non-EBs) will be interviewed one time during class.
- You will provide basic background information about yourself (e.g., sex, years teaching, degrees/certifications) and your students (e.g., sex, grade, age, language background and proficiency level).

**What happens if I do not want to be in this research?**
You can leave the research at any time and it will not be held against you.

**What happens if I say yes, but I change my mind later?**
You can leave the research at any time it will not be held against you.

If you decide to leave the research, the data that has been collected prior to your withdrawal will still be used for analysis. If you decide to leave the research, contact the investigator so that the investigator can find another teacher to include in the study.

**Is there any way being in this study could be bad for me?**
There may be discomfort when initially being recorded, but the discomfort usually disappears after a few observations.

**Will being in this study help me any way?**
We cannot promise any benefits to you or others from your taking part in this research. However, possible benefits include:

- Increased awareness about the semiotic resources you use in your own teaching
- Increased awareness about how your students’ (EBs and non-EBs) reasoning is influenced by the presence of different semiotic resources
• Increased awareness about the semiotic resources used in the SBAC mathematics practice assessment and the semiotic resources that might be used in the PARCC EOY mathematics assessments

**What happens to the information collected for the research?**

Efforts will be made to limit the use and disclosure of your personal information, including research study and medical records, to people who have a need to review this information. We cannot promise complete secrecy. Organizations that may inspect and copy your information include the IRB and other representatives of this organization.

**What else do I need to know?**

If you agree to take part in this research study, you will receive a $20 gift card for your time and effort.

If requested, a copy of the study will be provided once completed.

Your signature documents your permission to take part in this research.

________________________________________  _______________________________________
Signature of subject                        Date

________________________________________
Printed name of subject

________________________________________  _______________________________________
Signature of person obtaining consent       Date

________________________________________
Printed name of person obtaining consent
STUDENT ASSENT FORM—English

We want to learn how your teacher uses semiotic resources (pictures, graphs, gestures, computers, etc.) to help support your learning in mathematics. We also want to learn how the semiotic features of large-scale assessments affect your understanding of and response to mathematics items. Up to 10 teachers and 300 students are participating in this study.

We are asking you to participate in our project.

If you accept to participate,

- Your class will be observed two times between January and March.
- Your class will be video and/or audio recorded.
- You may be asked to answer interview questions about 4 items on the Smarter Balanced Grade 8 Math Practice Test.
- You will provide basic background information about yourself (e.g., sex, grade, age, and language background and proficiency level).

There are no risks in participating in this study. Your participation will not affect your grades and only project staff will have access to your information and the recordings. You can stop your participation at any time. At any time, you can ask the project’s staff any questions about the project or your participation.

Your participation is totally voluntary. No one will be mad at you if you say no.

☐ I want to be in the study at this time.
☐ I do not want to be in the study at this time.

Student’s Printed Name:____________________________________________
Student’s Signature:________________________________________________
Date:____________________________

I have explained the research at a level that is understandable by the child and believe that the child understands what is expected during this study.

Signature of Person Obtaining Assent:______________________ Date:__________
Queremos saber cómo usa tu maestro sus recursos semióticos (imágenes, gráficos, gestos, computadoras, etc.) para ayudar a apoyar su aprendizaje en matemáticas. También queremos aprender las características semióticas de evaluaciones a gran escala afectan su comprensión y respuesta a las preguntas de matemáticas. Hasta 10 profesores y 300 alumnos están participando en este estudio.

Queremos pedirte que participes en nuestro proyecto.

**Si aceptas participar,**
- Tu clase se observará dos veces entre enero y marzo.
- Tu clase estará grabado de video y/o audio.
- Se le puede pedir a responder a preguntas de la entrevista cerca de 4 de las preguntas de la prueba de práctica de SBAC.
- Se le dará información básica sobre sí mismo (por ejemplo, sexo, grado, edad, idioma y nivel de competencia).

No hay riesgos por participar en este estudio. Tu participación no afectará tus calificaciones y solamente el personal del proyecto tendrá acceso a tu información y a las videograbaciones. Puedes interrumpir tu participación en cualquier momento. En cualquier momento, puedes hacerle preguntas al personal del proyecto sobre tu participación.

Tu participación es totalmente voluntaria. Nadie se enojará si dices no.

☐ Quiero participar en el estudio.
☐ No quiero participar en el estudio.

**Nombre del estudiante (letra de molde):** __________________________________________

**Firma del estudiante:** __________________________________________________________

**Fecha:** __________________________

_I have explained the research at a level that is understandable by the child and believe that the child understands what is expected during this study._

**Signature of Person Obtaining Assent:** __________________________ Date: __________
Appendix E

Parent/Legal Guardian Consent Form – English

Title of research study: Mathematics Assessment in the Race to the Top Era: An Exploratory Study on the Role of Semiotic Resources in the Teaching, Learning, and Assessment of Emergent Bilingual Students

Investigator: Chelsey Shade

Why is my child being invited to take part in a research study?
We invite your child to take part in a research study because they are currently taking a high school mathematics class with mostly juniors.

What should I know about a research study?
- Someone will explain this research study to your child.
- Whether or not your child takes part is up to you and your child.
- Your child can choose not to take part.
- Your child can agree to take part and later change their mind.
- Your child’s decision will not be held against them.
- Your child can ask all the questions they want before they decide.

Who can I talk to?
If you have questions, concerns, or complaints, or think the research has hurt your child, talk to the research team: Chelsey Shade; Chelsey.shade@colorado.edu; 303-551-5106

This research has been approved by an Institutional Review Board (“IRB”). You may talk to them at (303) 735-3702 or irbadmin@colorado.edu if:
- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You have questions about your child’s rights as a research subject.
- You want to get information or provide input about this research.

Why is this research being done?
This research is being done to see how different resources used in math class and on the SBAC test, such as pictures, graphs, and computer features, affect your child’s interpretation of and ability to solve math problems. This study will help teachers and test developers better understand how to use different resources to teach and test math skills.

How long will the research last?
We expect that your child will be in this research study from the beginning of January until the end of March (approximately 3 months).

How many people will be studied?
We expect about 10 teachers and 300 students will be in this research study.
What happens if my child say yes, I want to be in this research?
If your child says yes, you can expect the following:

- Your child’s class will be observed two times between January and March.
- Your child’s class will be video and/or audio recorded.
- Your child may be asked to answer interview questions about 4 of the items on the SBAC practices test. Students involved in the interviews will receive a $10 gift card.
- Your child will provide basic background information about themselves (e.g., sex, grade, age, and language background and proficiency level).
- One class will be randomly selected to receive a pizza party at the end of the study.

What happens if I do not want my child to be in this research?
Your child can leave the research at any time and it will not be held against them.
If you do not want your child to be in the study, then they will not be recorded (or identified in recordings) or interviewed.

What happens if I say yes, but I change my mind later?
Your child can leave the research at any time it will not be held against them.

Is there any way being in this study could be bad for my child?
There may be discomfort when initially being recorded, but the discomfort usually disappears after a few observations.

Will being in this study help my child in any way?
We cannot promise any benefits to your child or others from taking part in this research. However, possible benefits include:
- Improving teaching and testing practices for the benefit of all students.

What happens to the information collected for the research?
Efforts will be made to limit the use and disclosure of your child’s personal information, including research study and medical records, to people who have a need to review this information. We cannot promise complete secrecy. Organizations that may inspect and copy your child’s information include the IRB and other representatives of this organization.

What else do I need to know?
If your child is selected to participate in the interviews, your child will receive a $10 gift card for their time and effort. Additionally, one class will be randomly selected to receive a pizza party.
Signature Block for Parents

Your signature documents your permission for the named child to take part in this research.

Printed name of child

Signature of parent or individual legally authorized to consent to the child’s general medical care

Date

☐ Parent

☐ Individual legally authorized to consent to the child’s general medical care (See note below)

Printed name of parent or individual legally authorized to consent to the child’s general medical care

Note: Investigators are to ensure that individuals who are not parents can demonstrate their legal authority to consent to the child’s general medical care. Contact legal counsel if any questions arise.

Signature of parent

Date

Printed name of parent

If signature of second parent not obtained, indicate why: (select one)

☐ The IRB determined that the permission of one parent is sufficient.

☐ Second parent is deceased

☐ Second parent is unknown

☐ Second parent is incompetent

☐ Second parent is not reasonably available

☐ Only one parent has legal responsibility for the care and custody of the child

Signature of person obtaining consent and assent

Date

Printed name of person obtaining consent
**Forma de Consentimiento de Padres/Tutores Legales—Español**

**Título del estudio de investigación:** Evaluación de matemáticas en la época de Race to the Top: una investigación exploratoria sobre la función de los recursos semióticos en la enseñanza, el aprendizaje, y la evaluación de los estudiantes bilingües emergentes

**Investigador:** Chelsey Shade

¿Por qué mi niño sea invitado a participar en un estudio de investigación?
Invitamos a su hijo a participar en un estudio de investigación porque está tomando una clase de matemáticas con la mayoría estudiantes del grado 11.

¿Qué debo saber acerca de un estudio de investigación?
Alguien le explicará este estudio de investigación para su hijo. Sea o no que su hijo tome parte depende de usted y su hijo. Su hijo puede optar por no participar. Su hijo puede estar de acuerdo en participar y luego cambiar de opinión. La decisión de su hijo no se llevará a cabo en contra de ellos. Su hijo puede hacer todas las preguntas que quieran antes de decidirse.

¿Con quién puedo hablar?
Si usted tiene preguntas, preocupaciones, quejas, o piensa la investigación que ha hecho daño a su hijo, hable con el equipo de investigación: Chelsey Shade; Chelsey.shade@colorado.edu; 303-551-5106

Esta investigación ha sido aprobada por una Junta de Revisión Institucional ("IRB"). Usted puede hablar con ellos al (303) 735-3702 o irbadmin@colorado.edu si:
- Sus preguntas, inquietudes, o quejas no están respondidas por el equipo de investigación.
- No se puede contactar el equipo de investigación.
- Quieres hablar con alguien además del equipo de investigación.
- Usted tiene preguntas sobre los derechos de su hijo como sujeto de investigación.
- Quiere recibir información o dar información sobre esta investigación.

Por qué se está haciendo esta investigación?
Esta investigación se está haciendo para ver cómo los diferentes recursos utilizados en clase de matemáticas y en la prueba de SBAC, como imágenes, gráficos y características de la computadora, afectan a la interpretación de su hijo y habilidad para resolver problemas matemáticos. Este estudio ayudará a los profesores y los diseñadores de pruebas a entender mejor cómo utilizar diferentes recursos para enseñar y evaluar habilidades matemáticas.

¿Cuánto tiempo durará la investigación?
Esperamos que su hijo estará en este estudio de investigación desde principios de enero hasta finales de marzo (aproximadamente 3 meses).
¿Cuántas personas se estudiarán?
Esperamos alrededor de 10 profesores y 300 estudiantes estarán en este estudio de investigación.

¿Qué pasa si mi hijo dice que sí, yo quiero estar en esta investigación?
Si su hijo dice que sí, se puede esperar lo siguiente:

- La clase de su hijo se observó dos veces entre enero y marzo.
- La clase de su hijo será de video y / o audio grabado.
- Su hijo tendrá que contestar a preguntas de la entrevista sobre 4 de los preguntas de la prueba práctica de SBAC SBAC. Los estudiantes que participan en las entrevistas recibirán una tarjeta de regalo de $ 10.
- Su hijo le dará información básica sobre sí mismos (por ejemplo, sexo, grado, edad, idioma y nivel de competencia).
- Una clase serán seleccionados al azar para recibir una fiesta de pizza al final del estudio.

¿Qué pasa si no quiero que mi hijo participe en esta investigación?
Su hijo puede dejar la investigación en cualquier momento y no se llevará a cabo en contra de ellos. Si usted no desea que su hijo participe en el estudio, entonces ellos no se grabarán (o identificados en grabaciones) o entrevistados.

¿Qué pasa si digo que sí, pero cambio de opinión más adelante?
Su hijo puede dejar la investigación en cualquier momento que no se llevará a cabo en contra de ellos.

¿Hay alguna manera en este estudio podría ser malo para mí hijo?
Puede haber malestar cuando se está grabando en un principio, pero la molestia generalmente desaparece después de unas observaciones.

¿Será en este estudio de ayudar a mi hijo de alguna manera?
No podemos prometer ningún beneficio a su hijo u otras personas de su participación en esta investigación. Sin embargo, los posibles beneficios incluyen:

- Mejorar las prácticas de enseñanza y prueba para el beneficio de todos los estudiantes.

¿Qué sucede con la información recogida por la investigación?
Se harán esfuerzos para limitar el uso y divulgación de la información personal de su hijo, incluyendo estudio de investigación y expedientes médicos, a las personas que tienen una necesidad de revisar esta información. No podemos prometer completo secreto. Organizaciones que pueden inspeccionar y copiar la información de su hijo incluyen la IRB y otros representantes de esta organización.

¿Qué más necesito saber?
Si su niño es seleccionado para participar en las entrevistas, su hijo recibirá una tarjeta de regalo de $ 10 por su tiempo y esfuerzo. Además, una clase serán seleccionados al azar para recibir una fiesta de pizza.
Firma del bloque para los Padres

Firma de la persona que obtiene el consentimiento y asentimiento

Fecha

Imprima el nombre de la persona que obtiene el consentimiento

Su firma documenta su permiso para que el niño llamado a participar en esta investigación.

Imprima el nombre de niño

Firma del padre o persona autorizada legalmente para dar su consentimiento a la atención médica del niño

Fecha

☐ Padre
☐ La persona autorizada legalmente para dar su consentimiento a la atención médica del niño (ver nota abajo)

Imprima el nombre del padre o persona autorizada legalmente para dar su consentimiento a la atención médica del niño

Nota: Los investigadores deben asegurarse de que las personas que no están los padres pueden demostrar su autoridad legal para consentir a la atención médica general del niño. Póngase en contacto con un asesor legal si surge alguna pregunta.

Firma de padre

Fecha

Imprima el nombre del padre

Si la firma del segundo padre no obtuvo, indique por qué: (selecione uno)

☐ El IRB determina que el permiso de uno de los padres es suficiente.
☐ El otro padre está muerto
☐ El otro padre es desconocido
☐ El otro padre es incompetente
☐ El otro padre no está razonablemente disponible
☐ Sólo uno de los padres tiene la responsabilidad legal por el cuidado y la custodia del niño
Appendix G

Items from The 2015 SBAC Grade 8 Math Practice Test

1. Drag each number to its correct position on the number line.

2. For each number, indicate whether it is rational or irrational.

<table>
<thead>
<tr>
<th></th>
<th>Rational</th>
<th>Irrational</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\frac{4}{7})</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>(\sqrt{30})</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>(\frac{21}{\sqrt{4}})</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>(x)</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>(-27)</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>
3
A square with side length $s$ has an area of 324 square centimeters. This equation shows the area of the square.
$s^2 = 324$
What is the side length of the square in centimeters?

4
Approximately $7.5 \times 10^5$ gallons of water flow over a waterfall each second. There are $8.6 \times 10^4$ seconds in 1 day. Select the approximate number of gallons of water that flow over the waterfall in 1 day.

A. $6.45 \times 10^{21}$
B. $6.45 \times 10^{20}$
C. $6.45 \times 10^{10}$
D. $6.45 \times 10^9$

5
The school is 100 meters from Jason’s house. The following describes his most recent trip:

- He walked 50 meters toward school in 2 minutes. He realized that he left a book at home.
- He turned around and walked home at the same speed.
- He spent 1 minute looking for his book.
- He walked all the way to school at twice his original speed.

Use the Connect Line tool to finish a graph that accurately represents Jason’s trip.
6. Use the Connect Line tool to draw the image of the figure after the following transformations:
   - a reflection over the x-axis
   - a horizontal translation 7 units to the left

7. Consider this graph of a line.

   Enter an equation for the line.
Coffee costs $2.00 per pound at a coffee shop.

Use the Add Arrow tool to draw a ray that shows the proportional relationship between the number of pounds of coffee purchased and the total cost.

A 13-foot ladder is leaning on a tree. The bottom of the ladder is on the ground at a distance of 5 feet from the base of the tree. The base of the tree and the ground form a right angle as shown.

Enter the distance, in feet, between the ground and the top of the ladder.
The points show different locations in Joe's town. Each unit represents 1 mile.

**Places in Joe’s Town**

Enter the shortest distance, in miles, between Joe's home and the park. Round your answer to the nearest tenth.
11
A cone with radius 4 feet is shown. Its approximate volume is 165 cubic feet.

Enter the height of the cone, in feet. Round your answer to the nearest hundredth.

12
Use the Add Arrow tool to graph a system of two equations that has a single solution of \((-2, -3)\).
13
Joe solved this linear system correctly.
\[6x + 3y = 6\]
\[y = -2x + 2\]
These are the last two steps of his work.
\[6x - 6x + 6 = 6\]
\[6 = 6\]
Which statement about this linear system must be true?
- [ ] \(x\) must equal 6
- [ ] \(y\) must equal 6
- [ ] There is no solution to this system.
- [ ] There are infinitely many solutions to this system.

14
Drag a number into each box to create an equation that has no solution.

\[8x - 3x + 2 - x = \square x + \square\]

15
Segment \(FG\) begins at point \(F(-2, 4)\) and ends at point \(G(-2, -3)\). The segment is translated by \(<x - 3, y + 2>\) and then reflected across the \(y\)-axis to form segment \(F'G'\).

How many units long is segment \(F'G'\)?
- [ ] 0
- [ ] 2
- [ ] 3
- [ ] 7
A sequence of transformations is applied to a polygon. Select all statements which indicate a sequence of transformations where the resulting polygon has an area greater than the original polygon.

- Reflect over the x-axis, dilate about the origin by a scale factor of $\frac{1}{2}$, translate up 5 units.
- Rotate 90° counterclockwise around the origin, dilate about the origin by a scale factor of $\frac{3}{2}$.
- Dilate about the origin by a scale factor of $\frac{2}{3}$, rotate 180° clockwise around the origin, translate down 2 units.
- Dilate about the origin by a scale factor of 2, reflect over the y-axis, dilate about the origin by a scale factor of $\frac{2}{3}$.

Consider this graph of a line.

Which equation has a rate of change **greater than** the rate of change for the line shown?

- $y = 3x - 1$
- $y = \frac{x}{2} + 4$
- $y = 2x + 2$
- $y = \frac{x}{3} - 3$
Consider this equation.

\[ c = ax - bx \]

Joseph claims that if \( a, b, \) and \( c \) are non-negative integers, then the equation has exactly one solution for \( x \).

Select all cases that show Joseph's claim is incorrect.

- \( a - b = 1, c = 0 \)
- \( a = b, c \neq 0 \)
- \( a = b, c = 0 \)
- \( a - b = 1, c \neq 1 \)
- \( a \neq b, c = 0 \)

John and Kim wrote down two different functions that have the same rate of change.

John's function is represented by the table shown.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>-5</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Use the Add Arrow tool to graph a function that could be Kim's function.
Two figures are shown on the coordinate grid.

Show that Figure A and Figure B are congruent by describing a sequence of basic transformations that maps Figure A onto Figure B. In your response, be sure to identify the transformations in the order they are performed.
21

The base of triangle $ABC$ and the base of triangle $DEF$ lie on line $m$, as shown in the diagram.

The measure of $\angle 4$ is less than the measure of $\angle B$.

For each comparison, select the symbol ($<$, $>$, $=$) that makes the relationship between the first quantity and the second quantity true.

<table>
<thead>
<tr>
<th>First Quantity</th>
<th>Comparison</th>
<th>Second Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m \leq 3$</td>
<td>$&lt;$</td>
<td>$m \leq 7$</td>
</tr>
<tr>
<td>$m \angle 1 + m \angle 2$</td>
<td>$&gt;$</td>
<td>$m \angle 5 + m \angle 6$</td>
</tr>
</tbody>
</table>

22

Line $a$ is shown on the graph. Use the Add Arrow tool to construct line $b$ on the graph so that:

- Line $a$ and line $b$ represent a system of linear equations with a solution of $(7, -2)$.
- The slope of line $b$ is greater than $-1$ and less than $0$.
- The $y$-intercept of line $b$ is positive.
The table shows the relationship between the average number of hours students study for a mathematics test and their average grade.

