

THE USE OF ILLUSTRATIONS IN LARGE-SCALE SCIENCE ASSESSMENT:

A COMPARATIVE STUDY

by

CHAO WANG

B.A., Anhui Normal University, 2002

M.Ed., Capital Normal University, 2005

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado at Boulder in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of Education
2012

This thesis entitled:
The Use of Illustrations in Large-Scale Science Assessment: A Comparative Study
Written by Chao Wang
has been approved for the Department of Education, University of Colorado, Boulder

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

Kathy Escamilla, Ph.D.

Janette Klingner, Ph.D.

Lucinda Soltero-Gonzalez, Ph.D.

Maria Thomas-Ruzic, Ph.D.

The final copy of this thesis as 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.

Wang, Chao (Ph.D., Educational Equity and Cultural Diversity)

The Use of Illustrations in Large-Scale Science Assessment: A Comparative Study

Thesis directed by Professor Guillermo Solano-Flores

This dissertation addresses the complexity of test illustrations design across cultures. More specifically, it examines how the characteristics of illustrations used in science test items vary across content areas, assessment programs, and cultural origins. It compares a total of 416 Grade 8 illustrated items from the areas of earth science, life science, and physics used in an international assessment, four U.S. state assessments, and China's middle school exit examinations. It also examines the relationship of student performance and item illustrations' formal properties.

Results from both statistical analyses and analyses of exemplar illustrations show that although illustrations across cultures tended to have similar complexities, the sets of features that make them complex can be different. Illustrations from the Chinese tests provide more frequently than illustrations from the American tests visual information connecting the content of the item with the country's national, cultural, and historical contexts. Results also show that the contextual features in test illustrations tend to be positively related to student performance on the items.

This study is among the first approaches to analyzing the use of illustrations in the context of testing across cultures. Its results speak to the need for systematically examining the design and effectiveness of illustrations in science tests. Identifying cultural differences in the features of item illustrations can inform the process of item development and contribute to ensuring more valid assessment of culturally diverse populations in both the U.S. and the context of international test comparisons.

DEDICATION

This dissertation is dedicated to both my former students in China and in the U.S., from whom I have always learned more than I have been able to teach, and to all of my teachers in the hope that I can help others as they have helped me. They are my everlasting inspiration to improve education and make a difference in the world. Thank you!

ACKNOWLEDGEMENTS

I am forever grateful for the faculty, fellow colleagues, and staff in the Education Equity Cultural Diversity program at CU Boulder. Their guidance, encouragement, and knowledge have provided me strength, courage, and wisdom necessary to continue my professional endeavors.

First, I would like to express my deepest gratitude to Dr. Solano-Flores, who has spent numerous hours in guiding me throughout every single step in my academic journey. Without his continuous encouragement and support, I would not be able to continue to strive for excellence in my educational career. I am also much obliged to the members of the committee, Dr. Kathy Escamilla, Dr. Janette Klingner, Dr. Lucinda Soltero-González, Dr. Maria Thomas-Ruzic, for their mentoring on this dissertation and for the constant support and encouragement they have given me over the years. I am indebted also to Dr. Marilyn Blackmon for her support in my work.

In addition, I want to thank Lisa, Katy, Magda, Khanh, Rachel, Karla, Ashley, Juli, Jackie, Zhe, Sandra, Wendy, Patrick, Nathan, Suzanne, Annette, Sara, and Sandy for their invaluable friendship and being my never-ending source of motivation since the very first days we met.

I also owe a great deal of gratitude to my parents, Yefang and Yongguang Wang, for their unwavering support and unconditional love, for stimulating my interest in learning more and giving me every opportunity in the world. I would also like to thank my older sister, Min Wang, for being my best friend and role model. Finally, I am tremendously grateful to my incredibly wonderful husband Chong. Thank you for your love, support, patience, and numerous sacrifices throughout my academic career.

Funding for this dissertation was provided by the National Science Foundation, Award No. DRL 0822362, and the University of Colorado at Boulder through the Leadership Education for

Advancement and Promotion Association Professor Growth Grant Program. The opinions expressed here do not necessarily reflect those of the funding agencies.

CONTENTS

TABLES	xi
FIGURES.....	xiii
CHAPTER	
I. INTRODUCTION.....	1
Background of the Study	1
Limitations of Current Approaches to Examining Illustrations.....	5
Definition	7
Statement of the Problem.....	8
Research Objectives and Questions.....	10
Importance of the Study.....	11
II. CONCEPTUAL FRAMEWORK	13
Overview.....	13
Proposed Conceptual Framework.....	13
Cognitive Perspective.....	14
Sociocultural Perspective.....	18
Semiotic Perspective.....	20
Conclusion	22
III. LITERATURE REVIEW	23
Overview.....	23
Effects of Illustrations on Student Performance	24
Image-Based Accommodations for ELL Students.....	24

Effectiveness of Using Illustrations in Educational Contexts.....	26
Cultural Differences in Illustrations Used in Educational Contexts.....	29
Analysis of Illustrations	33
Analysis of Test Illustrations.	33
Analysis of Instructional Illustrations	37
IV. METHOD	42
Design	42
Data Collection	42
Original Item Corpus.	42
Item Selection.	44
Coding.....	45
IDVs and Examples.	48
Coding Examples.....	56
Coding Process.....	63
Data Analysis	65
V. RESULTS	67
Overview.....	67
Research Question 1: Test Illustration Complexity	68
Descriptive Analyses.	68
t-Tests.....	79
ANOVAs.....	83
Summary of Findings for Research Question One.....	84
Research Question 2: Test Illustration Complexity and Student Performance.....	86

Item <i>p</i> -value.....	86
Correlational Analyses.....	87
Summary of Findings for Research Question Two.....	94
Illustration Complexity at the Individual IDV Level.....	94
Exemplar Illustrations in Sub-Content Areas Across Assessment Systems.....	100
VI. DISCUSSION.....	103
Overview.....	103
Summary of Results.....	104
Current Findings as They Relate to the Conceptual Framework.....	13
Cognitive Science.....	106
Sociocultural Theory.....	108
Semiotics.....	109
Implications.....	110
Implications for Test Item Development and Large-Scale Assessment.....	110
Implications for International Test Comparisons.....	111
Implications for Classroom Teaching and Assessment.....	112
Implications for Cross-Cultural Research.....	113
Limitations.....	113
Future Research.....	115
Conclusion.....	115
REFERENCES.....	117
APPENDIX.....	132
A. Historical and Contextual Information of Science Assessments.....	132

B. Definitions and Attributes of Visual Image Dimensions	135
C. Example Illustrations and Definitions for the IDVs.....	137
D. Illustrated Items With the Most Frequent Outstanding IDVs.....	212
E. Example of Flagged Illustrated Items Across Sub-Content Areas.....	226

TABLES

Table

1. The Grid of Analysis for the Visual Code of Images in Science Tests and Textbooks ...	35
2. Sources of Illustrated Items in the Corpus	43
3. Sample of Illustrated Items Analyzed in This Study.....	45
4. Illustration Dimensions, Categories, and Variables in the Coding System for Designing and Examining Science Test Item Illustrations	46
5. Definitions of IDVs in Coding Dimension on Representation of Objects and Backgrounds	48
6. Definitions of IDVs in Coding Dimension on Metaphorical Visual Language	50
7. Definitions of IDVs in Coding Dimension on Text in Illustration.....	52
8. Definitions of IDVs in Coding Dimension on Variables and Constants.....	53
9. Definitions of IDVs in Coding Dimension on Interaction between Illustration and Item .	55
10. Definitions of IDVs in Coding Dimension on Context in Illustration	56
11. Analysis of the Representation of Objects and Background Dimension for the Illustration of the Item Shown in Figure 5	59
12. Analysis of the Variables, Constants, and Functions Dimension for the Illustration of the Item Shown in Figure 6.....	61
13. Analysis of the Structure Category of the Illustration of the Item Shown in Figure 7....	63
14. Number and Frequency of Coding Events	64
15. Example of Flagged Illustrated Items Across Sub-Content Areas in Life Science	66
16. Number and Percentage of IDVs Coded 1 by Illustration Dimension	69
17. Mean Number and Range of IDVs Coded 1 by Test Illustration Dimension	69
18. Mean on Total Number of IDVs Coded 1 Across Assessment System	70
19. Number (n) and Percentage (%) of IDVs Coded 1 by Assessment System and Test Illustration Dimension.....	72

20.	Number (n) and Percentage (%) of IDVs Coded 1 in Categories Under the Dimension Context in Illustration.....	73
21.	Mean Number of IDVs Coded 1 by Content Area and Assessment System	74
22.	Mean Number of IDVs Coded 1 by Item Format Type, Assessment System, and Illustration Dimension.....	76
23.	Mean Number of IDVs Coded 1 by Item Commonality Type, Assessment System, and Illustration Dimension.....	78
24.	Independent-Samples <i>t</i> -test of Number of IDVs Coded 1 for Item Country of Origin...	80
25.	Independent-Samples <i>t</i> -test of Number of IDVs Coded 1 for Item Format Type.....	82
26.	One-way Fixed-effects Analysis of Variance for Illustration Complexity by Dimension	83
27.	Two-way Fixed-effects ANOVAs for Illustration Complexity by Dimension, Item Origin, Item Format Type, and Content Area	84
28.	Range and Mean of Item p-Value: TIMSS	86
29.	Pearson Correlations Between Student Performance and Illustration Complexity by Illustration Dimension and Student Country of Origin	89
30.	Pearson Correlations Between Student Performance and Illustration Complexity by Content Area and Illustration Dimension.....	91
31.	Pearson Correlations Between Student Performance and Illustration Complexity by Item Commonality and Dimension.....	93
32.	Outstanding IDVs Identified for China	96
33.	Outstanding IDVs Identified for the U.S.....	97
34.	Illustrations Identified With the Highest Number of Outstanding IDVs	98
E1.	Example of Flagged Illustrated Items Across Sub-Content Areas in Earth Science	226
E2.	Example of Flagged Illustrated Items Across Sub-Content Areas in Life Science	236
E3.	Example of Flagged Illustrated Items Across Sub-Content Areas in Physical Science	242

FIGURES

Figure

1.	Three illustrations from large-scale science tests in the U.S. and China.	3
2.	Four illustrations of car from large-scale science tests in the U.S..	3
3.	Two different illustrations representing sequence: U.S. and China.	4
4.	Representation of the proposed conceptual framework on test illustration from three theoretical perspectives.	13
5.	An illustrated science item from NAEP 2009	59
6.	An illustrated science item from NAEP 2005	60
7.	An illustrated science item from TIMSS 2007	62
8.	Number of IDVs coded 1 shown in percentage of number of items generated correspondingly in China and the U.S.	71
9.	Percentage of IDVs coded 1 by test illustration dimension in three assessment systems: China, the U.S., and TIMSS.	72
10.	Number of IDVs coded 1 in the dimension, Context in Illustration from three assessment systems: China, U.S. and TIMSS.	73
11.	Percentage of IDVs coded 1 by IDV and illustration country of origin. China as reference country.	95
12.	Two comparable illustrations from large-scale Life Science tests in the U.S. and China	99
13.	Three comparable illustrations on the Ring of Fire from large-scale science tests in China, the U.S., and TIMSS.	102

CHAPTER I

INTRODUCTION

Background of the Study

As a primary means of communication, visual images have a considerably longer past than written language. Cave paintings, the earliest known images, are about 35,000 years old (Conkey, 1994), while the oldest writing system is only 7,000 years old (Ho, 1976, cited in Mulcahy & Samuels, 1987). Over the centuries, images have blossomed and affected the way in which individuals and societies as a whole develop and function. Today, among 21st century global cultures, images are considerably more prevalent than in earlier times (Katz, 1994; Kosslyn, 2006), appearing in journals, newspapers, TV, and the internet.

Some studies show that the use of images in instructional materials has increased in the last decades. For example, over 20 years ago, Evans and his colleagues (Evans, Watson, & Willows, 1987) examined 60 Canadian textbooks in reading, mathematics, and science, and found that illustrations occupied 60% to 80% of the pages for elementary level textbooks. Over time, instructional materials underwent substantial transformations in illustrations use in terms of quantity, size, and color (Berglund, 1991, cited in Pettersson, 2002). A recent review (Pettersson, 2002) further shows that “approximately one-third to one-half of the space in science textbooks is devoted to illustrations” (p.245).

In spite of the frequent use of images as resources in academic contexts, many fundamental issues are yet to be investigated. For example, Is any illustration always worth a thousand words in any context? What factors affect the design and use of an illustration in a textbook? How visually literate are students and how able are they to interpret visual information in textbooks?

Do students with different cultural backgrounds interpret images generated in one culture in the same ways?

These questions, which sparked my interest in the topic of my dissertation, came about from work that I have done as a graduate research assistant for the Illustrations in the Testing of English Language Learners (ITELL) project. The project explores how illustrations can be used as visual supports for English language learners (ELLs)—students who are developing English as a second language while being schooled and tested in English (Abedi, 2004)—to understand the content of test items (Solano-Flores, 2010a; 2010b; 2011). I have participated in the examination of illustrated items in current large-scale science assessments and the development of procedures and criteria for evaluating their features. These experiences have allowed me to become cognizant of the complexity of illustrations used in tests.

Test illustrations vary considerably in contents and complexities. For example, Figure 1 shows illustrations originated from current large-scale science tests in the U.S. and China. Illustration A depicts a groundhog inside a burrow covered with weed. Illustration B shows a ball attached to the end of a string and spun in a circle and a hand holding the string. Illustration C shows Huanhuan, one of the 2008 Beijing Olympic Games mascots, lifting his left leg, putting his left hand backward, and kicking a football forward.

Although the three illustrations are line drawings, they use different representational techniques to depict actions. For example, Illustration A and B are realistic line drawings, while Illustration C is a cartoon. Furthermore, the content of Illustration C has an emblematic load, as it represents the Chinese national spirit for the 2008 Beijing Olympic Games.

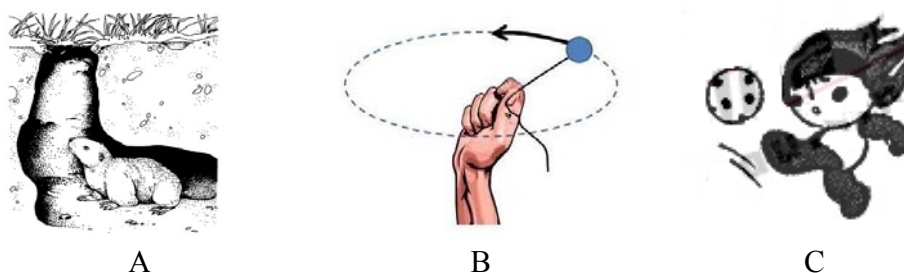


Figure 1. Three illustrations from large-scale science tests in the U.S. and China. Sources: (A) Texas Education Agency. (2004). Texas Assessment of Knowledge and Skills, 2004 released science test for Grade 5; (B) Arizona Department of Education. (2009). Arizona's instrument to measure standards AIMS Science, 2009 sample test for Grade 8; (C) Ningxia Department of Education. (2008). Ningxia Middle School Exit Examination, 2008 physics test for Grade 9.

At a superficial level, we may think about test illustrations in terms of a small set of broad visual representation categories, such as *photograph*, *drawing*, *table*, *chart*, *graph*, or hybrids of these forms of illustration. In the context of large-scale science assessment, these broad categories do not help much to appreciate the complexity in form, style, and the concreteness-abstractness with which objects, actions, variables, and textual information are represented. For example, Figure 2 shows car in three forms of drawing, realistic line drawing (A), silhouette (B), and cartoon (C). Compared to a car in a photo or a realistic drawing, a silhouette car is less concrete. A cartoon of a car is even less concrete and may elicit emotional responses.

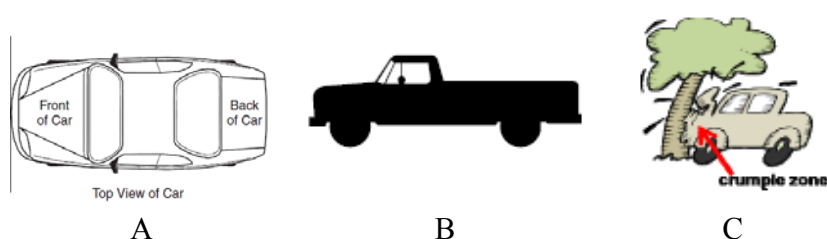


Figure 2. Four illustrations of car from large-scale science tests in the U.S. Sources: (A) Arizona Department of Education. (2009). Arizona's instrument to measure standards AIMS Science, 2009 sample test for Grade 8; (B) Georgia Department of Education. (2004). Georgia End-Of-Course Tests Physical Science, released test booklet; (C) Maryland State Department of Education. (2008). Maryland School Assessment Science, 2008 Public Release Grade 8.

There may be more variations than commonalities in the characteristics of illustrated items originated from different countries. In Figure 3, Illustration A, used in a large-scale science test in the U. S., shows the key steps in taking a dive. The six snapshots of the positions of a diver are used as a graphic device with the intent to ensure that viewers interpret a sequence. Likewise, Illustration B, a Chinese test illustration, shows a girl on a trampoline jumping upward and doing a back flip. The outline of the girl shows that she jumped upward before doing a backflip. The lines below the outline of the girl show the impact on the trampoline when she jumped upward.

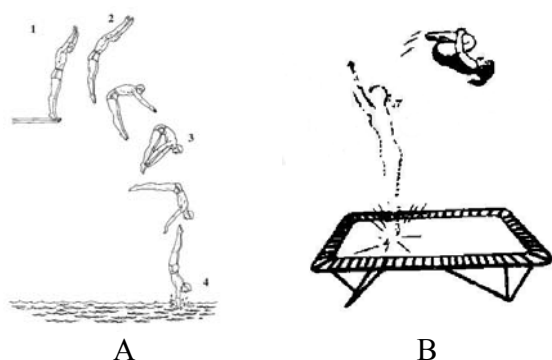


Figure 3. Two different illustrations representing sequence: U.S. and China. Sources: (A) Massachusetts Department of Elementary & Secondary Education. (2005). Massachusetts Comprehensive Assessment System Science, 2005 sample test for Grade 8; (B) Heilongjiang Department of Education. (2001). Heilongjiang Middle School Exit Examination, 2001 physics test for Grade 9.

These two illustrations utilize different styles and techniques to depict similar processes. Both of them show a process in which a person conducts an athletic type of activity and represent the scientific concept of energy transformation. However, the series of actions are shown with different kinds of graphic devices. Whereas motion stages are shown with numbers and as different positions of the same human figure in the U.S. illustration, motion is represented as a shadow and a solid line drawing of the same human figure in the Chinese illustration.

My dissertation addresses the need for effective approaches to analyzing illustrations in science assessment. I will examine the similarities and differences in the features of illustrations used in large-scale science tests in different cultures.

Limitations of Current Approaches to Examining Illustrations

Multiple Terms and Definitions of Illustrations. Terms such as *picture, drawing, image, visual, diagram, graphics, graphic display, and visual organizer* tend to emphasize different forms of representation. However, they are used in multiple ways; some of them may refer to the same form of illustration, or the same term may be used to refer to different forms of illustration. In addition, the use of these terms implies the assumption of clear-cut distinctions of different forms of illustration.

Despite this multitude of terms, many researchers tend to regard an illustration as a self-evident entity, and therefore pay little attention to defining its nature, scope, or meaning. The followings are examples used in books, official documents, and academic articles:

- “Pictures ...are what practitioners of semiotics call signs— representations whose meaning depends on a repertoire of learned strategies.” (Nodelman & Reimer, 2003, p.275)
- “Visual organizers are graphic representations of different kinds of thinking processes.” (Clarke, 1991 p. 526)
- “Visual displays can be thought of as spatial images of concepts that can also be presented in verbal, propositional form.” (Hunter, Crismore, & Pearson, 1987, P. 116)

- “A picture is a surface so treated that a delimited optic array to a point of observation is made available that contains the same kind of information that is found in the ambient optic arrays of an ordinary environment.” (Gibson, 1971, P. 31)
- “Scientific illustration (...) is a complex compound of information, craftsmanship, and cooperation between the artist and the scientist. It is one of many means of communication between various scientists. It is a visual explanation of scientific studies and findings. It is a mixture of centuries’ proven techniques of observing, drawing, and painting combined with inquisitiveness into nature.” (Jastrzebski, 1985, p. 5)

Illustrations come in a number of genres, which does pose a conceptual challenge to both naming them and defining what they are. For example, the term *visual* can be used to refer to a picture, a piece of film, or a process relating to seeing or sight. We should be aware of the fact that terms like this actually reveal a compilation of ideas from a wide variety of theoretical perspectives and can be interpreted in multiple ways.

Despite the fact that many scholars have a general idea of what illustration means and are aware of the communicative nature of an illustration, there is no clear way of operationalizing the definitions. Many of the definitions emphasize one ideal type of illustration (e.g. picture, table, map etc.).

Assumptions About the Illustrations Examined. It is worth to ponder why practitioners in the field of testing use illustrations in tests. The reason may lie on two assumptions about illustrations (see Nodleman & Reimer, 2003). One is that students are better at understanding information through illustrations than verbally. However, there is no available research that

proves that students have a better understanding of test illustrated items than non-illustrated items.

The other assumption is that illustrations are intrinsically understandable and more concrete than written texts, or that they convey meaning to young or untrained individuals in the same way they do with older or trained individuals. However, there is evidence that the content of illustrations representing objects is not automatically apparent to viewers from certain cultures (Levie, 1978). Sometimes, an object, such as a car (see Figure 2 in Chapter 1) can be shown in different styles, such as a realistic line drawing, a silhouette, or a cartoon. Students who have not developed some understanding of certain drawing styles may not be able to interpret the illustrations as intended. In addition, illustrations containing graphic devices (see Figure 3) assume in the viewer a more sophisticated understanding of the visual representation conventions used (Rankin & White, cited by Winn, 1987; Weidenmann, 1994).

Definition

Unlike conventional approaches that focus primarily on population differences, my study focuses on tests as cultural artifacts. For the purpose of this work, *illustration* in science assessment is defined as any device that provides information mainly in non-textual form, and whose characteristics can be described as a set of features relevant to the representation of object and background, the use of contextual features in the illustration, the use of metaphorical visual language, the use of text in the illustration, the representation of variables, constants, and functions, and the interaction between text and illustration (Solano-Flores & Wang, 2011). *Mainly non-textual* addresses the fact that illustrations commonly used in science printed materials contain text (e.g. phrases used as labels).

This definition focuses on illustrations used in a specific context—large-scale science tests—not in other contexts, such as textbooks, science magazines, and scientific publications. It reflects the fact that written text can be part of a test illustration (as when labels, captions, or numbers are included). It also allows us to examine illustrations without depending on terms such as *table*, *chart*, or *map*, which, as mentioned before, have multiple meanings and are not sensitive to the complexity of features of test illustrations. The illustration complexity in my study refers to number of features that an illustration of a science test item contains.

Statement of the Problem

Both conceptually and empirically, our knowledge of illustrations is much less advanced than our knowledge of written and spoken language (Freedberg, 1989; Marriott & Meyer, 1998). The misconception that illustrations are a universal language and are interpreted by everybody in the same way appears to influence the thinking of many researchers and practitioners in the field of testing. An implicit assumption derived from this misconception is that illustrations are always effective in promoting understanding of text in textbooks or tests. Unfortunately, only scant research has examined the factors that shape the effectiveness of illustrations used in printed materials. This scant research has focused on the relationship between illustrations and text (e.g., Bloomer, 1960; Miller, 1936, 1938; Zimet, 1966), the influence of illustrations on cognitive tasks (e.g., Pressley, 1977; Wohlwill, 1968), and the characteristics of illustrations according to their intended functions (e.g., Fleming, 1966; Goldsmith, 1987; Svensen, 1993).

This body of research has not considered the use of illustrations in testing and the possibilities that representational conventions rooted in cultures can lead to alternative interpretations of visual language (Boling, Eccarius, Smith, & Frick, 2004; Boling, Smith, Frick,

& Eccarius, 2007; Knight, Gunawardena, & Aydin, 2009; Schiffman, 1996). To my knowledge, no studies have been conducted that compare the characteristics of illustrations in science tests from different cultures.

Likewise, there has been scant research on the validity of illustrated tests for culturally and linguistically diverse students. In the context of national and local assessment, it is important to examine the accuracy, clarity, simplicity, soundness, and cultural appropriateness of illustrations (Solano-Flores, 2010a). In the context of international assessment, it is, in addition, important to examine the soundness of test illustrations for students from different educational systems and sociocultural contexts. Cultural bias may arise from the fact that the features of illustrations in an international test are more similar to the features of illustrations commonly used in some countries but rarely used in others.

An examination of current normative documents of large-scale assessment programs indicates that there is a serious lack of conceptual clarity in the design and use of illustrations in testing contexts. Documents of assessment frameworks or item specifications from assessment systems such as the Trends in International Mathematics and Science Study (TIMSS; see Mullis, Martin, Ruddock, O'Sullivan, Preuschoff, 2009), the Programme for International Student Assessment (PISA; see Organisation for Economic Co-operation and Development, 2003), the National Assessment of Educational Progress (NAEP; see WestEd & Council of Chief State School Officers, 2007), and some state assessment programs (e.g., the Colorado Student Assessment Program—CSAP and the Florida Comprehensive Assessment Test—FCAT) provide, at best, superficial information for using illustrations in tests. For instance, the NAEP 2009 science item specifications document provides overly simple and scant information about the

characteristics of illustrations that can be included in certain test items. This document states that “drawings, images, charts, and graphs [...] can be used for paper-and-pencil items” (WestEd & Council of Chief State School Officers, 2007, p.88), but does not provide enough guidance for test developers to design those illustrations systematically (e.g., by determining or controlling for visual complexity).

Research Objectives and Questions

Taken together, these facts speak to the long-standing need for an objective and systematic examination of the characteristics of illustrations used in large-scale science assessment. In the investigation reported here, I compared municipal, state, and international assessment programs as to the characteristics and effectiveness of the illustrations used in their science test items. The study was guided by the following two research questions:

1. *How do the characteristics of illustrations in large-scale science assessments vary across type of assessment system, item cultural origin, content area, and item type?*
2. *How is test illustration complexity related to student performance on TIMSS items?*

More specifically, I 1) used an existing coding system (Solano-Flores & Wang, 2011; Wang & Solano-Flores, 2011) for analyzing test illustrations to examine the characteristics of illustrations of samples of science items from large-scale science assessment programs in the U.S., China, and from TIMSS; 2) examined the frequency of the characteristics of test illustrations across item types (e.g., multiple-choice vs. constructed-response), grade levels (Grade 4, 8/9, and 12), science subject areas (e.g., physical science), types of assessment system (e.g., municipal, state, international), and cultural origins; and 3) investigated, for a subset of items for which data on student performance was available, the relationship between the

complexity of features of illustrations and the performance of students from different cultural/national origins.

Ultimately, the goal of this dissertation study is to contribute to enhanced research and practice concerning the design and use of illustrations in current large-scale science assessments for culturally and linguistically diverse students in the U.S. and in the context of international test comparisons.

Importance of the Study

This study is critical in an era of a global economy and communications, in which international test comparisons such as PISA and TIMSS play a key role in many countries' educational policies—including the U.S. (see Murphy, 2010). This cultural perspective is important for examining the extent to which test illustrations reflect the characteristics of the cultures in which they are generated.

I focus on large-scale science tests as cultural artifacts, which are concerned with “the means, modes, and modalities by which we take in, transform, and transmit information” (Greenfield, 1994, p. 3). Unlike conventional cross-cultural studies that examine population differences based on stimulus materials generated in one culture, this study compares the similarities and differences in stimulus materials generated in different cultures.

While illustrations play a critical part in science education (e.g., in textbooks) and scholars and teachers tend to agree that illustrations can benefit students by supporting their learning, their use in the context of assessment is yet to be investigated systematically. Because of the scant research on the design of illustrations, there is a serious gap between the extensive use of test illustrations and the lack of methods and reasonings for examining their features and

functions. In addition, existing normative documents of large-scale assessment programs tend to neglect the heterogeneity of test illustrations.

The study aims at providing empirical evidence of the status quo of illustrations used in large-scale science tests. My findings will contribute to the development of a systematic approach for analyzing and evaluating illustrations in science tests. It will provide information that will help to improve science test development practices. Knowledge gained will help test developers and educational researchers achieve an in-depth understanding of the characteristics of illustrations currently used in large-scale science assessments.

CHAPTER II

CONCEPTUAL FRAMEWORK

Overview

In this chapter, I propose and discuss a conceptual framework for investigating the use and design of test illustrations in a comprehensive manner.

Proposed Conceptual Framework

The multidisciplinary conceptual framework of this study is intended to provide a comprehensive lens that captures the key aspects of test illustrations. It postulates different approaches to illustrations, drawing from three theoretical perspectives: cognitive science, sociocultural theory, and semiotics (Figure 4).

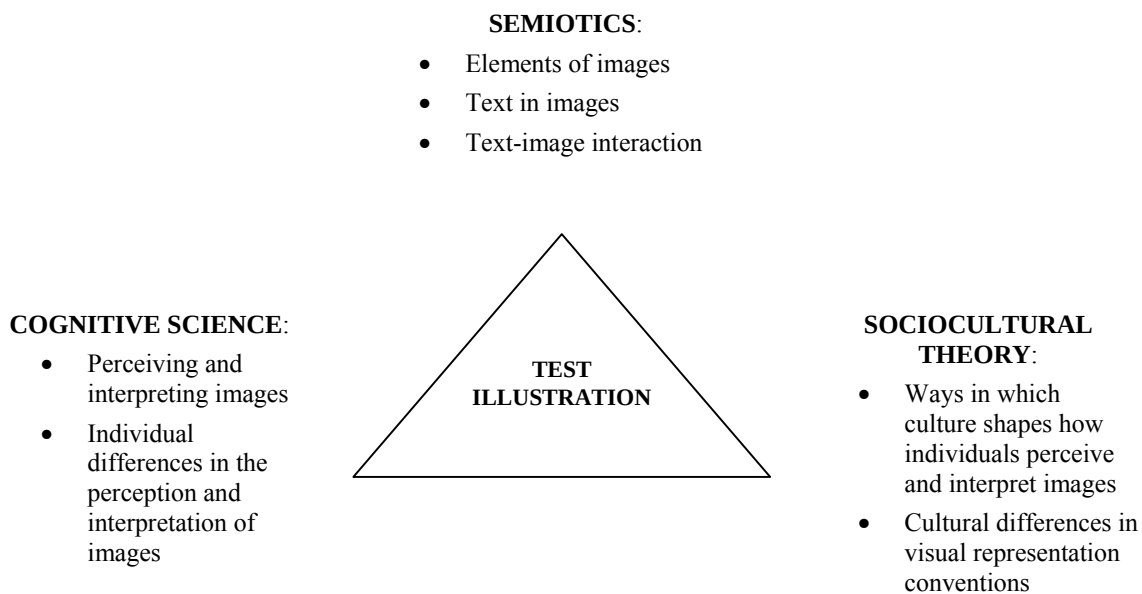


Figure 4. Representation of the proposed conceptual framework on test illustration from three theoretical perspectives.

Cognitive theories on visual perception and individual differences in perceiving visual information help explain the intricacies of processing visual information. Sociocultural views of

visual language provide support to two notions: that images generated in different cultures can be interpreted in multiple ways by individuals of diverse cultural backgrounds, and that images can be a reflection of the cultural backgrounds of their creators. Semiotics provides reasonings about the multiplicity of meanings that visual images may carry.

None of these theoretical approaches is better than the others. Since they examine different aspects of visual information, they are not mutually exclusive. Drawing from these multiple perspectives is critical to gaining a more comprehensive and in-depth understanding of the use and design of illustrations in science tests.

Cognitive Perspective. The first strand for examining test illustrations is the variety of theories on visual perception and individual differences in perceiving visual information. They help to better understand the intricacies of visual processing of both visual image producers and viewers. Visual perception—“how viewers see picture-object similarities” (Rollins, 1999, p.391), seems to be an automatic and effortless ability. But it is actually very complex, affected by a myriad of factors, especially individual differences. Among several theoretical approaches concerning visual perception, I select those that can directly inform my study. They include the constructivist theory of visual perception, the Gestalt theory of visual perception, the constructs on cognitive styles and types of perception, and the dual coding theory.

Constructivist theory of visual perception. According to constructivists (Gombrich, 1969; 1972; Gregory, 1997), picture perception is primarily a response to changing conventions; it is cognitively mediated. Gombrich distinguishes seeing from knowing. Seeing refers to “the awareness of sensations and/or retinal images” (Hagen, 1980, p. 7), while knowing is the concentrate of object hypotheses based on past experience. Pictures do not automatically and

explicitly show what they intend to represent; it is the viewer that has to decipher and construct a meaning.

In science assessment, many line drawings depict objects, such as material things, plants, animals, human beings, and natural phenomena. The shapes of the objects are represented through the use of lines. However, the objects in nature are not enclosed within lines. People have to learn about the conventions to interpret these lines. Similarly, the use of graphic devices, like those shown in Figure 3, are interpreted as intended only with extensive exposure and explicit instruction (Levie, 1978).

Gombrich (1969) points out that technology and culture play a critical role in changing people's understanding of realism in drawings. Gombrich (1972) also emphasizes the significance of viewers' being able to interpret paintings in the ways intended by their creators. Painters should be aware of the viewers' experiences with and expectations of what a picture is. The implication of this notion for science tests is that, ideally, the design of illustrations in science tests for a given target student population should be supported by knowledge of that population.

Gestalt theory of visual perception. Gestalt psychologists, Wertheimer, Koffka, and Köhler, are the first who attempted to investigate perceptual segregation systematically (Eysenck & Keane, 2005). Arnheim applies the principles of Gestalt theory to visual art, and argues that meaning exists in the organized whole “that is greater than the sum of its parts” (Anglin, Vaez, & Cunningham, 2004, p. 868). In addition, Arnheim (1986) points out that the conventions of Western art since the Renaissance are accepted as the norm, which negatively influences the interpretation and evaluation of art originated from other cultures or from historical periods.

Likewise, the drawing styles of test illustration in one culture should not be regarded as more important or more effective than styles generated in other cultures.

Another relevant issue raised by Arnheim (1974) regards the nature of abstraction in representational art. He points out the challenges in determining a proper level of abstraction of illustrations and selecting effective illustrations in teaching and learning. According to Arnheim, pictures tend to be more abstract than their referents in the real world, but may be less abstract than the concept they indicate or symbolize. For example, the cars shown in Figure 1 demonstrate different levels of abstraction: the silhouette of a car can be more abstract than a line drawing of a car, but less abstract than the word “car” in the text of the test item.

Thus, in the context of assessment, it is not valid to assume that there is only one set of universal standards on the ways in which illustrations can and should be designed. Even in the field of science and its connotations as an objective field of knowledge, the design and the interpretation of the object designed are mediated by culture.

Cognitive style theory and levels of perception. The term, *cognitive style* was proposed by Allport (1937) and further developed by Witkin and Asch (1948a; 1948b). It is defined as “individual information processing habit representing the learner’s typical mode of perceiving, thinking, problem solving and remembering” (Messick, 1976, cited in Sitz, 1996, p. 280). Differences in cognitive style can affect the way in which a student comprehends a picture in educational contexts.

According to the cognitive style theory, individuals can be grouped as either field dependent or field independent. The former are persons who tend to process visual information in a more holistically way without consciously analyzing separate visual elements; the latter are persons

who are able to identify the most important visual information out of the overall representation. Some people may require more intensive and direct instruction in developing the ability to better comprehend visual information.

In addition to cognitive styles, individuals differ at the levels of picture perception. According to Goldsmith (1984), persons at the lowest level of picture perception are able to recognize the existence of a picture; those at the second level are able to recognize what object a picture represents; those at the highest level are able to name the picture.

It may not be an easy task for a young child to reach the highest perception level. This can create challenges for students to interpret a picture developed in a culture other than theirs. Approaches to designing science test illustrations that are not sensitive to individual differences in the target population may be biased or ineffective.

Dual coding theory. The dual coding theory, an empirically-based theory of cognition, explains how information is processed and made sense of by individuals. It takes into account the interaction between verbal and nonverbal systems. This theory was first proposed and advanced by Paivio (1971; 1986); it has been applied to a range of domains, including bilingualism, literacy, neuropsychology, and cognitive robotics (Sadoski & Paivio, 2001).

Two notions from the dual coding theory are relevant to examining the relationship between text and illustration in tests. The first notion is that verbal information and visual information are processed by separate subsystems, one that specializes in verbal language (such as spoken or written words) and another that specializes in visual images (such as objects and illustrations). The second notion is that, although these two subsystems are independent of one another, they

are interconnected. Information delivered through both the verbal and visual modes is assumed to be better understood and remembered.

A well-known study by Mayer and Sims (1994) reveals that students perform better on interpreting information from science texts when both the visual and verbal information are presented concurrently rather than successively. More recently, Mayer and his colleagues (Mayer, Heiser & Lonn, 2001; Mayer, Hagarty, Mayer & Campbell, 2005) investigated the effectiveness of using static illustrations as compared to animation. The results indicate that despite the good intention of including multimedia input in learning scientific content, the use of animation may overload the individual's cognitive processes. In contrast, using static illustrations with text can have positive effects on student understanding. Mayer and Moreno (2003) note that the combination of auditory input and visual input can lead to cognitive overloading, and provide strategies to help decrease the cognitive load.

Although my study does not explore multimedia input, it is critical that, in evaluating the features of test illustrations, we should scrutinize the relationship between the written text of a test item and its accompanying illustration. The dual coding theory provides the conceptual foundation for understanding the balance that needs to exist between the text of test items and the complexity of the illustrations that accompany them.

Sociocultural Perspective. The second perspective informing my study is a sociocultural view of visual language. Culture is defined as “the actual practices and customs, languages, beliefs, forms of representation, and system of formal and informal rules that tell people how to behave most of the time” (O'Donnell, 2005, p. 554). Central to this view is that meaning is always mediated socially and culturally (Vygotsky, 1978; Wertsch, 1985; Wertsch, del Rio, &

Alvarez, 1995). Applied to this study, the view supports the notion that the creation, perception, and interpretation of test illustrations are mediated by a myriad of social and cultural factors.

An important aspect of Vygotsky's thinking (Vygotsky, 1978) is the concept of mediation. It conveys that "individual developmental change is rooted in society and culture" (Cole & Scribner, 1978, p. 7). According to Vygotsky, language, drawings, algebraic symbol systems, and all kinds of conventional signs are examples of cultural and psychological tools. He believes that, in order for higher order mental activities to form and develop, cultural tools and signs have to be acquired and sociocultural influences have to be internalized.

Hall (1980) has provided an encoding/decoding model that allows examination of the power relations embedded in images. The meaning of an image exists "in the articulation between viewer and viewed, between the power of the image to signify and the viewer's capacity to interpret meaning" (Evans & Hall, 1999, p. 4). The meanings rely on how viewers make sense of a visual representation based on shared meanings within a community or culture. An illustration can reflect what is valued in a dominant culture, which in turn determines what is and what is not represented. A viewer who accepts the world of meanings shared within a culture interprets illustrations as intended by that culture (O'Donnell, 2005).

According to Hall, an illustration is produced through a medium, within a historical period and a specific sociocultural context. Take Figure 1C as an example: The medium is the test booklet, the historical period is the present—a time of globalization, and the sociocultural context is the support for 2008 Beijing Olympic Games in China. Students in China bring their understanding of aspects of their culture to interpret this image, while students from other cultures may have difficulty making sense of the representation. In addition, the image is

associated with power relations. For example, Huanhuan is a common Han Chinese name, which tends to represent the dominant Han culture in China. Students from other 55 minority ethnic groups in China may resist the dominant intended meaning.

In sum, culture plays a critical role in perceiving and interpreting illustrations. An illustration creator should be well aware of the culture of the intended viewers; the elements of and the conventions used in the illustrations should be understandable by the intended viewers (Lester, 2003). Ideally, in developing and reviewing illustrations in science tests for culturally and linguistically diverse students, test developers, illustrators, and illustration reviewers should be well aware of how social and cultural factors affect the meaning-making process, and therefore develop a systematic approach to examining the illustrations to ensure that they are interpreted by these students as intended.

Semiotic Perspective. The third perspective informing this study draws on semiotics, the theory of signs (Eco, 1976; Kress, 1997; Peirce, 1966, Rowe, 1994; Siegel, 1995), which is mainly concerned with “the meaning that humans associate with the images they see” (Lester, 2003, p. 257). In semiotic theory, a sign is anything that “stands for an object or concept” (Hoopes, 1991, p. 141; Eco, 1986, p.15). It stresses that all knowledge is mediated by signs (Peirce, 1966) and communication is made through multiple forms (cited by Saint-Martin, 1990; Kress & van Leeuwen, 2006). It embraces and values the rich complexity of human experience and its role in the process of meaning-making.

Both Saussure and Peirce contribute much to the development of contemporary semiotics. Saussure focuses on meaning in text while Peirce focuses on visual communication (Moriarty, 2005). Peirce (1966) maintains that any image has complex cultural meanings. He uses a

semiotic triangle—iconic (Note 1), indexical, and symbolic—to describe the relationship between the sign and the object. Iconic signs are images that have a direct resemblance to the object they represent, such as a photograph or a realistic drawing showing objects in real life. Indexical signs are indicators of the object or concept they represent in a logical manner, such as an arrow used to indicate the direction of movement. Symbolic signs are connected to its object or concept by convention (such as the mascot, Huanhuan in Figure 1). There is no logical or representational connection between the signs and the thing or concept they represent (Lester, 2003).

Halliday (1975; 1978) proposes a socio-semiotic theory of language. He explains the nature of language from three aspects: 1) language as a social process; 2) language as systematic; and 3) language as functional. He posits that meaning-making of illustrations is a social practice. The implication of this notion for science tests is that in order to make sense of an illustrated test item in a science test, students need to be aware of and be able to interpret both the linguistic and non-linguistic symbols coexisting in a test item (Halliday, 1978; Lemke, 2003).

More recently, researchers in semiotics have attempted to provide a structural analysis of pictorial language. For example, Bertin (1983) examined in detail the properties of graphic representational systems. He examined over 1,000 maps and diagrams to elucidate graphic techniques, such as shape, orientation, color, texture, volume, and size. Saint-Martin (1990) identified six distinct categories in visual language: color/tonality, texture, dimension, vectoriality, boundaries, and position in the plane. She described the development of this visual language and its application in contemporary media. Bogdan (2002) proposed to analyze the semantic, lexical, and syntactic dimensions of pictorial language. She distinguished between

semantic and pragmatic signification. The semantic signification refers to using pictures to represent what the illustrators have in mind; the pragmatic signification refers to interpreting the meanings embedded in the pictures considering emotional, cultural, habitual components.

Lemke (2003) adopts both formal and social semiotic perspectives to explore mathematics functions considering historical and social contexts. He identifies two types of elements in mathematical meaning: typological meaning and topological meaning. Typological meaning refers to “meaning-by-kind, in which natural language specializes” while topological meaning refers to meaning-by-degree, presented by body gestures or visual illustrations (Lemke, 2003, p. 4).

Conclusion

In general, this three-part conceptual framework provides a comprehensive lens to examining the use and design of illustrations in large-scale science assessment. According to these reasonings, a test illustration is created by a test developer with his or her own understanding about illustrations and is perceived and interpreted by test-takers of various understandings and experiences. A number of sociocultural factors affect how a test illustration is created, used, and interpreted. These reasonings also allow us to be aware of the number of explicit and implicit meanings conveyed in a test illustration.

CHAPTER III

LITERATURE REVIEW

Overview

Due to the fact that scant research exists on the design and use of illustrations in science assessment, this literature review draws from several related fields. Particularly critical to this study are research and theory on: 1) the effects of using illustrations on student visual cognition and performance, 2) the cultural differences in developing, perceiving, and interpreting illustrations in educational contexts, and 3) the approaches to analyzing and classifying illustrations in test booklets, textbooks, and other instructional-related materials.

The literature in the first section consists of studies on the creation and implementation of illustrations as testing accommodations, and the role of cognition in visual interpretation. After providing an overview of how illustrations have been considered and treated in the testing of English language learners (ELLs), I discuss the few empirical studies that have designed and used illustrations as visual support for ELL students. In addition, I review studies that have investigated the effectiveness of using visual information on student performance in various educational contexts, and discuss the relations between internal and external visualizations. This body of literature supports a stance on test illustrations as dynamic and continuous visual communication between test developers and test takers.

In the second section, I discuss studies on differences between illustrations from different cultures, and review studies on illustration perception and interpretation used in educational settings by viewers of diverse cultural backgrounds. All these studies confirm that culture plays a critical role in generating illustrations and in processing information embedded in illustrations.

Finally, I review the few studies that have attempted to use different approaches to analyze illustrations in educational contexts. I discuss the empirical research studies that have attempted to use a semiotic perspective to analyze illustrations used in test items. In addition, I review literature on different approaches to analyzing and classifying illustrations in instructional materials.

Effects of Illustrations on Student Performance

Image-Based Accommodations for ELL Students. Some researchers strongly recommend incorporating visual supports in test items to facilitate information processing in ELL students. For example, Fichtner, Peitzman, and Sasser (1994) suggest modifying traditional tests by including more illustrations in them. Also, O'Malley and Pierce (1996) advocate allowing ELL students to use graphical organizers, such as diagrams and word maps to demonstrate their understanding of concepts and the relationships between concepts. Our ITELL project explores the effectiveness of using vignette illustrations as a form of testing accommodations for ELLs. Vignette illustrations are defined as images added to test items as a form of visual support without changing the text of the test item (Solano-Flores, 2010a; 2010b; 2011).

Shanahan (2006) examined the effects of adding illustrations on student performance in a 20 multiple-choice item life science test. She tested urban Grade 5 ELL and non-ELL students in three testing situations: test items with graphical illustrations, test items with graphic organizers, and test items with no illustrations. The results showed no statistically significant differences on student performance between items with and without illustrations when the two groups of students were compared. However, she found that ELL students with higher English proficiency were more likely than other students to be able to score higher on the items accompanied with

pictures. In conclusion, she suggested that both test designers and teachers should incorporate more illustrations in tests and instruction as contextual clues for ELL students.

Some important information is missing in this study. The author did not describe how the illustrations were developed and how the soundness of these illustrations was evaluated. This makes her findings difficult to generalize. In addition, her study was purely quantitative and did not qualitatively analyze the characteristics of the item illustration and student interpretation of the illustrations.

More recently, Martiniello (2009) explored the relationship between linguistic complexity, schematic representations, and differential item functioning (an approach to examining item bias based on item response theory) in students' solving math word problems. She employed think-aloud protocols to examine how Grade 4 Spanish-speaking ELLs utilize equations, diagrams, and figures to make meaning of the text of the items. She found that syntax, lexicon, cultural references, and layout can prevent ELLs from understanding test items, while schematic representations can mitigate the linguistic and cultural demand of the items.

Despite the interesting results, the method she adopted to categorize test illustrations is too simple. For example, all test items were first coded based on three categories: text only, primarily pictorial, and primarily schematic, and then the first two categories were grouped together as non-schematic representation. This way of categorizing test illustrations tends to neglect the complexity of features of illustrations used in tests.

Kopriva (2008) asserts that visual cues could support ELL students in comprehending test items while they acquire English and become proficient. She suggests that in ELL testing, content words such as nouns, verbs, adjectives, and adverbs were the most beneficial to illustrate.

She offers seven recommendations for effective use of illustrations for ELLs: 1) use relevant visuals, 2) use an effective visual format, 3) use illustrations to mirror text, 4) use illustrations to replace text, 5) use first person visuals with speech bubbles, 6) use visuals to organize events in time, and 7) use visuals to clarify textual meaning (Kopriva, 2008, p. 127-132).

Although these recommendations give an outline of key points that test designers should be cognizant of when incorporating illustrations in tests, they are too broad and may not be possible to apply in certain contexts.

Effectiveness of Using Illustrations in Educational Contexts. While many researchers in ELL testing argue vigorously in favor of adding illustrations to promote students' understanding of the test items, some research evidence cautions against their misuse in education. For example, Filippatou and Pumphrey (1996) examined 30 empirical studies on the effects of pictures and titles on reading performance. They maintained that not all graphics are equally helpful; if graphics do not mirror or replace text, they tend to be distracting. Consistent with this notion, evidence from the field of cognitive psychology shows that adding interesting, but conceptually irrelevant visual stimuli, such as pictures and videos, can hamper the sense-making process in learners (Harp & Mayer, 1997, 1998; Mayer, Heiser, & Lonn, 2001). To complicate matters, individual differences such as prior knowledge and spatial ability, may affect students' capability to build connections between visually presented material and its mental representation (Mayer & Sims, 1994).

Other researchers (Cohen & Hegarty, 2007; Hegarty, 2004) conducted a series of studies to examine relations between internal and external visualizations. They use the term "external visualization" to refer to external visual displays "that occur in the world and comprise both

static images, such as drawings, graphs, charts, and diagrams, and dynamic representations, such as animations” (Cohen & Hegarty, 2007, p. 701). They also define “internal visualization” as “the ability to mentally store and manipulate visual-spatial representations in the mind” (p. 701). Hegarty (2004, p. 3-10) proposed three possible relations between internal and external visualizations: 1) external visualization *substitutes* for internal visualization, 2) external visualization *depends* on the ability to internally visualize, and 3) external visualization *augments* internal visualization. She claimed that each of these three possibilities can be true for different types of visualizations, content, and people. She concluded that the interplay between internal and external visualizations was complex and dynamic, and an in-depth understanding of internal visualization abilities would be helpful in the design and use of effective external visualizations.

More recently, Cohen and Hegarty (2007) investigated individual differences in the use of external visualization to perform an internal visualization task. Thirty college students participated individually in a task that consisted of two spatial ability tests, twelve drawing trials, and manipulating interactive animations. The findings revealed large individual differences in how frequently and effectively the participants used the external visualizations. They suggest that spatial ability can affect the use of these visualizations.

In general, these studies indicate the complex relationship between textual material and its visual representation. They have far-reaching implications for developing appropriate illustrations as a whole. However, the stimulus materials they used were developed in cognitive laboratories, not taken from real textbooks or test booklets students use. Thus, there may be

limits to the extent to which these findings can be applied to the visual design of instructional and testing materials.

In the context of testing, Crisp and Sweiry (2006) explored how and when visual resources, such as pictures, diagrams, photographs, and tables affect understanding of a test item. They used six science illustrated test questions presented in two versions. In Version 1, most of the questions contained graphical elements; while in Version 2, certain graphical elements were changed intentionally. For example, in a question asking about healthy choice for a student's meal, the illustration used in Version 1 showed a plate of chicken nuggets, while the illustration in Version 2 showed unrealistically larger chicken nuggets. More than 500 high school students were randomly assigned to respond to one of the two versions of test questions. Student performance was then compared across the two versions of test items. Crisp and Sweiry (2006) found that both the characteristics of students and the features of test questions can influence how an illustrated test item is interpreted. They concluded that students tended to "have sensible expectations regarding what is important in a visual resource" (p. 151) and suggest that student ability impacts students' use of and attitudes towards test illustrations.

However, they did not provide procedures and reasonings to explain and justify how and why they changed certain elements of testing illustrations. In addition, they did not describe the participants in terms of their cultural or linguistic backgrounds. It is not clear whether and how student cultural and linguistic experiences play a role in their interpretation of certain graphical elements in the test questions.

As the studies presented in this section reveal, most of the research has focused on how students or viewers use and interpret illustrations, not on how illustrations have been developed,

selected, or structured. Diverse backgrounds of test developers and illustration designers can influence the process of creating test illustrations. Neglecting this part will result in incomplete understanding of the visual communication between designers and viewers.

Cultural Differences in Illustrations Used in Educational Contexts

Some empirical studies reveal that culture plays a critical role in existing illustrations (Horton, 1993; Tang, 1994), and in processing visual information of illustrations in educational contexts (Chua, Boland, & Nisbett, 2005; Nisbett & Miyamoto, 2005). Other studies (Boling, Eccarius, Smith, & Frick, 2004; Boling, Smith, Frick, & Eccarius, 2007; Knight, Gunawardena, & Aydin, 2009; Schiffman, 1996; Tang, 1991) suggest that conventions rooted in culture can lead to alternative interpretations of visual language.

Tang (1994) adopted three existing models developed in Western cultures to compare textbook illustrations generated in different countries. In her first part of the study, she examined social studies textbooks from Hong Kong, Japan, and Mexico. She found that the books were generally highly illustrated, contained mostly representational pictures, and appeared to use similar graphic forms and conventions. She also used a framework to code the knowledge structure of the textbook illustrations, and found that most of the illustrations were descriptive. In the second part of her study, she compared illustrations in two Grade 7 social studies textbooks generated in Hong Kong and Canada. She found that the textbooks from Hong Kong contained significantly more illustrations than the textbooks from Canada.

I caution against generalizing the results from this investigation without further considering the study design. The categories she adopted to define forms, functions, and knowledge structures of textbook illustrations were very general. They tended to be insensitive to the

possibility of hybrid varieties of features in illustrations. For example, *representational pictures* were defined in her study as “ordinary drawings and photographs that show what things look like” (Levie & Lentz, 1982, p. 214). However, drawings may include a multiplicity of representational forms, such as realistic-line drawings, silhouettes, and cartoons. It is likely that textbooks from countries with widespread use of cartoons in public media would contain more cartoony drawings than those from other countries. Neglecting these differences at a more specific level could result in misguided interpretation of similarities and differences in illustrations from different cultures.

Other studies have been conducted to investigate the cultural differences in perceptual processes. For example, Chua, Boland, and Nisbett (2005) used an eye-movement tracker to record how people from different cultures perceived visual stimuli. Twenty-five Americans and 27 Chinese participants first viewed pictures of a focal object on a complex background. After being involved in some distracting tasks for ten minutes, they were asked to look at a new set of pictures and judge whether they had seen the objects before in the study session. The results showed that Americans tended to focus on the forefront object, and they noticed the object sooner, while Chinese tended to focus more on the background. This study suggests that people from different cultures make sense of what they see in very different manners, even at the perceptual level.

However, it is very difficult to infer the findings in real educational contexts. The visual stimuli used in this study are mostly “culturally neutral photos” (p. 12630), instead of illustrations from existing textbooks or test booklets.

Boling and her colleagues (Boling, Eccarius, Smith, & Frick, 2004; Boling, Smith, Frick, & Eccarius, 2007) have explored whether the presence and absence of graphic devices in instructional illustrations affect how illustrations can be interpreted. They defined graphic devices as visual element(s) not representing objects, plants, animals, or natural phenomena which can be seen in a real life setting, but intended to extend the meaning of graphic devices, such as arrows for direction, or motion lines for action. They created and refined a 16-item survey containing illustrations with a variety of graphic devices. College students from the U.S., Malaysia, and Taiwan participated in their study. They found that in general, American students performed better than students from the other two cultures interpreting the meaning of the illustrations. In particular, U.S. college students scored significantly higher than Taiwanese college students in their interpretation of two illustrations, and they scored significantly higher than Malaysian college students in the interpretation of five illustrations. Taiwanese college students also scored significantly higher than their Malaysian counterparts on four illustrations.

The authors did not provide an explanation for these findings. But given the evidence discussed above, it could be argued that those illustrations were developed by an American illustrator, who knows better about the American culture than the other two cultures. It is likely that the American college students are more familiar with the form and style of the illustrations and it is therefore easier for them than their international counterparts to make the expected interpretations.

Tang (1991) conducted an ethnographic study to investigate the role and value of graphic representations of knowledge structures in two multicultural Grade 7 classes. The majority of participants were ELL students, identified as limited English proficient (LEP) with a wide range

of abilities. Except for art, music and physical education, classes in all subject areas were intensively observed for 163 class periods. Besides the observations, she examined textbooks, instructional materials, social studies projects, science projects, reading assignments, test papers, and journals. She also conducted interviews with students and teachers.

Tang's findings revealed that, despite exposure to large quantities of illustrations in instruction and curriculum, students generally showed negative attitudes towards the illustrations and were unable to extract valuable information from illustrations to facilitate their learning. Teachers tended to use only a small proportion of the illustrations in class and spent little time guiding students to interpret and use them. Most students tended not to recognize the relationship between text and accompanying illustrations; when information was presented in both text and graphic form, students focused primarily more on the written texts.

This study shows that it is important for teachers to become aware of the potential power of textbook illustrations and utilize the illustrations in everyday instruction to promote active and in-depth learning. Mostly likely, without being well informed of the role of illustrations in textbooks, students cannot effectively interpret the information embedded in the illustrations. Therefore, it is very likely that they will not feel comfortable and will not be able to interpret test illustrations as expected.

In general, the results from these investigations speak to the notion that culture plays a role in visual perception and interpretation. However, most of them fail to control for the effect of the cultural origin of the illustrations used as a factor that may bias the results. An illustrator "must know the culture of the intended audience and the symbols in the images must be understood by that culture" (Lester, 2003, p. 108). In developing and reviewing illustrations in science tests for

culturally and linguistically diverse students, test developers, illustrators, and illustration reviewers should know how the social and cultural factors affect the meaning-making process, and therefore develop a systematic approach to examining the illustrations to ensure that they are valid for these students.

Analysis of Illustrations

Analysis of Test Illustrations. Very few empirical research studies have attempted to examine and analyze illustrations used in test items. A common thread running through these few attempts is the lack of a systematic framework that is sensitive to the complexities of features of test illustrations.

Sharrocks -Taylor and Hargreaves (1999) outlined three types of illustrations in testing materials: decorative elements that are not related to the test item, relevant illustrations that present and support the text of the item to some extent, and essential illustrations that are not repeated in the text of test items but are referred to and expected to be understood to answer the questions. Their attempt was still on an initial stage for analyzing test illustrations. The three types of illustrations they identify are very general, and do not seem to be sensitive to the complexity of features of test illustrations. It is not clear how they defined the second type of illustration as presenting the text of the item to some extent, and how they differentiated between the latter two types.

Ginther (2001) examined the effects of the presence and absence of illustrations on student performance on computer-administered version of the Test of English as a Foreign Language (TOEFL). She first classified illustrations used in the TOEFL listening sub-test, such as pictures, diagrams, and still photos of the speakers, and then further divided them into two types: context

visuals and content visuals. She defined context visuals as those which provide “information about the context in which the verbal exchanges occur” (p. 2), and content visuals as those are “related to the content of the verbal stimulus” (p. 2).

Apparently, Ginther’s definitions of illustrations are directly based on their relations with the written text of the test items. However, this approach is too general and not sensitive to the intricacies of features of test illustrations. For example, the context and content of a verbal stimulus can be either simple or complicated, which can affect the level of complexities of the accompanying illustrations and, therefore, students’ interpretations. This lack of conceptual clarity makes interpretations of her study results difficult.

In her doctoral dissertation, Shanahan (2006) treated illustrations in life science as graphic illustrations and graphic organizers, and broke down the latter into tables and graphs. Based on my observation of illustrations from samples of large-scale science tests, test item illustrations vary considerably in terms of form and style. Thus, as with other investigations (see Sharrocks-Taylor and Hargreaves, 1999; and Ginther, 2001), Shanahan’s approach to analyzing test illustrations is too general and fails to capture the complexities of nuanced key aspects of illustrations in test items.

More recently, Hatzinikita, Dimopoulos, and Christidou (2008) analyzed and compared the linguistic and visual modes of both the PISA released science test items and Greek school science textbooks. They used an analytic grid (previously used by Dimopoulos, Koulaidis, and Sklaveniti, 2003) for examining illustrations along three dimensions: classification, framing, and formality. As shown in Table 1, these three dimensions consist of operational variables at three levels: strong, moderate, and weak.

Table 1

The Grid of Analysis for the Visual Code of Images in Science Tests and Textbooks Based on Dimensions of Classification, Framing, and Formality. Sources: Hatzinikita, Dimopoulos, & Christidou. (2008). PISA test items and school textbooks related to science: A textual comparison. *Science Education*, 92(4), 664-687. Dimopoulos, Koulaidis, & Sklaveniti. (2003). Towards an analysis of visual images in school science textbooks and press articles about science and technology. *Research in Science Education*, 33, 189-216.

Dimension of analysis		Level		
		Strong/ High	Moderate	Weak
Classification marker	Type of visual image	Conventional	Hybrid	Realistic
	Function of visual image	Classificational/ analytical		Narrative/Metaphorical
	Overall classification	Conventional/hybrid and classificational/ analytical	Other combinations	Realistic and Narrative/Metaphorical
Framing marker	Vertical angle of shot	Low angle	Eye-level	High angle
	Distance of shot	Distant	Medium	Close
	Horizontal angle of shot	Oblique		Frontal
	Overall framing	All markers are strong	Other combinations	All markers are weak
Formality marker	Elements of techno-scientific code	Geometrical shapes and alphanumeric strings	Geometrical shapes or alphanumeric strings	Absence of any geometrical shapes or alphanumeric strings
	Color differentiation	Monochrome	2-4 colors (black included)	More than four colors
	Color modulation	No shade	1-3 shades	All the shades picked up by the photographic lens
	Contextualization	Background of the same color as the rest of the page	Mono or bi-chromatic background	Background according to the photographic realism
	Overall formality	At least three markers take a high value, or two markers take a moderate and two take a high value	At least three markers take a moderate value, or two markers take a high and two take a low value	At least three markers take a low value, or two markers take a moderate and two take a low value

According to Dimopoulos, Koulaidis, and Sklaveniti (2003), classification “concerns all the elements involved in their syntactic construction that denote the degree of their content specialization” (p. 192), the framing “corresponds to the social-pedagogic relationships that tend to be established between the message of the images and their viewers” (p. 192), and the formality “corresponds to the degree of abstraction characterizing the techno-scientific images” (p. 193).