<table>
<thead>
<tr>
<th>Hours Studying</th>
<th>Average Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>62</td>
</tr>
<tr>
<td>1</td>
<td>78</td>
</tr>
<tr>
<td>2</td>
<td>85</td>
</tr>
<tr>
<td>5</td>
<td>74</td>
</tr>
</tbody>
</table>

Which type of function is most likely to model these data?

- linear function with positive slope
- linear function with negative slope
- non-linear function that decreases then increases
- non-linear function that increases then decreases

This table shows the linear relationship of the water level in a tank and time.

<table>
<thead>
<tr>
<th>Time (hr)</th>
<th>Water Level (ft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
</tr>
</tbody>
</table>

Enter the rate of change of the water level, in feet per hour.
25

An empty corn silo in the shape of a cylinder is being filled with corn.

<table>
<thead>
<tr>
<th>Number of Hours</th>
<th>Amount of Corn (cu ft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2475</td>
</tr>
<tr>
<td>5</td>
<td>4125</td>
</tr>
<tr>
<td>8</td>
<td>6600</td>
</tr>
</tbody>
</table>

Enter the **percent** of the silo that is filled with corn at 10 hours.

26

Justin's car can travel $77 \frac{1}{2}$ miles with $3 \frac{1}{10}$ gallons of gas.

Kim's car can travel $99 \frac{1}{5}$ miles with $3 \frac{1}{5}$ gallons of gas.

Drag the cars to the number line to show the number of miles each car can travel with 1 gallon of gas.
Kyle was given the following problem to solve.

A company sells baseball gloves and bats. The gloves regularly cost $30 and the bats regularly cost $90. The gloves are on sale for $4 off, and the bats are on sale for 10% off. The goal is to sell $1200 worth of bats and gloves each week. Last week, the store sold 14 gloves and 9 bats.

Did the store meet its goal?

The steps Kyle used to solve the problem are shown. Select the first step that shows an error.

☐ **Step 1:**

\[
\begin{align*}
$30 & - $4 \\
$26 &
\end{align*}
\]

☐ **Step 2:**

\[
\begin{align*}
$26 & \times 14 \\
$364 &
\end{align*}
\]

☐ **Step 3:**

\[
\begin{align*}
$90 & \div 0.9 \\
$100 &
\end{align*}
\]

☐ **Step 4:**

\[
\begin{align*}
$100 & \times 9 \\
$900 &
\end{align*}
\]

☐ **Step 5:** Yes, the store met its goal.

\[
\begin{align*}
$900 & + $364 \\
$1264 &
\end{align*}
\]

---

All 8th-grade students at a school answered Yes or No to the two survey questions shown.

- Do you have a cell phone?  
  - Yes  
  - No
- Do you have an MP3 player?  
  - Yes  
  - No

The same students responded to both questions. Complete the two-way frequency table to show the correct totals for the given data. You must complete all five cells of the table for a full credit response.

<table>
<thead>
<tr>
<th></th>
<th>MP3 Player</th>
<th>No MP3 Player</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Phone</td>
<td>57</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>No Cell Phone</td>
<td>30</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>87</strong></td>
<td><strong>187</strong></td>
<td><strong>274</strong></td>
</tr>
</tbody>
</table>
This graph shows a proportional relationship between the amount of money in Jack's savings account and the number of weeks Jack has been saving money.

Select the statement that correctly reflects what is shown in the graph.

a. The slope of the line is \( \frac{6}{1} \), so Jack's savings rate is $6 every week.

b. The slope of the line is \( \frac{6}{1} \), so Jack's savings rate is $1 every 6 weeks.

c. The slope of the line is \( \frac{1}{6} \), so Jack's savings rate is $6 every week.

d. The slope of the line is \( \frac{1}{6} \), so Jack's savings rate is $1 every 6 weeks.

---

## Coding Framework for Item Analysis

### Natural language (NL)

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Printed language.</td>
<td></td>
</tr>
</tbody>
</table>

### Interpersonal - NL

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directive</td>
<td>The NL tells the reader what to do.</td>
</tr>
<tr>
<td>Factual</td>
<td>The NL is concerned with what is actually the case rather than interpretations of or reactions to it.</td>
</tr>
<tr>
<td>Inquisitive</td>
<td>The NL asks the reader questions they are supposed to answer.</td>
</tr>
<tr>
<td>Rational</td>
<td>The NL is based on or in accordance with reason or logic.</td>
</tr>
<tr>
<td>False</td>
<td>The NL is not true.</td>
</tr>
</tbody>
</table>

### Ideational - NL

The purpose of this metafunction is to interpret our external and internal experiences with the world and to describe the purpose and/or context for the use of NL.

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text purpose</td>
<td>Features that relate to the intended purpose of the NL.</td>
</tr>
<tr>
<td>Directions</td>
<td>NL is used to give an order.</td>
</tr>
<tr>
<td>Relational Processes</td>
<td>NL is used to characterize or identify.</td>
</tr>
<tr>
<td>Question</td>
<td>NL is used to word a sentence in a way to elicit information.</td>
</tr>
<tr>
<td>Non-mathematical context/background</td>
<td>NL is used to describe the circumstances that form the setting in which the problem is placed.</td>
</tr>
<tr>
<td>Nominalization</td>
<td>NL is used to create a noun group from process verbs and quality adjectives.</td>
</tr>
<tr>
<td>Complex Chain of Reasoning</td>
<td>NL is used to show a system of logic.</td>
</tr>
<tr>
<td>Response options</td>
<td>NL is used as one of the response options (e.g., multiple choice options).</td>
</tr>
<tr>
<td>Label schematic visual components</td>
<td>NL is used to label the components of schematic visuals (e.g., label axes).</td>
</tr>
<tr>
<td>Label concrete visual components</td>
<td>NL is used to label the components of concrete visuals (e.g., name the object in the visual).</td>
</tr>
<tr>
<td>Label response options</td>
<td>NL is used to label response options (i.e., label response bubbles in multiple choice items – A, B, C, D).</td>
</tr>
<tr>
<td>Label response tools</td>
<td>NL is used to label parts of the tools students use to respond to the item (e.g., the different components of the graphing tool).</td>
</tr>
</tbody>
</table>

### Text functional categories

The different categories belonging to the ideational metafunction of NL.

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>The represented constituents.</td>
</tr>
<tr>
<td>Activity</td>
<td>The action that is occurring between the actors, recipients, and/or objects of the action.</td>
</tr>
<tr>
<td>Attributes</td>
<td>The participants’ characteristics.</td>
</tr>
<tr>
<td>Circumstances</td>
<td>The elements concerned with the setting.</td>
</tr>
</tbody>
</table>

### Textual - NL

The purpose of this metafunction is to describe the structural
<table>
<thead>
<tr>
<th>Text unit</th>
<th>The type of text.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbol</td>
<td>A character, or sign used instead of a word to represent a quantity, position, relationship, direction, or something to be done (outside of the regular use of punctuation).</td>
</tr>
<tr>
<td>Letter</td>
<td>(Self-explaining)</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>A shortened form of a word or phrase.</td>
</tr>
<tr>
<td>Word</td>
<td>(Self-explaining)</td>
</tr>
<tr>
<td>Phrase</td>
<td>A word or a group of 2 or more words without a verb.</td>
</tr>
<tr>
<td>Sentence</td>
<td>A set of words that is complete in itself, typically containing a subject and predicate, conveying a statement, question, exclamation, or command, and consisting of a main clause and sometimes one or more subordinate clauses.</td>
</tr>
<tr>
<td>Multiple sentences</td>
<td>More than one sentence.</td>
</tr>
<tr>
<td>Bullet points/lists</td>
<td>A number of connected items written consecutively, typically one below the other.</td>
</tr>
<tr>
<td>Text emphasis</td>
<td>Special font style of text in an illustration to accentuate the importance of something.</td>
</tr>
<tr>
<td>Bolding</td>
<td>(Self-explaining)</td>
</tr>
<tr>
<td>Text location</td>
<td>The place/position of the text in relation to the response space, visual, and/or MSE.</td>
</tr>
<tr>
<td>Left of response space</td>
<td>The text is located to the left of the response space.</td>
</tr>
<tr>
<td>Above response space</td>
<td>The text is located above the response space.</td>
</tr>
<tr>
<td>Separated by visual</td>
<td>The text is separated by the visual so that there is text above and below the visual.</td>
</tr>
<tr>
<td>Separated by MSE</td>
<td>The text is separated by the MSE so that there is text above and below the MSE.</td>
</tr>
<tr>
<td>Embedded in schematic visual</td>
<td>The text is embedded in the schematic visual.</td>
</tr>
<tr>
<td>Embedded in concrete visual</td>
<td>The text is embedded in the concrete visual.</td>
</tr>
<tr>
<td>Embedded in response space</td>
<td>The text is embedded in the response space (e.g., the letters that are provided to students to form their answers).</td>
</tr>
<tr>
<td>Embedded in response tool</td>
<td>The text is embedded in the response tool (i.e., the text that labels the different buttons of the tools, such as the Add Point button).</td>
</tr>
<tr>
<td>Text complexity</td>
<td>The features that are related to the difficulty of the text.</td>
</tr>
<tr>
<td>Has multi-word noun phrases</td>
<td>Text consists of nouns comprised of multiple words (e.g., the sum of 3 + 2).</td>
</tr>
<tr>
<td>Has multi-word verb constructions</td>
<td>Text consists of a basic verb plus another word(s), like a preposition or adverb (e.g., look up, look after, look forward to).</td>
</tr>
<tr>
<td>Present tense</td>
<td>The text is in the present tense.</td>
</tr>
<tr>
<td>Past tense</td>
<td>The text is in the past tense.</td>
</tr>
<tr>
<td>Text constituents</td>
<td>Components of the text.</td>
</tr>
<tr>
<td>Simple subject</td>
<td>What or whom the sentence is about (e.g., Judy).</td>
</tr>
<tr>
<td>Compound subject</td>
<td>Simple subject with more than 1 noun or pronoun (e.g., Judy and her dog).</td>
</tr>
<tr>
<td>Simple predicate</td>
<td>Tells something about the subject; verb(s) that link up with the subject (e.g., run on the beach).</td>
</tr>
</tbody>
</table>
### Compound predicate
Predicate that includes more than 1 verb pertaining to the same object (e.g., They walked and admired the artwork).

### Pronoun
A word that can function by itself as a noun phrase and that refers either to the participants or to someone/thing mentioned elsewhere (e.g., I, you, she, it, this).

### Adjective
A word or phrase naming an attribute that is added to or grammatically related to a noun to modify/describe it (e.g., She is pretty).

### Adverb
A word or phrase that modifies or qualifies an adjective, verb, or other adverbs, expressing a relation of place, time, circumstance, manner, cause, degree, etc. (e.g., gently, quite, then, there).

### Direct object
A noun phrase denoting a person/thing that is the recipient of the action of a transitive verb (a verb that needs a direct object to complete its meaning) (e.g., Jimmy fed THE DOG).

### Indirect object
A noun phrase referring to someone/thing that is affected by the action of a transitive verb, but is not the primary object (e.g., give HIM the book).

### Text register
A variety of language used for a particular purpose or in a particular social setting.

- **Natural language**
  The register of social language.
- **Testing**
  The register of testing (e.g., how to respond).

### Glossary components
Features related to the embedded glossaries.

- **English**
  Glossary is in English.
- **Spanish**
  Glossary is in Spanish.
- **Natural language register**
  The register of social language.
- **Math register**
  The register of math.
- **Testing register**
  The register of testing (e.g., how to respond).

### Math Symbols/Expressions (MSE)
Letters, characters, or signs used to represent a quantity, position, relationship, direction, or math process that are the focal point of the item or a specific part of the item.

### Interpersonal - MSE
The purpose of this metafunction is to describe how MSE is used to socially position the people involved (e.g., reader and writer).

- **Factual**
  The MSE is concerned with what is actually the case rather than interpretations of or reactions to it.
- **Neutral**
  The MSE does not help, hurt, or really interact with the student (e.g., numbers used to label the axes of a coordinate graph).
- **False**
  The MSE is not true.
- **Non-negotiable**
  The MSE cannot be changed (e.g., the student cannot change the properties of the transformation they need to do in a problem).
- **Helpful**
  The MSE helps students interpret or solve the item.

### Ideational - MSE
The purpose of this metafunction is to interpret our external and internal experiences with the world and to describe the purpose and/or context for the use of MSE.

- **MSE purpose**
  Features that relate to the intended purpose of the MSE.
- **Represent discrete numbers/values (not on**
  To represent data/numbers that can be counted and are not on a graph.
<table>
<thead>
<tr>
<th>Activity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph</td>
<td>To represent data/numbers that can be counted and are on a graph.</td>
</tr>
<tr>
<td>Represent a measurement</td>
<td>To represent a unit of measurement.</td>
</tr>
<tr>
<td>Represent an equation</td>
<td>To show that two mathematical expressions are equal.</td>
</tr>
<tr>
<td>Represent a transformation</td>
<td>To show that a point, line, or shape has been manipulated.</td>
</tr>
<tr>
<td>Represent a coordinate point</td>
<td>To show a position on a coordinate math.</td>
</tr>
<tr>
<td>Represent a function</td>
<td>To show a relationship involving one or more variables.</td>
</tr>
<tr>
<td>Represent linear system</td>
<td>To represent a set of linear equations with the same variables.</td>
</tr>
<tr>
<td>Represent a symbol</td>
<td>To represent a mark/character that has mathematical meaning.</td>
</tr>
<tr>
<td>Represent an operation</td>
<td>To represent a mathematical process.</td>
</tr>
<tr>
<td>Represent a variable</td>
<td>To represent an unknown number.</td>
</tr>
</tbody>
</table>

**MSE functional categories**
- The different categories belonging to the ideational metafunction of MSE.
- Participants: The represented constituents.
- Activity: The action that is occurring between the actors, recipients, and/or objects of the action.
- Attributes: The participants’ characteristics.
- Circumstances: The elements concerned with the setting.

**Textual - MSE**
- The purpose of this metafunction is to describe the structural rituals of the MSE.

**MSE unit**
- The type of MSE used.
  - Math/scientific symbol: A symbol that has mathematical meaning (e.g., pi).
  - Operations: A symbol indicating a mathematical process (e.g., +).
  - Arabic numeral/value: Any of the numerals 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 (i.e., constant)
  - Variable: A letter or symbol representing a quantity that can change.
  - Scientific notation: A method of writing or displaying numbers in terms of a decimal between 1 and 10, multiplied by a power of 10.

**MSE emphasis**
- Special font style of text in an illustration to accentuate the importance of something.

**MSE location**
- The place/position of the MSE in relation to the response space, visual, and/or NL.
  - Embedded in NL: The MSE is embedded in the text.
  - Between stem/prompt: The MSE is located between the stem/prompt.
  - Left of response space: The MSE is left of the response space.
  - Above response space: The MSE is above the response space.
  - Embedded in schematic visual: The MSE is embedded in the schematic visual (e.g., numbering intervals on an axis).
  - Embedded in concrete visual: The MSE is embedded in a concrete visual (e.g., measuring a component of the visual).
  - Embedded in response space: The MSE is embedded in the response space (e.g., the numbers that are provided to students to form their answers).
  - Embedded in response tool: The MSE is embedded in the response tool.

**Reference to MSE**
- Mention of the existence of the MSE in the text of the test item.
  - Explicit: The text of the item mentions the MSE explicitly.
  - Not stated: The text of the item does not mention the MSE.
### MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

<table>
<thead>
<tr>
<th>Stated actions to perform w/ MSE</th>
<th>Explicit guidance for examinees about actions relating to the use of the MSE in the test item.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drag</td>
<td>The text of the item asks examinees to drag parts of the MSE to another location.</td>
</tr>
<tr>
<td>Indicate/select</td>
<td>The text of the item asks examinees to point out, show, or select different parts or functions of the MSE.</td>
</tr>
</tbody>
</table>

### Visual Representation Constituents (VRC)

- **Non-schematic visuals** include concrete images that depict objects described in the math problem (e.g., a picture of a store); Schematic visuals are more abstract and represent several elements and their relationships (e.g., linear graphs) (Martiniello, 2009).

### Interpersonal - VR

- **Indicative**
  The VR serves as a sign or indication of something (e.g., how the students are supposed to answer).

- **Neutral**
  The VR does not help, hurt, or really interact with the student (e.g., an empty coordinate graph).

- **Factual**
  The VR is concerned with what is actually the case rather than interpretations of or reactions to it.

- **Helpful**
  The VR helps students interpret or solve the item.

### Ideational - VR

- The purpose of this metafunction is to interpret our external and internal experiences with the world and to describe the purpose and/or context for the use of the VR.

### VR purpose

- **Response**
  To provide an area to respond.

- **Identity**
  To visually represent an identity or relational process.

- **Measurement**
  To visually represent a measurement (known or unknown).

- **Discrete number**
  To visually represent a discrete number.

- **Category**
  To visually represent categories.

- **Infinite**
  To visually represent infinity.

- **Distance**
  To visually represent distance.

- **Time**
  To visually represent time.

- **Function**
  To visually represent a function.

- **Concrete image**
  To visually represent a real-life object.

- **Shape**
  To visually represent a shape.

- **Map**
  To visually represent a map and some/all properties belonging to maps.

### VR functional categories

- The different categories belonging to the ideational metafunction of the VR.

### Participants

- The represented constituents.

### Activity

- The action that is occurring between the actors, recipients, and/or objects of the action.

### Attributes

- The participants’ characteristics.

### Circumstances

- The elements concerned with the setting.

### Textual - VR

- The purpose of this metafunction is to describe the structural rituals of the VR.

### Illustration type

- Features that are related to the category in which the illustration
### MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

<table>
<thead>
<tr>
<th>Concept</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concrete</td>
<td>Images that depict objects described in the item (e.g., a picture of a store).</td>
</tr>
<tr>
<td>Schematic</td>
<td>Abstract images that represent several elements and their relationships (e.g., linear graphs).</td>
</tr>
<tr>
<td>Symbolic attribute</td>
<td>An arrow/line that points to something to identify an attribute (e.g., a line between 4 ft. and the radius of a cone indicating that the radius is 4 ft.).</td>
</tr>
<tr>
<td>Symbol</td>
<td>A mark/character used as a conventional representation of an object, function, or process (e.g., an x used to represent delete).</td>
</tr>
<tr>
<td>Response space</td>
<td>A visual frame used to highlight/point out the response space.</td>
</tr>
<tr>
<td>Composition</td>
<td>How many visuals comprise the image.</td>
</tr>
<tr>
<td>Single image</td>
<td>Visual comprises one image.</td>
</tr>
<tr>
<td>Compound image</td>
<td>Visual comprises more than one image.</td>
</tr>
<tr>
<td>VR location</td>
<td>Position of the visual relevant to the question stem/prompt, text, and/or response space.</td>
</tr>
<tr>
<td>Below text</td>
<td>The VR is below the text.</td>
</tr>
<tr>
<td>Embedded in stem/prompt</td>
<td>The VR is embedded in the text.</td>
</tr>
<tr>
<td>Embedded in response space</td>
<td>The VR is embedded in the response space (i.e., visual frame of the response space).</td>
</tr>
<tr>
<td>Embedded in response tool</td>
<td>The VR is embedded in the response tool (e.g., a graph the students must complete).</td>
</tr>
<tr>
<td>On the right of the text</td>
<td>The VR is to the right of the text.</td>
</tr>
<tr>
<td>Reference to illustration</td>
<td>Mention of the existence of an illustration in the text of the test item.</td>
</tr>
<tr>
<td>Explicit</td>
<td>The text of the item mentions the illustration explicitly.</td>
</tr>
<tr>
<td>Not stated</td>
<td>The text of the item does not mention the illustration.</td>
</tr>
<tr>
<td>Stated actions to perform with the illustration</td>
<td>Explicit guidance for examinees about actions relating to the use of an illustration in the test item.</td>
</tr>
<tr>
<td>Drag</td>
<td>The text of the item asks examinees to drag the illustration or parts of it to another location.</td>
</tr>
<tr>
<td>Enter</td>
<td>The text of the item asks examinees to enter their answer in the visual frame of the response space.</td>
</tr>
<tr>
<td>Observe or examine</td>
<td>The text of the item asks examinees to look at or examine the illustration.</td>
</tr>
<tr>
<td>Select/indicate</td>
<td>The text of the item asks examinees to select or indicate their answer in the response space.</td>
</tr>
<tr>
<td>Graph/draw/make/complete schematic visual</td>
<td>The text of the item asks examinees to graph, draw, mark, or complete the illustration.</td>
</tr>
<tr>
<td>Image concreteness</td>
<td>The level/degree of realism in representing object(s) or event(s).</td>
</tr>
<tr>
<td>Realistic</td>
<td>Image drawn so that it represents the object, event, and/or background the way it looks in real life.</td>
</tr>
<tr>
<td>Unrealistic</td>
<td>Image drawn so that it does not represent an object as it looks in real life (e.g., a cartoon image).</td>
</tr>
<tr>
<td>Scheme</td>
<td>Image that shows the essential components of the object, event, or the physical structure of something in a simplified manner (e.g., a cone or prism).</td>
</tr>
<tr>
<td>Does not apply</td>
<td>Does not apply because the image is a schematic VR.</td>
</tr>
<tr>
<td>Background</td>
<td>Visual includes non-essential background information that is</td>
</tr>
</tbody>
</table>
intended to provide a context for the proper interpretation of the focal visual.