To my knowledge, this analytic framework is one of the most comprehensive socio-semiotic approaches for analyzing visual images in test. Still, it has some serious limitations. The main one is that it does not enable for analyzing test illustrations such as tables and graphs. The authors did not define what illustrations they analyzed, but it appears that their analysis mainly included photos and line drawings, which depicted objects or phenomenon observable in real life contexts. However, there are a considerable number of tables, graphs, and hybrid of tables, graphs with realistic drawings. Failing to analyzing these test illustrations produces an incomplete account of what test illustrations are and how they are designed and implemented.

Another disadvantage is that this framework does not allow for analysis of mutually inclusive cases. With most variables at three levels, each visual image is assumed to have only one level. However, in some illustrations used in tests several levels co-exist, which would not be captured using this tool. Moreover, some terms in the framework overlap and are not well-defined. For example, *classification* is used to refer to a dimension, and simultaneously to the high level of function of visual images. This creates conceptual confusion and makes interpretation and use of this tool difficult.

Analysis of Instructional Illustrations. A growing body of research has been conducted to examine instructional illustrations in a variety of subject areas. Some researchers (Fleming, 1966; Goldsmith, 1987; Weidenmann, 1994) focus on line drawings and photographs in textbooks, while others (Kosslyn, 2006; Winn, 1987) attend to graphic organizers, such as diagrams, graphs, charts, and tables.

Fleming (1966) created and refined a taxonomy for analyzing characteristics of illustrations in textbooks in subject areas such as science, mathematics, and social sciences. He identified eleven physical attributes as scales for classifying samples of instructional illustrations.

Goldsmith (1987) created an analytic model for analyzing illustrations in educational books for students, including textbooks from other cultures. He identified twelve elements resulting from four visual codes: unity, location, emphasis, and text parallels. Also, he identified three semiotic levels: semantic, syntactic, and pragmatic. He strongly suggested that designers, illustrators, publishers, and teachers should work cooperatively to make illustrations more effective. Their attempts are very enlightening for my study in terms of the detailed analysis of key attributes of illustrations. However, some of the terms and concepts are very difficult to be operationalized.

In his article “Classification and Analysis of Instructional Illustrations,” Malcolm Fleming (1966) developed a classification system for analyzing textbook illustrations. Although the study was conducted over 40 years ago, how he defined illustration and what physical attributes of textbook illustrations he identified give me hints to develop operational definitions around basic elements of test illustration for my investigation.

He designed a taxonomy of illustrations used in textbooks, tested and refined the taxonomy in regard to a sample of textbook illustrations, and then conducted a systematic analysis of the

sample of textbook illustrations based on the revised taxonomy. In particular, he defined illustration as “a picture plus interior titles and labels as well as adjacent caption” (Fleming, 1966, p. 247), and identified three basic elements of illustrations: pictorial, verbal, and design. He considered “number lines, geometric figures, structural chemical formulae, curves, graphs, and time lines (Fleming, 1966, p. 247)” as borderline cases of pictorial elements. He then described four taxonomies used to analyze instructional illustrations: physical type taxonomy, objective type taxonomy, and verbal modifier taxonomy. Within physical type taxonomy, he used eleven observable physical attributes as scales to classify instruction illustrations. Those scales were: “(1) Area, (2) Framing, (3) Shape, (4) Position, (5) Elements, (6) Chroma, (7) Achroma, (8) Encoding Style, (9) Encoding Medium, (10) Information level, (11) Unification (Fleming, 1966, p. 247).” From the population, 40 textbooks with 10 from each of four subject areas as English, history, mathematics, and science at Grade 8 level, a 6% proportional random sample was taken which yielded 787 illustrations. He performed a series of Chi-square analyses and found significant associations between illustrations, educational objectives, and the subject matter of the illustrations.

This study uses an intriguing approach to identifying and analyzing basic elements of textbook illustrations. But the description of the procedure for selecting the sample and of the taxonomy of illustrations is not clear. For example, the author did not provide a detailed step-by-step description of the procedure for selecting the sample. It is not clear how the sample was selected. His definitions of the population and the sample are not easy to follow. Without a clear explanation of how proportional random sampling was carried out, it is not possible for other researchers to replicate the procedure and obtain similar results.

In the Methods section, Fleming (1966) only provided superficial definitions regarding the elements common to textbook pages, including pictorial elements, verbal elements, and design elements. Illustration was defined by picture: “an illustration was what would commonly be considered a picture plus interior titles and labels as well as adjacent caption.” Furthermore, his description of taxonomy of physical type, objective type, and verbal modifier is not clear. For example, 11 observable physical attributes of instructional illustrations were identified without giving operational definitions. Also, only the fifth attribute was described with instructional illustrations to demonstrate the coding process, and the other 10 attributes were not discussed.

Garrick (1978) discussed the design of instructional illustrations in medicine based on a selected review of literature. She attempted to explore issues such as the importance of illustrations to aid learning, effectiveness of illustrations, best visual medium, style of illustrations, and color and labeling in illustrations. She suggested that five factors should be borne in mind in the design and use of instructional illustrations in medicine. These five factors are: “1) the nature of the subject matter to be learnt, 2) the purpose of the instruction, 3) the function of the visual materials, 4) basic design principles, and 5) methods of presentation (Garrick, 1978, p. 168-171).” She concluded that findings from research studies were generally inconsistent, and suggested that a strong theoretical framework for interpreting the design and use of instructional illustrations was needed.

More recently, Alley (1994) analyzed the use of illustrations in 18 university introductory Spanish textbooks over three decades, from 1960 to 1989, and discussed implications for proper use of illustrations in teaching. *Illustration* in his study includes “photographs, drawings, diagrams, maps, graphs, tables, and charts as well as reproductions of authentic documents or

realia” (p. 489). He examined illustration from five aspects, including the percentage of illustrated pages within a text, size of illustrations, percentage of color illustrations, frequency of use of different types of illustrations, and the relation between the illustration and the corresponding written text. The results showed that illustrations in contemporary Spanish as a foreign language textbooks increased significantly in both quality and quantity with respect to those in past textbooks, and confirmed the integral role of illustrations in teaching Spanish as a foreign language. He also pointed out that the use of illustrations often was guided by the intuition of the textbook publishers rather than on empirical research findings. He concluded that many questions regarding the composition and arrangement of instructional illustrations remain unanswered.

Alley’s (1994) analysis provides some insights into the appropriate design and organization of instructional illustrations. However, his approach still rests on general categories of grouping illustrations, and thus fails to account for the multiple physical attributes existing in illustrations from textbooks and test booklets.

As mentioned before, Tang (1994) explored and compared textbook illustration across languages and cultures, and identified common types and forms of graphic representations. In the first part of her study, she examined the quantity and form of illustrations in history or social studies textbooks used in Hong Kong, Japan, and Mexico. In her second part of the study, she compared Grade 7 social studies textbook in Hong Kong and Canada from four aspects: quantity, form, clarity of reference to illustrations, and functions of the illustrations in relation to the text. To compare forms of illustrations, she used Levie and Lentz’s (1982) framework to classify forms of illustrations into two types: representational pictures and nonrepresentational pictures,

with the latter including maps, diagrams, and graphic organizers. To compare functions of illustrations, she employed Hunter's (Hunter, Crismore, & Pearson, 1987) model for categorizing graphic functions into five types: embellishment, reinforcement, elaboration, summarization, and comparison. She also used Mohan's (1986) knowledge framework to analyze the illustration along six dimensions: description, classification, principles, evaluation, temporal sequence, and choice/decision making.

Her study provides us with a cross-cultural perspective for scrutinizing instructional illustrations. However, the approaches she adopted were originally used to analyze illustrations generated in the Western culture, and might not be able to capture the subtle differences in illustrations generated in other cultures.

Overall, these studies reveal that among the multiple approaches to analyzing illustrations in educational contexts, little research has systematically examined the features of test illustrations originated in different cultures. Therefore, investigation on the characteristics of illustrations used in large-scale science tests from different countries and their relation to student performance is necessary to further understanding of their use and design. Identifying cultural differences in the features of item illustrations will inform the process of item development and will contribute to ensuring more valid assessment of culturally diverse populations in both the U.S. and the context of international test comparisons.

CHAPTER IV

METHODS

Design

The study primarily utilizes quantitative techniques to analyze characteristics of illustrations used in large-scale science tests for secondary school students in China, the U.S., and TIMSS. More specifically, it investigates the differences in the frequencies of different features of illustrations across item types, subject areas, types of assessment system, and national origin. It also examines, where appropriate data are available, the correlations between the complexities of features of test illustrations and students' performance on the corresponding science test items. Additionally, it uses symmetry graphs to investigate the complexity of illustration features at the individual IDV level. It also compares and contrasts exemplar illustrated items from the same sub-content areas across the three assessment systems.

Data Collection

Original Item Corpus. I collected illustrated items from state-, municipal-, national-, and international-level assessment systems. For the state or municipal level of assessments, I intended to obtain most comparable data across the two countries. Thus, I included tests from the most populated areas in China and the U.S. The Chinese tests were generated in four municipalities directly under the control of the central government, including Shanghai, Tianjin, Chongqing, and the capital city, Beijing. These four cities are regarded as equal to provinces in terms of political status, and are the top four most populated cities in China (National Bureau of Statistics of China, 2009). The U.S. tests were generated in the capital city, Washington D.C., and the three most populated states, California, Texas, New York (U.S. Census Bureau, 2010).

The corpus of items included over 1,500 released illustrated multiple-choice and constructed-response items from state-, municipal-, national-, and international-level assessment systems (see Table 2). These items were across four content areas: earth science, life science, physics, and chemistry, and at three grade levels: Grades 4/5, 8/9, and 12.

Table 2

Sources of Illustrated Items in the Corpus: Assessment Programs From China Four Municipalities, U.S. Four States, and TIMSS.

Origin	Assessment program	Year
Chinese, Municipal	Shanghai Middle School Exit Examination	2008-2009
	Tianjin Middle School Exit Examination	2008-2009
	Chongqing Middle School Exit Examination	2008-2009
	Beijing Middle School Exit Examination	2008-2009
	Shanghai High School Exit Examination	2008-2009
	Tianjin High School Exit Examination	2008-2009
	Chongqing High School Exit Examination	2008-2009
	Beijing High School Exit Examination	2008-2009
American, State	Texas Assessment of Knowledge and Skills	2006, 2009
	The University of the State of New York science Test	2007-2010
	District of Columbia Comprehensive Assessment System	2009
	California Standards Test	2006-2008
American, National	National Assessment of Educational Progress (NAEP)	2000, 2005, and 2009
International	Trends in International Mathematics and Science Study (TIMSS)	1995, 1999, 2003, and 2007

The assessment program in this corpus contained items that were representative of science testing respectively at state or municipal, national, and international levels. For instance, CMSEE and China's High School Exit Examination (CHSEE) are designed, administered, and scored at the municipal level in China, while in the U.S. the state tests are developed, administered, scored at the state level. TIMSS is an international test of knowledge and skills in Mathematics and Science administered every four years to Grade 4 and Grade 8 students in over 50 countries in the world, and whose results play a critical role in the education systems of many countries, including the U.S. (see National Center for Education Statistics, 2009). For TIMSS, I also had access to information on item *p*-value—a measure of item difficulty—for the released illustrated science test items (Note 2). See Appendix A for detailed historical and contextual information on the science assessments included in the dissertation study.

Item Selection. To ensure statistical power, I had to eliminate from my investigation certain combinations of grade level, and content area, for which only few illustrated items were available. For example, Chinese large-scale assessment programs do not have items for Grade 4/5, while TIMSS does not have items for Grade 12. Likewise, items in chemistry are not available in many assessment systems. In the case of NAEP, few released illustrated items in the four content areas at the three grade levels are available. Therefore, I focused on illustrated items for Grade 8/9 in earth science, life science, and physics (Table 3). Altogether, 416 released illustrated items were analyzed in my study.

Table 3

Sample of Illustrated Items to be Analyzed in This Study: Number of Released Items Across Three Content Areas at Grade 8/9^a From China Four Municipalities, U.S. Four States, and TIMSS.

Assessment system	Content area			
	Earth science	Life science	Physics	Total
CN 4 Municipalities	46	34	72	152
US 4 States	43	74	74	191
TIMSS	17	16	40	73
Total	106	124	186	416

^a While in the U.S. large-scale science tests on all science content areas are administered in Grade 8, in China large-scale tests are administered in Grade 8 for Earth and Space Science and Life science, and in Grade 9 for Physics and Chemistry.

Coding

The coding of the features of the item illustrations used and elaborated on the coding system for examining the characteristics of illustrations used in tests that we (Solano-Flores & Wang, 2011; Wang & Solano-Flores, 2011) have developed. Unlike other approaches for examining images (e.g., Evans, Watson, & Willows, 1987; Hunter, Crismore, & Peason, 1987), this coding approach is not based on characterizing images simply by giving them general labels such as *chart*, *diagram*, or *graph*, whose meanings may lead to different interpretations and which are not sensitive to the complexity of features of item illustrations. Rather, test illustrations are examined in terms of multiple dichotomous variables grouped in six dimensions that refer to the form, style, and complexity with which information is presented to the examinee. This coding system have developed based on knowledge from relevant research studies, discussion with science teachers and bilingual teachers, and examination about 600 illustrated items from TIMSS, PISA, and national and state assessment programs from China and the U.S.

In this study, illustration dichotomous variables (IDVs) were regarded as units of analysis. IDVs are defined as features whose combined presence or absence determines the characteristics of a given illustration (Solano-Flores & Wang, 2011). *Illustration* is defined as a mainly non-textual device whose characteristics can be described according to sets of IDVs. *Mainly non-textual* addresses the fact that illustrations commonly used in science printed materials include various forms of text, such as signs, labels, and legends.

Item illustrations were coded according to sets of IDVs, which describe whether and how an illustration: “1) represents objects and background (e.g., image concreteness, presence of a background, relative scale of objects), 2) uses metaphorical visual language (e.g., through visual description or visual narration), 3) uses text (e.g., through labels, legends, comments, and directions), 4) encrypts information (i.e., by representing variables/constants with values or symbols, by representing functions with graphs or formulas, by representing processes as linear or cyclical structures), 5) interacts with the text of the item (e.g., the text of the item makes explicit reference to the illustration, the student is required to respond to the item by drawing or writing on an illustration)” (Solano-Flores & Wang, 2011, p. 4-5) and 6) provides context (e.g., physical context, socio-historical context). Table 4 shows the six major dimensions, corresponding subcategories, and sets of IDVs. See Appendix B for definitions and attributes of each of the coding dimensions.

Table 4

Illustration Dimensions (bold letters), Categories (underlines), and Variables (italics) Identified in the Conceptual Framework for Designing and Examining Science Test Item Illustrations. Sources: Solano-Flores, G., & Wang, C. (2011); Wang, C., & Solano-Flores, G. (2011)

1. Representation of objects and background

- 1.1. Image concreteness: 1. *photograph*, 2. *scanned document*, 3. *realistic line drawing*, 4. *schematic*, 5. *map*, 6. *silhouette*, 7. *cartoon*, 8. *logo/icon/ metonymy*, 9. *emblem*, 10. *symbol*, 11. *reference*, 12. *entity*, 13. *X-ray, MRI, sonogram, thermography, spectrometry*, 14. *image in an object*, 15. *other*
- 1.2. Background: 1. *with background*, 2. *without background*, 3. *Other*
-

Table 4 continues

Table 4(continued)

1.3. <u>Zooming</u> : 1. zoom-in, 2. zoom-out, 3. zero zooming naked eye, 4. other
1.4. <u>View</u> : 1.external, 2.internal, 3. from above object, 4. from below object, 5. from side of object, 6. other
1.5. <u>Dimension</u> : 1. three dimensional, 2. two dimensional, 3. other
1.6. <u>Relative scale of objects</u> : 1. proportionate, 2. disproportionate, 3. Other
1.7. <u>Color</u> : 1.black & white, 2. multicolor, 3. gray scale, 4. Other
1.8. <u>Composition</u> : 1.single image, 2. compound image, 3. other
6. Context in illustration^a
6.1. <u>Physical context</u> :1. undefined person(s), 2. Peer(s)/teacher(s), 3. media figures (e.g. cartoon), 4. family/home, 5. school/class/lab/gym, 6. community/neighborhood, 7. state/province, 8. identified country, 9. unidentified/fictitious country, 10. world/global
6.2. <u>Socio-historical context</u> : 1. events in domestic affairs, 2. events in international affairs, 3. traditions and customs
6.3 <u>Human figure</u> : 1. complete human figure, 2. partial human figure (e.g. a head), 3. one human figure, 4. two or more human figures
2. Metaphorical visual language
2.1. <u>Space, time, and motion</u> : 1. space and location, 2. time and sequence, 3. dynamics and flow, 4. Magnifying glass/lens, 5. other
2.2. <u>Matter and energy</u> : 1. states of matter (e.g., gas), 2. temperature (e.g., coldness), 3. light and electricity (e.g., brightness), 4. sound (e.g., noise) , 5. force and impact, 6. chemical condition, 7. Light reflection, 8.other
2.3. <u>Human state</u> : 1. senses (e.g., seeing), 2. speech and cognition/thinking (e.g., utterance), 3. physical condition (e.g., freezing), 4. emotion (e.g., sadness), 5. other
3. Text in illustration
3.1. <u>Text unit</u> : 1. non-scientific/mathematical sign, 2. scientific/mathematical sign, and notation, 3. abbreviation, 4. Roman numeral, 5. Arabic numeral, 6. letter, 7. word/词, 8. phrase /短语, 9. sentence/句子, 10. paragraph/段落, 11. acronym/缩略语, 12. other
3.2. <u>Text function</u> : 1. label, 2. legend/code, 3. title/caption/heading, 4.elaborate/explain/state/describe, 5. opinions/comment/note, 6.instructions, 7. data, 8. text on object, 9.other
3.3. <u>Text emphasis</u> : 1.capitalization, 2.bolding, 3.italicizing, 4.underlying, 5.circling, 6.other
3.4. <u>Text direction</u> : 1. from left to right or vice versa, 2. from top and bottom or vice versa, 3. oblique direction, 4. other
4. Representation of variables, constants, and functions
4.1. <u>Variables and constants</u> : 1. cases/facts/events, 2.categories, 3. conditions/treatment/properties/locations, 4. levels, 5. analogic line, 6. value, 7.scale, 8. unit, 9.name, 10.other
4.2. <u>Function</u> : 1. discrete: graphic (e.g., bar graph), 2. discrete: numeric (e.g., table), 3. discrete: textual, 4. nodes/arcs, 5. formula/equation, 6.scientific/mathematical symbol, 7. other
4.3. <u>Structure</u> : 1. sequential, 2. tree, 3. cycle, 4. network, 5. composition, 6. Inclusion, 7. other
5. Illustration-text interaction
5.1. <u>Location of illustration</u> : 1. above stem/prompt, 2. between stem/prompt and options/response format, 3. embedded in stem/prompt, 4. embedded in options/response format, 5. on the left of the text, 6. on the right of the text, 7. under the text, 8. other
5.2. <u>Reference to illustration</u> : 1. explicit, 2. not stated, 3. other
5.3. <u>Stated actions to perform with the illustration</u> : 1. observe or examine, 2. write, draw, or mark on illustration provided, 3. generate an illustration, 4. no action stated, 5. other
5.4. <u>Commonality</u> : 1. part of a stand-alone item, 2. same illustration for several items, 3. illustration for one item with a series of related items, 4. other

^a Based on the development of understanding of nuanced aspects of test illustration, the dimensions in this coding system have gone through a number of changes. The dimension on Context in illustration has been added to the former five-dimensional coding system as a reflection of suggestions from the dissertation committee members after the prospectus defense. This dimension fits well as the second dimension. However, to ensure the comparability of findings in several relevant studies, I decided to keep all the number used originally to designate the dimension, and designated this dimension as the sixth dimension.

IDVs and Examples. Tables 5-10 contain the definitions of the variables used in this study (See Appendix C for a complete list of examples, definitions, and coding rules for each IDV). The first dimension concerns style, complexity, and level of concreteness or abstraction with which objects and background are represented. Table 5 shows the definitions of IDVs belonging to the first dimension.

Table 5

Definitions of IDVs in Coding Dimension on Representation of Objects and Backgrounds

Subcategory	Definition
Image concreteness	Level of realism in representing object(s) or event(s).
1. Photograph	Image captured by a camera.
2. Scanned document/ Text clip	Digital form of a written or printed material.
3. Realistic line drawing	Image drawn with lines which represent object, event, and/or background the way they look in real life.
4. Schematic	Image that shows essential components of object, event, or the physical structure of something in a simplified manner.
5. Map	Representation of land or a place, as seen from above.
6. Silhouette	Shadow image of object filled in with solid black.
7. Cartoon	Image that represents object and event in a humorous or stereotypical manner, typically exaggerating some feature of the object.
8. Logo/Icon/Metonymy	Simplified or stylized image of an object that represents a concept, abstract idea, event, or class of objects.
9. Emblem	Image of an object with a cultural-historical meaning embedded as the representation of a concept, abstract idea, event, or class of objects.
10. Symbol	Image that conventionally represents a concept, abstract idea, event, or class of objects and whose appearance does not have any resemblance with the concept, idea, or object.
11. Reference	Image that represents the existence of object by showing some of its elements, or its elements in a simplified manner.

Table 5 continues

Table 5 (continued)

Subcategory	Definition
12. Entity	Image that represents the existence of object in an abstract manner.
13. X-ray, MRI, sonogram, thermography, spectrometry, etc.	Image of an object obtained through technology that is sensitive to physical properties or changes the human eye cannot see.
14. Image in an object	Image being illustrated as part of the nature of an object to help interpret the related object, but not necessarily intended for the viewer to know what exactly it is (e.g., an image of an active computer screen).
Background	Visual element intended to provide a context for the proper interpretation of the focal object or event.
1. With background	Illustration that depicts not only the focal object or event, but also the set of visual elements intended to provide a visual perspective.
2. Without background	Illustration that depicts only the focal object or event.
Zooming	Shot distance of object or event.
1. Zero zooming	Naked-eye view.
2. Zoom-in	Magnified, detailed view of object as seen by a microscope.
3. Zoom-out	View of an object from a distance in a way that allows to see it in its entirety as seen by a telescope.
View	Position from which an object is observed.
1. External	Object seen from outside.
2. Internal	Object seen from inside.
3. From above object	Object seen from above.
4. From below object	Object seen from below.
5. From side of object	Object seen from the side.
Dimension	Measurement of length, width, and/or depth of object in an illustration.
1. Three dimensional	Object shown in length, width, and depth.
2. Two dimensional	Object shown only in length and width.

Table 5 continues

Table 5 (continued)

Subcategory	Definition
Relative scale of objects or components	Degree of size proportionality between objects or components.
1. Proportionate	Objects shown at a consistent scale.
2. Disproportionate	Objects not shown to scale.
Color	Visual spectrum range.
1. Black and white	(Self-explaining).
2. Multicolor	(Self-explaining).
3. Gray scale	(Self-explaining).

The second dimension concerns visual components included with the intent to clarify and enhance understanding of objects and actions in an illustration. These visual components have been previously referred to as graphic devices (Boling, Eccarius, Smith, & Frick, 2004). For the purpose of this work, *graphic device* is defined as any visual component included in test illustrations with the intent to ensure proper understanding of illustrations. Graphic devices usually consist of arrows, motion lines, speech balloons, and many other conventional forms to enhance representation of actions and states. Table 6 shows the definitions of IDVs belonging to this dimension.

Table 6

Definitions of IDVs in Coding Dimension on Metaphorical Visual Language

Subcategory	Definition
Space, time, and motion	Visual device intended to ensure clarity in the interpretation of an event or property of an object that is difficult to show in a static image.
1. Space and location	Visual device intended to distinguish an object from others, or to indicate location or detail.

Table 6 continues

Table 6 (continued)

Subcategory	Definition
2. Time and sequence	Visual device intended to show related movements in a particular order in time.
3. Dynamics and flow	Visual device intended to show movement and action, or direction of movement and action.
4. Magnifying glass/lens	Visual device intended to show an enlarged part of an object in detail as seen through a magnifying glass or a microscope lens (e.g. showing the frame of a magnifying glass or a microscope lens).
Matter and energy	Visual device intended to clarify and enhance interpretation of states of matter and energy.
1. States of matter	Visual device intended to show liquid, gaseous, or solid state.
2. Temperature	Visual device intended to show that something is hot or cold.
3. Light and electricity	Visual device intended to show the source or existence of illumination or electric current.
4. Sound	Visual device intended to show different kinds of sound, such as noise, snoring, and music.
5. Force and impact	Visual device intended to show the intensity of force and/or its impact.
6. Chemical condition	Visual device intended to show the condition of something caused by a chemical reaction.
7. Light reflection	Visual device intended to show shading or the reflection of light.
Human state	Visual device intended to show an internal condition that is difficult to show in a static image.
1. Senses	Visual device intended to show perception through any of the five senses: seeing, smelling, hearing, tasting, or touching.
2. Speech and cognition/thinking	Visual device intended to show people's talking or thinking.
3. Physical condition	Visual device intended to show the state of a person's body, such as freezing or warming up.
4. Emotion	Visual device intended to show a strong feeling, such as being sad or angry.

The third dimension concerns form and complexity of textual information in test illustration.

Table 7 shows the definitions of IDVs belonging to this dimension.

Table 7

Definitions of IDVs in Coding Dimension on Text in Illustration

Subcategory	Definition
Text unit	Written textual component in an illustration.
1. Non-scientific/mathematical sign	Mark, figure, or symbol conventionally used to represent something not relating to science and/or mathematics.
2. Scientific/mathematical sign, and notation	Mark, figure, or symbol conventionally used to represent something in science and/or mathematics.
3. Abbreviation	Shortened form of a word.
4. Roman numeral	(Self-explaining).
5. Arabic number	(Self-explaining).
6. Letter	(Self-explaining).
7. Word/词	(Self-explaining).
8. Phrase/短语	Group of two or more words without a verb.
9. Sentence/句子	Group of words containing a subject and a verb.
10. Paragraph/段落	Section of a written text consisting of several related sentences that convey meaning on the same topic.
11. Acronym/缩略语	Abbreviation of a set of words formed with the initial letters or characters of those words.
Text function	Special purpose of the text used in an illustration.
1. Label	Text of an illustration intended to point at a component in an illustration.
2. Legend/ code	Text of an illustration intended to explain symbol used in an illustration, especially in a map.
3. Title/caption/heading	Text of an illustration intended to provide a concise explanation of the illustration as a whole.
4. Elaborate/explain /state/describe	Text of an illustration intended to give a detailed explanation of a component in the illustration.

Table 7 continues

Table 7 (continued)

Subcategory	Definition
5. Opinions/ comment/note	Text of an illustration intended to express an opinion or comment.
6. Instructions	Text of an illustration intended to provide direction and information on what is expected to do.
7. Data	Text of an illustration intended to provide data.
8. Text on object	Text that is part of an object (as in a ruler or a scale), included in the illustration with the intent to represent the object accurately but not necessarily with the intent to be read by the viewer.
Text emphasis	Special font style of text in an illustration to accentuate the importance of something.
1. Capitalization	(Self-explaining).
2. Bolding	(Self-explaining).
3. Italicizing	(Self-explaining).
4. Underlying	(Self-explaining).
5. Circling	(Self-explaining).
Text direction	Angle in which the text in an illustration is shown.
1. From left to right or vice versa	Text in an illustration shown horizontally.
2. From top to bottom or vice versa	Text in an illustration shown vertically.
3. Oblique direction	Text in an illustration shown slant.

The fourth dimension concerns form and complexity with which variables and constants are represented. Table 8 shows the definitions of IDVs belonging to this dimension.

Table 8

Definitions of IDVs in Coding Dimension on Variables and Constants

Subcategory	Definition
Variables and constants	Level of concreteness or abstraction with which variables and constants are represented.

Table 8 continues

Table 8 (continued)

Subcategory	Definition
1. Cases/facts/events	Concrete examples of particular objects, persons, or situations representing different categories or levels of a variable.
2. Categories	Concrete representation of generic classes or groups of things in a system.
3. Conditions/treatment /properties/locations	Concrete representation of objects or persons under different conditions that result from experimental manipulation.
4. Levels	Attributes of objects or persons rank-ordered on a continuum of values of variables.
5. Analogic line	Continuous straight or curved line in a Cartesian (x-y axes) coordinate system.
6. Value	Numerical values of variables and/or constants.
7. Scale	Values of a variable showing different magnitude values.
8. Unit	Quantity of a variable regarded as a standard.
9. Name	Name of a variable.
Functions	Representation of the relationship between variables.
1. Discrete: Graphic	Representation of the relationship between variables for a limited number of values.
2. Discrete: Numeric	Numeric data in an orthogonal arrangement.
3. Discrete: Textual	Textual data in an orthogonal arrangement.
4. Discrete: Nodes/Arcs	Representation of the representation between discrete (1-0) variables as dots (or nodes) and lines (arrows or arcs).
5. Formula/equation	Representation of a variable as a formula or as the function between other variables according to a system of notation conventions.
6. Scientific/mathematical symbol	Symbol used to represent a variable or constant commonly used in a scientific and/or mathematical discipline.
Structure	Representation of discrete, dichotomous variables.
1. Sequential	Single path, linear order of a series of steps or stages in a process or procedure.
2. Tree	Branching (hierarchical, breaking down, genealogical) structure of relationship between components.

Table 8 continues

Table 8 (continued)

Subcategory	Definition
3. Cycle	Process or procedure in which stages or steps repeat regularly or iteratively in the same sequence.
4. Network	Multi-path interaction among different components in a complex system.
5. Composition	Components of a whole.
6. Inclusion	Representation of a relationship of components in which some components are part of or include other components.

The fifth dimension concerns ways in which test illustration is used in combination with the text of the item to provide information to examinees or to capture their responses. Table 9 shows the definitions of IDVs belonging to this dimension.

Table 9

Definitions of IDVs in Coding Dimension on Interaction between Illustration and Item Text

Subcategory	Definition
Location	Position of an illustration relevant to the text of the test item.
1. Above stem/prompt	Illustration located above stem (in a multiple-choice item) or prompt (in a constructed-response item).
2. Between stem/prompt and options/response format	Illustration located between the stem/prompt and the options/response format.
3. Embedded in stem/prompt	Illustration embedded in a stem or prompt.
4. Embedded in options/response format	Illustration embedded in options or response format.
5. On the left of the text	Illustration located on the left of the text of item.
6. On the right of the text	Illustration located on the right of the text of item.
7. Under the text	Illustration located under the text of item.
Reference to illustration	Mention of the existence of an illustration in the text of the test item.

Table 9 continues

Table 9 (continued)

Subcategory	Definition
1. Explicit	The text of the item mentions the illustration explicitly.
2. Not stated	The text of the item does not mention the illustration.
Stated actions to perform with the illustration	Explicit guidance for examinees about actions relating to the use of an illustration in the test item.
1. Observe or examine	Text of the test item asks examinees to look at or examine the illustration.
2. Write, draw or mark on illustration provided	Text of the test item asks examinees to write, draw, or mark on illustration provided.
3. Generate an illustration	Text of the test item asks examinees to draw an original illustration.
4. No actions stated	Text of the test item requests no actions from examinees to perform with the illustration.
Commonality	Attribution of illustration relating to the independence of the test item.
1. Part of a stand-alone item	Illustration depicted specifically for one independent test item.
2. Same illustration for several items	Illustration depicted for a bundle of related test items.
3. Illustration for one item within a series of related items	Illustration depicted specifically for one item which belongs to a bundle of related test items.

The sixth dimension concerns form and complexity with which physical or socio-historical contexts are represented. Table 10 shows the definitions of IDVs belonging to this dimension.

Table 10

Definitions of IDVs in Coding Dimension on Context in Illustration

Subcategory	Definition
Physical context	Presence of individual and place in a physical context.
1. Undefined person(s)	Individual referred to as fictitious person(s) in the text of the item.

Table 10 continues

Table 10 (continued)

Subcategory	Definition
2. Peer(s)/teacher(s)	Individual referred to as peer(s) or teacher(s) in the text of the item.
3. Media figures	Well-known individual or character in public media (e.g., a cartoon, an actor).
4. Family/home	Reference, in the text of the item, to an individual in the family or a place at home.
5. School/class/lab/ gym	Reference, in the text of the item, to the school, class, laboratory, or gym.
6. Community or neighborhood	Reference, in the text of the item, to a place other than school in the community or neighborhood.
7. State/province	Reference, in the text of the item, to a place in a state or province.
8. Identified country	A specific country in the world.
9. Unidentified/ fictitious country	A fictitious country.
10. World/global	Two or more countries or a geographic region in the world.
Socio-historical context	Presence of individual and place in relation to a social event.
1. Events in domestic affairs	Context provided based on a place where an event is occurring currently.
2. Events in international affairs	Context provided based on an event taking place in a non-fictitious country that is not of the examinees.
3. Traditions and customs	Context provided based on long-established ways of doing things in the society or culture of the examinees.
Human figure	Human form in photo or drawing.
1. Complete human figure	Person body shown in its entirety.
2. Partial human figure	Part of the body of a person.
3. One human figure	Only one person depicted in the illustration.
4. Two or more human figures	Two or more persons depicted in the illustration.

Coding Examples. In this coding approach, the selected test illustrations are coded dichotomously as *present* (1) or *absent* (0). Below, I provide three examples to illustrate the coding process.

Figure 5 shows an illustrated science item in physics and Table 11 shows the analysis of the illustration of that item for the first dimension, *1.1. Representation of Objects and Background*. In order to analyze the features of illustrations at the dimension level, the subcategories of IDVs are treated as vectors (Solano-Flores & Wang, 2011; Wang & Solano-Flores, 2011). Since the image concreteness of the test illustration shown in Figure 5 is regarded as realistic line drawing (see the third variable in Category 1.1, Table 5), the coding for the category, *1.1. Image Concreteness* for that illustration is described by the vector:

$$C1.1 = [0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0].$$

The variables within a given category are not necessarily mutually exclusive. Some test illustrations can have two or more IDVs under the same subcategory. For example, the illustration of the item shown in Figure 5 contains words (*Upward*, *Downward*) used to designate direction and phrases (*Spring Scale*, *Gym Floor*, *Sneaker 1*) to designate objects. “1” in *Sneaker 1* is not coded as Arabic numerals, because it does not indicate quantity, and is part of the phrase. Thus, the coding for the category, *3.1 Text Unit* for that illustration is described by the vector:

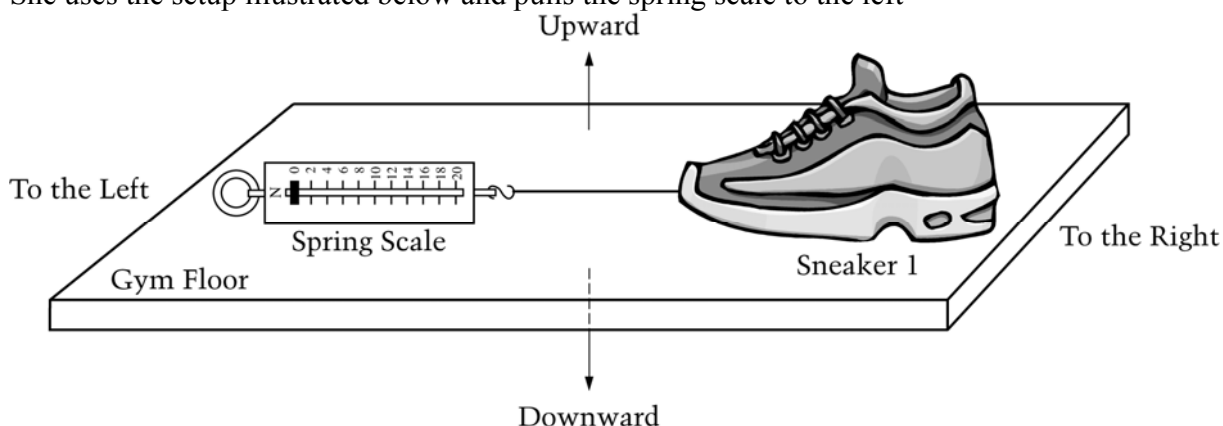
$$C3.1 = [0\ 0\ 0\ 0\ 0\ 0\ 1\ 1\ 0\ 0\ 0\ 0].$$

Questions 11-12 refer to the following information.

Meg designs an experiment to see which of three types of sneakers provides the most friction. She uses the equipment listed below.

1. Sneaker 1
2. Sneaker 2
3. Sneaker 3
4. Spring scale

She uses the setup illustrated below and pulls the spring scale to the left



11. In what direction does the force of friction act?

- A. To the left
- B. To the right
- C. Upward
- D. Downward

Figure 5. An illustrated science item. Source: National Assessment of Educational Progress. (2009). *NAEP 2009 Science Assessment Public Released Items for Grade 8*. Retrieved March 18, 2011, from <http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=science>

Table 11

Analysis of the Representation of Objects and Background Dimension for the Illustration of the Item Shown in Figure 5

1. Representation of objects and background

1.1. Image concreteness: Realistic line drawing—the representation drawn with lines, shows objects that is commonly seen in real life

1.2. Background: Without background—the sneaker is shown without reference to the gym wall

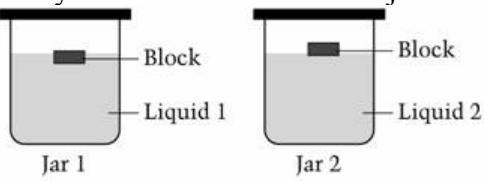
Table 11 continues

Table 11(continued)

1.3. Zooming: Zero zooming	the sneaker and gym floor are shown as seen with the naked eye
1.4. View: External	the illustration does not show the inside of any component
1.5. Dimension: three dimensional	the length, width, and the depth of the objects are shown
1.7. Relative scale of objects: Disproportionate	the sneaker appears to be unrealistically bigger than it really is, in relation to other components in the illustration, such as the gym floor
1.8. Color: grey scale	
6. Context in illustration	
6.1. Physical context: Gym	the illustration shows the gym floor
2. Metaphorical visual language	
2.1. Space, time, and motion: Dynamics	the arrows show the moving direction
3. Text in illustration	
3.1. Text unit: Word	Upward, Downward, and Phrase—Sneaker1, Spring Scale
5. Illustration-text interaction	
5.1. Location of illustration:	Between the stem and the options
5.2. Reference to illustration:	Explicit
5.3. Stated actions to perform with the illustration:	No actions stated
5.4. Commonality:	Same illustration for several items

Figure 6 below shows the analysis for another physics item. I use this example to show how to code the fourth dimension, which is on variables, constants, and functions (see Table 12).

1. Look at the two pictures below. They show what happened when two solid blocks were each put in a jar containing a liquid. Based just on what you can see in the pictures, what can you say about the blocks and the jars?



A. The liquid in the jars must be water.
 B. The block in jar 1 weighs more than the block in jar 2.
 C. The block in jar 1 is floating lower in its liquid than is the block in jar 2.
 D. The block in jar 1 must be made of metal and the block in jar 2 must be made of wood.

Figure 6. An illustrated science item. Source: National Assessment of Educational Progress. (2005). *NAEP 2005 Science Assessment Public Released Items for Grade 8*. Retrieved September 18, from <http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=science>

Table 12

Analysis of the Variables, Constants, and Functions Dimension for the Illustration of the Item

Shown in Figure 6

1. Representation of objects and background	
1.1. Image concreteness:	Schematic—the representation shows essential components of the jars and blocks in a simplified manner
1.2. Background:	Without background—the jars are shown without reference to a lab table
1.3. Zooming:	Zero zooming—the jars and blocks are shown as seen with the naked eye
1.4. View:	External—the illustration does not show the inside of any component and From the side of object—the objects are seen from the side
1.5. Dimension:	Two dimensional—only the length and width of the objects are shown
1.7. Relative scale of objects:	Proportionate—the jars and blocks appear to be at a consistent scale.
1.8. Color:	grey scale
2. Metaphorical visual language	
2.1. Space, time, and motion:	Space and location—the arrows indicate locations
2.2. Matter and energy:	States of matter—the grey surface shows water
3. Text in illustration	
3.1. Text unit:	Arabic number—1, Word—Block, and Phrase—Liquid1, Jar 1
4. Variables, constants, and functions	
4.1. Variables and constant:	Cases—two jars, two blocks, and treatment—the two blocks are placed in two different liquids
5. Illustration-text interaction	
5.1. Location of illustration:	Between the stem and the options
5.2. Reference to illustration:	Explicit
5.3. Stated actions to perform with the illustration:	Look at
5.4. Commonality:	Part of a stand-alone item

Figure 7 below shows an example of coding for a biology item with compound image. I use this example to show how to code the category on structure (see Table 13).

Look at the list of organisms (plants and animals) below, they all live in the Neritic Zone.

Organism	Description
Phytoplankton	Microscopic plants that photosynthesize
Zooplankton	Microscopic animals that feed on phytoplankton
Tuna	Medium sized fish that feeds on small fish
Herring	Small fish that feeds on Zooplankton
Shark	Large fish that feeds on other fish
Whale	Large mammal that feeds on zooplankton

- A. Complete the food web on the chart below to include all the organisms listed in the table. Write the name of one organism in each circle.

The information given about each organism will help you. Three organisms have been placed on the chart for you. The arrows show the direction that energy flows through the food web.

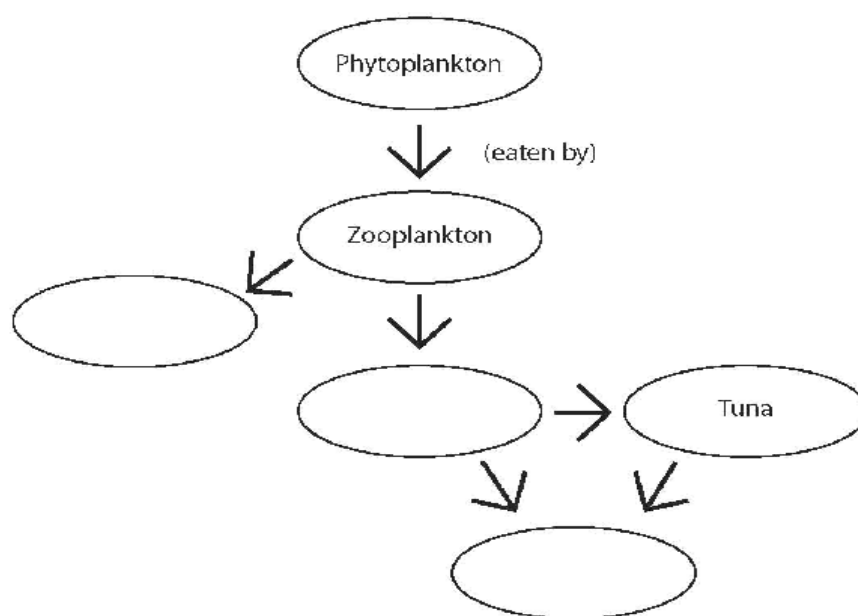


Figure 7. An illustrated science item. Source: TIMSS. (2007). *TIMSS 2007 Science Assessment Public Released Items for Grade 8*. Retrieved March 18, 2011, from <http://timss.bc.edu/TIMSS2007/items.html>

Table 13

Analysis of the Structure Category of the Illustration of the Item Shown in Figure 7

1. Representation of objects and background	
1.1. Image concreteness:	Schematic—the illustration uses oval shapes to represent different organisms
1.2. Background:	Without background—the objects are on the same foreground
1.3. Zooming:	Other
1.4. View:	Other
1.5. Projection:	Two dimensional—the depth of the objects are not shown
1.7. Relative scale of objects:	Other
1.8. Color:	Black & White
2. Metaphorical visual language	
2.1. Space, time, and motion:	Dynamics—the arrows show the relations between organisms
3. Text in illustration	
3.1. Text unit:	Non-math/scientific symbol—parenthesis in (eaten by), Word—Tuna, Herring, phrase—eaten by, Microscopic plants that photosynthesize
3.2. Text function:	Provide label—Tuna, Zooplankton, title/caption/heading—Organism, Description, elaborate/explain/state—Medium sized fish that feeds on small fish
3.3. Text emphasis:	Bolding— Organism, Description
3.4. Text direction:	From left to right—the direction of text in the table, from top to bottom—the direction of text in the diagram (‘Phytoplankton eaten by zooplankton’)
4. Variables, constants, and functions	
4.1. Variables and constant:	Categories—different organisms
4.2. Function:	Table
4.3. Structure:	Network—the food web
5. Illustration-text interaction	
5.1. Location of illustration:	Embedded in stem/prompt, embedded in options/ response format
5.2. Reference to illustration:	Explicit (“Complete the food web on the chart below to include all the organisms listed in the table...”)
5.3. Stated actions to perform with the illustration:	Draw or mark or write on illustration provided (“Write the name of one organism in each circle...”)
5.4. Commonality:	Part of a stand-alone item

Coding Process. Two bilingual, Chinese-English coders with knowledge of middle school science participated in this study. Both coders spoke Chinese as their native language, had received formal higher education in linguistics and literature in China, and had pursued a graduate degree in the U.S. Their background and knowledge were expected to help them

accurately interpret the Chinese illustrated items, particularly the text of test items and the text used as an integral part of test illustrations.

I was one of the coders. Since I had extensive experience in using the coding system for coding illustrated items, I trained the other coder in the use of this coding system. As part of the training, both coders examined definitions and sample illustrations, and coding rules for each IDV, and coded 60 sample item illustrations.

After the training was completed, the two coders coded the 416 illustrations in a random order intended to control for the effect of fatigue or practice. For each illustration, the coders independently examined each illustration and coded the presence or absence of each IDV respectively as 1 or 0. Then they compared and discussed their coding, and resolved any discrepancies to create an agreed version that reflected their consensus-based coding. As shown in Table 14, the two coders reached agreement for approximately 89% of all the coding events. This indicates that the two coders tended to have consistency in most coding events.

Table 14

Number and Frequency of Coding Events

Coding event	Number	Percentage
Coded by both coders as 1	8,131	89
Coded differently and after discussion coded 1	859	9
Coded differently and after discussion coded 0	173	2
Total	9,163	100

As a side note, although the absence of IDV was regarded as 0 in the database, the coders only checked the IDVs shown in each of the illustrations. They did not give 0s to the IDVs not shown in the illustrations on the coding form. As a consequence, if the coding by both coders as 0 was considered as one type of coding event, that would inflate the level of coding agreement.

Data Analysis

Upon completing the coding, I examined the absolute and relative frequencies of illustration variables coded 1 and compared the selected samples of items as to the percentages of items in which those variables were coded 1.

To examine any differences in illustration complexity, the illustrations used in the items from the four U.S. states combined, CMSEE, and TIMSS were compared as to the total number of illustration dichotomous variables (NIDVs) observed for each of the categories mentioned above. A series of ANOVAs were performed to determine the statistical significance of NIDVs measures by content area (e.g., earth science, physics), type of assessment system (e.g., state, national, international), and cultural origin (China, U.S.). The categories of IDVs in which statistically significant differences were observed were used to indicate the characteristics of illustrations that are relevant to culture and, therefore, to systematically designing item illustrations.

I performed a series of correlation analyses of the correlations between test illustration NIDVs and item difficulty (p -value) for TIMSS science test items—the set of items for which measures of item difficulty were available. Due to limited access to information on item p -values, I only examined the relationship for students from the U.S., Chinese Taipei, and Hong Kong, as a Special Administrative Region (SAR) of the People's Republic of China, who participated in TIMSS science tests.

Also, I constructed symmetry graphs to examine similarities and differences between China and the U.S. in the frequency of illustration features at the individual IDV level, and to identify the most outstanding IDVs—that is, the IDVs with most discrepant relative frequencies between

the two countries. Based on that information, I was able to identify and examine the features in which the illustrations from China and the U.S. differed the most.

Additionally, I adopted a strategy to control for variance due to the difference in topic areas under sub-content areas within each of the three main content areas. For example, as a sub-content area in physics, *Forms and Transfer of Energy* includes topic areas such as “identify different forms of energy,” “explain how energy can be transferred between objects and forms, mainly non-textual device” and “understand the nature and usage of electricity and light.” Test illustrations in different sub-content areas or topic areas may differ considerably in terms of the content and the complexity of forms. Thus, I flagged illustrated items under the same topics accompanying similar illustrations. This provided further evidence regarding the similarities and differences of the complexity of individual illustrations across assessment systems. Table 15 shows how I noted down the ID numbers of the individual items in life science across different assessment systems.

Table 15

Example of Flagged Illustrated Items Across Sub-Content Areas in Life Science From China Four Municipalities, U.S. Four States, and TIMSS.

Sub-content area	Assessment system		
	China	U.S.	TIMSS
Topic 1	Item #		Item #
Topic 2	Item #	Item #	
Topic 3		Item #	Item #
...			

CHAPTER V

RESULTS

Overview

This chapter reports on the characteristics of illustrations used in large-scale science test items. More specifically, the chapter examines the patterns of complexity of features in illustrations across assessment systems originated in China, the U.S., and TIMSS and, in the case of TIMSS, their relation to student performance. I calculated frequencies of illustration dichotomous variables (IDVs) to examine the complexity of features presented in test illustrations. Also, I computed correlations to investigate the relationship between the complexity of illustration features and student performance.

To address the first research question, I examined the frequencies of features coded according to IDVs across four factors: assessment system, content area, item format type, and item commonality type. First, I present the results of descriptive statistics analysis comparing the distribution of features across for each factor. Second, I present the results of a series of *t*-tests and ANOVAs performed to determine the statistical significance of differences in the frequencies of features between those factors.

To address the second research question, I discuss the results of a series of Pearson correlation analyses performed to assess the relation between the number of features and student performance. As noted in Chapter 4, due to the limited availability of data on student performance (percentage of students answering an item correctly), this study had access to data on student performance only for TIMSS. Given a myriad of factors that impact student performance in tests, it would not be realistic to expect significant correlations. Therefore, rather

than focusing on statistical significance, I focus on the patterns of direction and magnitude differences of those correlations.

Finally, I use symmetry graphs to present similarities and differences of the complexity of illustration features across item country of origin at the individual IDV level. Additionally, I discuss exemplar illustrated items to demonstrate possible differences in the complexity of illustration features across the three assessment systems.

Research Question 1: Test Illustration Complexity

This section presents the descriptive statistics and the results of a series of analyses of variance performed to answer the first research question: *How do the characteristics of illustrations in large-scale science assessments vary across type of assessment system, item cultural origin, content area, and item type?*

Descriptive Analyses. The construction of a measure on the complexity of illustration features, indexed as the number of IDVs coded 1, plays a critical role in this study. For each illustration, I calculated the absolute and relative frequencies of features (IDVs coded 1) for each of the six illustration dimensions and for all the dimensions combined.

Table 16 shows the absolute and relative frequencies with which features belonging to the six dimensions and all the dimensions combined were observed. Out of a total 8,973 of features coded 1 in the 416 test illustrations, 2,956 (33%) features observed belonged to the dimension OAB, and 2,762 (31%) features observed belonged to the dimension TII. The lowest number of illustration features belonged to the dimensions MVL and CII (respectively 4% and 3%).

Table 16

Number and Percentage of IDVs Coded 1 by Illustration Dimension

	Test illustration dimension						Total
	Objects and background illustration (OAB)	Context in illustration (CII)	Metaphorical visual language (MVL)	Text in illustration (TII)	Variables, constants, and functions (VCF)	Illustration-text interaction (ITI)	
Number	2,956	231	368	2,762	956	1,700	8,973
Percentage	33	3	4	31	11	19	100

Table 17 shows the mean number and range of features observed in the different test illustration dimensions. On average, 21.56 distinguishable features were observed in an item. The mean number of features identified ranged from .56 for the dimension CII to 7.11 for the dimension OAB. The dimension TII has the greatest variation in the mean number of features, while the dimension ITI has the smallest variation.

Table 17

Mean Number and Range of IDVs Coded 1 by Test Illustration Dimension. Standard Deviations in Parentheses.

	Test illustration dimension						Total
	Objects and background illustration (OAB)	Context in illustration (CII)	Metaphorical visual language (MVL)	Text in illustration (TII)	Variables, constants, and functions (VCF)	Illustration-text interaction (ITI)	
Mean	7.11 (3.08)	.56 (1.15)	.88 (0.92)	6.63 (3.51)	2.29 (2.29)	4.09 (0.37)	21.56 (5.52)
Range	2-15	0-5	0-4	0-17	0-9	3-7	7-48

Table 18 shows the mean total numbers of different features observed in the three assessment systems—China's 4 municipalities, the U.S. 4 states, and TIMSS. On average, 22.42 features

were observed in a Chinese test illustration, 21.26 in a U.S. test illustration, and 20.53 in a TIMSS test illustration. The differences in the number of features of test illustrations from the three assessment systems appear to be minimal.

Table 18

Mean on Total Number of IDVs Coded 1 Across Assessment Systems. Standard Deviations in Parentheses.

Assessment system	<i>n</i>	Number of features
China	152	22.42 (6.09)
U.S.	191	21.26 (5.15)
TIMSS	73	20.53 (4.94)
Total	416	21.56 (5.52)

The histograms in Figure 8 compare the frequencies of features observed in items generated in China and the U.S. As the shape of the distributions show, Chinese illustrations tend to have a slightly wider range of features than their U.S. counterparts. However, the distributions are very similar.

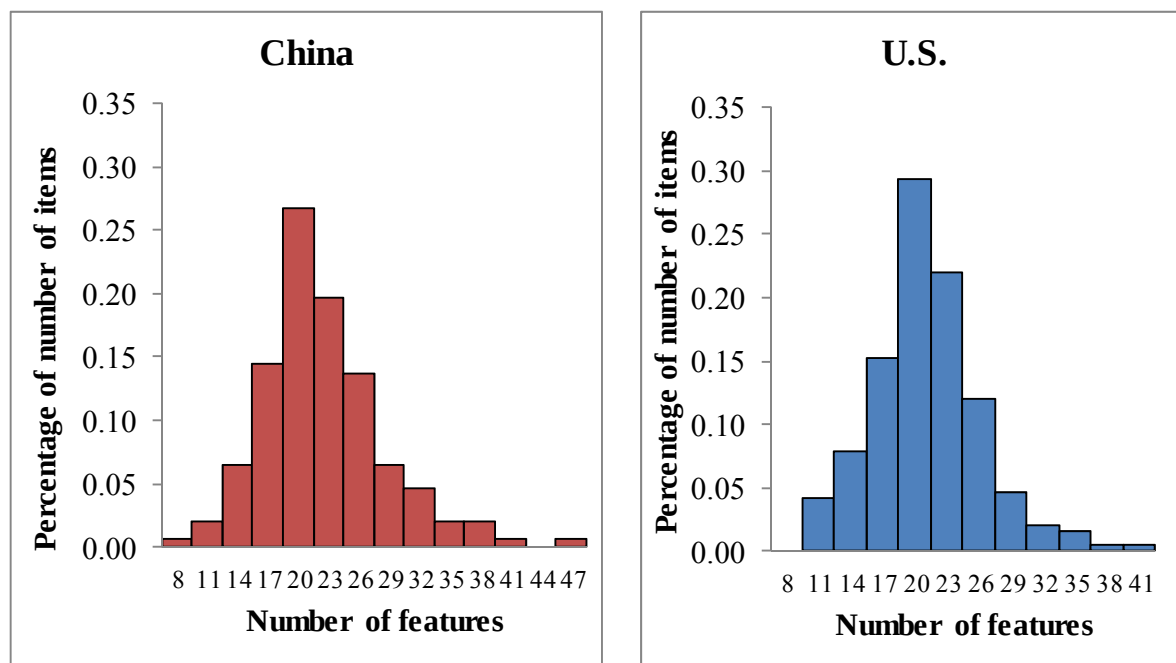


Figure 8. Number of IDVs coded 1 shown in percentage of number of items generated correspondingly in China and the U.S.

I also examined the relative frequencies of features of items from China, the U.S., and TIMSS across the six illustration dimensions. As shown in Table 19 and Figure 9, illustrations generated from these three assessment systems tend to have similar percentages of features across dimensions. The highest percentage of features observed belonged to the dimension OAB and TII. The lowest number of features belonged to the dimension CII and MVL.

Table 19

Number (n) and Percentage (%) of IDVs Coded 1 by Assessment System and Test Illustration

Dimension

Assessment system		Test illustration dimension						
		Variables, constants, and functions					Total	
		Objects and Context in background illustration		Metaphorical visual language	Text in illustration	Illustration -text interaction		
		(OAB)	(CII)	(MVL)	(TII)	(VCF)		(ITI)
China	<i>n</i>	1,124	142	127	1,054	347	619	3,413
	%	33	4	4	31	10	18	100
U.S.	<i>n</i>	1,308	67	180	1,301	432	773	4,061
	%	32	2	4	32	11	19	100
TIMSS	<i>n</i>	524	22	61	407	177	308	1,499
	%	35	1	4	27	12	21	100

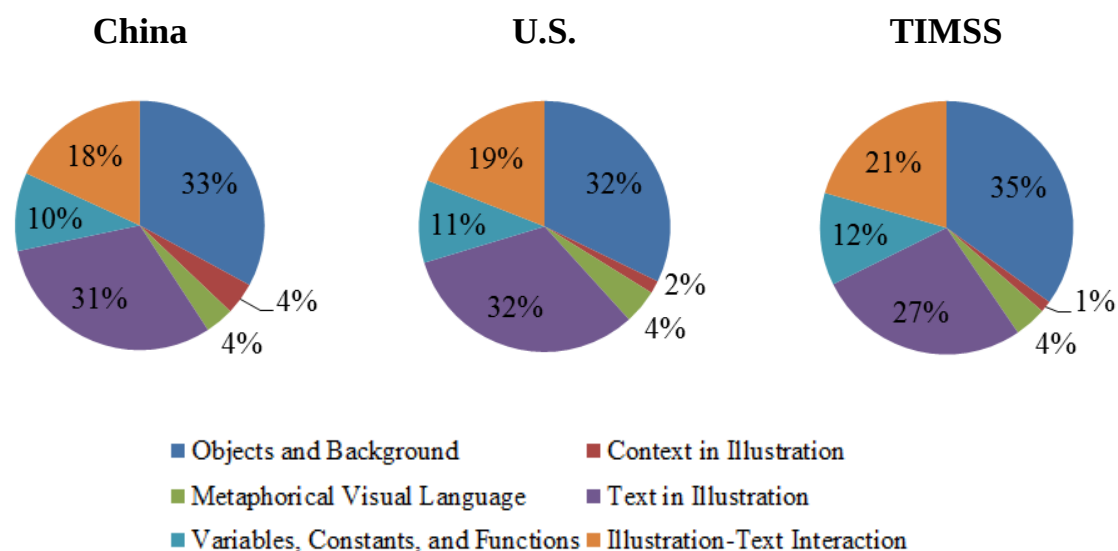


Figure 9. Percentage of IDVs coded 1 by test illustration dimension in three assessment systems: China, the U.S., and TIMSS.

To further investigate how the complexity of illustration features varies across assessment systems, I calculated the number and percentage of features for categories under each of the six dimensions. The results show that, except for dimension CII, the frequency of features in

illustrations originated in China, the U.S., and TIMSS tend to have similar patterns. As Table 20 and Figure 10 show, for dimension CII, more features showing socio-historical context were observed in Chinese illustrations than in the illustrations originated in the U.S. and TIMSS.

Table 20

Number (n) and Percentage (%) of IDVs Coded 1 in Categories Under the Dimension Context in Illustration

Assessment system		Context in Illustration			Total
		Physical context	Socio-historical context	Human figure	
China	<i>n</i>	73	18	51	142
	%	51	13	36	100
U.S.	<i>n</i>	29	1	37	67
	%	43	1	55	100
TIMSS	<i>n</i>	10	0	12	22
	%	45	0	55	100

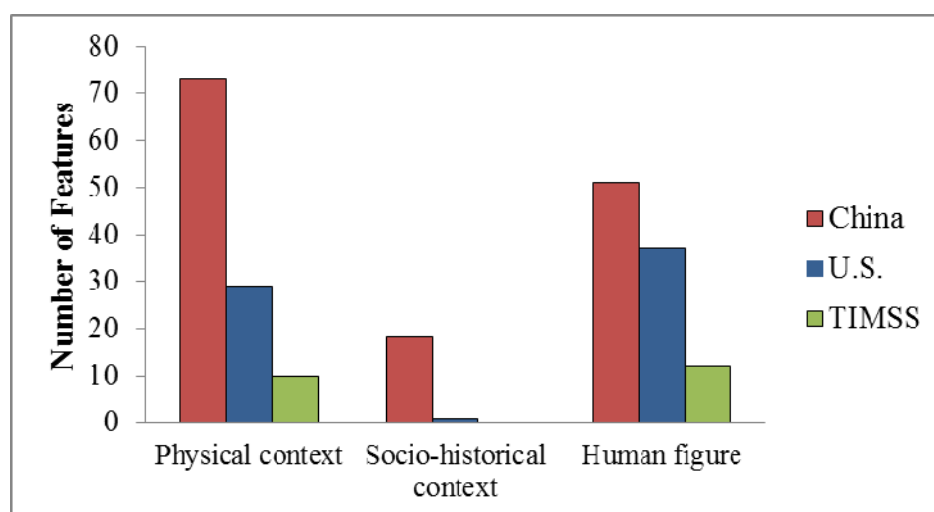


Figure 10. Number of IDVs coded 1 in the dimension, Context in Illustration from three assessment systems: China, U.S. and TIMSS.