**With background**
Visual depicts not only the focal object/event, but also the set of visual elements intended to provide a visual perspective.

**Without background**
Visual only depicts the focal object or event and is void of background images.

**View**
Position from which an object is observed.

- **From the side/external** Object seen from the outside or side.
- **Internal** Object seen from the inside.
- **Does not apply** Does not apply because the item does not contain a concrete or schematic visual representation.

**Dimension**
The measurement of length, width, and/or depth of an object or the visual as a whole.

- **2D** The length and width of the object/visual are depicted.
- **3D** The length, width, and depth of the object/visual are depicted.

**Relative position of objects**
Degree to which the position of objects are relevant to each other.

- **Relevant** The position of the objects in the visual is relevant to one another.
- **Irrelevant** The position of the objects in the visual is not relevant to one another.

**Relative scale of objects**
Degree of size proportionality between objects or components.

- **Proportionate** Objects are shown at a consistent scale.
- **Disproportionate** Objects are not shown to scale.

**Color**
Visual spectrum range.

- **Black and white** Visual is in black and white.
- **Gray scale** Visual uses a gray scale.
- **Color** Visual uses color.

**Schematic Type**
The different categories of schematic visuals.

- **Number line** The visual is or contains a line on which numbers are marked at intervals, often used to illustrate simple numerical operations.
- **Categorical table** The visual is or contains a table that represents types of data into different groups. Two-way tables are frequently used to present categorical data by counting the number of observations that fall into each group for two variables, one divided into rows and the other into columns.
- **Function** The visual represents a relationship or expression involving one or more variables (e.g., mx + b).
- **Coordinate graph** The visual is or contains a grid that has 2 perpendicular lines, or axes, labeled like number lines.
- **Points** The visual is or contains a geometric figure formed at the intersection of 2 distinct lines.
- **Line/segment** The visual is or contains a line, which is the geometric figure formed by two points. A line is the straight path connecting 2 points and extending beyond the points in both directions. A line segment includes all points between 2 given points.
- **2D shape** The visual is or contains the form of a 2D object – how it is laid out in space (e.g., circles, squares, triangles).
- **3D shape** The visual is or contains the form of a 3D object – how it is laid
<table>
<thead>
<tr>
<th>Axes</th>
<th>The visual contains a fixed reference line for the measurement of coordinates.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>X-axis</strong></td>
<td>The visual contains the horizontal axis of a system of coordinates, points along which have a value of zero for all other coordinates.</td>
</tr>
<tr>
<td><strong>Y-axis</strong></td>
<td>The visual contains the vertical axis of a system of coordinates, points along which have a value of zero for all other coordinates.</td>
</tr>
<tr>
<td>Intervals</td>
<td>The visual contains intervals, which are the distances between 1 number and the next on the scale of a graph.</td>
</tr>
<tr>
<td><strong>Equal between axes</strong></td>
<td>The visual contains intervals that are equal distances between all numbers on the scale of the graph.</td>
</tr>
<tr>
<td><strong>Unequal between axes</strong></td>
<td>The visual contains intervals that are not equal distances between all numbers on the scale of the graph.</td>
</tr>
<tr>
<td><strong>Does not apply</strong></td>
<td>Does not apply because the image does not have axes or there is no image.</td>
</tr>
<tr>
<td>Coordinate plane</td>
<td>These variables represent the different components of a coordinate graph/plane, which is determined by a horizontal number line (x-axis) and a vertical number line (y-axis) intersecting at a point called the origin.</td>
</tr>
<tr>
<td><strong>Grid</strong></td>
<td>The coordinate plane contains a framework of spaced bars that are parallel to or cross each other.</td>
</tr>
<tr>
<td><strong>1 quadrant</strong></td>
<td>The upper right-hand corner of a coordinate (Cartesian) plane where both x and y are positive.</td>
</tr>
<tr>
<td><strong>2 quadrants</strong></td>
<td>The upper left-hand corner of a coordinate (Cartesian) plane where x is negative and y is positive.</td>
</tr>
<tr>
<td><strong>3 quadrants</strong></td>
<td>The lower left-hand corner of a coordinate (Cartesian) plane where both x and y are negative.</td>
</tr>
<tr>
<td><strong>4 quadrants</strong></td>
<td>The lower right-hand corner of a coordinate (Cartesian) plane where x is positive and y is negative.</td>
</tr>
<tr>
<td><strong>No coordinate plane</strong></td>
<td>Visual does not contain a coordinate plane.</td>
</tr>
</tbody>
</table>
Appendix I
Examples of Item Coding

I present four examples of the coding system in this appendix. Due to the length of the system, I do not demonstrate all the codes. Rather, I demonstrate the codes for the interpersonal metafunction for natural language (NL), the ideational metafunction for mathematical symbols and expressions (MSE), the textual metafunction for visual representations (VR), and the process of intersemiotic ideational analysis.

Example 1: Interpersonal Metafunction for Natural Language

Figure II shows Item 1 from the SBAC Grade 8 Math Practice Test. The interpersonal metafunction only has one dimension: interpersonal. The IDVs for this dimension for natural language are directive, factual, inquisitive, rational, and false. The NL of this item was coded as directive, as it tells students what to do (Drag each number to its correct position on the number line.) The text does not give facts, ask questions, follow logic, or give false information/answers. Table II contains the coding for this metafunction for Item 1.

Table I1

*Interpersonal Analysis for the Natural Language Used in Item 1*

<table>
<thead>
<tr>
<th>NL.1. Interpersonal metafunction</th>
<th>NL.1.1. Interpersonal: I. directive</th>
</tr>
</thead>
</table>

**Example 2: Ideational Metafunction for Mathematical Symbols and Expressions**

Figure I2 shows Item 5 from the practice test. The ideational metafunction for MSE has two dimensions: MSE purpose and MSE functional categories. The MSE purpose of this item was coded as *discrete numbers on a graph* and *measurement* because the numbers are used to label the x- and y-axis and to describe how far Jason’s house is from the school and how far and long he walked, respectively. The MSE functional categories of this item were coded as *attributes* and *circumstances* because the numbers are used to describe the attributes of the coordinate graph and the circumstances of Jason’s trip, respectively. Table I2 contains the codes for this metafunction for Item 5.
Table I2

_Ideational Analysis for the Mathematical Symbols and Expressions Used in Item 5_

MSE.2. Ideational metafunction

MSE.2.1. MSE purpose: 2. discrete numbers on graph – numbers are used to label the x- and y-axis, 3. measurement – numbers are used to describe how far Jason’s house is from the school, how far he walked, and how long he walked

MSE.2.2. MSE functional categories: 3. attributes – numbers are used to describe the attributes of the coordinate graph, 4. circumstances – numbers are used to describe the circumstances of Jason’s trip

Example 3: Textual Metafunction for Visual Representations

Figure I3 shows Item 24 from the practice test. The textual metafunction for VR has several dimensions: VR type, composition, location, reference to VR, stated actions to perform with VR, image concreteness, background, view, dimension, relative position of objects, relative scale of objects, color, schematic type, axes, intervals, and coordinate plane. VR type was coded as _schematic_ because there is a table containing the function,
symbols because there are symbols (e.g., delete and undo) in the text box, and \textit{response space} because there is a visual frame around the response space. Composition was coded as \textit{compound image} because there is one image for the table and another for the response space. Table I3 shows all the codes for the textual metafunction of VR in Item 24.

\begin{table}
\centering
\begin{tabular}{|c|c|}
\hline
\textbf{Time (hr)} & \textbf{Water Level (ft)} \\
\hline
0 & 50 \\
2 & 40 \\
4 & 30 \\
6 & 20 \\
\hline
\end{tabular}
\caption{This table shows the linear relationship of the water level in a tank and time.}
\end{table}

\begin{itemize}
\item Enter the rate of change of the water level, in feet per hour.
\end{itemize}

\textit{Figure I3.} Item 24 from the 2015 SBAC Grade 8 Math Practice Test. Source: \textit{SBAC Grade 8 Math Practice Test. Taken from Smarter Balanced Assessment Consortium.} (2014). Retrieved from \url{http://sbac.portal.airast.org/practice-test/}. 
Mathematics Assessment: Semitocs and Emergent Bilinguals

Table I3

Textual Analysis for the Visual Representations Used in Item 24

**VR.3. Textual metafunction**

- **VR type:** 2. schematic – there is a table for the function, 4. symbols – there are symbols (e.g., delete and undo) in the text box, 5. RS – there is a visual frame around the response space
- **Composition:** 2. compound image – there is one image for the table and another for the response space
- **VR location:** 2. embedded in stem/prompt – the table is between the stem and the item prompt, 3. embedded in RS – the visual frame of the response space is embedded in the response space, 4. embedded in response tool – the symbols, such as delete, are in the response tool
- **Reference to VR:** 1. explicit – the text refers to the table
- **Stated actions to perform w/ VR:** 2. enter – the text tells students that they will be entering their answer in the response space
- **Image concreteness:** 4. NA – does not apply as it is a schematic visual
- **Background:** 2. without background – there is no background visual
- **View:** 1. side/external – the table is viewed from the side
- **Dimension:** 1. 2D – the table and response space are in 2D
- **Relative position of objects:** 1. relevant – the positions of the numbers in the table are relevant, 2. irrelevant – the positions of the symbols in the response tool are irrelevant
- **Relative scale of objects:** 1. proportionate – none of the constituents are larger or smaller than each other
- **Color:** 1. black and white – the item is in black and white
- **Schematic type:** 3. function – the table is modeling a linear function
- **Axes:**
- **Intervals:** 3. NA – does not apply because there is no coordinate graph
- **Coordinate plane:** 5. no coordinate plane – does not apply because there is no coordinate graph

**Example 4: Intersemiotic Ideational Analysis**

I provide only one example of the intersemiotic ideational coding process using Item 24 (Figure I3). There are four types of functional categories (participants, activity, attributes, circumstances) and seven types of intersemiotic relations (Repetition, Synonymy, Antonymy, Hyponymy, Metonymy, Collocation, Identity). The process of intersemiotic ideational analysis is the same for each functional category and each pair of
semiotic resources. Table I4 shows the first part of the process, which is to identify the functional category constituents belonging to NL, MSE, and VR.

Table I4

*Functional Category Constituents for Item 24*

<table>
<thead>
<tr>
<th>Semiotic Resource</th>
<th>Participants</th>
<th>Activity</th>
<th>Attributes</th>
<th>Circumstances</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL</td>
<td>table, relationship of the water level in a tank and time, the rate of change of the water level</td>
<td>shows, enter</td>
<td>linear relationship, feet per hour</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>time increases by 2, water level decreases by 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VR</td>
<td>categorical table (time/water level), text box</td>
<td>enter answer</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table I5 shows the second part of the process, which is to identify the intersemiotic relationship between the three pairs of semiotic resources. In this example, there is intersemiotic Repetition and Identity between NL and VR for participants because both semiotic resources have a table, and the NL identifies the type of relationship (linear) that is found in the information contained in the table, respectively. There is intersemiotic Collocation between the NL and VR for activity because the text tells students to enter the rate of change of the water level, thus it is expected that the visual contains a response space for students to enter their answers. Lastly, there is intersemiotic Repetition and Identity between NL and MSE for attributes because the text and MSE both represent *feet*.
per hour and the text identifies the constant rate of change of the time and water level as a linear relationship, respectively.

Table I5

*Intersemiotic Ideational Relations for Item 24*

<table>
<thead>
<tr>
<th>Semiotic Resources Pair</th>
<th>Participants</th>
<th>Activity</th>
<th>Attributes</th>
<th>Circumstances</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL/MSE</td>
<td>Repetition, Identity</td>
<td></td>
<td>Repetition, Identity</td>
<td></td>
</tr>
<tr>
<td>NL/VR</td>
<td>Repetition, Identity</td>
<td>Collocation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE/VR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix J
Pilot Stage of the Development of Instruments

The original version comprised five phases: Introduction, Sample Item, Concurrent Reporting, Retrospective Reporting, and Follow-Up Questions. The first phase (Introduction) allowed me to introduce myself and my study. In this phase, I explained the purpose for and procedures of the interview. I then asked the students for the following demographic information: name, age, grade, and language(s) spoken.

The second phase (Sample Item) allowed the students to become familiar with the interview protocol, think-aloud procedure, and the SBAC interface and accessibility features. In this phase, students solved the sample item, complete the think-aloud procedure, and answered the retrospective reporting questions. An item from the 6th grade math practice test was chosen for this phase to allow students to focus their cognitive efforts on understanding how to use the SBAC computer-based system and resources. The practice item was also chosen because it contains multiple semiotic resources and one of the most common answer types—using the keypad to enter one’s response.

The third phase (Concurrent Reporting) allowed me to observe the students’ cognitive processes, indirectly, as they interpreted, solved, and responded to four items from the 2015 SBAC Grade 8 Math Practice Test. In this phase, students read each problem out loud and told me everything they were doing and thinking about as they solved each problem. As recommended by Ericsson and Simon (1993), I minimized my influence on students’ thinking by limiting my interactions with them. For example, I limited my speech to “keep talking” or “tell me more” when students fell silent, so as not to interfere
with their thought processes. Items 19, 20, 21, and 25 were chosen because they were the most different between each other, in terms of their semiotic resources, and, altogether, contained most of the features used in the practice test.

The fourth phase (Retrospective Reporting) allowed me to ask students questions that were directly related to how they interacted with the semiotic resources of each item as they interpreted, solved, and responded to them. In this phase, I asked students scripted and spontaneous probes. The scripted probes began by asking students to rate the level of mental effort used for each item to gauge the intrinsic cognitive loads of the items (Mayer, 2005). I then asked the students to recall the steps they took to solve each item. The succeeding probes contained questions concerning the semiotic resources that I wanted the students to specifically address. For instance, how the students used words or numbers and symbols to help them answer the item, or how the user interface affected the problem-solving process. The protocol allowed for spontaneous probes to emerge as the interviews developed with each student so as not to lose any important information that might not have been addressed in the scripted probes (Beatty & Willis, 2007; Willis, 1999). Items 19, 20, 21, and 25 were also used for the fourth phase.

The fifth phase (Follow-Up Questions) allowed me to ask additional questions that were not necessarily related to the semiotic composition of the items or the problem-solving process. I asked the follow-up questions after the students had completed the concurrent and retrospective reporting for Items 19, 20, 21, and 25. These questions asked the students to rate the difficulty of the items as a whole (to measure the germane cognitive load of the assessment), if and how their teachers prepared them for the PARCC assessment, if they were ready to take the PARCC assessment, and if they
preferred computer-based or paper-and-pencil assessments and why. While the items I used were SBAC items, the follow-up questions were equally applicable for SBAC and PARCC.

The original intent was to pilot test and refine the first version of the protocol using a separate sample of 10 EB and 10 non-EB 11th grade students from two classes that were not in the sample for classroom observations and field interviews. However, these numbers were not attained due to the small number of EBs in these classrooms and the low return rate of student assent and parent permission forms. I was only able to interview three EB and three non-EB students.

Table J1 shows the sociodemographic on the students who participated in the pilot interviews. One of the students was in Teacher 7’s Contemporary Mathematics in Context III class and five of the students were in Teacher 5’s Trigonometry and Math Analysis class. All three of the EBs were female Latinas and spoke Spanish as their first language; two of them were classified as limited English proficiency 4 and one as fluent English proficiency 5. The pilot interviews took place between January 28, 2015 and March 5, 2015.
## Pilot Interview Student Sociodemographic Information

<table>
<thead>
<tr>
<th>School</th>
<th>Teacher &amp; Section</th>
<th>Class</th>
<th>Student</th>
<th>Sex</th>
<th>Grade</th>
<th>Age</th>
<th>Race/Ethnicity</th>
<th>ELL</th>
<th>ELP Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>T5.3</td>
<td>Trig/Math Anal</td>
<td>PS1</td>
<td>F</td>
<td>11</td>
<td>16</td>
<td>African American</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>T5.3</td>
<td>Trig/Math Anal</td>
<td>PS2</td>
<td>M</td>
<td>11</td>
<td>16</td>
<td>Latino</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>T5.3</td>
<td>Trig/Math Anal</td>
<td>PS3</td>
<td>F</td>
<td>11</td>
<td>16</td>
<td>Latino</td>
<td>1</td>
<td>LEP 4</td>
</tr>
<tr>
<td>2</td>
<td>T5.3</td>
<td>Trig/Math Anal</td>
<td>PS4</td>
<td>F</td>
<td>11</td>
<td>17</td>
<td>Latino</td>
<td>1</td>
<td>LEP 4</td>
</tr>
<tr>
<td>2</td>
<td>T5.3</td>
<td>Trig/Math Anal</td>
<td>PS5</td>
<td>F</td>
<td>11</td>
<td>16</td>
<td>Latino</td>
<td>1</td>
<td>FEP 5</td>
</tr>
<tr>
<td>2</td>
<td>T7.1</td>
<td>CMIC III</td>
<td>PS6</td>
<td>M</td>
<td>11</td>
<td>16</td>
<td>White</td>
<td>0</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Note: Trig/Math Anal is the acronym for the Trigonometry and Math Analysis class. CMIC is the acronym for the *Contemporary Mathematics in Context* curriculum. LEP is the acronym for limited English proficiency and FEP is the acronym for fluent English proficiency.

Table J2 shows whether the students answered the item correctly, the percentage of total points received, the average score for each item, and the overall average. The scores were calculated using the rubrics provided by SBAC (Smarter Balanced Assessment Consortium, 2014). Items 19 and 21 were worth one point, so students received either full or no credit. Items 20 and 25 were worth two points, and students could receive no credit, partial credit, or full credit for their answers. The overall average was 30.6%; none of the students answered Item 19 correctly, three students received partial credit on Item 20, one student answered Item 21 correctly, and four students received full credit on Item 25. While Item 25 had a high success rate, it does not truly represent their content knowledge, as I helped these students reach the correct solution, which will be discussed in more detail.
Table J2

*Pilot Interview Scores*

<table>
<thead>
<tr>
<th>Student</th>
<th>Interview Date</th>
<th>Item</th>
<th>19 (1 pt)</th>
<th>20 (2 pts)</th>
<th>21 (1 pt)</th>
<th>25 (2 pts)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS1</td>
<td>2/12/15</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2*</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>PS2</td>
<td>2/19/15</td>
<td>0</td>
<td>1*</td>
<td>0</td>
<td>2*</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>PS3</td>
<td>3/5/15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2*</td>
<td>16.7</td>
<td></td>
</tr>
<tr>
<td>PS4</td>
<td>3/2/15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PS5</td>
<td>3/3/15</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2*</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>PS6</td>
<td>1/28/15</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>16.7</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>N/A</td>
<td>0</td>
<td>0.25</td>
<td>0.17</td>
<td>0.67</td>
<td>30.6</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Cells with a * indicate that the student received help and/or advice from me.