To further explore the complexity of features in test illustrations, I also examined the mean number of features across content area, item format type, and item commonality type respectively. Table 21 shows that the number of illustration features tends to vary across the three content areas: Earth Science, Life Science, and Physical Science. On average, illustrations in Earth Science tend to have slightly more features than those in Physical Science and Life Science, while Life Science has the lowest total number of features. Apparently, the average number of features of test illustrations varies slightly by content area and assessment system. For example, illustrations in Earth Science from China had the highest number of features, while those in Life Science from China displayed the lowest number of features. In addition, those in Earth Science from the U.S. and TIMSS tend to have similar illustration complexities.

Table 21

Mean Number of IDVs Coded 1 by Content Area and Assessment System. Standard Deviations in Parentheses.

Content area	<i>n</i>	China	U.S.	TIMSS	Total
Earth science	106	24.68 (7.54)	21.65 (4.34)	21.18 (5.20)	22.91 (6.22)
Life science	124	19.09 (4.65)	20.04 (4.37)	21.25 (6.54)	19.94 (4.78)
Physical science	186	22.51 (4.89)	22.26 (6.05)	19.98 (4.10)	21.87 (5.31)

I also examined the mean number of features per dimension across both content area and assessment system. The results show that test illustrations from different assessment system in a certain content area tend to have similar levels of complexity across dimensions. In general, the number of features for a certain dimension tends to vary by content area and assessment system. The mean number of features in dimension CII in Earth Science in Chinese illustrations is

always higher than in the U.S. and TIMSS illustrations. The mean number of features in dimension TII in Earth Science and Physical Science in the Chinese illustrations is slightly higher than in their U.S. and TIMSS counterparts, while the mean number of features in dimension TII in Life science in the Chinese illustrations is smaller than in the U.S. and TIMSS illustrations. Dimensions OAB and TII have the largest number of features across content area and assessment system. Dimensions CII and MVL have the smallest numbers of features. Dimension ITI has the least variation in the number of features.

Table 22 shows the mean number of features by item format type and assessment system. The mean number of features per dimension across assessment systems for each of the item type tends to show similar patterns—similar means and standard deviations. This indicates that test illustrations of the same item type from different assessment systems tend to have similar levels of complexity across dimensions. Of all the multiple choice items, the mean number of features in dimension CII in Chinese illustrations is higher than that in the U.S. and TIMSS illustrations. Dimensions OAB and TII have the largest number of features across item types and assessment systems. Dimensions CII and MVL have the smallest number of features. Dimension ITI has the least variation in the number of features.

Table 22

Mean Number of IDVs Coded 1by Item Format Type, Assessment System, and Illustration Dimension. Standard Deviations in Parentheses.

Item format type	Assessment system	n	Test illustration dimension						All dimensions combined
			Objects and background (OAB)	Context in illustration (CII)	Metaphorical visual language (MVL)	Text in illustration (TII)	Variables, constants, and functions (VCF)	Illustration- text interaction (ITI)	
Multiple choice	China	54	7.57 (3.31)	1.22 (1.72)	.83 (.97)	6.06 (3.51)	1.74 (1.66)	4.02 (.14)	21.44 (7.35)
	U.S.	132	6.91 (3.17)	.41 (1.00)	.86 (.90)	6.43 (3.73)	2.28 (2.34)	4.04 (.19)	20.93 (4.91)
	TIMSS	44	7.41 (2.88)	.41 (.97)	.89 (1.02)	5.07 (3.31)	2.05 (2.06)	4.14 (.35)	19.95 (4.13)
	All items	230	7.16 (3.15)	.60 (1.25)	.86 (.93)	6.08 (3.63)	2.11 (2.15)	4.05 (.22)	20.87 (5.45)
Constructed response	U.S.	30	7.80 (2.41)	.30 (.79)	1.20 (.93)	7.27 (3.32)	1.83 (1.98)	4.07 (.25)	22.47 (6.38)
	All items	58	7.16 (2.79)	.38 (.90)	.91 (.90)	7.02 (3.55)	2.22 (2.42)	4.03 (.26)	21.72 (5.92)
Fill in the blank	China	55	7.47 (2.53)	.69 (1.20)	.84 (.76)	7.04 (3.12)	2.49 (2.27)	4.04 (.33)	22.56 (4.82)
	All items	67	7.33 (2.58)	.61 (1.15)	.93 (.89)	7.18 (2.97)	2.42 (2.18)	4.03 (.30)	22.49 (4.71)
Draw on illustration	All items	30	6.43 (3.54)	.37 (.93)	.97 (1.03)	7.20 (3.24)	2.50 (2.75)	4.40 (.77)	21.87 (5.61)
Two/more formats	All items	32	6.81 (3.61)	.63 (.98)	.81 (.82)	8.19 (3.18)	3.31 (2.67)	4.25 (.67)	24.00 (6.00)

Table 23 shows the mean number of features across item commonality type and assessment system. In general, test illustrations of different item commonality type tend to have a similar mean number of features across dimensions. Illustrations commonly used for several related items tend to have more features than those for one item within a series of related items and as part of a stand-alone item, with those as part of a stand-alone item having the smallest total number of features. The mean number of features in dimension CII in Chinese illustrations is always higher than that in the U.S. or TIMSS illustrations. The mean number of features in dimension OAB and TII in the second commonality type (same illustration for several items) across assessment systems is slightly higher than that in the first commonality type (part of a stand-alone item). Dimensions OAB and TII have the highest number of features across commonality types and assessment systems. Dimensions CII and MVL have the smallest number of features. Dimension ITI has the least variation in the number of features.

Table 23

Mean Number of IDVs Coded 1 by Item Commonality Type, Assessment System, and Illustration Dimension. Standard Deviations in Parentheses.

Item commonality type	Assessment system	n	Test illustration dimension						
			Objects and background (OAB)	Context in illustration (CII)	Metaphorical visual language (MVL)	Text in illustration (TII)	Variables, Constants, and Functions (VCF)	Illustration-text interaction (ITI)	All dimensions combined
Part of a stand-alone item	China	81	7.16 (3.01)	.89 (1.53)	.73 (.92)	6.41 (2.88)	2.02 (1.99)	4.05 (.27)	21.26 (5.39)
	U.S.	141	6.90 (3.18)	.39 (.97)	.91 (.94)	6.37 (3.60)	2.24 (2.30)	4.05 (.25)	20.86 (4.95)
	TIMSS	58	7.12 (3.08)	.36 (.93)	.81 (.95)	5.21 (3.52)	2.22 (2.34)	4.17 (.46)	19.90 (4.26)
	All items	280	7.02 (3.10)	.53 (1.17)	.84 (.94)	6.14 (3.41)	2.18 (2.22)	4.08 (.31)	20.78 (4.95)
Same illustration for several items	China	55	8.07 (2.81)	1.13 (1.36)	1.04 (.69)	7.51 (3.51)	2.36 (2.28)	4.04 (.33)	24.15 (6.35)
	U.S.A.	43	7.14 (2.71)	.28 (.80)	1.07 (.96)	8.09 (3.18)	2.19 (2.16)	4.05 (.21)	22.81 (5.55)
	All items	110	7.65 (2.87)	.68 (1.17)	1.01 (.82)	7.67 (3.41)	2.41 (2.36)	4.10 (.43)	23.53 (6.04)
Illustration for one item within a series of items	All items	25	5.60 (3.34)	.32 (.69)	.84 (1.11)	7.68 (3.93)	3.24 (2.70)	4.20 (.58)	21.88 (7.11)

***t*-Tests.** In order to examine potential differences in the complexity of features associated to item country of origin, I compared the mean number of features of illustrations from China and the U.S. Table 24 presents the results from a series of independent samples *t*-test analyses that show the statistical significance of differences in the number of features based on item country of origin. The *t*-test results for all dimensions combined show that there was barely a statistically significant effect for item country of origin, $t(343)=1.96$, $p=.05$, with Chinese illustrations having more features than their U.S. counterparts. The magnitude of the differences in the means for all dimensions combined was small (Cohen's $d=.21$). The *t*-tests across dimensions show no statistically significant mean number of differences attributable to individual illustration dimensions. However, there was a significant effect for item country of origin in dimension CII, $t(343)=4.41$, $p<.01$, with Chinese illustrations having more features than the U.S. illustrations. The magnitude of these differences in the mean number of features was medium (Cohen's $d=.56$). This indicates that, for dimension CII, considerable differences exist in the mean number of features between China and the U.S. illustrations.

Table 24

Independent-Samples t-Test of Number of IDVs Coded 1 by Item Country of Origin

Illustration dimension	Item origin	<i>n</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>Sig.</i>	<i>d</i>
Objects and background (OAB)	China	152	7.39	3.00	1.64	341	.10	.18
	U.S.	191	6.85	3.13				
Context in illustration (CII)	China	152	.93	1.41	4.41**	247	.00	.56
	U.S.	191	.35	.92				
Metaphorical visual language (MVL)	China	152	.84	.85	-1.08	341	.28	-.12
	U.S.	191	.94	.96				
Text in illustration (TII)	China	152	6.93	3.30	.33	341	.74	.04
	U.S.	191	6.81	3.57				
Variables, constants, functions (VCF)	China	152	2.28	2.22	.09	341	.93	.01
	U.S.	191	2.26	2.29				
Illustration-text interaction (ITI)	China	152	4.07	.37	.74	246	.46	.08
	U.S.	191	4.05	.24				
All dimensions combined	China	152	22.45	6.10	1.96	341	.05	.21
	U.S.	191	21.26	5.15				

**t* value is significant at the 0.05 level (2-tailed)

** *t* value is significant at the 0.01 level (2-tailed)

Another series of independent-samples *t*-tests were performed to examine the differences in the mean number of features between multiple choice and constructed response items (Table 25). In general, the results show that there were no statistically significant differences in the mean number of features for these two types of test items. The magnitude of the differences in the means for dimensions TII and CII was small (Cohen's *d* are -.26 and .23 respectively).

Table 25

Independent-Samples t-Test of Number of IDVs Coded 1 by Item Format Type

Illustration Dimension	Item format type	<i>n</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>Sig.</i>	<i>d</i>
Objects and background (OAB)	Multiple choice (MC)	230	7.16	3.15	.01	286	.99	.00
	Constructed response (CR)	58	7.16	2.79				
Context in illustration (CII)	Multiple choice (MC)	230	.60	1.25	1.54	119	.13	.23
	Constructed response (CR)	58	.38	.90				
Metaphorical visual language (MVL)	Multiple choice (MC)	230	.86	.93	-.39	286	.70	-.06
	Constructed response (CR)	58	.91	.90				
Text in illustration (TII)	Multiple choice (MC)	230	6.08	3.63	-1.76	286	.08	-.26
	Constructed response (CR)	58	7.02	3.55				
Variables, constants, functions (VCF)	Multiple choice (MC)	230	2.11	2.15	-.36	286	.72	-.05
	Constructed response (CR)	58	2.22	2.42				
Illustration-text interaction (ITI)	Multiple choice (MC)	230	4.05	.22	.52	286	.60	.08
	Constructed response (CR)	58	4.03	.26				
All dimensions combined	Multiple choice (MC)	230	20.87	5.45	-1.05	286	.29	-.16
	Constructed response (CR)	58	21.72	5.92				

t* value is significant at the 0.05 level (2-tailed) *t* value is significant at the 0.01 level (2-tailed)

ANOVAs. A series of one-way fixed-effects ANOVA analyses were performed to examine the statistical significance of differences in the number of features by illustration dimension based on content area (earth science, life science, and physical science) and assessment system (China, the U.S., and TIMSS) respectively (Table 26). For content area, the ANOVA results show a statistically significant effect for all dimensions combined and for dimension CII, although with small effect sizes ($\eta^2 = .04$ and $.05$, respectively). For assessment system, there was a statistically significant effect for dimension CII, with a medium effect size ($\eta^2 = .06$). This reveals that the number of features on dimension CII vary with assessment system.

Table 26

One-way Fixed-effects Analysis of Variance for Illustration Complexity by Dimension

Source	Dimension	<i>df</i>	<i>F</i>	<i>Sig.</i>	η^2
Content area	Objects and background (OAB)	2	2.91	.06	.01
	Context in illustration (CII)	2	10.40**	.00	.05
	Metaphorical visual language (MVL)	2	1.95	.14	.01
	Text in illustration (TII)	2	3.00	.05	.01
	Variables, constant, functions (VCF)	2	.09	.91	.00
	Illustration-text interaction (ITI)	2	2.63	.07	.01
	All dimensions combined	2	9.45**	.00	.04
Origin of assessment system	Objects and background (OAB)	2	1.36	.26	.01
	Context in illustration (CII)	2	13.92**	.00	.06
	Metaphorical visual language (MVL)	2	.70	.50	.00
	Text in illustration (TII)	2	4.19*	.02	.02
	Variables, constant, functions (VCF)	2	.14	.87	.00
	Illustration-text interaction (ITI)	2	6.02**	.00	.03
	All dimensions combined	2	3.58*	.03	.02

* $p < .05$, ** $p < .01$

Table 27 shows the results of a series of two-way fixed-effects ANOVA analyses performed to examine the statistical significance of any differences in the numbers of features observed for different combinations of factors, including illustration dimension, item origin, content area, and

item format type. A series of 6 (illustration dimension) x 3 (assessment system) ANOVA analyses showed that the mean number difference due to illustration dimension was statistically significant, with a large effect size (partial $\eta^2 = .58$). This indicates that there is a considerable difference in the number of features observed across the six illustration dimensions. Another series of 3 (content area) x 2 (item country of origin) ANOVA analyses showed that the mean number difference due to the interaction of the two factors was statistically significant, with a medium effect size (partial $\eta^2 = .07$).

Table 27

Two-way Fixed-effects ANOVAs for Illustration Complexity by Dimension, Item Origin, Item Format Type, and Content Area

Design	Source	<i>df</i>	<i>F</i>	<i>Sig.</i>	<i>Partial η^2</i>
6 x 3	Illustration dimension	5	577.87**	.00	.58
Illustration dimension	Assessment system	2	3.43*	.03	.02
x Assessment system	Dimension x Assessment system	10	2.71**	.00	.01
3 x 2	Content area	2	12.05**	.00	.07
Content area x	Item country of origin	1	1.86	.17	.01
Item country of origin	Content x Item country of origin	2	3.59*	.03	.02
2 x 3	Item format type	1	.21	.65	.00
Item format type x	Content area	2	1.27	.28	.01
Content area	Item format type x Content area	2	1.41	.25	.01
2 x 3	Item format type	1	.57	.45	.00
Item format type x	Assessment system	2	1.82	.16	.01
Assessment system	Item format type x Assessment system	2	.37	.69	.00

* $p < .05$, ** $p < .01$

Summary of Findings for Research Question One. On average, a test illustration contains about 22 distinguishable features. Although Chinese illustrations tend to have a slightly wider range of features than their U.S. and TIMSS counterparts, the differences in the number of features of test illustrations from the three assessment systems seem to be trivial.

Test illustrations from different assessment systems tend to have a similar level of complexity across dimensions. This tendency is also evident when considering content area (e.g., Earth Science, Physical Science, and Life Science), item format type (e.g., multiple-choice, constructed-response, and fill-in-the-blank etc.), and item commonality type (e.g., as part of a stand-alone item, same illustration for several items, and for one item within a series of items).

A series of independent-sample *t*-tests conducted to explore the differences in the mean number of features across dimensions revealed that the complexity of illustration features does not to vary by item country of origin, except for dimension CII. More specifically, Chinese illustrations on dimension CII have significantly more features than U.S. illustrations at the $p < .05$ level, with a large effect size (Cohen's $d = .56$). Another series of independent-sample *t*-tests revealed no statistically significant differences in the mean number of features on the six dimensions attributable to item format type.

Different dimensions vary considerably in the number of features. The two dimensions, OAB and TII, display the greatest number of features, while the other two dimensions, CII and MVL, display the smallest number of features. The dimension ITI shows the least variation in terms of number of features. An ANOVA for illustration complexity by dimension and illustration origin shows that there was a substantial difference in the number of features observed across dimensions.

Some content area appear to contain more features than others. On average, test illustration in Earth Science tend to have slightly more features than those in Physical Science and Life Science. The average number of features in test illustrations varies across the combination of content area and test item origin. Test illustrations in Earth Science from China display the largest number of

features, while those in Life Science from China display the smallest number of features. A two-way, between-groups ANOVA conducted to explore the impact of content area and item country of origin on total number of features revealed statistically significant differences due to the main effect content area, with a medium effect size (partial eta squared=.07).

Research Question 2: Relation between Illustration Complexity and Student Performance

This section presents the results of correlation analyses performed to answer the second research question: *How is test illustration complexity related to student performance on TIMSS items?*

Item *p*-Value. Although I expected to collect item *p*-values for all the 416 illustrated items, I only had access to item *p*-values in TIMSS. I also expected to have item *p*-values in TIMSS for students from China and the U.S.A. However, the People's Republic of China does not participate in TIMSS. As a consequence, I collected *p*-values of TIMSS items for students from Chinese Taipei and Hong Kong, as a Special Administrative Region (SAR) of the People's Republic of China. As Table 28 shows, students from Chinese-Taipei had the largest mean, while students from the U.S. have the smallest mean.

Table 28

Range and Mean of Item p-value: TIMSS. Standard Deviations in Italics

Student origin	Range	Mean
U.S.A.	16-92	53.38 <i>19.33</i>
Hong Kong SAR ^a	15-93	55.81 <i>21.18</i>
Chinese Taipei	21-93	61.03 <i>19.55</i>

Table 28 continues

Table 28 (continued)

Student origin	Range	Mean
International average ^b	11-85	45.94 18.39

^a Hong Kong became a colony of the British Empire after the First Opium War (1839-42) and controlled under the British until 1997, when the People's Republic of China resumed sovereignty. It now participates in TIMSS as a subnational education system.

^b The international average refers to the average for all the students who participated in TIMSS science tests in the year 1995, 1999, 2003, and 2007.

Two things should be noted and kept in mind in order to better interpret student performance and its relation to illustration complexity. First, due to historical and political influences, Hong Kong SAR and Chinese Taipei may adopt education systems substantially different than that of mainland China. Second, students who participated in TIMSS may vary considerably in terms of gender, race or ethnicity, and socioeconomic status etc. both across and within countries or regions.

Correlational Analyses. A series of Pearson correlation analyses were performed to examine the relationship between test illustration complexity and item difficulty in TIMSS, considering dimension, content area, and item type. Given the small correlation coefficients, the results should be interpreted with caution. Rather than focusing on statistical significance, I focus on the patterns of direction and magnitude differences of those correlations.

Table 29 shows the correlation between number of features across dimensions and item *p*-value for each student group. Overall, the total number of features tends to correlate negatively with student performance. This indicates that, in general, increased complexity appears to be negatively associated with item difficulty. Different dimensions tend to have different correlation with student performance. For dimension OAB, the correlation of illustration complexity and

student performance is minimal for China Hong Kong SAR ($r=.003$) and Chinese Taipei ($r=-.063$), and larger for the U.S. ($r=.181$). For dimension CII, the correlation of the number of features coded as IDVs are higher. The correlations between number of features of dimension CII and student performance for the U.S. ($r=.286, p<.05$) and all participating countries/regions together ($r=.305, p<.01$) were statistically significant. Except for dimension MVL, the number of features seemed to have a stronger relationship with student performance for the U.S. than for Chinese Taipei.

Table 29

Pearson Correlations Between Student Performance and Illustration Complexity by Illustration Dimension and Student Country of Origin

Illustration Dimension	Chinese Taipei (n=71)	Hong Kong SAR (n=72)	U.S. (n=72)	International average (n=72)
Objects and background (OAB)	-.063	.003	.181	.189
Context in illustration (CII)	.186	.189	.286*	.305**
Metaphorical visual language (MVL)	-.105	-.095	-.026	.002
Text in illustration (TII)	-.041	-.020	-.193	-.210
Variables, constant, functions (VCF)	.017	.031	-.137	-.161
Illustration-text interaction (ITI)	.055	.037	-.069	-.066
All dimensions combined	-.046	.023	-.059	-.068

*Correlation is significant at the 0.05 level (2-tailed)

**Correlation is significant at the 0.01 level (2-tailed)

Table 30 shows the results of a series of Pearson correlation analyses performed to examine the relationship by content area and illustration dimension. Overall, the number of features for dimension CII was positively correlated with student performance across content area. This indicates that, in general, increased complexity on dimension CII appears to make items easier. Also, the number of features on different dimensions tended to correlate differently with student performance in different content area. In Earth Science, for dimension OAB, the illustration complexity correlated negatively with student performance for China Hong Kong SAR ($r=-.170$) and Chinese Taipei ($r=-.212$), while it correlated positively for the U.S. ($r=.077$)—although this correlation is very low. In Life Science, for dimension OAB, illustration complexity consistently correlated positively with student performance for China Hong Kong SAR ($r=.373$), Chinese Taipei ($r=.118$), and the U.S. ($r=.425$). In Earth Science, for dimension TII, illustration complexity correlated negatively with student performance for the U.S. ($r=-.388$), while the magnitude of this correlation was minimal for Chinese Taipei ($r=.007$) and China Hong Kong SAR ($r=-.075$). In Life Science, for dimension TII, illustration complexity consistently correlated negatively with student performance for students from China Hong Kong SAR ($r=-.308$), Chinese Taipei ($r=-.475$), and the U.S. ($r=-.437$).

Table 30

Pearson Correlations Between Student Performance and Illustration Complexity by Content Area and Illustration Dimension

Content	Dimension	Chinese-Taipei	Hong Kong SAR	U.S.	International average
Earth science (n=17)	Objects and background (OAB)	-.212	-.170	.077	.156
	Context in illustration (CII)	.352	.319	.342	.402
	Metaphorical visual language (MVL)	-.158	-.165	-.018	.103
	Text in illustration (TII)	.007	-.075	-.388	-.291
	Variables, constant, functions (VCF)	.129	.056	-.164	-.141
	Illustration-text interaction (ITI)	.265	.221	.213	.305
	All dimensions combined	-.017	-.113	-.328	-.143
Life science (n=15)	Objects and background (OAB)	.118	.373	.425	.401
	Context in illustration (CII)	.238	.238	.106	.336
	Metaphorical visual language (MVL)	.256	.444	.407	.452
	Text in illustration (TII)	-.475	-.308	-.437	-.518 *
	Variables, constant, functions (VCF)	-.472	-.361	-.459	-.533 *
	Illustration-text interaction (ITI)	.100	.229	.216	.137
	All dimensions combined	-.336	.027	-.085	-.150
Physical science (n=39)	Objects and background (OAB)	-.086	-.089	.131	.105
	Context in illustration (CII)	.093	.112	.312	.260
	Metaphorical visual language (MVL)	-.177	-.205	-.136	-.156
	Text in illustration (TII)	.082	.105	-.078	-.105
	Variables, constant, functions (VCF)	.148	.169	-.058	-.049
	Illustration-text interaction (ITI)	.005	-.032	-.137	-.161
	All dimensions combined	.064	.081	.020	-.036

*Correlation is significant at the 0.05 level (2-tailed)

**Correlation is significant at the 0.01 level (2-tailed)

Table 31 shows the results of a series of Pearson correlation analyses performed to examine the relationship by item commonality type and illustration dimension. Overall, the number of features for dimension CII correlated positively with student performance on stand-alone items, and negatively with student performance on common illustrations for several items. This indicates that, in general, increased complexity on dimension CII for stand-alone items appears to make items easier, while increased complexity on this dimension for common illustrations for several items appears to make items more difficult.

Different dimensions tend to correlate differently with student performance on items of different commonality types. Among stand-alone items, for dimension OAB, illustration complexity had low negative correlations with student performance for China Hong Kong SAR ($r=-.023$) and Chinese Taipei ($r=-.061$), but positive and higher correlations for the U.S. ($r=.199$). Among common illustrations for several items, for dimension OAB, illustration complexity correlated positively with performance for students from Chinese Taipei ($r=.125$), China Hong Kong SAR ($r=.293$), and the U.S. ($r=.309$). Among stand-alone items, for dimension TII, illustration complexity had a negative and small correlation with student performance for the U.S. ($r=-.094$), while it had a small positive correlation for students from Chinese Taipei ($r=.069$) and China Hong Kong SAR ($r=.083$). Among common illustrations for several items, for dimension TII, illustration complexity consistently correlated negatively with student performance for China Hong Kong SAR ($r=-.188$), Chinese Taipei ($r=-.361$), and the U.S. ($r=-.320$).

Table 31

Pearson Correlations Between Student Performance and Illustration Complexity by Item Commonality and Dimension

Commonality	Dimension	Chinese-Taipei	Hong Kong SAR	U.S.	International average
Part of a stand-alone item (<i>n</i> =58)	Objects and background (OAB)	-.061	-.023	.199	.236
	Context in illustration (CII)	.235	.189	.310*	.330*
	Metaphorical visual language (MVL)	-.057	-.104	-.021	.037
	Text in illustration (TII)	.069	.083	-.094	-.075
	Variables, constant, functions (VCF)	.135	.138	-.051	-.056
	Illustration-text interaction (ITI)	-.004	.044	-.050	-.051
	All dimensions combined	.125	.151	.096	.152
Same illustration for several items (<i>n</i> =10)	Objects and background (OAB)	.125	.293	.309	.295
	Context in illustration (CII)	-.403	-.099	-.216	-.273
	Metaphorical visual language (MVL)	-.188	.231	.334	.217
	Text in illustration (TII)	-.361	-.188	-.320	-.470
	Variables, constant, functions (VCF)	-.379	-.240	-.320	-.406
	Illustration-text interaction (ITI)	.209	.170	-.060	.023
	All dimensions combined	-.410	-.008	-.167	-.335

*Correlation is significant at the 0.05 level (2-tailed)

**Correlation is significant at the 0.01 level (2-tailed)

Summary of Findings for Research Question Two. Compared with students from China Hong Kong SAR, the correlations for all dimensions combined appeared to be different for those from Chinese Taipei, the U.S., and students from all participating countries/regions. The patterns and directions of the correlations across dimensions indicate that illustration complexity seems to correlate differently with the performance of students from Chinese Taipei and China-Hong Kong SAR than with that of students from the U.S. and all students together.

Statistically significant correlations were observed only for dimension CII, for students from the U.S. and students from all participating countries/regions. For students from all participating countries/regions, there was a statistically significant positive correlation between illustration complexity and student performance. This indicates that, for an average student participating in TIMSS, the more contextual features a test illustration contains, the higher the percentage of students who respond to it correctly.

Illustration Complexity at the Individual IDV Level

To investigate similarities and differences of how the number of features varied at a fine grain level, I constructed a symmetry graph to present the relative frequencies (in percentages) for the number of features across IDVs for China and the U.S. For each country, I first calculated the relative frequency of each IDV coded 1. Then I ranked the relative frequencies on values from the lowest to the highest frequent for illustrations originated from China. Then for each of these IDVs ranked, I showed corresponding percentages of IDVs for illustrations from the U.S. As shown in Figure 11, the graph tends to show a symmetrical shape, with irregularities in some IDVs. Although, in general, the two countries tend to generate illustrations with similar complexities, differences exist at the individual IDV level.

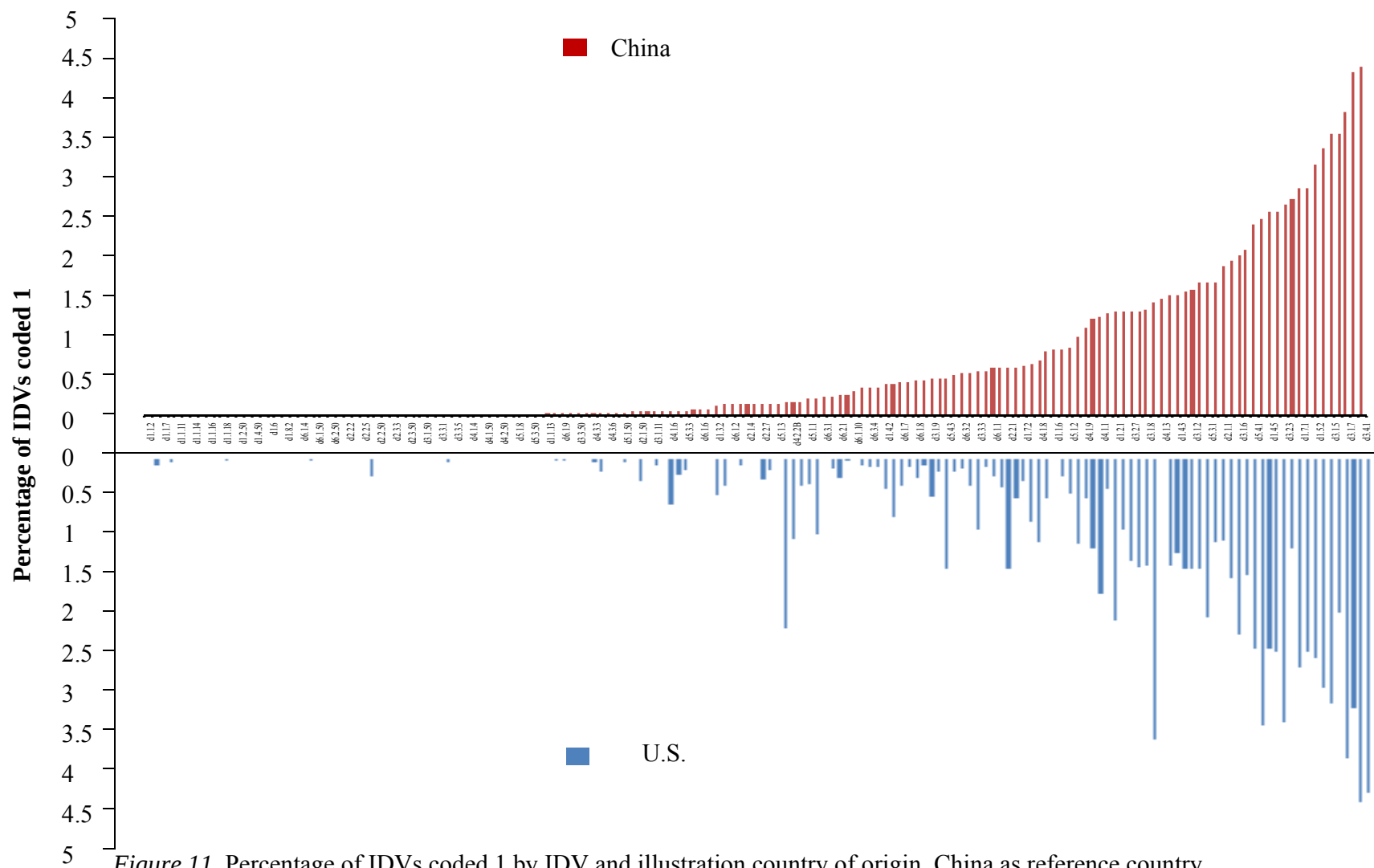


Figure 11. Percentage of IDVs coded 1 by IDV and illustration country of origin. China as reference country.