From the findings presented above, it is obvious that the items were too challenging. All but one of the students could graph the linear function that was represented in the table shown in Item 19, but they did not realize that was not the function they were supposed to graph. Thus, I replaced this item with Item 24. This item still addresses linear functions, but the directions are more explicit.

All the students stated that they did not remember the information necessary to solve Items 20 (transformation), 21 (angle properties), and 25 (volume of cylinders). Given that these are eighth grade items, this is a legitimate reason for not remembering the content as all three areas are only briefly covered in the CMIC eighth grade curriculum. Additionally, Items 20 and 21 have their own set of vocabulary that does not, generally, apply to mathematics in a broader sense. For example, *translate* (i.e., move up/down or right/left) and *reflect* (i.e., turn/flip) are words commonly used when discussing transformations and *alternate interior angles* and *vertical angles* are commonly used when discussing angle properties. Item 25 was a multi-step problem requiring students to
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

know the equation used to find the volume of a cylinder. I provided the students with the equation and reminded them what else they needed to do to solve the problem, as they usually grew tired around the second or third step.

Items 20, 21, and 25 were replaced with Items 1 (number sense), 5 (representing distance and rate on a graph), and 9 (Pythagorean Theorem). I chose these items for several reasons: (a) students are required to master the content of each of these items in eighth grade; (b) they are exposed to the material several times in eighth grade and again in at least one, if not all, high school math classes; (c) the vocabulary is not as specialized as that used in Items 20 and 21; and (d) they do not require as many steps as does Item 25. Appendix M contains descriptions of each selected item (1, 5, 9, and 24) and a table of their English and/or Spanish glossaries.

The pilot interviews were also coded qualitatively. The process began by taking notes as I listened to the tape recordings of the interviews. I then looked for patterns among my notes and used the patterns to make changes to the interview protocol. One of the findings that emerged from the data was that the students did not need to see a sample item. For example, the students did not try when answering this question because they knew it was just a sample item. Additionally, the average interview time was 40 minutes, which was not enough time to collect the students from their class, walk them to the interview location, and answer all the questions before they had to go to their next class. Thus, I removed Phase 2 from the second version and replaced it with a short description of how to use my computer, where the items came from, and a few of the major differences between PARCC and SBAC (e.g., the glossary feature).
The remaining changes were made to the retrospective reporting questions and will be explained briefly. I first removed the question that asked students to recall their steps because they thought it was redundant after finishing the think-aloud procedure. Instead of asking, *How did you use the words to figure out what the question was asking you to do?*, I asked the students which words they paid attention to or which words really helped them figure out what they were supposed to do. This change was necessary because the students often did not understand the purpose of the original question and/or thought the answer was self-explanatory.

The next set of original questions asked the students how they used language, numbers and symbols, and visuals to solve the question and how they were similar or different. I changed these questions to be more specific to each semiotic resource and question because, once again, the students thought the answers were self-explanatory. For example, I asked the students why there were arrows under each fraction in Item 1. I also found that students would not recommend any changes to the items unless I gave them an example. Thus, I added an example to the end of this question to help the students gauge what I was looking for. I also found that students would not elaborate on their responses unless asked to do so, so I added questions such as *How?*, *Why?*, and statements such as *Explain why*, to encourage more detailed responses.

Table J3 shows the sociodemographic information for the observed classes. Ten classes from ten different teachers were randomly selected for classroom observations. Each class was observed and video recorded two times for a total of 20 observations. Two classes were chosen from both School 1 and School 3 and the remaining six classes were chosen from School 2. The average number of students in each class was 25. The
average percentage of sophomores in each class was 39.0%, followed by 47.7% juniors, and 11.9% seniors. The average number of EBs in each class was 35% with a standard deviation of 18.6%.

The classroom observations were used to identify the semiotic resources that were pertinent in the classroom but not yet identified in the interview protocol. To capture this information, I watched the recordings of each observation and created inductive codes as they emerged from the data (Strauss & Corbin, 1994). The codes identified the types of semiotic resources that were used (e.g., color, gestures, and underlining), by whom and
# Table J3

**Sociodemographic Variables for Students Who Participated in Classroom Observations (%)**

<table>
<thead>
<tr>
<th>School</th>
<th>Teacher &amp; Section</th>
<th>Class</th>
<th>Number of Students</th>
<th>EBs (%)</th>
<th>Grade 10</th>
<th>Grade 11</th>
<th>Grade 12</th>
<th>NA</th>
<th>Male</th>
<th>Female</th>
<th>Caucasian</th>
<th>Hispanic</th>
<th>Asian</th>
<th>Black</th>
<th>Native American</th>
<th>Biracial</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T1</td>
<td>CMIC 3</td>
<td>26</td>
<td>15.4</td>
<td>0.0</td>
<td>84.6</td>
<td>15.4</td>
<td>53.8</td>
<td>46.2</td>
<td>42.3</td>
<td>34.6</td>
<td>3.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>19.2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>T2</td>
<td>CMIC 3</td>
<td>19</td>
<td>21.1</td>
<td>15.8</td>
<td>68.4</td>
<td>15.8</td>
<td>57.9</td>
<td>42.1</td>
<td>63.2</td>
<td>21.5</td>
<td>10.5</td>
<td>0.0</td>
<td>5.3</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>T3</td>
<td>CMIC 2</td>
<td>27</td>
<td>51.9</td>
<td>0.0</td>
<td>96.2</td>
<td>0.0</td>
<td>3.7</td>
<td>15.0</td>
<td>85.0</td>
<td>18.5</td>
<td>70.4</td>
<td>7.4</td>
<td>3.7</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>T4</td>
<td>CMIC 2</td>
<td>25</td>
<td>60.0</td>
<td>88.0</td>
<td>12.0</td>
<td>0.0</td>
<td>36.0</td>
<td>64.0</td>
<td>28.0</td>
<td>64.0</td>
<td>4.0</td>
<td>4.0</td>
<td>0.0</td>
<td>4.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>T5.1</td>
<td>CMIC 2</td>
<td>26</td>
<td>57.7</td>
<td>73.1</td>
<td>19.2</td>
<td>0.0</td>
<td>7.7</td>
<td>38.5</td>
<td>61.5</td>
<td>11.5</td>
<td>80.8</td>
<td>0.0</td>
<td>0.0</td>
<td>3.8</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>T6</td>
<td>CMIC 4</td>
<td>20</td>
<td>20.0</td>
<td>60.0</td>
<td>25.0</td>
<td>15.0</td>
<td>65.0</td>
<td>35.0</td>
<td>35.0</td>
<td>50.0</td>
<td>5.0</td>
<td>5.0</td>
<td>0.0</td>
<td>5.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>T7.2</td>
<td>CMIC 3</td>
<td>27</td>
<td>48.1</td>
<td>59.3</td>
<td>40.7</td>
<td>0.0</td>
<td>59.3</td>
<td>40.7</td>
<td>22.2</td>
<td>70.4</td>
<td>7.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>T8.2</td>
<td>CMIC 2</td>
<td>26</td>
<td>42.3</td>
<td>26.9</td>
<td>53.8</td>
<td>19.2</td>
<td>38.5</td>
<td>61.5</td>
<td>46.2</td>
<td>50.0</td>
<td>0.0</td>
<td>3.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>T9</td>
<td>CMIC 2</td>
<td>30</td>
<td>16.7</td>
<td>66.7</td>
<td>26.7</td>
<td>3.3</td>
<td>3.3</td>
<td>53.3</td>
<td>46.7</td>
<td>56.7</td>
<td>33.3</td>
<td>10.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>T10</td>
<td>Algebra II</td>
<td>24</td>
<td>16.7</td>
<td>0.0</td>
<td>50.0</td>
<td>50.0</td>
<td>66.7</td>
<td>33.3</td>
<td>66.7</td>
<td>29.2</td>
<td>4.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

with whom (i.e., teacher to class, teacher to student, student to teacher, student to student, student work, and class materials), and the frequency with which each resource was used in each context.

Across all teachers, observations, and contexts, 77 semiotic resources were identified during coding. The codes were then compiled by teacher and summed across teachers. Any semiotic resource that was used less than twice across all situations was removed, leaving 30 resources. I analyzed the list and removed three types of semiotic resources: (a) those that were not relevant to the SBAC practice test, (b) those that were not relevant to the four items used in the protocol, and (c) those that I had already addressed in the original or semi-modified protocol. For example, while pointing to key terms and expressions was the most prevalent semiotic resource \( (n = 53) \), test designers cannot use this feature in an online interface, and thus I did not create a question pertaining to pointing. Many of the teachers were teaching or reviewing how to expand expressions (e.g., \((x+5)(x-3)\)) and used lines to represent the multiplication of each term \( (n = 16) \). While this was a common resource and can be used in an online interface, I did not create a question concerning it, however, because none of the items in the protocol addressed this content. Additionally, many of my original questions or modifications already addressed commonly used resources, such as the use of fraction lines \( (n = 42) \) or coordinate points \( (n = 16) \).

Three codes remained at the end of this process: using a calculator \( (n = 44) \), using color \( (n = 16) \), and boxing/circling key information and/or the final answer \( (n = 15) \). Questions were created to address calculator use such as *Did you like the calculator tool? Was it easy to use? Do you prefer this calculator or the one on the PARCC test?* I also
added two questions to the end of the scripted probes for each item to address the remaining two semiotic resources: (a) *Would you use color somehow in the question? How and why?*, and (2) *Would you use bold words, italicized words, or box/circle any information? Which ones? Why?* Additionally, I added the question *Do you wish there was a place to show your work? Why or why not?* While this question does not address one semiotic resource per se, all the teachers told multiple students several times in all the observations to, *show their work*.

The last changes that I made to the protocol were due to the timing of the pilot version versus that of the modified version; the pilot interviews occurred before the PARCC assessment, while the field interviews occurred after the PARCC assessment. Naturally, this allowed me to ask more questions about PARCC, such as the similarities and differences between PARCC and the SBAC practice items. The additional follow-up questions include the following:

1. How do you think you did on the PARCC test? Why?
2. What did you like about the PARCC test?
3. What did you not like about the PARCC test?
4. Are there any similarities or differences between these problems and the problems on the PARCC test? If so, what are they?
5. Are these questions and the questions on PARCC similar or different to the types of questions you see in math class? Why?

The final, modified version of the original interview protocol can be found in Appendix N (English) and Appendix O (Spanish).
Student Cognitive Interview Protocol—English

Phase 1—Introduction

• Hi my name is Chelsey and I am from the University of Colorado. I’m investigating how different resources used on the new standardized assessments, such as pictures, graphs, and computer features, affect your interpretation of and ability to solve math problems in order to improve these tests.
• We will start with a few background questions. Remember, that your answers will not affect your grade in class.

1. What’s your name?
2. How old are you?
3. What grade are you in?
4. Do you speak any other languages in addition to English?

Phase 2—Sample Item

• We are going to start with a sample item just so you can become familiar with the interview process. The sample item is from the 6th grade math test.

Sample Item: Concurrent Reporting

- Once I give you the problem, I want you to read it out loud and tell me everything you are thinking about as you solve the problem. For example, what you’re doing, why you’re doing it, what you’re looking at, etc.

Sample Item: Retrospective Reporting

- Now I’m going to ask you some questions about the item. Answer as best you can with as much detail as you can.

1. On a scale of 1 to 10, 1 being extremely low and 10 being extremely high, rate the level of mental effort that you used to solve this problem.
2. Can you recall the steps you took to solve the question? Start with the first thing you thought and did and continue with the next step.
3. What is the question asking you to do?
4. How did you use the words to figure out what the question was asking you to do?
5. How did you use the words to solve the problem?
6. What is this picture for?
7. How does the picture relate to the question?
8. How did you use the picture to help you figure out what the question was asking you to do?
9. Did you use the picture to solve the problem? If so, how did you use the picture?
10. How are the words and this picture similar/related?
11. How are the words and this picture different?
12. Did you use any of the glossaries? If so, which glossaries did you use?
13. Were the glossaries helpful? Why or why not?
14. Are there words or terms in the question that you wished had glossaries? Why or why not?
15. How did you enter your answer?
16. Was it easy or hard to figure out how to enter your answer? Why?
17. Are there numbers, symbols, buttons, or keys that you don’t understand on the keypad? If so, which ones?
18. Was there anything confusing about the item? If so, what and why?
19. Would you change anything about this item? If so, what, how, and why?
20. Is there anything else about this item that you would like to talk about?

Grade 8 Math Practice Test Selected Items

- Thank you for answering the practice item. We are now going to move to the 8th grade items. I will be asking you similar questions for each of the 4 items.
Sample Item 19: Concurrent Reporting

- Once I give you the problem, I want you to read it out loud and tell me everything you are thinking about as you solve the problem. For example, what you’re doing, why you’re doing it, what you’re looking at, etc.

**Figure K2. 8th Item 19 from the 2015 SBAC Grade 8 Math Practice Test. Source: Grade 8 Math Practice Test. Taken from Smarter Balanced Assessment Consortium. (2014). Retrieved from http://sbac.portal.airast.org/practice-test/.**

Sample Item 19: Retrospective Reporting

- Now I’m going to ask you some questions about the item. Answer as best you can with as much detail as you can.

1. On a scale of 1 to 10, 1 being extremely low and 10 being extremely high, rate the level of mental effort that you used to solve this problem.
2. Can you recall the steps you took to solve the question? Start with the first thing you though and did and continue with the next step.
3. What is the question asking you to do?
4. How did you use the words to figure out what the question was asking you to do?
5. How did you use the words to solve the problem?
6. How are numbers and symbols used in this problem?
7. How did you use numbers and symbols to solve the problem?
8. How are the words and numbers and symbols similar or related?
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

9. How are the words and numbers and symbols different?
10. Please explain the graph using as much detail as possible.
11. How does the graph relate to the question?
12. How did you use the graph to help you figure out what the question was asking you to do?
13. How did you use the graph to solve the problem?
14. How are the words and graph similar/related?
15. How are the words and graph different?
16. How are the numbers and symbols and graph similar/related?
17. How are the numbers and graph different?
18. How are the words, numbers and symbols, and graph similar/related?
19. How are the words, numbers and symbols, and graph different?
20. Did you use any of the glossaries? If so, which glossaries did you use?
21. Were the glossaries helpful or not? Why or why not?
22. Are there words or terms in the question that you wished had glossaries? Why or why not?
23. How did you enter your answer?
24. Was it easy or hard to figure out how to enter your answer? Why?
25. Was there anything confusing about the item? If so, what and why?
26. Would you change anything about this item? If so, what, how, and why?
27. Is there anything else about this item that you would like to talk about?

Sample Item 20: Concurrent Reporting

- Once I give you the problem, I want you to read it out loud and tell me everything you are thinking about as you solve the problem. For example, what you’re doing, why you’re doing it, what you’re looking at, etc.
Sample Item 20: Retrospective Reporting

- Now I’m going to ask you some questions about the item. Answer as best you can with as much detail as you can.

1. On a scale of 1 to 10, 1 being extremely low and 10 being extremely high, rate the level of mental effort that you used to solve this problem.
2. Can you recall the steps you took to solve the question? Start with the first thing you though and did and continue with the next step.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

3. What is the question asking you to do?
4. How did you use the words to figure out what the question was asking you to do?
5. How did you use the words to solve the problem?
6. How are numbers and symbols used in this problem?
7. How did you use numbers and symbols to solve the problem?
8. How are the words and numbers and symbols similar or related?
9. How are the words and numbers and symbols different?
10. Please explain the graph using as much detail as possible.
11. How does the graph relate to the question?
12. How did you use the graph to help you figure out what the question was asking you to do?
13. How did you use the graph to solve the problem?
14. How are the words and graph similar/related?
15. How are the words and graph different?
16. How are the numbers and symbols and graph similar/related?
17. How are the numbers and graph different?
18. How are the words, numbers and symbols, and graph similar/related?
19. How are the words, numbers and symbols, and graph different?
20. Did you use any of the glossaries? If so, which glossaries did you use?
21. Were the glossaries helpful or not? Why or why not?
22. Are there words or terms in the question that you wished had glossaries? Why or why not?
23. How did you enter your answer?
24. Was it easy or hard to figure out how to enter your answer? Why?
25. Was there anything confusing about the item? If so, what and why?
26. Would you change anything about this item? If so, what, how, and why?
27. Is there anything else about this item that you would like to talk about?

Sample Item 21: Concurrent Reporting

- Once I give you the problem, I want you to read it out loud and tell me everything you are thinking about as you solve the problem. For example, what you’re doing, why you’re doing it, what you’re looking at, etc.
Sample Item 21: Retrospective Reporting

- Now I’m going to ask you some questions about the item. Answer as best you can with as much detail as you can.

1. On a scale of 1 to 10, 1 being extremely low and 10 being extremely high, rate the level of mental effort that you used to solve this problem.
2. Can you recall the steps you took to solve the question? Start with the first thing you thought and did and continue with the next step.
3. What is the question asking you to do?
4. How did you use the words to figure out what the question was asking you to do?
5. How did you use the words to solve the problem?
6. How are numbers and symbols used in this problem?
7. How did you use numbers and symbols to solve the problem?
8. How are the words and numbers and symbols similar or related?
9. How are the words and numbers and symbols different?
10. Please explain the diagrams using as much detail as possible.
11. How do the diagrams relate to the question?
12. How did you use the diagrams to help you figure out what the question was asking you to do?
13. How did you use the diagrams to solve the problem?
14. How are the words and diagrams similar/related?
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

15. How are the words and diagrams different?
16. How are the numbers and symbols and diagrams similar/related?
17. How are the numbers and symbols and diagrams different?
18. How are the words, numbers and symbols, and diagrams similar/related?
19. How are the words, numbers and symbols, and diagrams different?
20. Did you use any of the glossaries? If so, which glossaries did you use?
21. Were the glossaries helpful or not? Why or why not?
22. Are there words or terms in the question that you wished had glossaries? Why or why not?
23. How did you enter your answers?
24. Was it easy or hard to figure out how to enter your answer? Why?
25. Was there anything confusing about the item? If so, what and why?
26. Would you change anything about this item? If so, what, how, and why?
27. Is there anything else about this item that you would like to talk about?

Sample Item 25: Concurrent Reporting

- Once I give you the problem, I want you to read it out loud and tell me everything you are thinking about as you solve the problem. For example, what you’re doing, why you’re doing it, what you’re looking at, etc.

![Concurrent Reporting Diagram](image)

Sample Item 25: Retrospective Reporting

- Now I’m going to ask you some questions about the item. Answer as best you can with as much detail as you can.

1. On a scale of 1 to 10, 1 being extremely low and 10 being extremely high, rate the level of mental effort that you used to solve this problem.
2. Can you recall the steps you took to solve the question? Start with the first thing you though and did and continue with the next step.
3. What is the question asking you to do?
4. How did you use the words to figure out what the question was asking you to do?
5. How did you use the words to solve the problem?
6. How are numbers and symbols used in this problem?
7. How did you use numbers and symbols to solve the problem?
8. How are the words and numbers and symbols similar or related?
9. How are the words and numbers and symbols different?
10. Please explain the diagrams using as much detail as possible.
11. How do the diagrams relate to the question?
12. How did you use the diagrams to help you figure out what the question was asking you to do?
13. How did you use the diagrams to solve the problem?
14. How are the words and diagrams similar/related?
15. How are the words and diagrams different?
16. How are the numbers and symbols and diagrams similar/related?
17. How are the numbers and symbols and diagrams different?
18. How are the words, numbers and symbols, and diagrams similar/related?
19. How are the words, numbers and symbols, and diagrams different?
20. Did you use any of the glossaries? If so, which glossaries did you use?
21. Were the glossaries helpful or not? Why or why not?
22. Are there words or terms in the question that you wished had glossaries? Why or why not?
23. How did you enter your answers?
24. Was it easy or hard to figure out how to enter your answer? Why?
25. Was there anything confusing about the item? If so, what and why?
26. Would you change anything about this item? If so, what, how, and why?
27. Is there anything else about this item that you would like to talk about?

Follow-up Questions

1. On a scale of 1 to 10, 1 being extremely easy and 10 being extremely hard, rate the difficulty of these problems.
2. Do you feel prepared to take the PARCC test?
3. How has your teacher prepared you for the test?
4. Do you prefer taking tests on computers or on paper? Why?