To further determine the differences in illustration complexity across the two cultures, I examined the proportional frequency distributions of all the IDVs to identify the most frequent sets of variables. First, I obtained the proportional frequencies for each of the IDVs by dividing the percentages of IDVs for one country by the percentages of IDVs for another country. The average proportional frequency obtained was 0.83. I ranked all the proportional frequencies from the least to the largest respectively for China and the U.S. Identifying IDVs with the largest frequencies could help decide whether and how the most frequent sets of IDVs were different for China and the U.S. I decided to select the most outstanding variables, 10% of the total of 159 IDVs. This gave nine IDVs for China, ranging from 3.57 to 10.71, and eight IDVs for the U.S., ranging from 3.36 to 14.96. Table 32 and 33 shows the 17 most outstanding IDVs and their corresponding proportional frequencies. Apparently, there were no overlaps in terms of the most frequent sets of IDVs for the two countries. This indicates that illustrations generated in the two countries tend to have very different sets of features that make them complex.

Table 32

Outstanding IDVs Identified for China

IDV	Label	Proportional frequency of CN/US
d1.1.1	photograph	3.57
d6.3.4	two or more human figures	3.57
d1.1.6	map	3.70
d6.1.10	world/global	3.97
d3.2.8	text on object	4.05
d6.1.7	state/province	4.16
d3.3.3	italicizing	5.65
d6.1.8	identified country	5.95
d6.2.1	events in domestic affairs	10.71

Table 33

Outstanding IDVs Identified for the U.S.

IDV	Label	Proportional frequency of US/CN
d4.1.6	analogic line	3.36
d5.1.1	above stem/prompt	4.80
d2.1.2	time and sequence	5.04
d4.3.1	sequential discrete structure	5.32
d4.3.3	cycle	5.88
d3.2.5	opinions/comment/note	5.88
d3.4.2	text direction, from top to bottom	10.09
d5.1.3	embedded in stem/prompt	14.96

To examine the most different illustrations in China and the U.S., I identified six Chinese items and three U.S. items respectively with the most frequent outstanding IDVs. As Table 34 shows, on average, these illustrations contain four of the 17 most outstanding IDVs. All these nine illustrated items are in either Earth Science or Life Science. No illustrated items in Physical Science were identified with the most frequent outstanding IDVs. In addition, the six Chinese illustrations were generated in either Beijing or Shanghai, while the three U.S. illustrations were generated in either New York or Texas. See Appendix D for the original illustrated items and accompanying translation for Chinese items.

Table 34

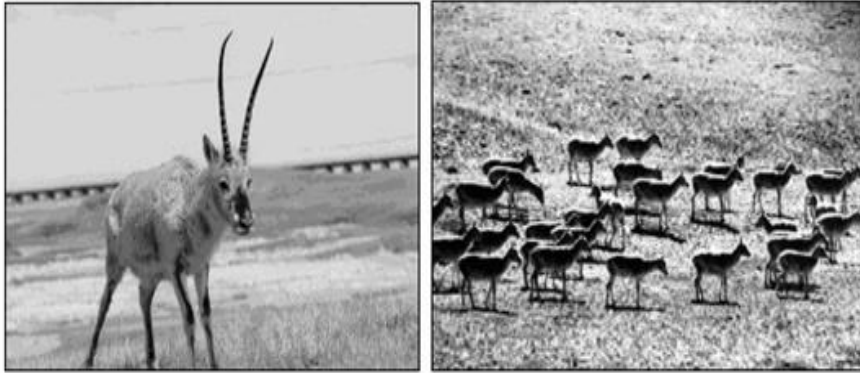
Illustrations Identified With the Highest Number of Outstanding IDVs

Item number	Assessment system	Content area	Year	Country of origin	Number of outstanding IDVs
1	Beijing Middle School Exit Examination	Earth Science	2011	China	4
2	Beijing Middle School Exit Examination	Life Science	2007	China	4
3	Shanghai Middle School Exit Examination	Earth Science	2008	China	4
4	Shanghai Middle School Exit Examination	Earth Science	2008	China	4
5	Shanghai Middle School Exit Examination	Earth Science	2008	China	4
6	Shanghai Middle School Exit Examination	Earth Science	2008	China	4
7	The University of the State of New York science Test	Life Science	2009	U.S.	4
8	Texas Assessment of Knowledge and Skills	Earth Science	2006	U.S.	4
9	Texas Assessment of Knowledge and Skills	Life Science	2009	U.S.	4

To examine the differences in the characteristics of illustrated items originated from the two countries, I selected one Chinese illustration and one American illustration from the above nine illustrations with the highest numbers of outstanding IDVs. In Figure 12, a Chinese test illustration in Life Science, presented as photographs, shows Tibetan antelopes along the Qinghai-Tibet railway line. The illustration from the U.S., also from a large-scale science test in Life Science, shows as realistic line drawing, adult giraffes.

China

Beautiful scenery along the Qinghai-Tibet railway



A *A Tibetan antelope is looking for food near the gigantic bridge of Qingshui River along the Qinghai-Tibet*

B *A group of Tibetan antelopes is migrating nearby Chumaer River along the Qinghai-Tibet railway*

U.S.

The diagram below shows a population of adult giraffes over time. Letter A, B, and C represent three time periods.

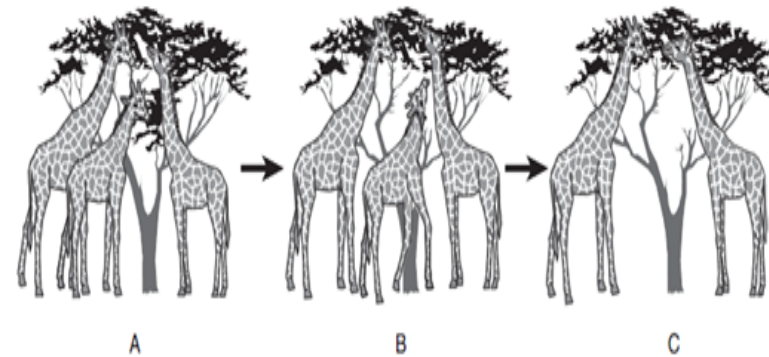


Figure 12. Two comparable illustrations from large-scale Life Science tests in the U.S. and China. Sources: (A) Beijing Xuanwu District Board of Education. (2007). Beijing Middle School Exit Examination, 2008 Life Science test for Grade 8. (B) New York State Education Department. (2009). New York Intermediate-Level Science test for Grade 8. The Chinese illustration shown in English translations. See the original illustrated items in Appendix D.

It appears that these two illustrations have similar complexity of features: The illustration from China contains 22 features, while the illustration from the U.S. contains 23 features. However, examining the individual IDVs coded 1 indicates that these two illustrations tended to possess different sets of features. For example, the illustration from China was coded as photograph, viewed from above, showing physical context (e.g. the bridge of Qingshui River, the Qinghai-Tibet railway), using text units such as sentence as heading and description, and as the same illustration for several related items (now shown). In contrast, Illustration B was coded as realistic line drawing, using metaphorical visual language to show time and sequence, using text units such as non-scientific/mathematical sign, abbreviation, Arabic numeral, and words as note, showing discrete structure, and as part of a stand-alone item (not shown).

Exemplar Illustrations in Sub-Content Areas Across Assessment Systems

I also identified illustrated items under the same topics accompanying similar illustrations across the three assessment systems. See Appendix E for all the flagged illustrated items and accompanying translation for Chinese items. Examining these illustrations provided further evidence regarding the differences of illustration complexity across assessment systems. Figure 13 provides an example of three illustrated items in the same sub-content area from the three assessment systems: China, the U.S., and TIMSS. These three items were from Earth Science, and asked students to use information about the most active earthquake areas along the Ring of Fire to solve problems. The illustration from China contains 27 features, the illustration from the U.S. contains 24 features, and the illustration from TIMSS contains 13 features. The level of complexity of the Chinese illustration is similar to that from the U.S., while the TIMSS illustration shows a considerably lower level of complexity.

The illustrations from China and the U.S. have features in common: the world map, without background, zoom-out, seen externally and from above, two dimensional, proportionate, using grey scale, using text unit such as phrase as label and heading, text direction from left to right, located between stem/prompt and options/response format, explicitly referring the illustration, asking students to observe or examine the illustration, and as the same illustration for several related items. However, some differences exist between these two illustrations. The Chinese illustration still contains the following features: using black and white (as opposed to gray scale), showing the physical context (e.g. name of identified countries or cities such as Chile, Haiti, and Yushu to make a connection between the Ring of Fire with recent strong earthquakes in these places), using text unit Arabic numeral and word as data, and showing variables as value and scale. In contrast, the U.S. illustration contains features such as using metaphorical visual language to show space and location, and using text unit such as scientific sign, abbreviation, and letter as legend.

In a word, the close examination of illustrations indicates that despite the similar complexities of illustrations across item country of origins, the sets of features that make them complex can be different.

China

Recently, earthquakes occur frequently in the entire earth. Answer Questions 15 to 20 according to the following materials.

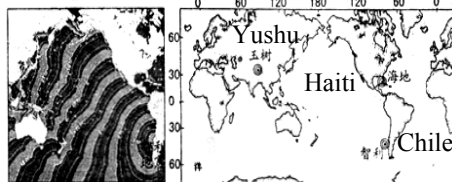


图1
Figure 1

图2
Figure 2

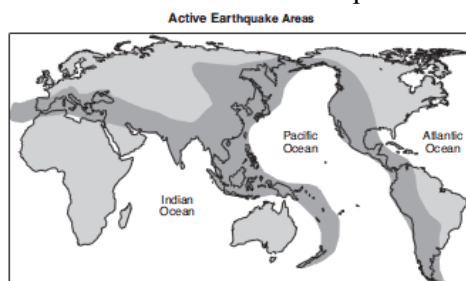
15 According to Figure 1, the major ocean affected by the tsunami caused by Chile earthquake is ()

- A. Pacific Ocean
- B. Indian Ocean
- C. Arctic Ocean
- D. Atlantic Ocean

A

U.S.

Base your answers to questions 51 and 52 on the map below and on your knowledge of science. The darker shading on the map shows the most active earthquake areas.

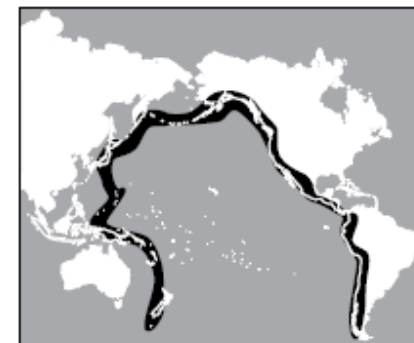


51 The most active earthquake areas are associated with the boundaries of lithospheric plates. Explain what happens to the lithospheric plates at these boundaries that causes an earthquake. [1]

52 Identify one geologic event, other than an earthquake, that may also occur in the darker shaded areas on the map. [1]

B

TIMSS



The diagram above shows the Pacific Ring of Fire. Earthquakes and volcanic activity occur along the Ring of Fire. Which of the following best explains why?

- a It is located at the boundaries of tectonic plates.
- b It is located at the boundary of deep and shallow water.
- c It is located where the major ocean currents meet.
- d It is located where ocean temperature is the highest.

C

Figure 13. Three comparable illustrations on the Ring of Fire from large-scale science tests in China, the U.S., and TIMSS. Sources: (A) Tianjin Municipal Education Commission. (n.d.). Tianjin Middle School Exit Examination, 2010 Earth Science test for Grade 8; (B) The University of the State of New York. (n.d.). 2009 Grade 8 Intermediate-level science test; (C) Trends in International Mathematics and Science Study. (n.d.). TIMSS 2003 Science Assessment Public Released Items for Grade 8. The Chinese item shown in English translation. See the original Chinese item in Appendix E.

CHAPTER VI

DISCUSSION

Overview

Frequently used in tests, illustrations vary considerably in contents and complexities. Over time, the reasoning and practices of many researchers and educators have been influenced by a prevalent premise—illustrations are a universal language and interpreted by everybody in the same way. However, scant research has examined factors that shape the effectiveness of illustrations used in printed materials, especially in tests. Additionally, existing normative documents of large-scale assessment programs do not provide a systematic method to develop and evaluate test illustrations. This lack of evidence and methodology reflects the great need for effective approaches to analyzing and evaluating illustrations in science assessment. This study is among the first approaches to analyzing the use of illustrations in the context of testing across cultures.

The purpose of this study was to provide empirical evidence for the similarities and differences of test illustration complexity in large-scale assessment systems in China, the U.S., and an international test comparison—TIMSS. Given the myriad of factors that affect the relationship between text and illustration and the lack of theories on the use of illustrations in tests, I adopted a multidisciplinary conceptual framework intended to provide a comprehensive lens through which to examine the key aspects of test illustrations. This framework is based on knowledge from three theoretical perspectives: cognitive science, sociocultural theory, and semiotics. Cognitive theories on visual perception and individual differences in perceiving visual information help explain the nuanced aspects on how a person processes visual information.

Sociocultural views of visual language provide support to the notion that social, cultural, and linguistic factors influence the interpretation of text with images. Semiotics provides a foundation for analyzing the multiplicity of meanings that visual images may carry.

In this study, I focused on two issues: how the numbers of different features in testing illustrations vary across test item origins, content areas, and item types, and the relationship between the number of features and student performance. This chapter summarizes the findings of this dissertation study and its relevance to the conceptual framework, its implications of the results as they relate to test item development, large-scale assessment, international test comparisons, classroom assessment and instruction, and cross-cultural research. In addition, I discuss this study's limitations and give suggestions for future research.

Summary of Results

Through quantitative analyses, I evaluated the characteristics of the test illustrations originated in China, the U.S., and TIMSS. An average test illustration tends to contain 22 features. Test illustrations from the three different assessment systems tended to have similar levels of complexity in terms of the total number of features and the number of features across dimensions. This tendency is also evident when considering content area (e.g., Earth Science, Life Science, Physics), item format type (e.g., multiple choice, constructed response, fill in blank etc.), and item commonality type (e.g., as part of a stand-alone item, same illustration for several items, and for one item within a series of related items). Examining symmetry graph of relative frequencies of features provides evidence of differences in test illustrations from China and the U.S. at the individual IDV level. Analyzing exemplar illustrations in the same sub-content areas

also indicates that although illustrations across cultures tended to have similar complexities, the sets of features that make them complex can be different.

Furthermore, differences (though not statistically significant) were observed in the number of features across the six illustration dimensions. The dimensions, Objects and Background and Text in Illustration, display the highest number of features, while the other two dimensions, Context in Illustration and Metaphorical Visual Language, display the least number of features. The dimension, Illustration-Text Interaction shows the least variation. A related study on investigating the semiotic characteristics of the PISA-2009 science illustrated items (see Solano-Flores & Wang, 2012) reported similar findings.

Another very important finding from this study was that significant differences in illustration complexity on the dimension, Context in Illustration were attributable to the item country of origin. More specifically, illustrations used in large-scale testing programs in China showed significantly more features than illustrations used in large-scale testing programs in the U.S. for Context in Illustration (form and complexity of visual information represented as physical context, socio-historical context, and/or appearance of human figure). China's large-scale science tests provide, more frequently than illustrations used in the U.S. large-scale science tests, visual information connecting the content of the item with the country's national, cultural, political, and historical contexts.

Further statistical analyses also revealed that content areas tended to contain different number of features. On average, test illustration in Earth Science tended to have slightly more features than those in Physical Science and Life Science. The average number of features in test illustrations varies across the combination of content area and test item origin. Test illustrations

in Earth Science from China display the highest number of features, while those in Life Science from China display the lowest number of features. Significant differences in illustration complexity were attributable to the content area of test illustrations. Additionally, examining illustrations with the most frequent outstanding IDVs across item country of origin indicates that those in Physical Science tended to have more similar sets of features than their counterparts in Earth Science or Life Science.

Overall, illustration complexity seems to have different impact on the performance of students of Chinese heritage (Chinese-Taipei and China Hong Kong SAR) and that of students from the U.S. and from all participating countries/regions. Statistically significant correlations were observed only for the dimension, Context in Illustration for students from the U.S. and from all participating countries/regions. This indicates that when the contextual features in a test illustration increase, the performance of students from the U.S. and from all participating countries/regions also increases.

Current Findings as They Relate to the Conceptual Framework

Cognitive Science. The empirical evidence of the complexity of features in test illustration contributes to cognitive science, especially the dual coding theory. Overall, illustrations used in test items tend to be complex and vary in their complexity of features across assessment systems, content areas, and test item types. However, the findings do not give a clear indication that illustrations necessarily support student performance. Increased complexity in the textual features (e.g. using words, phrases as labels or captions) tends to be negatively related to the performance of students regardless of their linguistic and cultural backgrounds. This indicates despite the good intention of including illustrations in instructional and testing materials, the

complexity of the textual features in illustrations may overload the process in which student perceive and interpret the visual information presented. The combination of complexity of features in illustration and complexity of features in accompanied text may even increase the overload. In contrast, another relevant study (Solano-Flores & Wang, 2012) found a positive correlation between the number of text-in-illustration features and the proportion of students who responded correctly to PISA 2009 items. The more textural features an illustrated science item contained, the higher the percentage of students who gave correct response to the items. Future research needs to examine this particular illustration dimension, and gauge the sets of illustration features which influence cognitive load experienced by student. The cognitive demands due to the interaction between the text of the test item and its illustration should also be further explored.

My findings are also consistent with the constructivist theory of visual perception that indicates that perceiving and interpreting an image is cognitively mediated by stored knowledge and experience of changing conventions (Gombrich, 1969; 1972; Gregory, 1997). For example, graphic devices, commonly seen as an integral part of test illustrations, consist of arrows, motion lines, speech balloons, and many other conventional forms of representation. In the present study, the more graphic devices were used in a test illustration, the less likely students were to give a correct response to the item in spite of their different cultural and linguistic backgrounds. This indicates that, in order to properly convey meaning, graphic devices need to be designed and used based on the notion that their meaning is mediated by culturally-bound conventions. Students need to have extensive exposure or explicit instruction to become familiar with these visual resources, if these resources are to support students to gain access to the content of the illustration and the text as intended.

Sociocultural Theory. The findings in the present study are consistent with sociocultural theory by showing that test illustrations generated from different countries tend to have different contextual features. The findings also speak to the importance of including sociocultural contextual features in test illustrations. According to Vygotsky (1978), human beings make meaning through participation in culturally and socially embedded activities. Chinese test illustrations tend to provide a more meaningful context to students than their U.S. counterparts. For students of Chinese origin (in my study, those from Chinese Taipei and China Hong Kong), the correlations between number of features and item p -value tend to show different patterns in terms of magnitude differences and directions from students from the U.S. This indicates that the level of accuracy with which students make sense of illustrations is mediated by their familiarity with the culturally and socially embedded conventions. In a study on students' interpretation of science test illustrations, college students from China and the U.S. tended to make more accurate interpretations of illustrations from items created in their own countries than those not created in their countries (Wang & Solano-Flores, 2011).

Hall believes that an image shows what is valued in a dominant culture, which in turn determines what is and what is not represented (1980). My findings are consistent with his thinking in that in the case of design and use of test illustration, a test illustration is developed within a historical point in time and a specific sociocultural context. Compared with their counterpart from the U.S., test illustrations from China contains considerably more contextual features, which reflect the cultural, historical, and political contexts unique and salient to contemporary China. Furthermore, the complexity of these contextual features in test illustrations is significantly correlated with student performance. Illustration designers and

creators need to be well aware of to the value of making sociocultural context in illustration explicit to the student, and develop test illustrations accordingly.

Semiotics. The findings in my study are consistent with the semiotic theory that communication is made through multiple forms (Kress & van Leeuwen, 2006). Communication and representation always draw on various semiotic modes. In science, meaning is conveyed not only from the lexis, grammar, and discourse commonly seen in science academic language, but also from multiple forms of representation of visual information. In the context of testing, to make meaning of an illustrated science test item, students need to be aware of and be able to interpret both the linguistic and non-linguistic features embedded in the text and illustration of the item. My study examined the features in test illustrations, analyzed their distributions, and examined their relationship with the performance of students from diverse cultural and linguistic backgrounds. It provides evidence that the features embedded in a test illustration can be considerably complex, and the increase of the complexity in the illustration may, in the case of text-in-illustration features, make student correct responses to test items less likely to occur.

The findings of the present study are also consistent with the notion that any image has complex cultural meanings (Peirce, 1966). For example, illustrations originated in large-scale testing programs in China tend to contain significantly more contextual features than those in the U.S. Students from China may be more familiar with visual symbols that convey meaning relating to the country's national, cultural, political, and historical contexts. Additionally, examination of illustrations by constructing a symmetry graph and comparing illustrations in the same sub-content areas across assessment systems showed that, despite the similarity in terms of

number of features exhibited, test illustrations of different origins can contain very different sets of features.

Implications

Implications for Test Item Development and Large-Scale Assessment. In order to produce more valid measures of science achievement, test developers need to follow systematic procedures to develop and evaluate illustrations used in large-scale assessment programs. However, both relevant existing empirical research and normative documents on item specifications or test frameworks do not provide such procedures and guidance. Results from this study show that the coding system used to code features of illustrations in science tests is sensitive to differences in illustration complexity. It can be adopted by test developers to design and examine the features of test illustration systematically. It can also be used to examine the ways in which illustration complexity shapes student performance.

This study also indicates that important differences exist in the ways that illustration complexity varies across dimensions and content areas. For example, regardless of item origin, the dimensions, Objects and Background and Text in Illustration, had proportionally the highest number of features observed in the items examined. The inclusion of illustrations in certain items and the control of their complexity should be justified on the sub-content area those items belong to. Test developers need to consider these differences in order to determine illustration design criteria and parameters and therefore standardize the characteristics of test illustrations and control their complexity. In another recent investigation, my colleagues and I (Wang, Chia, Kachchaf & Solano-Flores, 2012) found that we were able to systematically create test illustrations with similar complexities.

Implications for International Test Comparisons. This study contributes to informing the process of assessment design in international test comparison programs. One of the major findings of this study is that illustrations from China displayed considerably more features than their U.S. counterparts on the illustration dimension, Context in Illustration. China's illustrations provide more contextual information relating to social events, political affairs, holidays and festivals, etc. This fact and the fact that items with similar complexities owe their complexities to different sets of features have important implications for the design and evaluation of illustrations used in international test comparisons. It is critical to be aware of the differences in the degree of familiarity students from different countries or regions have with test illustrations of high complexity and with test illustrations of different sets of features.

Besides, current research studies on international test comparisons primarily focus on the written text. The findings obtained in this study provide empirical evidence regarding important cultural differences in visual information in international test comparisons, and may lead to systematic design and use of illustrations in such assessment context.

Additional findings indicate that illustration complexity has a different impact on the performance of students of diverse cultural backgrounds. For students of Chinese origin, illustration complexity was not correlated with student performance on any of the six dimensions. In contrast, the correlation was statistically positive on the dimension Context in Illustration for students from the U.S. This information may have important implications for improved test design involving the testing of culturally and linguistically diverse populations: The illustration features of science items may or may not correlate with student performance depending on the

population. Irrelevant construct variance can be minimized through proper visual contextualization.

Implications for Classroom Teaching and Assessment. Teachers need to know about features and functions of illustrations to better use visual aids to support and assess student learning. Results from this study show that an average test illustration displays more than 20 features. Those features varied across content area and item type. This finding alerts us about the cognitive demand in the complexity of illustration features placed on individual students. Simply adding illustrations may not always facilitate students' interpretation of test items. Teachers need to be well aware of the complex ways illustration features may affect student learning and understanding.

Gregory (1989), a perception psychologist, also argues:

Like language, pictures can project into the past or imagined future and create new or even impossible worlds. But they work only for observers with knowledge and intelligence to create meaning from the pictures, as one creates meaning from the words of a language. (p. 8)

Visual literacy in science is an essential precondition of effective science communication. Teachers should be aware of the urgent need to make visuals used in science instruction or tests more accessible for students by providing explicit instruction that help students develop expertise in understanding their nuanced aspects. The information on the illustration features should inform their teaching about systematically use and design illustrations in the context of classroom instruction and testing. Additionally, teacher professional development programs

should also be aware of the importance of examining illustrations systematically and provide opportunities to help teachers acquire corresponding knowledge and skills.

Implications for Cross-Cultural Research. Finally, this study contributes to the field of research involving cross-cultural testing with a more comprehensive approach to examining cultural differences. Unlike common cross-cultural research that focuses on cultural group as a factor based on stimulus materials produced in one culture, this study addresses the similarities and differences in stimulus materials produced in different cultures. This kind of design makes it possible to perform more thorough analyses of cultural issues in testing. This is critical in an era of a global dialogue between countries and cultures, in which cross-cultural communications and studies are in the forefront.

Limitations

One limitation of this study was the relatively small sample size. The illustrated items used in this study were selected from three states in the U.S., and the Capital city, Washington D.C., and four municipalities in China. For this reason, the findings cannot be generalized to large-scale tests originated from other U.S. states and China cities or provinces. Also, items in certain combinations of grade level and content area had to be eliminated due to limited number of illustrated items available. For example, illustrated items in chemistry were not available in many assessment systems. As a consequence, the item sample used in this study was limited to illustrated items for Grade 8/9 in Earth Science, Life Science, and Physical Science. In addition, the items from the U.S. and TIMSS were released items. Therefore, the sample of items selected in this study was not completely representative of the entire population of illustrated science items in China, the U.S., and TIMSS.

A second limitation of this study has to do with the restricted access that I had to data on student performance. For all the illustrated items in this study, I only had access to item p -values in TIMSS. Also, due to the fact that the People's Republic of China does not participate in TIMSS, I could only collect p -values for students from the U.S., Hong Kong, and Chinese Taipei. Although students from Hong Kong SAR and Chinese Taipei were of Chinese origin, because of historical reasons, their education systems are different in many ways from that of students from mainland China. Therefore, the findings regarding the relationship between illustration complexity and student performance cannot be generalized to students in the People's Republic of China.

Next, there are possible cultural biases in coders. The two coders have similar cultural and linguistic backgrounds. They speak Chinese as their native language, and were born, raised, and received formal college and masters education in China. Their background and knowledge might help them accurately interpret the Chinese illustrated items, but not necessarily help them achieve the similar level of accuracy in the interpretation of the U.S. illustrated items.

Finally, there are some limits of the coding system used to examine the features of testing illustrations. For example, the coding system was developed based on illustrated items from a relatively limited number of large-scale assessment systems in the U.S. and China. As a consequence, the extent to which the current coding system can capture all relevant illustration features is unknown. However, besides the 130 identified IDVs, this coding system intentionally included the term *other* at the end of each of the 24 subcategories, in order to capture any features not considered. In addition, since this coding system is generic, it may not be able to

capture content-specific features. For example, representation of animals and plants is commonly seen in life science, not in physics.

Future Research

Larger random samples of illustrated items would ensure sufficient numbers of items across all the content areas, item format types, and item commonality types should be used in future research on test illustration complexity and its relation to student performance. These larger samples would also provide more illustrations for qualitative analyses across different assessment systems. Next, future research needs to recruit additional coders who were born, raised, received formal training in science education and/or linguistics in the U.S., and bilingual in both English and Mandarin Chinese.

Conclusion

In sum, the evidence obtained speaks to the fact that test developers must be extremely cautious in their assumptions about the properties and effectiveness of illustrations in science testing. While it may be true that a picture is worth a thousand words, the ways in which it is designed, used, and interpreted depends on a variety of factors, especially the culture in which it originates. While many of these factors do not appear to influence student performance, others do, especially those of contextual nature, and especially for some populations. Test developers need to be well aware of the importance of using a systematic procedure to design and evaluate illustrations to ensure more valid assessment of culturally diverse populations in both the U.S. and the context of international test comparisons.

Notes

Note 1. The term *iconic* was used by Peirce to refer to resemblance to the object they represent.

However, in my study, it refers to a level of realism in representing specific objects or actions, and is interpreted as compared with photo, realistic line drawing, schematic, silhouette, cartoon, metonymy, symbol, reference, and entity.

Note 2. *p*-value refers to the percentage of students who answer a test item correctly.

REFERENCES

- Abedi, J. (2004). The No Child Left behind Act and English Language Learners: Assessment and accountability issues. *Educational Researcher*, 33(1), 4-14.
- Alley, D. C. (1994). Trends in the use of illustrations in university Spanish textbooks. *Hispania*, 77(3), 489-495.
- Allport, G.W. (1937). *Personality: A psychological interpretation*. New York: Holt & Co.
- Anglin, G. J., Vaez, H., & Cunningham, K. L. (2004). Visual representations and learning: The role of static and animated graphics. In D. H. Jonassen (Ed.), *Handbook of Research on Educational Communications and Technology* (pp. 865-916). Mahwah, NJ: Lawrence Erlbaum.
- Arizona Department of Education. (2009). Arizona's instrument to measure standards AIMS Science, 2009 sample test for Grade 8. Retrieved October 28, 2009, from <http://www.ade.state.az.us/standards/aims/sampletests/Gr8-AIMSSampletestscience.pdf>
- Arnheim, R. (1974). *Art and visual perception: A psychology of the creative eye*. Los Angeles, CA: University of California Press.
- Arnheim, R. (1986). *New essays on the psychology of art*. Berkeley, CA: University of California Press.
- Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps*. Madison, WI: University of Wisconsin Press.
- Bloomer, R. H. (1960). Students's Preferences and Responses as Related to Styles and Themes of Illustration. *The Elementary School Journal*, 60(6), 334-340.