Thank you so much for participating in this interview!
Estudiante Entrevista Cognitiva Protocolo—Español

Fase 1-Introducción

- Hola mi nombre es Chelsey y soy de la Universidad de Colorado. Estoy investigando cómo los diferentes recursos utilizados en las nuevas evaluaciones estandarizadas, como imágenes, gráficos y características del sistema, afectarán a su interpretación de y capacidad para resolver problemas de matemáticas con el fin de mejorar estas evaluaciones.

1. ¿Cómo te llamas?
2. ¿Cuántos años tienes?
3. ¿En qué grado estás?
4. ¿Hablas otros idiomas además del Inglés?

Fase 2—Pregunta de la Práctica

- Vamos a empezar con una pregunta de la práctica sólo para que pueda familiarizarte con el proceso de la entrevista. La pregunta es de la prueba de matemáticas de 6º grado.
Pregunta de la Práctica: Reporte Concurrente

- Cuando te doy el problema, quiero que lo leas en voz alta y me digas todo lo que estás pensando en cómo resolver el problema. Por ejemplo, lo que estás haciendo, por qué lo estás haciendo, lo que estás viendo, etc.

Pregunta de la Práctica: Reporte Retrospectivo

- Ahora voy a hacerte algunas preguntas sobre el problema. Respondas lo mejor que puedas con tanto detalle como sea posible.

1. En una escala de 1 a 10, 1 es muy bajo y 10 es muy alta, evaluar el nivel de esfuerzo mental que utilizaste para resolver este problema.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

2. ¿Puedes recordar los pasos a seguir para resolver el problema? Comiencnes con lo primero que pensaste y lo hiciste y continuar con el siguiente paso.
3. ¿Lo que es la pregunta está pidiendo que hagas?
4. ¿Cómo te usan las palabras para entender lo que la pregunta estaba pidiendo que hagas?
5. ¿Cómo te usan las palabras para resolver el problema?
6. ¿Qué es esta ilustración para?
7. ¿Cómo la ilustración se relacionan con la pregunta?
8. ¿Cómo te usa la ilustración para ayudarte a determinar cuál es la pregunta que estaba pidiendo a hacer?
9. ¿Usaste la ilustración para resolver el problema? Si es así, ¿cómo te usa la ilustración?
10. ¿Cómo son las palabras y la ilustración similares o relacionados?
11. ¿Cómo son las palabras y la ilustración diferentes?
12. ¿Usaste alguno de los glosarios? ¿Cuáles?
13. ¿Fueron útiles los glosarios? ¿Por qué o por qué no?
14. ¿Hay palabras o términos de la pregunta que te gustaría que tenían glosarios? ¿Por qué o por qué no?
15. ¿Cómo escribiste su respuesta?
16. ¿Fue fácil o difícil de encontrar la manera de entrar en tu respuesta? ¿Por qué?
17. ¿Existen los números, los símbolos, los botones, o las teclas que no entiendes en el teclado? Si es así, ¿cuáles?
18. ¿Hubo algo confuso sobre la pregunta? Si es así, qué y por qué?
19. ¿Le cambiarías algo de esta pregunta? Si es así, qué, cómo, y por qué?
20. ¿Hay algo más sobre esta pregunta que le gustaría hablar?

Grade 8 Math Practice Test Preguntas Seleccionadas

- Gracias por contestar la pregunta de práctica. Ahora vamos a pasar a las preguntas de 8° grado. Estaré haciéndote preguntas similares para cada una de las cuatro preguntas.

Pregunta 19: Reporte Concurrente

- Cuando te doy el problema, quiero que lo leas en voz alta y me digas todo lo que estás pensando en como resolver el problema. Por ejemplo, lo que estás haciendo, por qué lo estás haciendo, lo que estás viendo, etc.

**Pregunta 19: Reporte Retrospectivo**

- Ahora voy a hacerte algunas preguntas sobre el problema. Respondas lo mejor que puedas con tanto detalle como sea posible.

1. En una escala de 1 a 10, 1 es muy bajo y 10 es muy alta, evaluar el nivel de esfuerzo mental que utilizaste para resolver este problema.
2. ¿Puedes recordar los pasos a seguir para resolver el problema? Comiences con lo primero que pensaste y lo hiciste y continuar con el siguiente paso.
3. ¿Lo que es la pregunta está pidiendo que hagas?
4. ¿Cómo te usan las palabras para entender lo que la pregunta estaba pidiendo que hagas?
5. ¿Cómo te usas los números utilizados en este problema?
6. ¿Cómo te usan los números para resolver el problema?
7. ¿Cómo son las palabras y números similares o relacionados?
8. ¿Cómo son las palabras y números diferentes?
10. ¿Qué es este gráfico para?
11. ¿Cómo el gráfico se relacionan con la pregunta?
12. ¿Cómo te usa el gráfico para ayudarte a determinar cuál es la pregunta que estaba pidiendo a hacer?
13. ¿Usaste el gráfico para resolver el problema? Si es así, ¿cómo te usa el gráfico?
14. ¿Cómo son las palabras y el gráfico similares o relacionados?
15. ¿Cómo son las palabras y el gráfico diferentes?
16. ¿Cómo son los números y el gráfico similares o relacionados?
17. ¿Cómo son los números y el gráfico diferentes?
18. ¿Cómo son las palabras, los números, y el gráfico similares o relacionados?
19. ¿Cómo son las palabras, los números y el gráfico diferentes?
20. ¿Usaste alguno de los glosarios? ¿Cuáles?
21. ¿Fueron útiles los glosarios? ¿Por qué o por qué no?
22. ¿Hay palabras o términos de la pregunta que te gustaría que tenían glosarios? ¿Por qué o por qué no?
23. ¿Cómo escribiste su respuesta?
24. ¿Fue fácil o difícil de encontrar la manera de entrar en tu respuesta? ¿Por qué?
25. ¿Hubo algo confuso sobre la pregunta? Si es así, qué y por qué?
26. ¿Le cambiarías algo de esta pregunta? Si es así, qué, cómo, y por qué?
27. ¿Hay algo más sobre esta pregunta que le gustaría hablar?

Pregunta 20: Reporte Concurrente

- Cuando te doy el problema, quiero que lo leas en voz alta y me digas todo lo que estás pensando en como resolver el problema. Por ejemplo, lo que estás haciendo, por qué lo estás haciendo, lo que estás viendo, etc.
Pregunta 20: Reporte Retrospectivo

- Ahora voy a hacerte algunas preguntas sobre el problema. Respondas lo mejor que puedas con tanto detalle como sea posible.

1. En una escala de 1 a 10, 1 es muy bajo y 10 es muy alta, evaluar el nivel de esfuerzo mental que utilizaste para resolver este problema.
2. ¿Puedes recordar los pasos a seguir para resolver el problema? Comiences con lo primero que pensaste y lo hiciste y continuar con el siguiente paso.
3. ¿Lo que es la pregunta está pidiendo que hagas?
4. ¿Cómo te usan las palabras para entender lo que la pregunta estaba pidiendo que hagas?
5. ¿Cómo te usas las palabras para resolver el problema?
6. ¿Cómo son los números utilizados en este problema?
7. ¿Cómo te usan los números para resolver el problema?
8. ¿Cómo son las palabras y números similares o relacionados?
9. ¿Cómo son las palabras y los numeros diferentes?
10. ¿Qué es este gráfico para?
11. ¿Cómo el gráfico se relacionan con la pregunta?
12. ¿Cómo se usa el gráfico para ayudarte a determinar cuál es la pregunta que estaba pidiendo a hacer?
13. ¿Usaste el gráfico para resolver el problema? Si es así, ¿cómo se usa el gráfico?
14. ¿Cómo son las palabras y el gráfico similares o relacionados?
15. ¿Cómo son las palabras y el gráfico diferentes?
16. ¿Cómo son los números y el gráfico similares o relacionados?
17. ¿Cómo son los números y el gráfico diferentes?
18. ¿Cómo son las palabras, los números, y el gráfico similares o relacionados?
19. ¿Cómo son las palabras, los números y el gráfico diferentes?
20. ¿Usaste alguno de los glosarios? ¿Cuáles?
21. ¿Fueron útiles los glosarios? ¿Por qué o por qué no?
22. ¿Hay palabras o términos de la pregunta que te gustaría que tuvieran glosarios? ¿Por qué o por qué no?
23. ¿Cómo escribiste su respuesta?
24. ¿Fue fácil o difícil de encontrar la manera de entrar en tu respuesta? ¿Por qué?
25. ¿Hubo algo confuso sobre la pregunta? Si es así, qué y por qué?
26. ¿Le cambiarías algo de esta pregunta? Si es así, qué, cómo, y por qué?
27. ¿Hay algo más sobre esta pregunta que le gustarías hablar?

Pregunta 21: Reporte Concurrente

- Cuando te doy el problema, quiero que lo leas en voz alta y me digas todo lo que estás pensando en como resolver el problema. Por ejemplo, lo que estás haciendo, por qué lo estás haciendo, lo que estás viendo, etc.
La base del triángulo ABC y la base del triángulo DEF yacen sobre la línea m, como se muestra en el diagrama.

La medida del \( \angle 4 \) es menor que la medida del \( \angle 8 \).

Para cada comparación, selecciona el signo (\(<\), \(>\), \(=\)) que hace que la relación entre la primera cantidad y la segunda cantidad sea verdadera.

The base of triangle ABC and the base of triangle DEF lie on line \( m \), as shown in the diagram.

The measure of \( \angle 4 \) is less than the measure of \( \angle 8 \).

For each comparison, select the symbol (\(<\), \(>\), \(=\)) that makes the relationship between the first quantity and the second quantity true.


**Pregunta 21: Reporte Retrospectivo**

- Ahora voy a hacerte algunas preguntas sobre el problema. Respondas lo mejor que puedas con tanto detalle como sea posible.

1. En una escala de 1 a 10, 1 es muy bajo y 10 es muy alta, evaluar el nivel de esfuerzo mental que utilizaste para resolver este problema.

2. ¿Puedes recordar los pasos a seguir para resolver el problema? Comiences con lo primero que pensaste y lo hiciste y continuar con el siguiente paso.
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

3. ¿Lo que es la pregunta está pidiendo que hagas?
4. ¿Cómo te usan las palabras para entender lo que la pregunta estaba pidiendo que hagas?
5. ¿Cómo te usan las palabras para resolver el problema?
6. ¿Cómo son los números utilizados en este problema?
7. ¿Cómo te usan los números para resolver el problema?
8. ¿Cómo son las palabras y números similares o relacionados?
9. ¿Cómo son las palabras y los numeros diferentes?
10. ¿Qué es este diagrama para?
11. ¿Cómo el diagrama se relacionan con la pregunta?
12. ¿Cómo te usa el diagrama para ayudarte a determinar cuál es la pregunta que estaba pidiendo a hacer?
13. ¿Usaste el diagrama para resolver el problema? Si es así, ¿cómo te usa el diagrama?
14. ¿Cómo son las palabras y el diagrama similares o relacionados?
15. ¿Cómo son las palabras y el diagrama diferentes?
16. ¿Cómo son los números y el diagrama similares o relacionados?
17. ¿Cómo son los números y el diagrama diferentes?
18. ¿Cómo son las palabras, los números, y el diagrama similares o relacionados?
19. ¿Cómo son las palabras, los números y el diagrama diferentes?
20. ¿Usaste alguno de los glosarios? ¿Cuáles?
21. ¿Fueron útiles los glosarios? ¿Por qué o por qué no?
22. ¿Hay palabras o términos de la pregunta que tuviera que tener glosarios? ¿Por qué o por qué no?
23. ¿Cómo escribiste su respuesta?
24. ¿Fue fácil o difícil de encontrar la manera de entrar en tu respuesta? ¿Por qué?
25. ¿Hubo algo confuso sobre la pregunta? Si es así, qué y por qué?
26. ¿Le cambiarías algo de esta pregunta? Si es así, qué, cómo, y por qué?
27. ¿Hay algo más sobre esta pregunta que le gustaría hablar?

Pregunta 25: Reporte Concurrente

- Cuando te doy el problema, quiero que lo leas en voz alta y me digas todo lo que estás pensando en como resolver el problema. Por ejemplo, lo que estás haciendo, por qué lo estás haciendo, lo que estás viendo, etc.
Pregunta 25: Reporte Retrospectivo

- Ahora voy a hacerte algunas preguntas sobre el problema. Respondas lo mejor que puedas con tanto detalle como sea posible.

1. En una escala de 1 a 10, 1 es muy bajo y 10 es muy alta, evaluar el nivel de esfuerzo mental que utilizaste para resolver este problema.
2. ¿Puedes recordar los pasos a seguir para resolver el problema? Comiences con lo primero que pensaste y lo hiciste y continuar con el siguiente paso.
3. ¿Lo que es la pregunta está pidiendo que hagas?
4. ¿Cómo te usan las palabras para entender lo que la pregunta estaba pidiendo que hagas?
5. ¿Cómo te usan las palabras para resolver el problema?
6. ¿Cómo son los números utilizados en este problema?
7. ¿Cómo te usan los números para resolver el problema?
8. ¿Cómo son las palabras y números similares o relacionados?
9. ¿Cómo son las palabras y los numeros diferentes?
10. ¿Qué es este diagrama para?
11. ¿Cómo el diagrama se relacionan con la pregunta?
12. ¿Cómo te usa el diagrama para ayudarte a determinar cuál es la pregunta que estaba pidiendo a hacer?
13. ¿Usaste el diagrama para resolver el problema? Si es así, ¿cómo te usa el diagrama?
14. ¿Cómo son las palabras y el diagrama similares o relacionados?
15. ¿Cómo son las palabras y el diagrama diferentes?
16. ¿Cómo son los números y el diagrama similares o relacionados?
17. ¿Cómo son los números y el diagrama diferentes?
18. ¿Cómo son las palabras, los números, y el diagrama similares o relacionados?
19. ¿Cómo son las palabras, los números y el diagrama diferentes?
20. ¿Usaste alguno de los glosarios? ¿Cuáles?
21. ¿Fueron útiles los glosarios? ¿Por qué o por qué no?
22. ¿Hay palabras o términos de la pregunta que te gustaría que tenían glosarios? ¿Por qué o por qué no?
23. ¿Cómo escribiste su respuesta?
24. ¿Fue fácil o difícil de encontrar la manera de entrar en tu respuesta? ¿Por qué?
25. ¿Hubo algo confuso sobre la pregunta? Si es así, qué y por qué?
26. ¿Le cambiarías algo de esta pregunta? Si es así, qué, cómo, y por qué?
27. ¿Hay algo más sobre esta pregunta que le gustaría hablar?

Seguimiento Preguntas

1. En una escala de 1 a 10, 1 es muy fácil y 10 es muy duro, evaluar la dificultad de estos problemas.
2. ¿Estás preparado para tomar el examen PARCC?
3. ¿Cómo ha tu profesor te preparado para el examen PARCC?
4. Prefieres tomar pruebas en las computadoras o en papel? ¿Por qué?
Appendix M
Summary of Selected 2015 SBAC Grade 8 Math Practice Test Items

Item 1 asks students to drag each number to its correct position on the number line. The response box is to the right of the text and contains a number line and three numbers with arrows underneath the numbers that point down. The number line ranges from zero to 0.85 and every .10 increment is labeled. The numbers are all fractions and include: \( \frac{\sqrt{2}}{5}, \frac{\pi}{5}, \frac{3}{10} \). Students can drag each number and arrow anywhere on the number line. This problem has three constituents—a word, phrase, or term—with glossaries. Two of the constituents have Spanish glossaries and the remaining constituent has an English glossary. Table M contains the glossed constituents and their Spanish and/or English glossaries for each item.

Item 5 asks students to finish a graph that represents the information in the text. The text describes a hypothetical character’s, Jason, trip to school. The text that is directly related to the different parts of the graph is listed using bullet points. Students are to use the Connect Line tool to finish the graph that is to the right of the text. The first segment of the graph is completed for the students. The x-axis represents time and is numbered zero to eight, while the y-axis represents the distance from Jason’s house and is numbered zero to 100. This item has 15 glossed constituents—two with English glossaries, 10 with Spanish glossaries, and three with both.

Item 9 contains a picture of a ladder leaning on a tree, which forms a right triangle. The text is above the picture and describes the different parts of the illustration. The length of the ladder and the distance from the bottom of the ladder to the base of the tree are given. The item asks students to find the unknown side, the distance between the
ground and the top of the ladder, and enter their answer in the response box. This item contains seven glossed constituents, one with an English glossary, five with Spanish glossaries, and one with both.

Item 24 contains an x/y table representing time in hours and water level in feet, respectively. Using the data in the table, Item 24 asks students to enter the rate of change of the water level in feet per hour. The response box is the same format as that in Item 9. Item 24 has six glossed constituents, four Spanish glossaries and two with both English and Spanish glossaries.
Table M

*Each Item’s Glossed Constituents and their Spanish and/or English Glossaries*

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Glossed Constituent</th>
<th>Spanish</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>drag each position</td>
<td>arrastra cada</td>
<td>location</td>
</tr>
<tr>
<td>5</td>
<td>most recent trip</td>
<td>más reciente recorrido</td>
<td>not long ago journey, travel somewhere in the direction of understood</td>
</tr>
<tr>
<td></td>
<td>toward realized left turned around same spent all the way twice original use Connect Line Tool finish accurately</td>
<td>hacia dejó regresó misma pasó, tardó todo el trayecto doble usa herramienta Conectar líneas usa correctamente</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>ladder leaning bottom base of the tree enter between top</td>
<td>escalera recostada, inclinada los pies, parte de abajo parte de arriba, punta de arriba</td>
<td>bending from a straight position the bottom of a large plant</td>
</tr>
<tr>
<td>24</td>
<td>level tank enter rate change per</td>
<td>nivel pacera, tanque escribe razón cambio por</td>
<td>how high something comes large container for liquid</td>
</tr>
</tbody>
</table>
Appendix N
Cognitive Interview Protocol: Version 2—English

Student Cognitive Interview Protocol—English

Phase 1: Introduction

• Hi my name is Chelsey and I am from the University of Colorado. I’m investigating how the new tests, like PARCC, use different resources, such as pictures, graphs, and computer features. I’m also looking at how these things affect your interpretation of and ability to solve math problems in order to improve these tests.

• We will start with a few background questions. Remember, that your answers will not affect your grade in class.

1. What’s your name?
2. How old are you?
3. What is your race/ethnicity?
4. Do you speak any other languages in addition to English?

Phase 2: Grade 8 Math Practice Test Selected Items

• We are now going to move to the 8th grade items. You will model your thinking for each question and then I will ask you questions. Please show all of your work on a scratch piece of paper.

• Once I give you the problem, I want you to read it out loud and tell me everything you are thinking about as you solve the problem. For example, what you’re doing, why you’re doing it, and what you’re looking at.

Sample Item 1: Concurrent Reporting

Sample Item 1: Retrospective Reporting

- Now I’m going to ask you some questions about the item. Answer as best you can with as much detail as you can.

1. On a scale of 1 to 10, 1 being extremely low and 10 being extremely high, rate the level of mental effort that you used to solve this problem. Why?
2. What is the question asking you to do? How do you know that?
3. Which words did you pay attention to or which words really helped you to figure out what you were supposed to do?
4. Describe the number line in as much detail as possible.
5. How do the three fractions relate to the number line? (Ask questions about number line if they don’t talk about intervals, labeling, arrows, etc.)
6. Why are there arrows below each fraction?
7. What does the square root mean?
8. What is pi or what does pi mean?
9. What is a fraction? What are the different parts of the fraction?
10. Was it easy or hard to figure out how to enter your answer? Why?
11. Did you use any of the glossaries? If so, which glossaries did you use?
12. Were the glossaries helpful or not? Why or why not?
13. Are there words or terms in the question that you wished had glossaries? Why or why not?
14. Was there anything confusing about the item? If so, what, and why?
15. Would you change anything about this item? If so, what, how, and why? For example, would you change the number line, the dragging feature, or the wording of the problem?
16. Would you use color somehow in the question? How and why?
17. Would you use bold words, italicized words, or box/circle any information? Which ones? Why?
18. Do you wish there was a place to show all of your work? Why or why not?
19. Is there anything else about this item that you would like to talk about?
Sample Item 5: Concurrent Reporting

The school is 100 meters from Jason’s house. The following describes his most recent trip:

- He walked 50 meters toward school in 2 minutes. He realized that he left a book at home.
- He turned around and walked home at the same speed.
- He spent 1 minute looking for his book.
- He walked all the way to school at twice his original speed.