- Bogdan, C. (2002). *The semiotics of visual languages*. New York, NY: Columbia University Press.
- Boling, E., Eccarius, M., Smith, K., & Frick, T. (2004). Instructional illustrations: Intended meanings and learner interpretations. *Journal of Visual Literacy*, 24(2), 185–204.
- Boling, E., Smith, K., Frick, T. & Eccarius, M. (2007). *Graphic devices in instructional illustrations: Designers' intentions and viewers' interpretations*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Carletta, J. (1996) Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2), pp. 249–254.
- Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America*, 102(35), 12629–12633.
- Clarke, J. H. (1991). Using visual organizers to focus on thinking. *Journal of Reading*, 34(7), 526-534.
- Cohen, C. A. & Hegarty, M. (2007). Individual differences in use of an external visualization while performing an internal visualization task. *Applied Cognitive Psychology*, 21, 701-711.
- Cohen, J. (1968). "Weighed kappa: Nominal scale agreement with provision for scaled disagreement or partial credit". *Psychological Bulletin* 70: 213–220.
- Conkey, M. W. (1994). A century of Palaeolithic cave art. In B. Katz (Ed), *A history of book illustration: 29 points of view* (pp. 1-22). Metuchen, N.J., & London: The Scarecrow Press, Inc.

- Crisp, V. & Sweiry, E. (2006). Can a picture ruin a thousand words? The effects of visual resources in exam questions. *Educational Research*, 48(2), 138-154.
- Dimopoulos, K., Koulaidis V., & Sklaveniti, S. (2003). Towards an analysis of visual images in school science textbooks and press articles about science and technology. *Research in Science Education*, 33(2), 189-216.
- Eco, U. (1976). *A theory of semiotics*. Bloomington: Indiana University Press.
- Eco, U. (1986). *Semiotics and the philosophy of language*. Bloomington: Indiana University Press.
- Evans, J., & Hall, S. (1999). What is visual culture? In J. Evans & S. Hall. (Eds.) *Visual culture: A reader* (pp. 1-8). London: Sage.
- Evans, M. A., Watson, C., & Willows, D. M. (1987). A naturalistic inquiry into illustrations in instructional textbooks. In D. Willows & H. Houghton (Eds.), *The psychology of illustration: Volume 1, Basic research* (pp. 86-115). New York: Springer-Verlag.
- Eysenck, M. W., & Keane, M. T. (2005). *Cognitive psychology: A student's handbook*. New York, NY: Taylor & Francis.
- Fichtner, D., Peitzman, F., & Sasser, L. (1994). What's fair? Assessing subject matter knowledge of LEP students in sheltered classrooms. In F. Peitzman & G. Gadda (Eds.), *With different eyes: Insights into teaching language minority students across the disciplines* (pp. 143-159). Los Angeles: UCLA Publishing Center for Academic Interinstitutional Programs.
- Filippatou, P. and Pumphrey, P.D. (1996) Pictures, titles, reading accuracy and reading comprehension: A research review (1973–95). *Educational Research*, 38(3), 259–291.

- Fleming, M. L. (1966). *Instructional illustrations: A survey of types occurring in print materials for four subject areas*. Final Report, November, 1966, Project 1381, NDEA Title VII A, U.S. Office of Education.
- Florida Department of Education. (2005). *FCAT handbook: A resource for educators*. Retrieved June 22, 2009, from <http://fcat.fldoe.org/handbk/fcathandbook.asp>
- Freedberg, D. (1989). *The power of images: Studies in the history and theory of response*. Chicago and London: The University of Chicago Press.
- Garrick, C. E. (1978). Design of instructional illustrations in medicine. *Journal of Audiovisual Media in Medicine*, 1, 161-173.
- Georgia Department of Education. (2004). Georgia End-Of-Course Tests Physical Science, released test booklet. Retrieved September 10, 2010, from http://www.doe.k12.ga.us/ci_testing.aspx?PageReq=CI_TESTING_EOCT&SubPageReq=GUIDES
- Gibson, J. J. (1971). The information available in pictures. *Leonardo*, 4, 27-35.
- Ginther, A. (2001). *Effects of the presence and absence of visuals on performance on TOEFL CBT listening-comprehensive stimuli* (TOEFL Research Report No. 66). Princeton, NJ: Educational Testing Service.
- Goldsmith, E. (1984). *Research into illustration*. Cambridge: Cambridge University Press.
- Goldsmith, E. (1987). The analysis of illustration in theory and practice. In H. A. Houghton & D. M. Willows (Eds.). *The psychology of illustration, Volume 2, Instructional issues* (pp. 53-86). New York: Springer.
- Gombrich, E. (1969). *Art and illusion*. Princeton, NJ.: Princeton University Press.

- Gombrich, E. (1972). The “what” and the “how”: Perspective representation and the phenomenal world. In R. Rudner and I. Scheffler (Eds.), *Logic and art: Essays in honor of Nelson Goodman*. Indianapolis: Bobbs-Merrill.
- Greenfield, P.M. (1994). Video games as cultural artifacts. *Journal of Applied Developmental Psychology*, 15, 3–12.
- Gregory, R.L. (1989). Turning Minds on to Science by Hands-On Exploration: The Nature and Potential of the Hands-On Medium. In: *Sharing science: Issues in the development of the interactive science and technology centres* (pp. 1-9). (Key-Note essay). The Nuffield Foundation Interactive Science and Technology Project. On behalf of COPUS - Committee on the Public Understanding of Science.
- Gregory, R. L. (1997). *Eye and brain: The psychology of seeing* (5th ed.). Princeton, NJ: Princeton University Press.
- Hagen, M. (1980). Generative theory: A perceptual theory of pictorial representation. In M. Hagen (Eds.), *The perception of pictures* (pp. 3–46). New York: Academic Press.
- Hall, S. (1980). Cultural Studies: two paradigms. *Media, Culture and Society*, 2, 57–72.
- Halliday, M. A. K. (1975). Language as social semiotic: Towards a general sociolinguistic theory. In J. Webster (Ed.), *Volume 10 in the collected works of M.A. K. Halliday: Language and society*. London and New York: Continuum.
- Halliday, M. A. K. (1978). *Language as social semiotic*. Baltimore, MD: University Park Press.
- Harp, S. F. & Mayer, R. E. (1997). Role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology*, 89, 92-102.

- Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90, 414-434.
- Hatzinikita, V., Dimopoulos, K., & Christidou, V. (2008). PISA Test Items and School Textbooks Related to Science: A textual comparison. *Science Education*, 92(4), 664-687.
- Hegarty M. (2004). Diagrams in the mind and in the world: Relations between internal and external visualizations. In A. Blackwell, K. Mariott and A. Shimojima (Eds.). *Diagrammatic Representation and Inference. Lecture Notes in Artificial Intelligence* 2980 (pp. 1-13). Berlin: Springer-Verlag.
- Heilongjiang Department of Education. (2001). Heilongjiang Middle School Exit Examination, 2001 physics test for Grade 9. Retrieved September 10, 2010, from <http://roll.edu.sina.com.cn/more/zk/zkst/index.shtml>
- Hoopes, J. (1991). *Peirce on signs*. Chapel Hill, NC: The University of California Press.
- Horton, W. (1993). The almost universal language: Graphics for international documents. *Technical communication*, 40(4), 682–693.
- Hunter, B. Crismore, A., Pearson, P. D. (1987). Visual displays in basal readers and social studies textbooks. In D. Willows & H. Houghton (Eds.), *The psychology of illustration: Volume 2, The psychology of illustration* (pp. 116-135). New York: Springer-Verlag.
- Jastrzebski, Z. (1985). *Scientific Illustration: A guide for the beginning artist*. Englewood Cliffs: Prentice-Hall.
- Katz, B. (1994). *A history of book illustration: 29 points of view*. Metuchen, N.J., & London: The Scarecrow Press, Inc.

- Knight, E., Gunawardena, C. N., Aydin, C. H. (2009). Cultural interpretations of the visual meaning of icons and images used in North American web design. *Educational Media International*, 46 (1), 17-35.
- Kopriva, R. (2008). *Improving testing for English language learners*. New York: Routledge.
- Kosslyn, S. M. (2006). *Graph design for the eye and mind*. New York: Oxford University Press, Inc.
- Kress, G. (1997). *Before writing. Rethinking the paths to literacy*. London: Routledge.
- Kress G. , & van Leeuwen T. (2006). *Reading images*. London, New York: Routledge.
- Lemke, J. L. (2003). *Mathematics in the middle: Measure, picture, gesture, sign, and word*. Retrieved March 9, 2010, from <http://academic.brooklyn.cuny.edu/education/jlemke/papers/myrdene.htm>
- Lester, P. M. (2003). *Visual communication: Images with messages* (3rd ed.). Belmont, CA: Thomson/Wadsworth.
- Levie, W. H. (1978). A prospectus for research on visual literacy. *Educational Communication and Technology*, 26, 25–36.
- Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational communications and technology journal*, 30(4), 195-232.
- Marriott K., & Meyer, B. (1998). *Visual language theory*. New York: Springer-Verlag.
- Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. *Educational Assessment*, 14, 160-179.

- Maryland State Department of Education. (2008). Maryland School Assessment Science, 2008 Public Release Grade 8. Retrieved September 10, 2010, from http://mdk12.org/assessments/k_8/index_c.html
- Massachusetts Department of Elementary & Secondary Education. (2005). Massachusetts Comprehensive Assessment System Science, 2005 sample test for Grade 8. Retrieved September 10, 2010, from <http://www.doe.mass.edu/mcas/testitems.html>
- Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animation in multimedia instruction. *Journal of Experimental Psychology*, 11(4), 256-265.
- Mayer, R.E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93(1), 187-198.
- Mayer, R.E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 28(1), 43-52.
- Mayer, R. E. & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86, 389-401.
- Miller, W. A. (1936). The Picture Choices of Primary-Grade Students. *The Elementary School Journal*, 37(4), 273-282.
- Miller, W. A. (1938). Reading with and without Pictures. *The Elementary School Journal*, 38(9), 676-682.
- Mohan, B. (1986). Language and content. Reading, MA: Addison-Wesley.

- Moriarty, S. (2005). *Visual semiotics theory*. In K. Smith, S. Moriarty, G. Barbatsis & K. Kenney (Eds.). *Handbook of Visual Communication* (pp. 227-241). New Jersey: Lawrence Erlbaum Associates, Inc.
- Mulcahy, P., & Samuels, S. J. (1987). Three hundred years of illustrations in American textbooks. In H. Houghton, & D. Willows (Eds.), *The psychology of illustration*. New York, NY: Verlag.
- Mullis, I.V.S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 Assessment Frameworks*. Boston, MA: International Association for the Evaluation of Educational Achievement (IEA); TIMSS & PIRLS International Study Center.
- Murphy, S. (2010). The pull of PISA: Uncertainty, influence, and ignorance. *Revista Interamericana de Educación para la Democracia--Interamerican Journal of Education for Democracy*, 3(1), 27-44.
- National Assessment of Educational Progress. (2005). *NAEP 2005 Science Assessment Public Released Items for Grade 8*. Retrieved September 18, 2009, from <http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=science>
- National Assessment of Educational Progress. (2009). *NAEP 2009 Science Assessment Public Released Items for Grade 8*. Retrieved March 18, 2011, from <http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=science>
- National Bureau of Statistics of China. (2009). *China statistical yearbook*. Fengtai, Beijing: China Statistics Publishing House.
- National Center for Education Statistics. (2009). *NAEP Overview*. Retrieved November 20, 2009 from <http://nces.ed.gov/nationsreportcard/about/>

- National Center for Education Statistics. (2009). *Overview of Trends in International Mathematics and Science Study*. Retrieved November 20, 2009 from <http://nces.ed.gov/timss/> <http://nces.ed.gov/timss/>
- Ningxia Department of Education. (2008). Ningxia Middle School Exit Examination, 2008 physics test for Grade 9. Retrieved September 10, 2010, from <http://roll.edu.sina.com.cn/more/zk/zkst/index.shtml>
- Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: holistic versus analytic perception. *Trends in Cognitive Sciences*, 9(10), 467-473.
- Nodelman, P., & Reimer, M. (2003). *The pleasures of children's literature*. Boston: Allyn and Bacon.
- O'Donnell, V. (2005). *Cultural studies theory*. In K. Smith, S. Moriarty, G. Barbatsis & K. Kenney (Eds.). *Handbook of Visual Communication* (pp. 521-537). New Jersey: Lawrence Erlbaum Associates, Inc.
- O'Malley, J.M., & Pierce, L. V. (1996). *Authentic Assessment for English Language Learners: Practical Approaches for Teachers*. Reading, MA: Addison-Wesley Publishing Company.
- Organisation for Economic Co-operation and Development. (2003). *The PISA Assessment Framework. Mathematics, Reading, Science and Problem Solving Knowledge and Skills*. Paris: OECD.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.
- Paivio, A. (1986). *Mental representations: a dual coding approach*. Oxford. England: Oxford University Press.

- Peirce, C.S. (1966). *Collected papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.
- Pettersson, R. (2002). *Information design: An introduction*. Amsterdam, The Netherlands: John Benjamins Publishing Co.
- Pressley, M. (1977). Imagery and Students's Learning: Putting the Picture in Developmental Perspective. *Review of Educational Research*, 47(4), 585-622.
- Rollins, M. (1999). Pictorial representation: When cognitive science meets aesthetics. *Philosophical Psychology*, 12, 387-413.
- Rowe, D.W. (1994). *Preschoolers as authors: Literacy learning in the social world of the classroom*. Cresskill, NJ: Hampton Press.
- Sadoski, M., & Paivio, A. (2001). *Imagery and text: A dual coding theory of reading and writing*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Saint-Martin, F. (1990). *Semiotics of Visual Language*. Bloomington and Indianapolis: Indiana University Press.
- Schiffman, H.F. (1996). *Linguistic culture and language policy*. London, UK: Routledge.
- Shanahan, T. B. (2006). *A comparison of graphic and visual test accommodations in measuring urban fifth-grade English language learners' science content acquisition*. Doctoral dissertation, University of Southern California.
- Sharrocks-Taylor, D., & Hargreaves, M. (1999). Making it clear: A review of language issues in testing with special reference to the National Curriculum Mathematics Tests at Key Stage 2. *Educational Research*, 41(2), 123-136.

- Siegel, M. (1995). More than words: The generative power of transmediation for learning. *Canadian Journal of Education*, 20(4), 455-475.
- Sitz, R. (1996). Eyes on cognitive styles and the processing of visual information. In R.E. Griffin (Ed.), *Eyes on the future: Converging images, ideas, and instruction* (p. 279-288). Selected readings from the Annual Conference of the International Visual Literacy Association(27th, Chicago, Illinois October 18-22, 1995).
- Solano-Flores, G. (2010a). Vignette illustrations as a form of testing accommodation for English language learners: A design methodology for use in large-scale science assessment. Paper presented at the Annual Conference on the National Council of Measurement in Education, Denver, Colorado, April 29- May 3.
- Solano-Flores, G. (2010b). The use of pictorial supports as an accommodation for increasing access to test items for students with limited proficiency in the language of testing. Paper presented at the 7th Conference of the International Test Commission. Hong Kong, July 19-21, 2010.
- Solano-Flores, G. (2011). *Development of illustrations as image supports for English language learners in large-scale testing: A report on the procedure for designing vignette illustrations*. Paper to be presented at the CADRE ELL-STEM Roundtable session, “Advancing English Language Learners in Science and Math: Realizing the Promise,” at the annual meeting of the American Educational Research Association, April 7-11.
- Solano-Flores, G., & Wang, C. (2011). *Conceptual framework for analyzing and designing illustrations in science assessment: Development and use in the testing of linguistically*

- and culturally diverse populations*. Paper presented at the Annual Conference of the National Council on Measurement in Education, New Orleans, LA, April 7-11.
- Svensen, B. (1993). *Practical lexicography*. Oxford: Oxford University Press.
- Tang, G. (1991). The role and value of graphic representation of knowledge structures in ESL student learning: An ethnographic study. *TESL Canada Journal/Revue TESL Du Canada*, 9 (1), 29-41.
- Tang, G. (1994). Textbook illustrations: A cross-cultural study and its implications for teachers of language minority students. *The Journal of Educational Issues of Language Minority Students*, 13, 175-194.
- Texas Education Agency. (2004). Texas Assessment of Knowledge and Skills, 2004 released science test for Grade 5. Retrieved September 10, 2010, from <http://www.tea.state.tx.us/student.assessment/released-tests/>
- Trends in International Mathematics and Science Study. (2007). *TIMSS 2007 Science Assessment Public Released Items for Grade 8*. Retrieved March 18, 2011, from <http://timss.bc.edu/TIMSS2007/items.html>
- U.S. Census Bureau. (2010). *Interactive population map*. Retrieved May 18, 2011, from <http://2010.census.gov/2010census/popmap/>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, Massachusetts: Harvard University Press.
- Wang, C., & Solano-Flores, G. (2011). *The use of illustrations in large-scale science assessment: A comparative study*. Paper presented at the annual meeting of the American Educational Research Association, April 8-12.

- Wang, C., Chia, M., Kachchaf, R., & Solano-Flores, G. (2012). *Item complexity and the performance of English language learners in a science test*. Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, British Columbia, Canada, April 13-17.
- Weidenmann, B. (1994). Codes of instructional pictures. In W. Schnotz & W. Kulhavy (Eds.), *Comprehension of Graphics*. Amsterdam: North-Holland, pp.29-42.
- Wertsch, J.V. (1985). *Culture, Communication and Cognition: Vygotskian Perspectives*. Cambridge: Cambridge University Press.
- Wertsch, J., del Rio, P., & Alvarez, A. (1995). Sociocultural studies: History, action, and mediation. In J. Wertsch, P. del Rio, and A. Alvarez (Eds.), *Sociocultural studies of mind*, (pp. 1–34). New York: Cambridge University Press.
- WestEd and the Council of Chief State School Officers. (2007). Science assessment and item specifications for the 2009 National Assessment of Educational Progress. Retrieved June 22, 2010, from <http://www.nagb.org/publications/frameworks/naep2009-science-specs07.doc>
- Winn, B. (1987). Charts, graphs, and diagrams in educational materials. In D. Willows & H. Houghton (Eds.), *The psychology of illustration: Volume 1, Basic research* (pp. 152-198). New York: Springer-Verlag.
- Witkin, H.A., & Asch, S.E. (1948a). Studies in space orientation III: Perception of the upright in the absence of a visual field, *Journal of Experimental Psychology* **38**, pp. 603–614.

- Witkin, H.A., & Asch, S.E. (1948b). Studies in space orientation IV. Further experiments on perception of the upright with displayed visual field, *Journal of Experimental Psychology* **38**, pp. 762–782.
- Wohlwill, J. F. (1968). Responses to Class-Inclusion Questions for Verbally and Pictorially Presented Items. *Child Development*, 39(2), 449-465.
- Zimet, S. F. (1966). Children's Interest and Story Preferences: A Critical Review of the Literature. *The Elementary School Journal*, 67(3), 122-130.

APPENDIX

APPENDIX A

Historical and contextual information of the science assessments included in the dissertation study

Item origin	Assessment program	Assessment designer	Assessment purposes	Targeted population
China	Shanghai Middle School Exit Examination	Shanghai Municipal Educational Examination Authority	The examinations aim at comprehensively and accurately assessing students' proficiency levels in learning according to curriculum standard for each of the content areas. The results of the examinations are not only an important reflection of knowledge and skills learned from Junior high school, but also a critical criterion for senior secondary school enrollment.	Currently in Shanghai Grade 8 in Earth Science and Life Science, and Grade 9 in Physical Science and Chemistry
	Tianjin Middle School Exit Examination	Tianjin Education Examination Authority	The examinations aim at examining if students meet the curriculum requirements in Junior high school.	Currently in Tianjin Grade 8 in Earth Science and Life Science, and Grade 9 in Physical Science and Chemistry
	Chongqing Middle School Exit Examination	Chongqing Municipal Academy of Education Sciences	The content of the examinations should meet the requirements covered in curriculum standards in each subject areas, which are formulated by the Ministry of Education.	Currently in Chongqing Grade 8 in Earth Science and Life Science, and Grade 9 in Physical Science and Chemistry
	Beijing Middle School Exit Examination	Beijing Education Examination Authority	The scope of the examinations is based on the contents covered in curriculum standards in each subject areas. The examinations aim at examining students' knowledge, basic skills, and	Currently in Beijing Grade 8 in Earth Science and Life Science, and Grade 9 in Physical Science and Chemistry

U.S.	Texas Assessment of Knowledge and Skills (TAKS)	Texas Education Agency, Pearson, and Texas educators collaborate to make TAKS	ability to solve problems. TAKS aims at assessing students' attainment of reading, writing, math, science, and social studies skills required under Texas education standards.	Currently in Texas grade 9-11, and before 2012 grade 3-8
	The University of the State of New York science Test	The Office of Assessment Policy, Development and Administration (APDA) is responsible for the coordination, development, and implementation of the New York State Testing Program (NYSTP)	The Regulations of the Commissioner of Education provide that an intermediate-level science test is to be administered in Grade 8 to serve as a basis for determining students' need for academic intervention services in science. The Grade 8 Intermediate-Level Science Test is designed to measure the content and skills contained in the <i>Intermediate-Level Science Core Curriculum, Grades 5–8</i> . The core curriculum is based on the <i>New York State Learning Standards for Mathematics, Science, and Technology</i> .	New York State grade 8
	District of Columbia Comprehensive Assessment System (DC CAS)	DC Office of the State Superintendent of Education (OSSE)	These annual tests aim to measure student academic proficiency of DC Content Standards, and used to calculate whether a schools meets Adequate Yearly Progress (AYP).	DC grade 2-10 in English–language arts (ELA), grade 2-8 and grade 10 in Math, grade 5, 8, 10 in Science, and grade 5, 8, 10 in Health.
	California Standards Tests (CSTs)	The CSTs are developed by California educators and test	They measure students' progress toward achieving California's state-adopted academic content standards in English–language arts (ELA), mathematics,	Students in grades two through eleven take multiple-choice CSTs for various subjects. Students in grades

		developers specifically for California	science, and history—social science, which describe what students should know and be able to do in each grade and subject tested.	four and seven complete a writing assessment—the CST for Writing—as a part of the CST for ELA.
International Test	Trends in International Mathematics and Science Study (TIMSS)	International Association for the Evaluation of Educational Achievement (IEA)	These series of international assessments of student achievement are dedicated to improving teaching and learning in mathematics and science.	Grade 4 and 8 students around the world

APPENDIX B
Definitions and attributes of visual image dimensions

Visual image dimension	Definition	Attribute
Representation of objects and background	Style, complexity, and level of concreteness or abstraction with which objects and background are represented.	Image concreteness (photograph, scanned document, realistic line drawing, schematic, silhouette, cartoon, and others), background, zooming (zoom-in, zoom-out, and zero zooming naked eye), view (external, internal, from above object, from below object, from the side of object), dimension (3D, and 2D), relative scale of objects, color (black and white, multicolor, and gray scale), and composition (single image, and compound image)
Use of context in the illustration	Form and complexity of contextual information.	Physical context (undefined person, peers/teacher, media figure, family/home, school/class/lab/gym, community/neighborhood, state/province, identified country, unidentified/fictitious country, world/global), socio-historical context (events in domestic affairs, events in international affairs, traditions and customs), and human figure (complete human figure, partial human figure, one human figure, two or more human figures)
Use of metaphorical visual language	Visual components included with the intent to clarify and enhance understanding of objects and actions in an illustration.	Space, time, and motion; matter and energy (states of matter, temperature, light and electricity, and sound); human state (senses, speech and cognition, physical condition, emotion)
Use of text in the illustration	Form and complexity of textual information.	Text unit (number, letter, word/词, phrase/短语, sentence/句子, paragraph/段落) and text function (label, provide a code/legend, title/caption, elaborate, comment/note, and provide instructions)
Representation of variables, constants, and	Form, complexity, and level of concreteness or abstraction with which variables and constants are	Variables and constants (cases, stages, levels, line, value, scale, and symbol), functions (graph, table, nodes/ arcs, formula, and symbol), and discrete structure (sequential, tree, cycle, and network)

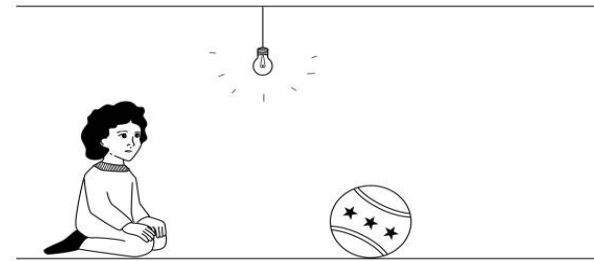
functions	represented.	
Interaction with the text of the item	Ways in which an illustration is used in combination with the text of the item to provide information to examinees or to capture their responses.	Location, reference to illustration (explicit, and not stated), and stated actions to perform with the illustration (observe or examine, draw or mark on illustration provided, generate an illustration, no action stated), constituents (subject, object, action, background), commonality (part of a stand-alone item, same illustration for several items)

APPENDIX C
Example illustrations and definitions for the IDVs for analyzing illustrations in science tests*

Category	Definition	Coding Rule	Example illustrated items
1.1. Image concreteness	Level of realism in representing object(s) or event(s).		
1.1.1 Photograph	Image captured by a camera.	Make sure to distinguish drawing or painting of people from photograph of people.	 法拉第 焦耳 安培 欧姆 Read the following newspaper article and answer the questions that follow. The History of Vaccination Mary Montagu was a beautiful woman. She survived an attack of smallpox in 1715 but she was left covered with scars. While living in Turkey in 1717, she observed a method called inoculation that was commonly used there. This treatment involved scratching a weak type of smallpox virus into the skin of healthy young people who then became sick, but in most cases only with a mild form of the disease. Mary Montagu was so convinced of the safety of these inoculations that she allowed her son and daughter to be inoculated. In 1796, Edward Jenner used inoculations of a related disease, cowpox, to produce antibodies against smallpox. Compared with the inoculation of smallpox, this treatment had less side effects and the treated person could not infect others. The treatment became known as vaccination.
1.1.2 Scanned document/ Text clip	Digital form of a written or printed material.		
1.1.3			

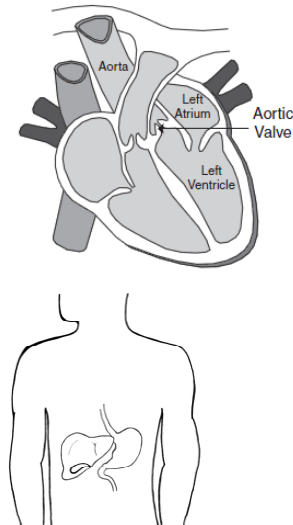
1.1.4 Realistic
line drawing

Image drawn with lines which represent object, event, and/or background the way they look in real life.



1.1.5
Schematic

Image that shows essential components of object, event, or the physical structure of something in a simplified manner.

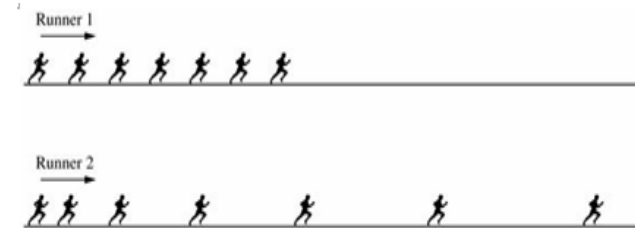


1.1.6 Map Representation of land or a place, as seen from above.



1.1.7 Silhouette Shadow image of object filled in with solid black.

Silhouette is always two-dimensional.



1.1.8 Cartoon Image that represents object and event in a humorous or stereotypical manner, typically exaggerating some feature of the object.

Smiley face is a particular case of cartoon.



1.1.9
Logo/Icon/Metonymy

Simplified or stylized image of an object that represents a concept, abstract idea, event, or class of objects.



1.1.10

1.1.11
Emblem

Image of an object with a cultural-historical meaning embedded as the representation of a concept, abstract idea, event, or class of objects.



1.1.12

1.1.13 Symbol

Image that conventionally represents a concept, abstract idea, event, or class of objects and whose appearance does not have any resemblance with the concept, idea, or object.



1.1.14
Reference

Image that represents the existence of object by showing some of its elements, or its elements in a simplified manner.



1.1.15 Entity

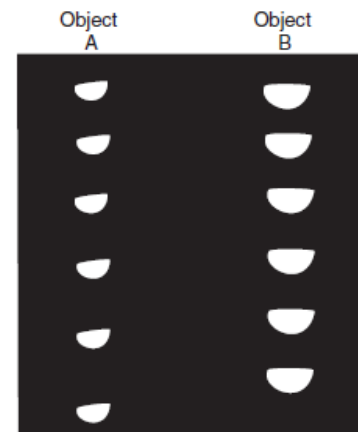
Image that represents the existence of object in an abstract manner.



1.1.16

1.1.17 X-ray, MRI, Sonogram, thermography, spectrometry, etc. Image of an object obtained through technology that is sensitive to physical properties or changes the human eye cannot see.

4



The picture shows two objects that were dropped and recorded with a stroboscopic camera. The *best* explanation for the results is that object A

- A has less air resistance.
- B was dropped from a greater height.
- C has a greater mass.
- D accelerated more slowly.

CSP00168

1.1.18 Image
in an object

Image being
illustrated as part of
the nature of an
object to help
interpret the related
object, but not
necessarily intended
for the viewer to
know what exactly it
is (e.g., an image of
an active computer
screen).



1.2. Background

Visual element
intended to provide a
context for the
proper interpretation
of the focal object or
event.

Map has no
background, is
two-dimensional,
and zoom-out.

1.2.1 With
background

Illustration that
depicts not only the
focal object or event,
but also the set of
visual elements
intended to provide a
visual perspective.

Background can be
removed without
altering the
essential
information
conveyed in the
illustration as a
whole.



If there is ceiling,
or a surface on

NAEP2009-8S10-11 –p

Questions 11-12 refer to the following information.

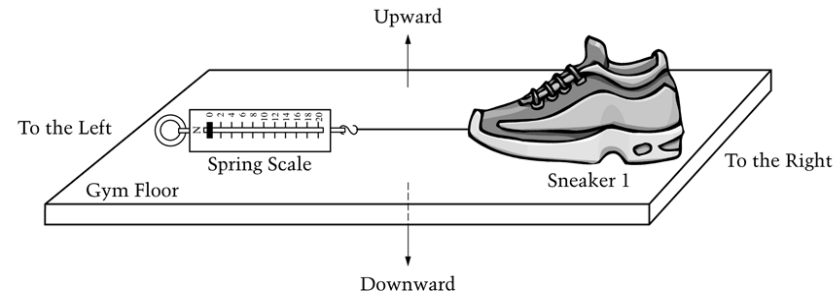
which objects are,
then there is
background.