Use the Connect Line tool to finish a graph that accurately represents Jason’s trip.


Sample Item 5: Retrospective Reporting

- Now I’m going to ask you some questions about the item. Answer as best you can with as much detail as you can.

1. On a scale of 1 to 10, 1 being extremely low and 10 being extremely high, rate the level of mental effort that you used to solve this problem. Why?
2. What is the question asking you to do? How do you know that?
3. Which words did you pay attention to or which words really helped you to figure out what you were supposed to do?
4. Describe the graph, before you answered the question, in as much detail as possible. (Ask questions about the axes, labels, line from x=0-2, arrows, why meters is spelled out and minutes is abbreviated, etc. if they don’t talk about them.)
5. Are there features in the graph that are also in the text? Why do you think this information was repeated?
6. Why are there bullet points?
7. Why do you think they gave you the first part of the graph from x=0 to x=2? Was this helpful? Why or why not?
8. Describe how your graph relates to each bullet point.
MATHEMATICS ASSESSMENT: SEMITOCs AND EMERGENT BILINGUALS

9. Why did they want you to create a graph and not a table? What kind of information does the graph give that a table wouldn’t? What kind of information would a table give that a graph wouldn’t?

10. Was it easy or hard to figure out how to enter your answer? Why?

11. Did you use any of the glossaries? If so, which glossaries did you use?

12. Were the glossaries helpful or not? Why or why not?

13. Are there words or terms in the question that you wished had glossaries? Why or why not?

14. Was there anything confusing about the item? If so, what, and why?

15. Would you change anything about this item? If so, what, how, and why? For example, would you change the graph, the Connect Line tool, or the wording of the problem?

16. Would you use color somehow in the question? How and why?

17. Would you use bold words, italicized words, or box/circle any information? Which ones? Why?

18. Is there anything else about this item that you would like to talk about?

Sample Item 9: Concurrent Reporting

*Figure N3.* Item 9 from the 2015 SBAC Grade 8 Math Practice Test. Source: *SBAC Grade 8 Math Practice Test. Taken from Smarter Balanced Assessment Consortium. (2014). Retrieved from http://sbac.portal.airast.org/practice-test/*.
Sample Item 9: Retrospective Reporting

- Now I’m going to ask you some questions about the item. Answer as best you can with as much detail as you can.

1. On a scale of 1 to 10, 1 being extremely low and 10 being extremely high, rate the level of mental effort that you used to solve this problem. Why?
2. What is the question asking you to do? How do you know that?
3. Which words did you pay attention to or which words really helped you to figure out what you were supposed to do?
4. Describe the picture in as much detail as possible. (Ask questions about the tree, shading, ladder, numbers, abbreviations, question mark, arrows, right angle sign, etc. if they don’t talk about them.)
5. Are there features in the picture that are also in the text? Why do you think this information was repeated?
6. Why does this question have a ladder and a tree instead of just a picture of a triangle? Which would you prefer? Is there any difference between the two? Which is more helpful? Why?
7. Which information in the picture did you use to solve the problem?
8. How did you know which equation to use?
9. Should there be a list of equations or should students know this equation? Why?
10. What do the variables, exponents, plus sign, and equal sign represent?
11. Do you know why $a^2 + b^2 = c^2$? If so, why? Can you draw a picture to explain it?
12. Was it easy or hard to figure out how to enter your answer? Why?
13. Do you know what all of the symbols mean in the space below the answer box? Which ones do you know and what do they mean? Which ones don’t you know? Why do they have all of those symbols? Do they need all of the symbols? Why or why not? Are they missing any symbols?
14. Did you use any of the glossaries? If so, which glossaries did you use?
15. Were the glossaries helpful or not? Why or why not?
16. Are there words or terms in the question that you wished had glossaries? Why or why not?
17. Was there anything confusing about the item? If so, what, and why?
18. Would you change anything about this item? If so, what, how, and why? For example, would you change the picture, answer box area/tools, or the wording of the problem?
19. Would you use color somehow in the question? How and why?
20. Would you use bold words, italicized words, or box/circle any information? Which ones? Why?
21. Do you wish there was a place to show all of your work? Why or why not?
22. Is there anything else about this item that you would like to talk about?
Sample Item 24: Concurrent Reporting


Sample Item 24: Retrospective Reporting

- Now I’m going to ask you some questions about the item. Answer as best you can with as much detail as you can.

1. On a scale of 1 to 10, 1 being extremely low and 10 being extremely high, rate the level of mental effort that you used to solve this problem. Why?
2. What is the question asking you to do? How do you know that?
3. Which words did you pay attention to or which words really helped you to figure out what you were supposed to do?
4. Describe the table in as much detail as possible. (Ask questions about the variables, rate of change, bold, abbreviations, etc. if they don’t talk about them.)
5. Are there features in the table that are also in the text? Why do you think this information was repeated?
6. Why does this question use water level in a tank instead of just numbers? Do you like or understand the idea of water level? Would you use a different situation or just an x/y table with numbers? Why?
7. Which information in the table did you use to solve the problem?
8. How did you know which equation to use?
9. Should there be a list of equations or should students know this equation? Why?
10. What do the variables, subscripts, minus sign, and equal sign represent?
11. How does the equation for slope relate to a linear graph?
12. What would the graph of this line look like?
13. How do the graph and table relate to each other?
14. Why did they use a table and not a graph? Which one do you prefer? Why?
15. Which part of a linear equation are they asking you to find? How do you know?
16. What is the equation of this line?
17. What is the relationship between a graph, table, and linear equation? Which one do you prefer and why?
18. Was it easy or hard to figure out how to enter your answer? Why?
19. Do you know what all of the symbols mean in the space below the answer box? Which ones do you know and what do they mean? Which ones don’t you know? Why do they have all of those symbols? Do they need all of the symbols? Why or why not? Are they missing any symbols?
20. If they used the calculator – Did you like the calculator tool? Why or why not? Was it easy to use? Why or why not? Do you prefer this calculator or the one on the PARCC test? Why?
21. Did you use any of the glossaries? If so, which glossaries did you use?
22. Were the glossaries helpful or not? Why or why not?
23. Are there words or terms in the question that you wished had glossaries? Why or why not?
24. Was there anything confusing about the item? If so, what, and why?
25. Would you change anything about this item? If so, what, how, and why? For example, would you change the table, answer box area/tools, or the wording of the problem?
26. Would you use color somehow in the question? How and why?
27. Would you use bold words, italicized words, or box/circle any information? Which ones? Why?
28. Do you wish there was a place to show all of your work? Why or why not?
29. Is there anything else about this item that you would like to talk about?

Phase 4: Follow-up Questions
1. On a scale of 1 to 10, 1 being extremely easy and 10 being extremely hard, rate the difficulty of these problems. Why?
2. Did you teacher prepare you for the PARCC test? If so, how? Was it helpful?
3. How do you think you did on the PARCC test? Why?
4. What did you like about the PARCC test?
5. What did you not like about the PARCC test?
6. Are there any similarities or differences between these problems and the problems on the PARCC test? If so, what are they?
7. Are these questions and the questions on PARCC similar or different to the types of questions you see in math class? Why?
8. Do you prefer taking tests on computers or on paper? Why?

- Thank you so much for participating in this interview!
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

Appendix O
Cognitive Interview Protocol: Version 2—Spanish

Student Cognitive Interview Protocol—Spanish

Phase 1: Introduction

- Hola mi nombre es Chelsey y estoy estudiando en la Universidad de Colorado. Estoy investigando cómo las nuevas pruebas, como PARCC, utilizan diferentes recursos, como imágenes, gráficos y características del sistema. Yo también estoy mirando cómo estas cosas afectan a su interpretación de y capacidad para resolver problemas de matemáticas con el fin de mejorar estas pruebas.
- Vamos a empezar con algunas preguntas de fondo. Recuerdes, que tus respuestas no afectarán tu grado en la clase.

5. ¿Cómo te llamas?
6. ¿Cuántos años tienes?
7. ¿Cuál es tu raza o etnicidad?
8. ¿Hablas otros idiomas además del Inglés?

Phase 2: Grade 8 Math Practice Test Selected Items

- Ahora vamos a pasar a las preguntas del octavo grado. Vas a modelar tu pensamiento para cada pregunta y luego voy a hacerte preguntas. Por favor muestra todo tu trabajo en un pedazo de papel rayado.
- Cuando te doy el problema, quiero que lo leas en voz alta y me digas todo lo que estás pensando/a como a resolver el problema. Por ejemplo, lo que estás haciendo, por qué lo estás haciendo, y lo que estás viendo.
Sample Item 1: Concurrent Reporting

![Diagram](image)

20. En una escala de 1 a 10, siendo 1 muy bajo y 10 es extremadamente alta, calificas el nivel de esfuerzo mental que usaste para resolver este problema. ¿Por qué?
21. ¿Cuál es la pregunta que te pide que hagas? ¿Cómo sabes eso?
22. ¿Qué palabras te prestas atención o cuales palabras te ayudaron a averiguar lo que se supone que hagas?
23. Describir la recta numérica con el mayor detalle posible.
24. ¿Cómo las tres fracciones se refieren a la recta numérica? (Haga preguntas sobre la recta numérica si no hablan de intervalos, el etiquetado, flechas, etc.)
25. ¿Por qué hay flechas debajo de cada fracción?
26. ¿Qué significa la raíz cuadrada?
27. ¿Qué es pi o ¿qué significa pi?
28. ¿Qué es una fracción? ¿Cuáles son las diferentes partes de una fracción?
29. ¿Fue fácil o difícil de encontrar la manera de escribir tu respuesta? ¿Por qué?
MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

30. ¿Hay palabras o términos de la pregunta que desearías tener glosarios o diccionarios? ¿Por qué o por qué no?
31. ¿Hubo algo confuso sobre la pregunta? Si es así, qué, y por qué?
32. ¿Cambiarías algo sobre esta pregunta? Si es así, qué, cómo, y por qué? Por ejemplo, ¿cambiarías la recta numérica, la función de arrastre, o la redacción del problema?
33. ¿Utilizarías el color de alguna manera en la pregunta? ¿Cómo y por qué?
34. ¿Utilizarías palabras en negrita, palabras en cursiva, o caja / círculo cualquier información? ¿Cuáles? ¿Por qué?
35. ¿Te gustaría que hubiera un lugar para mostrar todo tu trabajo? ¿Por qué o por qué no?
36. ¿Hay algo más sobre esta pregunta que te gustaría hablar?

Sample Item 5: Concurrent Reporting

La escuela está a 100 metros de la casa de Jason. En seguida se describe su recorrido más reciente:

- Él caminó 50 metros hacia la escuela en 2 minutos. Se dio cuenta de que dejó un libro en casa.
- Regresó y caminó a su casa a la misma velocidad.
- Pasó un minuto buscando su libro.
- Caminó todo el trayecto a la escuela al doble de su velocidad original.

Usa la herramienta Conectar líneas para completar una gráfica que represente correctamente el recorrido de Jason.

The school is 100 meters from Jason’s house. The following describes his most recent trip:

- He walked 50 meters toward school in 2 minutes. He realized that he left a book at home.
- He turned around and walked home at the same speed.
- He spent 1 minute looking for his book.
- He walked all the way to school at twice his original speed.

Use the Connect Line tool to finish a graph that accurately represents Jason’s trip.

Sample Item 5: Retrospective Reporting

- Ahora voy a hacerte algunas preguntas sobre la pregunta. Respondas lo mejor que puedas con tanto detalle como sea posible.

19. En una escala de 1 a 10, siendo 1 muy bajo y 10 es extremadamente alta, calificas el nivel de esfuerzo mental que usaste para resolver este problema. ¿Por qué?
20. ¿Cuál es la pregunta que te pide que hagas? ¿Cómo sabes eso?
21. ¿Qué palabras te prestas atención o cuales palabras te ayudaron a averiguar lo que se supone que hagas?
22. Describes la gráfica, antes de responder a la pregunta, con el mayor detalle posible. (Haga preguntas acerca de los ejes, las etiquetas, la línea de x = 0-2, flechas, ¿por metros se detallan y minuto se abrevian, etc. si no se habla de ellos.)
23. ¿Existen características de la gráfica que también están en el texto? ¿Por qué crees que esta información fue reiterada?
24. ¿Por qué hay puntos de bala?
25. ¿Por qué crees que te dieron la primera parte de la gráfica de x = 0 y x = 2? ¿Fue útil esta ¿Por qué o por qué no?
26. Describas cómo tu gráfico se refiere a cada punto de bala.
27. ¿Por qué ellos quieren que tú creas un gráfico y no una tabla? ¿Qué tipo de información te da el gráfico que no te da una tabla? ¿Qué tipo de información te da la table que no te da el gráfico?
28. ¿Fue fácil o difícil de encontrar la manera de escribir tu respuesta? ¿Por qué?
29. ¿Hay palabras o términos de la pregunta que deseabas tener glosarios o diccionarios? ¿Por qué o por qué no?
30. ¿Hubo algo confuso sobre la pregunta? Si es así, qué, y por qué?
31. ¿Cambiarías algo sobre esta pregunta? Si es así, qué, cómo, y por qué? Por ejemplo, ¿cambiarías la gráfica, la herramienta conectar líneas, o la redacción del problema?
32. ¿Utilizarías el color de alguna manera en la pregunta? ¿Cómo y por qué?
33. ¿Utilizarías palabras en negrita, palabras en cursiva, o caja / círculo cualquier información? ¿Cuáles? ¿Por qué?
34. ¿Hay algo más sobre esta pregunta que te gustaría hablar?
Sample Item 9: Concurrent Reporting

Ahora voy a hacerte algunas preguntas sobre la pregunta. Respondas lo mejor que puedas con tanto detalle como sea posible.

23. En una escala de 1 a 10, siendo 1 muy bajo y 10 es extremadamente alta, calificas el nivel de esfuerzo mental que usaste para resolver este problema. ¿Por qué?
24. ¿Cuál es la pregunta que te pide que hagas? ¿Cómo sabes eso?
25. ¿Qué palabras te prestas atención o cuales palabras te ayudaron a averiguar lo que se supone que hagas?
26. Describe la imagen con el mayor detalle posible. (Haga preguntas sobre el árbol, sombreado, escalera, números, abreviaturas, signo de interrogación, flechas, señal de ángulo recto, etc. si no se habla de ellos.)
27. ¿Existen características en la imagen que también se encuentran en el texto? ¿Por qué crees que esta información fue reiterada?
28. ¿Por qué esta pregunta tiene una escalera y un árbol en lugar de sólo una foto de un triángulo? ¿Qué prefieres? ¿Hay alguna diferencia entre los dos? ¿Qué es más útil? ¿Por qué?
29. ¿Qué información en la imagen usaste para resolver el problema?
30. ¿Cómo sabías cual ecuación a utilizar?
31. ¿Debería haber una lista de ecuaciones o deberían saber los estudiantes esta ecuación? ¿Por qué?
32. ¿Qué significan las variables, los exponentes, el signo más, y el signo igual?
33. ¿Sabes por qué $a^2 + b^2 = c^2$? Si es así, ¿por qué? ¿Puedes hacer un dibujo para explicar?
34. ¿Fue fácil o difícil de encontrar la manera de escribir tu respuesta? ¿Por qué?
35. ¿Sabes lo que todos los símbolos significan en el espacio debajo del cuadro de respuesta? ¿Cuáles sabes y que significan? ¿Cuáles no lo sabes? ¿Por qué tienen todos esos símbolos? ¿Necesitan todos los símbolos? ¿Por qué o por qué no? ¿Hay símbolos faltan?
36. ¿Hay palabras o términos de la pregunta que desearas tener glosarios o diccionarios? ¿Por qué o por qué no?
37. ¿Hubo algo confuso sobre la pregunta? Si es así, qué, y por qué?
38. ¿Cambiarías algo sobre esta pregunta? Si es así, qué, cómo, y por qué? Por ejemplo, ¿cambiaria la imagen, el área de la caja respuesta / herramientas, o la redacción del problema?
39. ¿Utilizarías el color de alguna manera en la pregunta? ¿Cómo y por qué?
40. ¿Utilizarías palabras en negrita, palabras en cursiva, o caja / círculo cualquier información? ¿Cuáles? ¿Por qué?
41. ¿Te gustaría que hubiera un lugar para mostrar todo tu trabajo? ¿Por qué o por qué no?
42. ¿Hay algo más sobre esta pregunta que te gustaría hablar?
Sample Item 24: Concurrent Reporting

![Image of a table showing the linear relationship of water level to time](http://sbac.portal.airast.org/practice-test/)

*Esta tabla muestra la relación lineal del nivel de agua en un tanque y el tiempo.*

<table>
<thead>
<tr>
<th>Tiempo (hr)</th>
<th>Nivel de agua (pies)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
</tr>
</tbody>
</table>

Escribe la tasa de cambio del nivel de agua, en pies por hora.

*This table shows the linear relationship of the water level in a tank and time.*

<table>
<thead>
<tr>
<th>Time (hr)</th>
<th>Water Level (ft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
</tr>
</tbody>
</table>

*En la escala de 1 a 10, siendo 1 muy bajo y 10 extremadamente alta, califica el nivel de esfuerzo mental que usaste para resolver este problema. ¿Por qué?*

30. En una escala de 1 a 10, siendo 1 muy bajo y 10 extremadamente alta, califica el nivel de esfuerzo mental que usaste para resolver este problema. ¿Por qué?

31. ¿Cuál es la pregunta que te pide que hagas? ¿Cómo sabes eso?

32. ¿Qué palabras te prestaste atención o cuales palabras te ayudaron a averiguar lo que se supone que hagas?

33. Describe la tabla con el mayor detalle posible. (Haga preguntas acerca de las variables, la tasa de cambio, negrita, abreviaturas, etc. si no se habla de ellos.)

34. ¿Existen características en la tabla que también están en el texto? ¿Por qué crees que esta información fue reiterada?

35. ¿Por qué usa el nivel de agua en un tanque en lugar de sólo los números? ¿Te gusta o entiendes la idea del nivel de agua? ¿Utilizarías una situación diferente o simplemente una tabla con los números? ¿Por qué?

36. ¿Qué información de la tabla usaste para resolver el problema?

37. ¿Cómo sabías cual ecuación a utilizar?

38. ¿Debería haber una lista de ecuaciones o deberían saber los estudiantes esta ecuación? ¿Por qué?

39. ¿Qué significan las variables, subíndices, signo menos, y signo igual?

40. ¿Qué es la relación entre una ecuación y un gráfico?

---


Sample Item 24: Retrospective Reporting

- Ahora voy a hacerte algunas preguntas sobre la pregunta. Respondas lo mejor que puedas con tanto detalle como sea posible.

30. En una escala de 1 a 10, siendo 1 muy bajo y 10 es extremadamente alta, califica el nivel de esfuerzo mental que usaste para resolver este problema. ¿Por qué?

31. ¿Cuál es la pregunta que te pide que hagas? ¿Cómo sabes eso?

32. ¿Qué palabras te prestaste atención o cuales palabras te ayudaron a averiguar lo que se supone que hagas?

33. Describe la tabla con el mayor detalle posible. (Haga preguntas acerca de las variables, la tasa de cambio, negrita, abreviaturas, etc. si no se habla de ellos.)

34. ¿Existen características en la tabla que también están en el texto? ¿Por qué crees que esta información fue reiterada?

35. ¿Por qué usa el nivel de agua en un tanque en lugar de sólo los números? ¿Te gusta o entiendes la idea del nivel de agua? ¿Utilizarías una situación diferente o simplemente una tabla con los números? ¿Por qué?

36. ¿Qué información de la tabla usaste para resolver el problema?

37. ¿Cómo sabías cual ecuación a utilizar?

38. ¿Debería haber una lista de ecuaciones o deberían saber los estudiantes esta ecuación? ¿Por qué?

39. ¿Qué significan las variables, subíndices, signo menos, y signo igual?

40. ¿Qué es la relación entre una ecuación y un gráfico?
MATHEMATICS ASSESSMENT: SEMITODS AND EMERGENT BILINGUALS

41. Cómo parece la gráfica de esta línea?
42. ¿Qué es la relación entre un gráfico y una tabla?
43. ¿Por qué se usa una tabla y no un gráfico? ¿Cuál te gusta más? ¿Por qué?
44. ¿Qué parte de una ecuación lineal se te pide que encontrar? Cómo lo sabes?
45. ¿Qué es la ecuación de esta línea?
46. ¿Qué es la relación entre una ecuación, un gráfico, y una tabla? ¿Cuál te gusta más y por qué?
47. ¿Fue fácil o difícil de encontrar la manera de escribir tu respuesta? ¿Por qué?
48. ¿Sabes lo que todos los símbolos significan en el espacio debajo del cuadro de respuesta? ¿Cuáles sabes y que significan? ¿Cuáles no lo sabes? ¿Por qué tienen todos esos símbolos? ¿Necesitan todos los símbolos? ¿Por qué o por qué no? ¿Qué símbolos faltan?
49. Si se utilizan la calculadora ¿Te gusta la herramienta calculadora? ¿Por qué o por qué no? ¿Fue fácil de usar? ¿Por qué o por qué no? ¿Prefieres esta calculadora o la de la prueba PARCC? ¿Por qué?
50. ¿Hay palabras o términos de la pregunta que deseesas tenido glosarios o diccionarios? ¿Por qué o por qué no?
51. ¿Hubo algo confuso sobre la pregunta? Si es así, qué, y por qué?
52. ¿Cambiarías algo sobre esta pregunta? Si es así, qué, cómo, y por qué? Por ejemplo, ¿cambiarías la tabla, el área de la caja respuesta / herramientas, o la redacción del problema?
53. ¿Utilizarias el color de alguna manera en la pregunta? ¿Cómo y por qué?
54. ¿Utilizarias palabras en negrita, palabras en cursiva, o caja / círculo cualquier información? ¿Cuáles? ¿Por qué?
55. ¿Te gustaría que hubiera un lugar para mostrar todo tu trabajo? ¿Por qué o por qué no?
56. ¿Hay algo más sobre esta pregunta que te gustaría hablar?

Phase 4: Follow-up Questions

9. En una escala de 1 a 10, 1 es muy fácil y 10 es muy difícil, valora la dificultad de estos problemas. ¿Por qué?
10. ¿Te preparó tu maestro para la prueba PARCC? Si es así, ¿cómo? ¿Fue útil?
11. ¿Cómo crees que lo hiciste en la prueba PARCC? ¿Por qué?
12. ¿Qué te gustó de la prueba PARCC?
13. ¿Qué no te gusta de la prueba PARCC?
14. ¿Hay similitudes o diferencias entre estos problemas y los problemas en la prueba PARCC? Si es así, ¿cuáles son?
15. ¿Son estas preguntas y las preguntas en la prueba de PARCC similares o diferentes a los tipos de preguntas que se ven en la clase de matemáticas? ¿Por qué?
16. ¿Prefieres tomar pruebas en las computadoras o en papel? ¿Por qué?

• Muchas gracias por participar en esta entrevista
### Appendix P

#### Coding Framework for Interview Analysis

**Table P**

**Interview Code Book**

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic</td>
<td>Components related to the intrinsic cognitive load – determined by the nature of the learning task, the number of elements that must be processed simultaneously, and/or the interaction between the complexity of the task and the learner’s level of expertise.</td>
</tr>
<tr>
<td>Item rating</td>
<td>Students rating of the individual item (1-10).</td>
</tr>
<tr>
<td>Germane</td>
<td>Components related to the germane cognitive load – determined by the amount of resources that learners invest in the construction of schemata and automation.</td>
</tr>
<tr>
<td>All rating</td>
<td>Students rating of all four problems (1-10).</td>
</tr>
<tr>
<td>Key SR</td>
<td>Semiotic resource (SR) identified as being important.</td>
</tr>
<tr>
<td>NL</td>
<td>Item NL identified as being important.</td>
</tr>
<tr>
<td>MSE</td>
<td>Item MSE identified as being important.</td>
</tr>
<tr>
<td>VR</td>
<td>Item VR identified as being important.</td>
</tr>
<tr>
<td>Emphasis</td>
<td>What form of emphasis students think should be used in the item.</td>
</tr>
<tr>
<td><strong>Bold</strong></td>
<td>(Self-explaining)</td>
</tr>
<tr>
<td><strong>Underline</strong></td>
<td>(Self-explaining)</td>
</tr>
<tr>
<td><strong>Color</strong></td>
<td>(Self-explaining)</td>
</tr>
<tr>
<td><strong>Italicize</strong></td>
<td>(Self-explaining)</td>
</tr>
<tr>
<td><strong>Large font size</strong></td>
<td>(Self-explaining)</td>
</tr>
<tr>
<td><strong>Not specified</strong></td>
<td>Student says that key information should be emphasized, but does not specify how.</td>
</tr>
<tr>
<td>Tools</td>
<td>Tools students think should be provided to help solve the problem.</td>
</tr>
<tr>
<td><strong>List of equations</strong></td>
<td>(Self-explaining)</td>
</tr>
<tr>
<td>Extraneous</td>
<td>Components related to the extraneous cognitive load – determined by the presentation of the task and extraneous features.</td>
</tr>
<tr>
<td>MSE answer box</td>
<td>Answer box contains irrelevant MSE.</td>
</tr>
<tr>
<td>Real-life example</td>
<td>Real-life example/context (word problem) used to confuse students and make them detangle math elements out of it to solve.</td>
</tr>
<tr>
<td>Decorative</td>
<td>Features are only there for decorative purposes.</td>
</tr>
<tr>
<td>NL – Ideational</td>
<td>The purpose of this metafunction is to interpret our external and internal experiences with the world and to describe the purpose and/or context for the use of NL.</td>
</tr>
<tr>
<td>Purpose</td>
<td>Features that relate to the intended purpose of the NL.</td>
</tr>
<tr>
<td><strong>Directions</strong></td>
<td>NL is used to give an order.</td>
</tr>
<tr>
<td><strong>Clarity</strong></td>
<td>NL is used to make the information clear and/or more organized.</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Visualize</strong></td>
<td>NL is used to help students visualize the problem/situation.</td>
</tr>
<tr>
<td><strong>Think more</strong></td>
<td>NL is used to make students think more.</td>
</tr>
<tr>
<td><strong>Textual cohesions</strong></td>
<td>How meaning is created across NL.</td>
</tr>
<tr>
<td><strong>Repetition</strong></td>
<td>Meaning is repeated across NL.</td>
</tr>
<tr>
<td><strong>NL – Interpersonal</strong></td>
<td>The purpose of this metafunction is to describe the role relations between the text (i.e., NL) and reader.</td>
</tr>
<tr>
<td><strong>Directive</strong></td>
<td>The NL gives an official or authoritative instruction.</td>
</tr>
<tr>
<td><strong>Factual</strong></td>
<td>The NL provides facts to the reader.</td>
</tr>
<tr>
<td><strong>Helpful</strong></td>
<td>The NL helps the student solve the problem.</td>
</tr>
<tr>
<td><strong>NL – Textual</strong></td>
<td>The purpose of this metafunction is to describe the structural rituals of the NL.</td>
</tr>
<tr>
<td><strong>Emphasis</strong></td>
<td>Special font style of text in an illustration to accentuate the importance of something.</td>
</tr>
<tr>
<td><strong>Bold</strong></td>
<td>(Self-explaining)</td>
</tr>
<tr>
<td><strong>Unit</strong></td>
<td>The type of text.</td>
</tr>
<tr>
<td><strong>Abbreviation</strong></td>
<td>A shortened form of a word or phrase.</td>
</tr>
<tr>
<td><strong>Word</strong></td>
<td>(Self-explaining)</td>
</tr>
<tr>
<td><strong>Phrase</strong></td>
<td>A word or a group of 2 or more words without a verb.</td>
</tr>
<tr>
<td><strong>MSE – Ideational</strong></td>
<td>The purpose of this metafunction is to interpret our external and internal experiences with the world and to describe the purpose and/or context for the use of MSE.</td>
</tr>
<tr>
<td><strong>Purpose</strong></td>
<td>Features that relate to the intended purpose of the MSE.</td>
</tr>
<tr>
<td><strong>Discrete numbers on graph</strong></td>
<td>To represent data/numbers that can be counted and are on a graph.</td>
</tr>
<tr>
<td><strong>Identity</strong></td>
<td>MSE used to create a relational process.</td>
</tr>
<tr>
<td><strong>Textual cohesion</strong></td>
<td>How meaning is created across MSE.</td>
</tr>
<tr>
<td><strong>Identity</strong></td>
<td>Meaning is repeated through relational processes.</td>
</tr>
<tr>
<td><strong>Repetition</strong></td>
<td>Meaning is repeated across MSE.</td>
</tr>
<tr>
<td><strong>Synonym</strong></td>
<td>Meaning is similar across MSE.</td>
</tr>
<tr>
<td><strong>MSE – Textual</strong></td>
<td>The purpose of this metafunction is to describe the structural rituals of the MSE.</td>
</tr>
<tr>
<td><strong>Unit</strong></td>
<td>The type of MSE used.</td>
</tr>
<tr>
<td><strong>Arabic numerals/values</strong></td>
<td>Any of the numerals 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 (i.e., constant).</td>
</tr>
<tr>
<td><strong>VR – Ideational</strong></td>
<td>The purpose of this metafunction is to interpret our external and internal experiences with the world and to describe the purpose and/or context for the use of the VR.</td>
</tr>
<tr>
<td><strong>Purpose</strong></td>
<td>Features that relate to the intended purpose of the VR.</td>
</tr>
<tr>
<td><strong>Infinity</strong></td>
<td>To visually represent infinity.</td>
</tr>
<tr>
<td><strong>Identity</strong></td>
<td>To visually represent an identity or relational process.</td>
</tr>
<tr>
<td><strong>Discrete numbers</strong></td>
<td>To visually represent a discrete number.</td>
</tr>
<tr>
<td><strong>Direction</strong></td>
<td>To visually represent direction.</td>
</tr>
<tr>
<td><strong>Inequality</strong></td>
<td>To visually represent an inequality.</td>
</tr>
<tr>
<td><strong>Measurement</strong></td>
<td>To visually represent a measurement (known or unknown).</td>
</tr>
<tr>
<td><strong>Function</strong></td>
<td>To visually represent a function.</td>
</tr>
<tr>
<td><strong>Categories</strong></td>
<td>To visually represent categories.</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td><strong>Decorative</strong></td>
<td>To decorate the item.</td>
</tr>
<tr>
<td><strong>Visualize</strong></td>
<td>To help students visualize the problem/context.</td>
</tr>
<tr>
<td><strong>Easy to look at</strong></td>
<td>VR only used because it is easy to look at (easier than a graph).</td>
</tr>
</tbody>
</table>

**VR – Interpersonal**
The purpose of this metafunction is to describe the role relations between the text (i.e., NL) and reader.

**Directive**
The VR gives an official or authoritative instruction.

**Factual**
The VR provides facts to the reader.

**Helpful**
The VR helps the student solve the problem.

**VR – Textual**
The purpose of this metafunction is to describe the structural rituals of the VR.

**Schematic type**
The different categories of schematic visuals.

**Number line**
The visual is or contains a line on which numbers are marked at intervals, often used to illustrate simple numerical operations.

**Categorical table**
The visual is or contains a table that represents types of data into different groups. Two-way tables are frequently used to present categorical data by counting the number of observations that fall into each group for two variables, one divided into rows and the other into columns.

**Line/Segment**
The visual is or contains a line, which is the geometric figure formed by two points. A line is the straight path connecting 2 points and extending beyond the points in both directions. A line segment includes all points between 2 given points.

**Axes**
The visual contains a fixed reference line for the measurement of coordinates.

**X-axis**
The visual contains the horizontal axis of a system of coordinates, points along which have a value of zero for all other coordinates.

**Y-axis**
The visual contains the vertical axis of a system of coordinates, points along which have a value of zero for all other coordinates.

**Intervals**
The visual contains intervals, which are the distances between 1 number and the next on the scale of a graph.

**Equal**
The visual contains intervals that are equal distances between all numbers on the scale of the graph.

**Context**
The circumstances that form the setting.

**Unfamiliar**
Students are not familiar with the context.

**Intersemiosis**
Creating meaning across SR (for each pair of SR).

**Repetition**
Same relations.

**Identity**
Meaning shows relational processes.

**Synonymy**
Similar relations.

**Metonymy**
Part-whole relations.

**Collocation**
Expectancy relations.

**Antonymy**
Opposite relations.

**Interpretation**
Components related to students’ interpretation of the item.
<table>
<thead>
<tr>
<th>Interpretation</th>
<th>Students’ interpretation of the item.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>Students interpreted the item correctly.</td>
</tr>
<tr>
<td>Correct with help</td>
<td>Students interpreted the item correctly with help.</td>
</tr>
<tr>
<td>Incorrect</td>
<td>Students interpreted the item incorrectly.</td>
</tr>
<tr>
<td>Process</td>
<td>Components related to the process of interpreting the item.</td>
</tr>
<tr>
<td>Struggles with word(s)</td>
<td>Students struggle saying/pronouncing the word, or reads it incorrectly.</td>
</tr>
<tr>
<td>Asks clarifying questions</td>
<td>Students ask clarifying questions about the task, what they are supposed to do.</td>
</tr>
<tr>
<td>Statements</td>
<td>Statements students made about the interpretation process.</td>
</tr>
<tr>
<td>Confusing</td>
<td>Students claim that the item is confusing (directions).</td>
</tr>
<tr>
<td>NL clear</td>
<td>Students claim that the NL is not clear.</td>
</tr>
<tr>
<td>NL unnecessary</td>
<td>Students claim that the use of NL is unnecessary.</td>
</tr>
<tr>
<td>VR unnecessary</td>
<td>Students claim that the use of VR is unnecessary.</td>
</tr>
<tr>
<td>Problem solving</td>
<td>Components related to the problem-solving process.</td>
</tr>
<tr>
<td>Schema</td>
<td>Students schematic recall.</td>
</tr>
<tr>
<td>Correct</td>
<td>Students use the correct schema.</td>
</tr>
<tr>
<td>Incorrect</td>
<td>Students use the incorrect schema.</td>
</tr>
<tr>
<td>Process</td>
<td>The process of problem-solving.</td>
</tr>
<tr>
<td>Logical estimation</td>
<td>Students use logical estimation (not correct answer, but correct reasoning).</td>
</tr>
<tr>
<td>Math error</td>
<td>Students make a math error (e.g., simplify incorrectly).</td>
</tr>
<tr>
<td>Math question</td>
<td>Students ask a question relating to the math of the item.</td>
</tr>
<tr>
<td>Hint/question</td>
<td>Students receive a hint in the form of a phrase or question from myself.</td>
</tr>
<tr>
<td>I don’t know/Guess</td>
<td>Students say they don’t know or that they are guessing.</td>
</tr>
<tr>
<td>Difficulty in English</td>
<td>Students have a hard time explaining their thoughts in English (e.g., How do I say…?)</td>
</tr>
<tr>
<td>Checks answer</td>
<td>Student makes sure their answer is correct/makes sense or goes back to answer to correct.</td>
</tr>
<tr>
<td>Semiotic resources</td>
<td>SR used to solve the problem (different from responding).</td>
</tr>
<tr>
<td>NL – verbal</td>
<td>Students solve the problem verbally.</td>
</tr>
<tr>
<td>NL – paper</td>
<td>Students solve the problem using NL on paper.</td>
</tr>
<tr>
<td>MSE – paper</td>
<td>Students solve the problem using MSE on paper.</td>
</tr>
<tr>
<td>Schematic – paper</td>
<td>Students solve the problem using a schematic VR on paper.</td>
</tr>
<tr>
<td>Concrete – paper</td>
<td>Students solve the problem using a concrete VR on paper.</td>
</tr>
<tr>
<td>Hands</td>
<td>Students solve the problem using hand-gestures (e.g., using fingers to count or think about slope).</td>
</tr>
<tr>
<td>Responding</td>
<td>Components related to responding to the item.</td>
</tr>
<tr>
<td>Representation</td>
<td>Representation of students’ answers.</td>
</tr>
<tr>
<td>Does</td>
<td>Students’ answers do represent their thinking.</td>
</tr>
<tr>
<td>Does not</td>
<td>Students’ answers do not represent their thinking.</td>
</tr>
<tr>
<td>Response tool</td>
<td>Components related to the response tool.</td>
</tr>
<tr>
<td>Easy</td>
<td>Students say the response tool was easy to use.</td>
</tr>
<tr>
<td>Hard</td>
<td>Students say the response tool was difficult to use.</td>
</tr>
<tr>
<td>Unfamiliar with parts</td>
<td>Students are unfamiliar with some of the components of the response tools.</td>
</tr>
</tbody>
</table>
### MATHEMATICS ASSESSMENT: SEMITOCS AND EMERGENT BILINGUALS

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Asks questions</strong></td>
<td>Students ask questions about some of the components of the response tools.</td>
</tr>
<tr>
<td><strong>Help</strong></td>
<td>Students receive help (directions or modeling) from me about how to use the response tool.</td>
</tr>
<tr>
<td><strong>Calculator</strong></td>
<td>Students think the response tool (i.e., text box) is a calculator.</td>
</tr>
<tr>
<td><strong>Suggestions</strong></td>
<td>Student suggestions pertaining to the response tool.</td>
</tr>
<tr>
<td><strong>Add NL</strong></td>
<td>NL should be added to the response tool (e.g., units).</td>
</tr>
<tr>
<td><strong>Add MSE</strong></td>
<td>MSE should be added to the response tool.</td>
</tr>
<tr>
<td><strong>Remove MSE</strong></td>
<td>MSE should be removed from the response tool.</td>
</tr>
<tr>
<td><strong>Item changes</strong></td>
<td>Students’ suggestions pertaining to the item as a whole.</td>
</tr>
<tr>
<td><strong>Add NL</strong></td>
<td>Add NL to make the problem clearer.</td>
</tr>
<tr>
<td><strong>NL clearer</strong></td>
<td>Modify the NL to make it clearer.</td>
</tr>
<tr>
<td><strong>NL location</strong></td>
<td>Change the location of the NL.</td>
</tr>
<tr>
<td><strong>Replace NL with MSE</strong></td>
<td>Replace NL with MSE.</td>
</tr>
<tr>
<td><strong>Add schematic</strong></td>
<td>Add a schematic VR.</td>
</tr>
<tr>
<td><strong>Add concrete</strong></td>
<td>Add a concrete VR.</td>
</tr>
<tr>
<td><strong>Remove concrete</strong></td>
<td>Remove concrete VR.</td>
</tr>
<tr>
<td><strong>Add color</strong></td>
<td>Add color.</td>
</tr>
<tr>
<td><strong>Change response tool</strong></td>
<td>Change/modify the response tool.</td>
</tr>
<tr>
<td><strong>Extra tools</strong></td>
<td>Components related to the extra tools (tools that are available but not necessary).</td>
</tr>
<tr>
<td><strong>Calculator</strong></td>
<td>Components related to the calculator.</td>
</tr>
<tr>
<td><strong>Likes</strong></td>
<td>Students like the calculator.</td>
</tr>
<tr>
<td><strong>Hard</strong></td>
<td>Students think the calculator is hard to use.</td>
</tr>
<tr>
<td><strong>Modify</strong></td>
<td>Students think the calculator should be changed/modified.</td>
</tr>
<tr>
<td><strong>Prefers TI</strong></td>
<td>Students prefer the TI-83/TI-84 calculators that they use in class and that was used on PARCC.</td>
</tr>
<tr>
<td><strong>Asked questions</strong></td>
<td>Students asked questions about how to use the calculator.</td>
</tr>
<tr>
<td><strong>Help</strong></td>
<td>Students receive help (directions/modeling) from myself for the calculator.</td>
</tr>
<tr>
<td><strong>Glossary</strong></td>
<td>Components related to the glossary.</td>
</tr>
<tr>
<td><strong>Used</strong></td>
<td>Students used the glossary.</td>
</tr>
<tr>
<td><strong>Made sense</strong></td>
<td>The glosses made sense to the students (didn’t have to use, just looked at).</td>
</tr>
<tr>
<td><strong>Helpful</strong></td>
<td>Students thought the gloss was helpful.</td>
</tr>
<tr>
<td><strong>More detail</strong></td>
<td>Students though the gloss needed more detail.</td>
</tr>
<tr>
<td><strong>Dialect</strong></td>
<td>Students did not like the dialect of the gloss.</td>
</tr>
<tr>
<td><strong>Add non-essential</strong></td>
<td>Students wanted to add glosses for non-essential vocabulary (e.g., contextual terms or terms related to tools).</td>
</tr>
<tr>
<td><strong>Add essential</strong></td>
<td>Students wanted to add glosses for essential vocabulary (e.g., define linear).</td>
</tr>
<tr>
<td><strong>Add VR</strong></td>
<td>Students thought the gloss should have a VR (e.g., video).</td>
</tr>
<tr>
<td><strong>Did not help</strong></td>
<td>Students claim that the gloss did not help them understand the item.</td>
</tr>
</tbody>
</table>
Appendix Q
Example of Interview Coding

I will model the coding process by showing at least one example of a variable from most of the categories (Item 9 does not contain codes for all the categories) using data from the audio recordings, transcriptions, and artifacts from Item 9 of the practice test (Figure Q). Each variable was coded dichotomously as present (1) or absent (0).


Table Q contains a selection of the variables from most of the categories and examples from the data. Students thought that the ideational purposes of the NL, MSE,
and VR used in this item was to make the problem harder, to show that it was a right triangle, and to give you the important information, respectively. Textually, students identified the values given in the problem (13 and 5) and were unfamiliar with the type of tree that was in the VR. Many of the students stated that the picture repeated all the information that was in the text (i.e., intersemiotic Repetition between NL and VR). The student in the item rating example gave the problem a three because she was familiar with the Pythagorean Theorem and used it a lot. One of the students said that he would underline the key information. Many students stated that all the symbols in the answer box were confusing because they did know how they were supposed to write their answers. In the interpretation stage, one of the students claimed that the instructions were helpful, but not needed because the VR contained all the essential information. The examples for problem-solving and responding demonstrate a mathematical error made by one of the students and how their final answer did not match their thinking (or how they wanted to write the answer), respectively. One of the students suggested that the directions should be moved from the bottom to the top of the VR so that it is located with the rest of the text. Lastly, many students stated that they preferred the calculator that was given to them on PARCC (TI-83) because that is the calculator they use in class so they are more familiar with it.
### Coding Examples for Item 9

<table>
<thead>
<tr>
<th>Code</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. NL – Ideational metafunction</td>
<td>The word problem makes it harder.</td>
</tr>
<tr>
<td>1.1. Text purpose: 5. think more</td>
<td></td>
</tr>
<tr>
<td>3. MSE – Ideational metafunction</td>
<td>The square means it’s a right angle so you can use the theorem.</td>
</tr>
<tr>
<td>3.1. MSE purpose: 2. identity</td>
<td></td>
</tr>
<tr>
<td>4. MSE – Textual metafunction</td>
<td>It shows the tree and it has the ladder leaning up against it with the 13 feet...</td>
</tr>
<tr>
<td>4.1. MSE unit: 1. Arabic numerals/values</td>
<td></td>
</tr>
<tr>
<td>5. VR – Ideational metafunction</td>
<td>The picture gives you all of the important information.</td>
</tr>
<tr>
<td>5.1. VR purpose: 12. important information</td>
<td></td>
</tr>
<tr>
<td>6. VR – Textual metafunction</td>
<td>There’s a really weird tree and a ladder leaning towards it.</td>
</tr>
<tr>
<td>6.4. Context: 1. unfamiliar</td>
<td></td>
</tr>
<tr>
<td>7. Intersemiosis (for each pair of semiotic resource)</td>
<td>The text tells you everything that’s in the picture.</td>
</tr>
<tr>
<td>7.1 Intersemiotic relations: 1. Repetition (NL/VR)</td>
<td></td>
</tr>
<tr>
<td>8. Intrinsic cognitive load</td>
<td>Three because it was easy, I remember it, and I use it a lot.</td>
</tr>
<tr>
<td>8.1. Item rating: 1. item rating</td>
<td></td>
</tr>
<tr>
<td>9. Germane cognitive load</td>
<td>I would underline 13 feet and the ‘distance of 5 feet from the base’.</td>
</tr>
<tr>
<td>9.3. Emphasis: 2. underline</td>
<td></td>
</tr>
<tr>
<td>10. Extraneous cognitive load</td>
<td>All of the options in the text box are confusing.</td>
</tr>
<tr>
<td>10.1. Extraneous features: 1. MSE answer box</td>
<td></td>
</tr>
<tr>
<td>11. Interpretation</td>
<td>The instructions were helpful, but you could tell what it wanted with just the picture.</td>
</tr>
<tr>
<td>11.3. Student statements: 2. NL clear, 3. NL unnecessary</td>
<td></td>
</tr>
<tr>
<td>12. Problem-Solving</td>
<td>So, you’re trying to find b so you would do 5 squared plus b squared equals 13 squared. Which is um 10 plus b squared equals 169.</td>
</tr>
<tr>
<td>9.2. Process: 2. math error</td>
<td></td>
</tr>
<tr>
<td>13. Responding</td>
<td>So b equals 12 feet (written as ‘b = 12 ft’ on his paper). Do I just type it? Where’s b at? Just put equals 12? Oh, just put 12?</td>
</tr>
<tr>
<td>13.1. Representation: 2. does not</td>
<td></td>
</tr>
<tr>
<td>14. Item changes</td>
<td>The directions should be with the rest of the text.</td>
</tr>
<tr>
<td>14.1. Suggested changes: 3. NL location</td>
<td></td>
</tr>
<tr>
<td>15. Extra tools</td>
<td>I like the one on PARCC better because it was helpful because we use it in our classes, so it’s easier to know how to use it.</td>
</tr>
<tr>
<td>15.1. Calculator: 4. prefers TI</td>
<td></td>
</tr>
</tbody>
</table>
### Table R1

*Descriptive Statistics and Two-Way ANOVA Results for NL by Item Number and Language Group*

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-EB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>2.80</td>
<td>0.81</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>1.80</td>
<td>0.61</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>2.30</td>
<td>0.88</td>
<td>30</td>
</tr>
<tr>
<td>EB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>2.80</td>
<td>1.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>1.80</td>
<td>0.61</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>2.00</td>
<td>0.83</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>0.34</td>
<td>1</td>
<td>0.34</td>
<td>0.70</td>
<td>0.40</td>
<td>0.00</td>
</tr>
<tr>
<td>Item Number</td>
<td>258.71</td>
<td>3</td>
<td>86.24</td>
<td>179.4*</td>
<td>0.00</td>
<td>0.70</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>1.01</td>
<td>3</td>
<td>0.34</td>
<td>0.70</td>
<td>0.55</td>
<td>0.01</td>
</tr>
<tr>
<td>Error</td>
<td>111.50</td>
<td>232</td>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. *p < .05.
Table R2

*Descriptive Statistics and ANOVA Results for MSE by Item Number and Language Group*

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.97</td>
<td>0.18</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.87</td>
<td>0.35</td>
<td>30</td>
</tr>
<tr>
<td><strong>EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.97</td>
<td>0.18</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.90</td>
<td>0.31</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>0.12</td>
<td>0.73</td>
<td>0.00</td>
</tr>
<tr>
<td>Item Number</td>
<td>51.55</td>
<td>3</td>
<td>17.18</td>
<td>492.13*</td>
<td>0.00</td>
<td>0.86</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>0.01</td>
<td>3</td>
<td>0.00</td>
<td>0.12</td>
<td>0.95</td>
<td>0.00</td>
</tr>
<tr>
<td>Error</td>
<td>8.10</td>
<td>232</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* $^*p < .05$. 


### Table R3

**Descriptive Statistics and ANOVA Results for VR by Item Number and Language Group**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>3.23</td>
<td>0.73</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>4.23</td>
<td>0.73</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.27</td>
<td>0.45</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.53</td>
<td>0.68</td>
<td>30</td>
</tr>
<tr>
<td><strong>EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>3.23</td>
<td>0.57</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>4.17</td>
<td>0.65</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.20</td>
<td>0.41</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.73</td>
<td>0.69</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>0.02</td>
<td>1</td>
<td>0.02</td>
<td>0.04</td>
<td>0.84</td>
<td>0.00</td>
</tr>
<tr>
<td>Item Number</td>
<td>679.65</td>
<td>3</td>
<td>226.55</td>
<td>582.27*</td>
<td>0.00</td>
<td>0.88</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>0.72</td>
<td>3</td>
<td>0.24</td>
<td>0.61</td>
<td>0.61</td>
<td>0.01</td>
</tr>
<tr>
<td>Error</td>
<td>90.27</td>
<td>232</td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. *$p < .05$.
### Table R4

**Descriptive Statistics and ANOVA Results for Ideational Metafunction by Item Number and Language Group**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>3.30</td>
<td>0.75</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.10</td>
<td>0.31</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.90</td>
<td>0.31</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.53</td>
<td>0.68</td>
<td>30</td>
</tr>
<tr>
<td><strong>EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>3.23</td>
<td>0.63</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.20</td>
<td>0.41</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.90</td>
<td>0.31</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.73</td>
<td>0.69</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>0.20</td>
<td>1</td>
<td>0.20</td>
<td>0.70</td>
<td>0.40</td>
<td>0.00</td>
</tr>
<tr>
<td>Item Number</td>
<td>346.75</td>
<td>3</td>
<td>115.58</td>
<td>394.92*</td>
<td>0.00</td>
<td>0.84</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>0.61</td>
<td>3</td>
<td>0.20</td>
<td>0.70</td>
<td>0.55</td>
<td>0.01</td>
</tr>
<tr>
<td>Error</td>
<td>67.90</td>
<td>232</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* $^*p < .05$. 
Table R5

*Descriptive Statistics and ANOVA Results for Interpersonal Metafunction by Item*

*Number and Language Group*

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>1.80</td>
<td>0.55</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>1.17</td>
<td>0.46</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.63</td>
<td>0.67</td>
<td>30</td>
</tr>
<tr>
<td><strong>EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>2.03</td>
<td>0.56</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>1.10</td>
<td>0.48</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.50</td>
<td>0.68</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>0.02</td>
<td>0.90</td>
<td>0.00</td>
</tr>
<tr>
<td>Item Number</td>
<td>120.55</td>
<td>3</td>
<td>40.18</td>
<td>163.26*</td>
<td>0.00</td>
<td>0.68</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>1.15</td>
<td>3</td>
<td>0.38</td>
<td>1.55</td>
<td>0.20</td>
<td>0.02</td>
</tr>
<tr>
<td>Error</td>
<td>57.10</td>
<td>232</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* *p < .05.
Table R6

*Descriptive Statistics and ANOVA Results for Textual Metafunction by Item Number and Language Group*

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.90</td>
<td>0.31</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>5.13</td>
<td>0.78</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>2.53</td>
<td>0.63</td>
<td>30</td>
</tr>
<tr>
<td><strong>EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.97</td>
<td>0.18</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>4.73</td>
<td>0.87</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>2.40</td>
<td>0.81</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>0.82</td>
<td>1</td>
<td>0.82</td>
<td>2.57</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>Item Number</td>
<td>835.93</td>
<td>3</td>
<td>278.64</td>
<td>877.54*</td>
<td>0.00</td>
<td>0.92</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>1.92</td>
<td>3</td>
<td>0.64</td>
<td>2.01</td>
<td>0.11</td>
<td>0.03</td>
</tr>
<tr>
<td>Error</td>
<td>73.67</td>
<td>232</td>
<td>0.32</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* $*p < .05$. 

Appendix S
Descriptive Statistics and Two-Way Analysis of Variance Results for Three-Part Sequence of Responding to an Item

Table S1

Descriptive Statistics and ANOVA Results for Student Struggled During Item Interpretation by Item Number and Language Group

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-EB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.33</td>
<td>0.48</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>1.00</td>
<td>0.79</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.13</td>
<td>0.35</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.53</td>
<td>0.78</td>
<td>30</td>
</tr>
<tr>
<td>EB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.27</td>
<td>0.52</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>1.33</td>
<td>0.84</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.23</td>
<td>0.57</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>1.13</td>
<td>1.04</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>3.50</td>
<td>1</td>
<td>3.50</td>
<td>7.07*</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Item Number</td>
<td>38.25</td>
<td>3</td>
<td>12.75</td>
<td>25.73*</td>
<td>0.00</td>
<td>0.25</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>3.78</td>
<td>3</td>
<td>1.26</td>
<td>2.54*</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>Error</td>
<td>114.97</td>
<td>232</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. *p < .05.
Table S2

*Descriptive Statistics and ANOVA Results for Glossary Helped by Item Number and Language Group*

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-EB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.03</td>
<td>0.18</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>EB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.23</td>
<td>0.43</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.13</td>
<td>0.57</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.07</td>
<td>0.37</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>0.60</td>
<td>1</td>
<td>0.60</td>
<td>7.08*</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Item Number</td>
<td>0.58</td>
<td>3</td>
<td>0.19</td>
<td>2.29</td>
<td>0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>0.33</td>
<td>3</td>
<td>0.11</td>
<td>1.31</td>
<td>0.27</td>
<td>0.02</td>
</tr>
<tr>
<td>Error</td>
<td>19.67</td>
<td>232</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. *$p < .05$.*
Table S3

*Descriptive Statistics and ANOVA Results for Student Suggested NL Change by Item*

**Number and Language Group**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.03</td>
<td>0.18</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.27</td>
<td>0.52</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.13</td>
<td>0.43</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.23</td>
<td>0.43</td>
<td>30</td>
</tr>
<tr>
<td><strong>EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.27</td>
<td>0.52</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.30</td>
<td>0.47</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.07</td>
<td>0.25</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.53</td>
<td>0.57</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>0.94</td>
<td>1</td>
<td>0.94</td>
<td>4.82*</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Item Number</td>
<td>2.98</td>
<td>3</td>
<td>0.99</td>
<td>5.10*</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>1.31</td>
<td>3</td>
<td>0.44</td>
<td>2.25</td>
<td>0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>Error</td>
<td>45.17</td>
<td>232</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. *p < .05.*
Table S4

*Descriptive Statistics and ANOVA Results for Student Did Not Try by Item Number and Language Group*

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.37</td>
<td>0.56</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.23</td>
<td>0.43</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.17</td>
<td>0.38</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.10</td>
<td>0.31</td>
<td>30</td>
</tr>
<tr>
<td><strong>EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.40</td>
<td>0.56</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.23</td>
<td>0.43</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.20</td>
<td>0.48</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.20</td>
<td>0.41</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>0.10</td>
<td>1</td>
<td>0.10</td>
<td>0.51</td>
<td>0.48</td>
<td>0.00</td>
</tr>
<tr>
<td>Item Number</td>
<td>1.91</td>
<td>3</td>
<td>0.64</td>
<td>3.12*</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>0.8</td>
<td>3</td>
<td>0.26</td>
<td>0.13</td>
<td>0.94</td>
<td>0.00</td>
</tr>
<tr>
<td>Error</td>
<td>47.37</td>
<td>232</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* $*p < .05.$
Table S5

**Descriptive Statistics and ANOVA Results for Student Struggled During Problem-Solving by Item Number and Language Group**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.90</td>
<td>0.92</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>1.07</td>
<td>1.02</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>1.10</td>
<td>1.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.50</td>
<td>0.51</td>
<td>30</td>
</tr>
<tr>
<td><strong>EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>1.13</td>
<td>1.25</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>1.53</td>
<td>0.97</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>1.03</td>
<td>1.03</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.63</td>
<td>0.72</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>2.20</td>
<td>1</td>
<td>2.20</td>
<td>2.44</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>Item Number</td>
<td>16.91</td>
<td>3</td>
<td>5.64</td>
<td>6.24*</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>2.21</td>
<td>3</td>
<td>0.74</td>
<td>0.82</td>
<td>0.49</td>
<td>0.01</td>
</tr>
<tr>
<td>Error</td>
<td>209.63</td>
<td>232</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* *p* < .05.
Table S6

Descriptive Statistics and ANOVA Results for Student Solved Using Something Other than the Computer by Item Number and Language Group

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-EB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.27</td>
<td>0.45</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.30</td>
<td>0.88</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>1.33</td>
<td>1.12</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.30</td>
<td>0.84</td>
<td>30</td>
</tr>
<tr>
<td>EB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.53</td>
<td>0.51</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.33</td>
<td>1.06</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>1.73</td>
<td>0.91</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.20</td>
<td>0.55</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>1.35</td>
<td>1</td>
<td>1.35</td>
<td>1.98</td>
<td>0.16</td>
<td>0.01</td>
</tr>
<tr>
<td>Item Number</td>
<td>66.68</td>
<td>3</td>
<td>22.23</td>
<td>32.65*</td>
<td>0.00</td>
<td>0.30</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>2.28</td>
<td>3</td>
<td>0.76</td>
<td>1.12</td>
<td>0.34</td>
<td>0.01</td>
</tr>
<tr>
<td>Error</td>
<td>157.93</td>
<td>232</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05.
### Table S7

*Descriptive Statistics and ANOVA Results for Student Suggested an MSE Change by Item and Language Group*

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.63</td>
<td>0.49</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td><strong>EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.63</td>
<td>0.49</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Item Number</td>
<td>18.05</td>
<td>3</td>
<td>6.02</td>
<td>100.18*</td>
<td>1.00</td>
<td>0.56</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>0.00</td>
<td>3</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Error</td>
<td>13.93</td>
<td>232</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* *p < .05.
Table S8

Descriptive Statistics and ANOVA Results for Response Does Not Represent Student’s Thinking by Item Number and Language Group

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-EB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.73</td>
<td>0.45</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.13</td>
<td>0.35</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.23</td>
<td>0.43</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.80</td>
<td>0.41</td>
<td>30</td>
</tr>
<tr>
<td>EB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.53</td>
<td>0.51</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.23</td>
<td>0.43</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.17</td>
<td>0.38</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.77</td>
<td>0.43</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>0.15</td>
<td>1</td>
<td>0.15</td>
<td>0.83</td>
<td>0.36</td>
<td>0.00</td>
</tr>
<tr>
<td>Item Number</td>
<td>16.70</td>
<td>3</td>
<td>5.57</td>
<td>30.85*</td>
<td>0.00</td>
<td>0.29</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>0.68</td>
<td>3</td>
<td>0.23</td>
<td>1.26</td>
<td>0.29</td>
<td>0.02</td>
</tr>
<tr>
<td>Error</td>
<td>41.87</td>
<td>232</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* *p* < .05.
### Table S9

**Descriptive Statistics and ANOVA Results for Student Says Response Tool was Hard to Use by Item Number and Language Group**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.07</td>
<td>0.25</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>1.57</td>
<td>0.90</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.60</td>
<td>0.68</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.27</td>
<td>0.45</td>
<td>30</td>
</tr>
<tr>
<td><strong>EB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.07</td>
<td>0.25</td>
<td>30</td>
</tr>
<tr>
<td>Item 5</td>
<td>1.40</td>
<td>0.89</td>
<td>30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.63</td>
<td>0.62</td>
<td>30</td>
</tr>
<tr>
<td>Item 24</td>
<td>0.40</td>
<td>0.50</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Item Number</td>
<td>68.02</td>
<td>3</td>
<td>22.67</td>
<td>60.09*</td>
<td>0.00</td>
<td>0.44</td>
</tr>
<tr>
<td>Language * Item Number</td>
<td>0.70</td>
<td>3</td>
<td>0.23</td>
<td>0.62</td>
<td>0.60</td>
<td>0.01</td>
</tr>
<tr>
<td>Error</td>
<td>87.53</td>
<td>232</td>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. *p < .05.