Meg designs an experiment to see which of three types of sneakers provides the most friction.

She uses the equipment listed below.

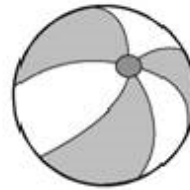
1. Sneaker 1
2. Sneaker 2
3. Sneaker 3
4. Spring scale

She uses the setup illustrated below and pulls the spring scale to the left



11. In what direction does the force of friction act?

1. To the left
2. To the right
3. Upward
4. Downward



1.2.2 Without
background

Illustration that
depicts only the
focal object or event.

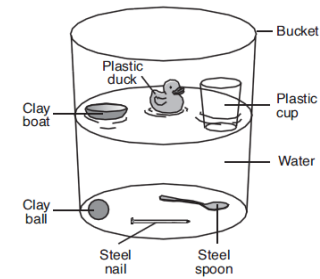
1.3. Zooming

Shot distance of
object or event.

Zooming is not
about
transparency, but
about closeness.

1.3.1 Zero
zooming
naked eye

Naked-eye view.

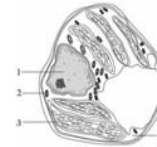


1.3.2 Zoom-in

Magnified, detailed
view of object as
seen by a
microscope.

Cell is zoom-in.

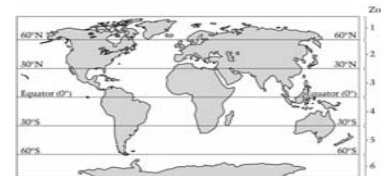
If you zoom in into
an abdomen, you
will see the cells of
the skin of the
abdomen, not what
is behind.



1.3.3 Zoom-
out

View of an object
from a distance in a
way that allows to
see it in its entirety
as seen by a
telescope.

Map is zoom-out.

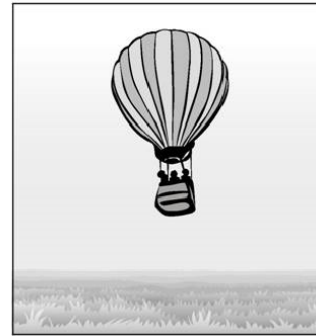


1.4. View

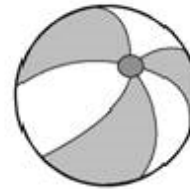
Position from which
an object is
observed.

Views are not
mutually
exclusive.

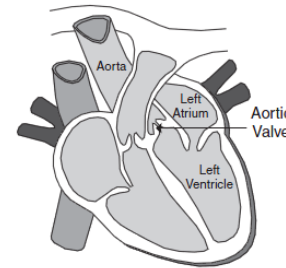
Make a holistic
decision.



1.4.1 External Object seen from
outside.



1.4.2 Internal Object seen from
inside.

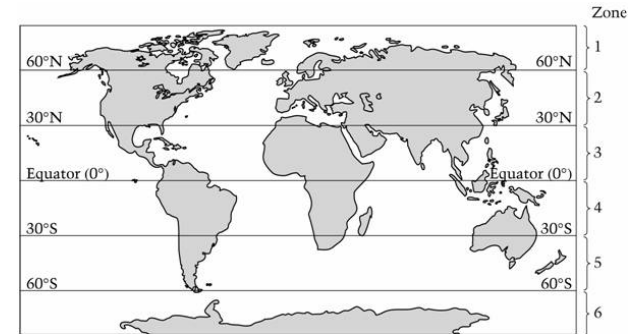


1.4.3 From above

Object seen from above.

Allow a range of angles.

Maps always coded as from above.



1.4.4 From below object

Object seen from below.

Allow a range of angles.

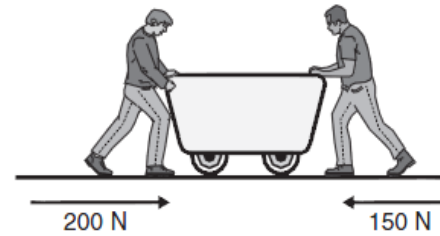
1.4.5 From the side of object

Object seen from the side.

1.4.5 must be completely from the eye level.

An object seen from the front (for example, the facade of a house), is coded as “from the side of object.”





1.5. Dimension

Measurement of length, width, and/or depth of object in an illustration.

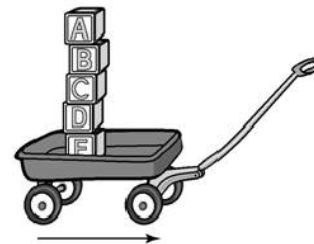
Map and table are two-dimensional.

Photos are three-dimensional.

If the dimension of certain object in one illustration cannot be determined, we regard its dimension as that of other objects in the same illustration.

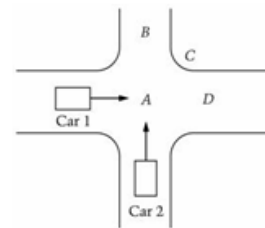
1.5.1 Three dimensional

Object shown in length, width, and depth.



1.5.2 Two dimensional

Object shown only in length and width.



1.7. Relative scale of objects or components

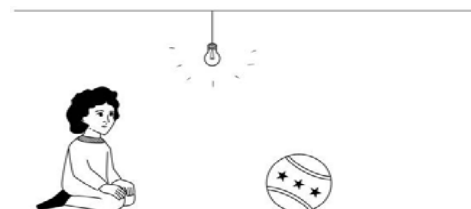
Degree of size proportionality between objects or components.

1.7 and 1.8 are mutually exclusive.

1.7.1 Proportionate

Objects shown at a consistent scale.

If there is doubt about the proportionality, code as proportionate.

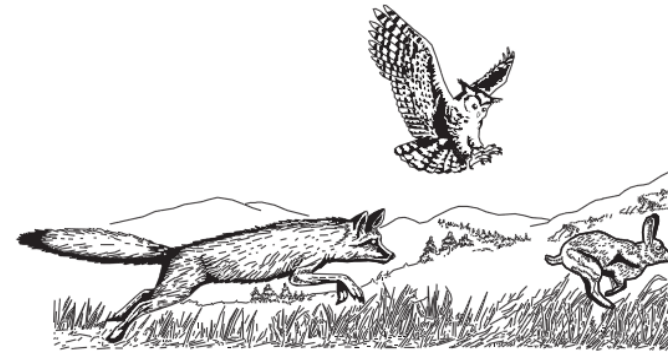


1.7.2
Disproportionate

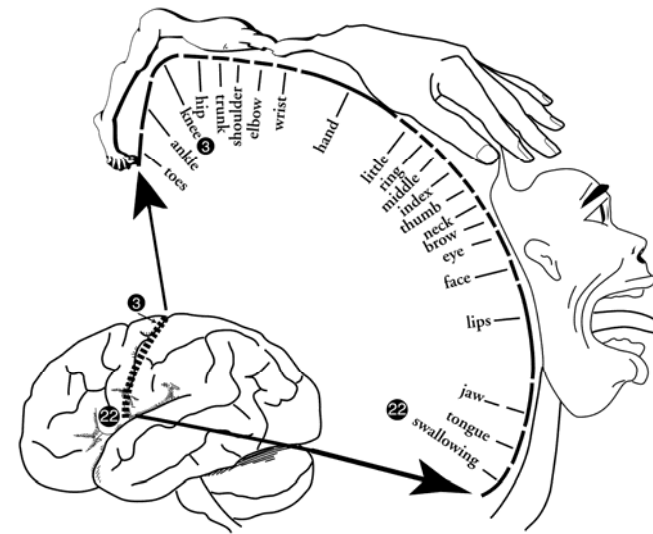
Objects not shown to scale.

If there is an element that is disproportionate in relation to other elements in an illustration, then just code the entire illustration as disproportionate.

If there is doubt about the proportionality, code as proportionate.



(Not drawn to scale)

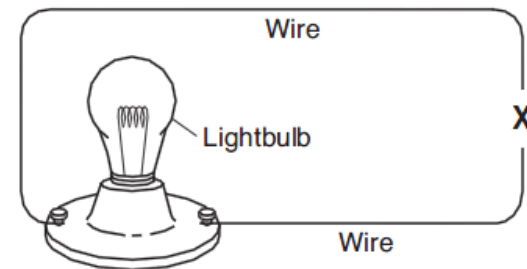


1.8. Color

Visual spectrum range.

1.8.1 Black and white

(Self-explaining).



1.8.2
Multicolor

(Self-explaining).



1.8.3 Gray
scale

(Self-explaining).

Gray scale refers
to the level of
inking of surface,
not visual effects.



Do not take into
account of color of
font.

**6.1. Physical
context**

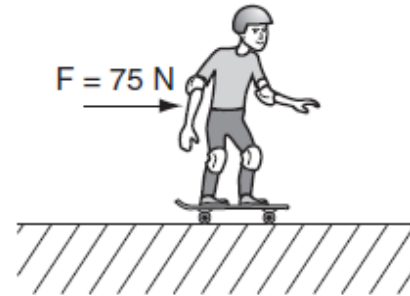
Presence of
individual and place
in a physical
context.

6.1.1
Undefined
person(s)

Individual referred
to as fictitious
person in the text of
the item.

If a body part of
the individual is
shown (e.g., a
hand), that counts
as an individual.

- 16** A 50-kg child on a skateboard experiences a 75-N force as shown.



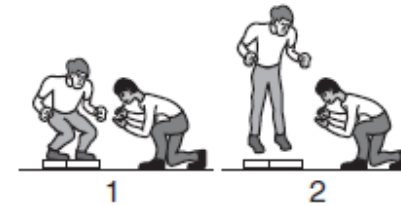
What is the expected acceleration of the child?

- A $0.67 \frac{\text{m}}{\text{s}^2}$
B $1.50 \frac{\text{m}}{\text{s}^2}$
C $6.70 \frac{\text{m}}{\text{s}^2}$
D $25.00 \frac{\text{m}}{\text{s}^2}$

CSP10450

6.1.2
Peers/teacher(s)
Individual referred
to as peers or teacher
in the text of the
item.

- 18** A student in a lab experiment jumps upward off a common bathroom scale as the lab partner records the scale reading.



What does the lab partner observe during the instant the student pushes off?

- A The scale reading will remain unchanged during the entire time the student is in contact with the scale.
- B The scale reading will increase momentarily then will decrease as the student is moving upward from the scale.
- C The scale reading will increase during the entire time the student is in contact with the scale.
- D The scale reading will decrease momentarily then will increase as the student is moving upward from the scale.

CSP10451

6.1.3 Media Well-known

figure individual or character in public media (e.g., a cartoon, an actor).

6.1.4 Family/home Reference, in the text of the item, to an individual in the family or a place at home.

6.1.5 School/class/lab/gym Reference, in the text of the item, to the school, class, laboratory, or gym. 6.1.5-6.1.10 only refer to places.

6.1.6 Community/neighborhood Reference, in the text of the item, to a place other than school in the community or neighborhood.

6.1.7 State/province Reference, in the text of the item, to a place in a state or province.

33. 热门话题你参与:

青藏铁路线上的美丽风景



A 一只藏羚羊在青藏铁路青水河特大桥附近觅食



B 一群藏羚羊在青藏铁路沿线的楚玛尔河一带迁徙

(Translation) 33. Please join the discussion of hot topics:

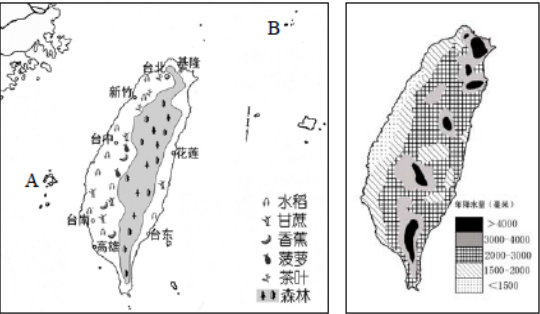
Beautiful scenery along the Qinghai-Tibet railway line

A A Tibetan antelope is looking for food near the gigantic bridge of Qingshui River along the Qinghai-Tibet railway

B A group of Tibetan antelope is migrating nearby Chumaer River along the Qinghai-Tibet railway

得分	评卷人

三、台湾是祖国的神圣领土，是富饶的宝岛。读图回答：（14 分）



台湾岛森林和主要农产品分布图

台湾岛年降水量分布图

(Translation) 3. Taiwan, a fertile island, is our homeland’s sacred territory. Read the illustration and answer questions: (14 points)

(Title for the illustration on the left)

Forest on Taiwan Island and distribution of agricultural products

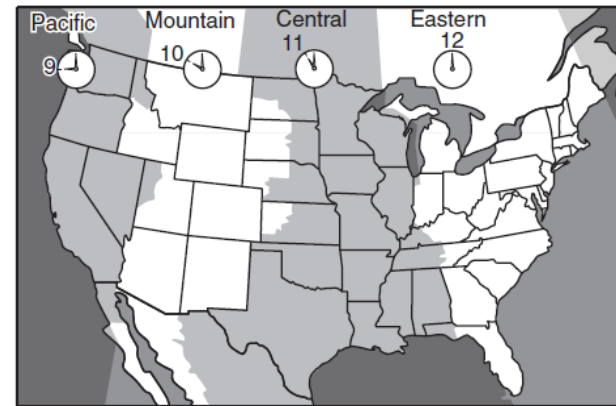
(Title for the illustration on the right)

Distribution of annual precipitation on Taiwan Island

6.1.8
Identified
country

A specific country in
the world.

45 The map below shows the four major time zones in the continental United States.



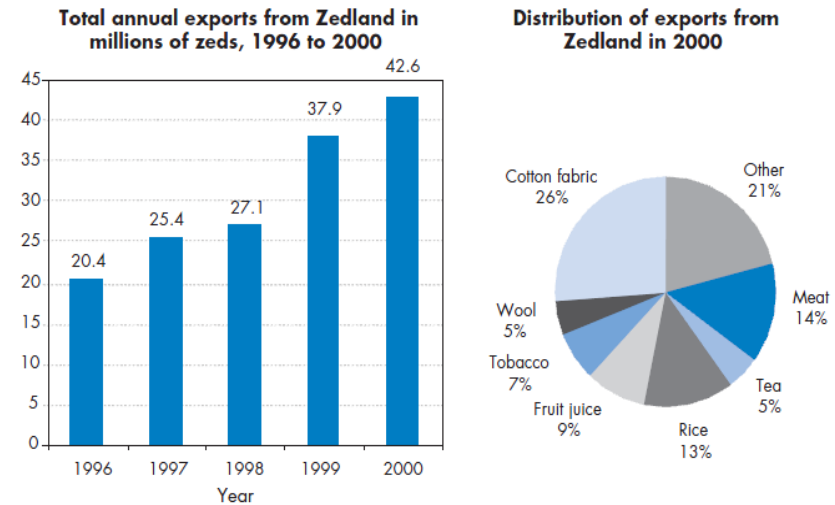
If it is 9 a.m. in the Eastern Time Zone, what time is it in the Pacific Time Zone?

- | | |
|------------|------------|
| (1) 3 a.m. | (3) 6 p.m. |
| (2) 6 a.m. | (4) 9 p.m. |

6.1.9 A fictitious country.
Unidentified/f
ictitious
country

MATHEMATICS UNIT 13: EXPORTS

The graphics below show information about exports from Zedland, a country that uses zeds as its currency.

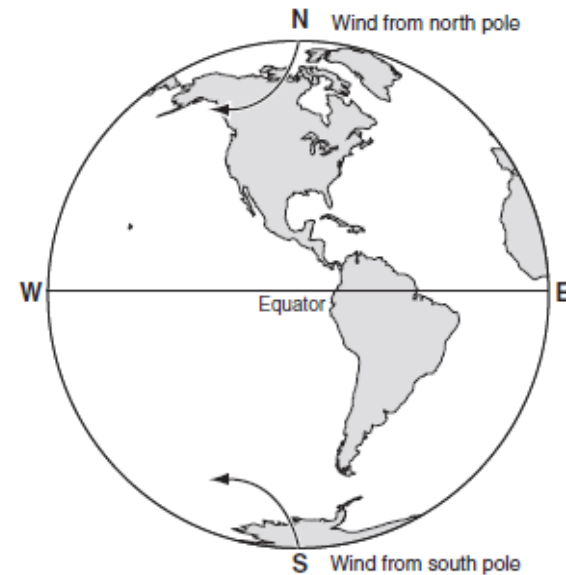


QUESTION 13.1

What was the total value (in millions of zeds) of exports from Zedland in 1998?

Answer:

6.1.10
World/global
Two or more
countries or a
geographic region in
the world.



What causes the wind deflection from the north and south poles?

- A the rotation of Earth on its axis
- B the oblate shape of Earth
- C the tilt of Earth's axis relative to its orbital plane
- D the difference in total land mass of the two hemispheres

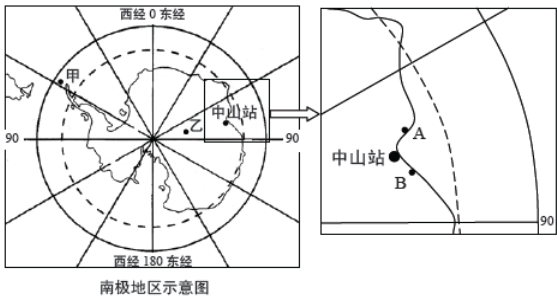
6.2. Socio- Presence of

historical context individual and place in relation to a social event.

6.2.1 Events in domestic affairs Context provided based on a place where an event is occurring currently.

得分	评卷人

七、我国第 27 次南极科学考察队于 2011 年 4 月 1 日乘“雪龙号”考察船载誉而归。读图回答：（12 分）



1. 受冰层等条件的限制，这次科考队选定了新的登陆点，它位于中山站的西北方向，即图中字母_____所表示的区域。
2. 此次科考队员还将一块镌刻有胡锦涛主席题写的“中国南极昆仑站”玉碑矗立在昆仑站，图中甲、乙两地中代表昆仑站的是_____。

Translation: 7. On April 1st, 2011, the 27th Antarctic scientific expedition of our country came back in a research vessel called “Snow Dragon.” Read the illustration and answer questions:

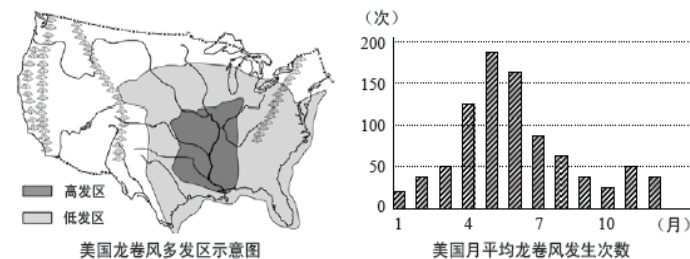
- 1.Limited by certain conditions such as ice conditions, this time the scientific expedition selected new landing point, which is located to the northwestern of Zhongshan Station, that is the area labelled as letter _____.
- 2.This time the expedition team members established a jade tablet with the words “China’s Antarctic Kunlun Station” written by Chairman Hu Jing-tao. In the left illustration, the Kunlun Station between location Jia and Yi is _____.

6.2.2 Events
in
international
affairs

Context provided
based on an event
taking place in a
non-fictitious
country that is not of
the examinees.

得分	评卷人

六、美国是“龙卷风之乡”。2011年4月至5月期间，美国遭受了几十年来最严重的龙卷风袭击。读图回答：（8分）



1. 美国龙卷风频发与其所处地理位置有一定关系。美国东临_____洋，南临_____湾，大量的水汽容易从东、西、南三面流入大陆中部，当水汽汇集成雷雨云并积聚到一定强度时，就会产生龙卷风。
2. 美国龙卷风高发地区以_____地形为主。
3. 美国平均龙卷风发生次数超过100次的月份有_____。

Translation: 6. The U.S. is “the home of tornadoes”. During the period of April to May in 2011, the U.S. suffered the worst tornado since couple of decades. Read the illustration and answer: (8 points)

Title for the left illustration: Schematic diagram of the U.S. tornado-prone areas

Title for the right illustration: Average number of U.S. tornadoes per month

1. The frequent occurrence of the U.S. tornadoes has some relations with its geographical locations. The east of the U.S. is next to _____ Ocean, and its south is next to _____ Ocean, so a large amount of water vapor easily flows into the center of the continent from the east, west, and south. When the water vapor forms thunder, rain, and cloud, and reaches certain intensity, tornadoes generate.

6.2.3
Traditions and
customs

Context provided
based on long-
established ways of
doing things in the
society or culture of
the examinees.

四、论述计算题(19题6分, 20题8分, 21题9分, 共23分。解答应写出必要的文字说明、步骤和公式, 只写出最后结果的不能给分)

19、在祖国60华诞, 除了阅兵这一重头戏外, 彩车游行也是一大亮点, 作为对祖国的献礼, 突出三峡符号的重庆彩车得到了一致好评(如图14所示), 游行时彩车重达29吨, 行进的速度为每分钟70米。



图15

(1) 彩车行进时受到的重力有多大? (g 取 10N/kg)

(2) 训练时彩车要经过500米的距离需要多长的时间? (结果保留整数)

Translation: 19. On the 60th anniversary of our motherland, besides military parade, parade floats was also a highlight. As a gift to our motherland, the parade float from Chongqing received unanimous praise (shown in Illustration 14). In the process of parade, that float weighted 29 tons with a speed of 70 meters per minute.

(1) How much gravity did the float have while in parade?

(2) How long should it take the float to pass 500 meters in practice?

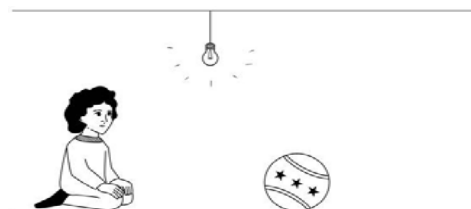
6.3 Human figure

Human form in
photo or drawing.

Including cartoon
figures.

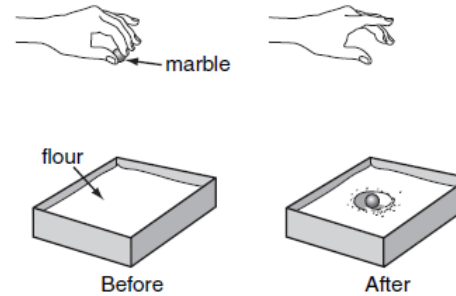
6.3.1
Complete
human figure

Person body shown
in its entirety.



6.3.2 Partial human figure Part of the body of a person.

5



A student models an impact crater on the Moon by dropping a marble from a known height onto a pan of smooth flour. Before reaching any conclusions about the results of this simple experiment, the student repeats the activity several times so that

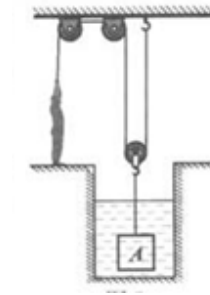
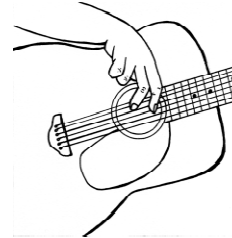
- A differences produced by standard variability in conditions become clear.
- B she can produce as large a crater as possible before measuring a diameter.
- C her ability to simulate a meteor impact becomes more realistic with practice.
- D she can illustrate a perfectly circular crater for her write-up of the experiment.

CSY10617

6.3.3 One human figure

Only one person depicted in the illustration.

Code body parts as human figure.



6.3.4 Two or more human figures

Two or more persons depicted in the illustration.



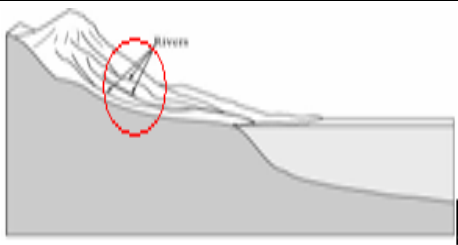
2.1. Space, time, and motion

Visual device intended to ensure clarity in the interpretation of an event or property of an object that is difficult to show in a static image.

2.1.1 Space and location

Visual device intended to

distinguish an object from others, or to indicate location or detail.



2.1.2 Time and sequence

Visual device intended to show related movements in a particular order in time.

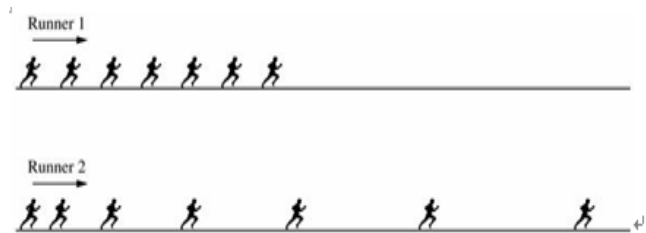


30 People pan for gold by scooping material from a streambed and shaking it in a pan as shown above. The gold nuggets in the pan can easily be separated from the sand and gravel because of their greater density. The material scooped from the streambed is an example of —

- F a solution
- G a mixture
- H an acid
- J an element

2.1.3 Dynamics and flow

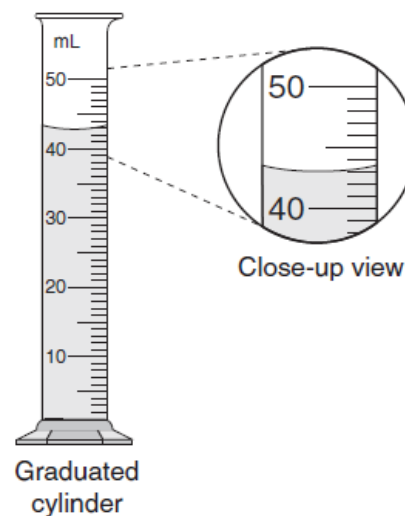
Visual device intended to show movement and action, or direction of movement and action.



The arrow shows direction, the multiple men show location of the same man at different points in time

2.1.4
Magnifying
glass/lens

Visual device intended to show an enlarged part of an object in detail as seen through a magnifying glass or a microscope lens (e.g. showing the frame of a magnifying glass or a microscope lens).



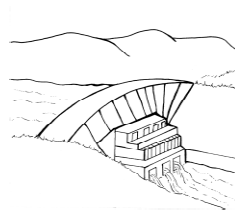
- 20 What is the volume of the liquid in the graduated cylinder? Record and bubble in your answer to the nearest milliliter on the answer document.

2.2. Matter and energy

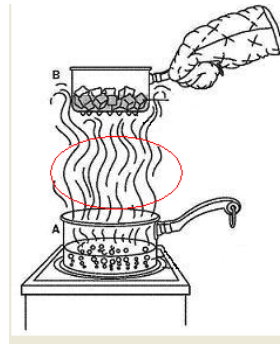
Visual device intended to clarify and enhance interpretation of states of matter and energy.

2.2.1 States of matter

Visual device intended to show liquid, gaseous, or solid state.

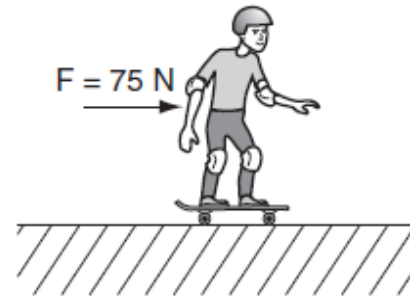


Water



Water and steam

-
- 16** A 50-kg child on a skateboard experiences a 75-N force as shown.

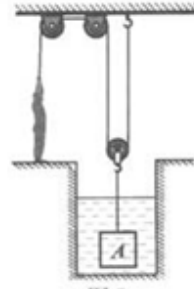


What is the expected acceleration of the child?

- A $0.67 \frac{\text{m}}{\text{s}^2}$
- B $1.50 \frac{\text{m}}{\text{s}^2}$
- C $6.70 \frac{\text{m}}{\text{s}^2}$
- D $25.00 \frac{\text{m}}{\text{s}^2}$

CSP10450

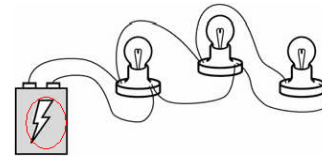
Solid ground



Ceiling and floor

2.2.2 Temperature
Visual device intended to show that something is hot or cold.

2.2.3 Light and electricity
Visual device intended to show the source or existence of illumination or electric current.

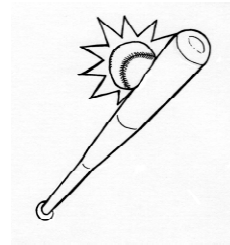


2.2.4 Sound
Visual device intended to show different kinds of sound, such as noise, snoring, and music.



2.2.5 Force
and impact

Visual device
intended to show the
intensity of force
and/or its impact.



2.2.6
Chemical
condition

Visual device
intended to show the
condition of
something caused by
a chemical reaction.



rust

2.2.7 Light
reflection

Visual device
intended to show
shading or the
reflection of light.



The parallel slanted lines show that it is a mirror.



2.3. Human state

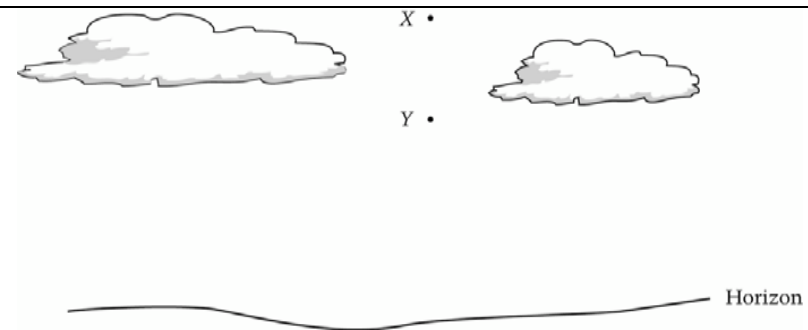
Visual device intended to show an internal condition that is difficult to show in a static image.

2.3.1 Senses

Visual device intended to show perception through any of the five senses: seeing, smelling, hearing, tasting, or touching.

2.3.2 Speech and cognition/thinking

Visual device intended to show people's talking or thinking.



Using speech bubbles, balloons or positioning the thoughts/ideas over the head of the person(s).



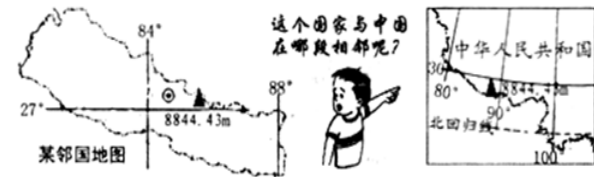
Translation:

Teacher asks: what does everybody know about solution?

Student answer: Solution is a mixture. Each component in a solution has the same density. Solution is transparent.

8. 回答下图中的问题
时,“某邻国地图”中用
不上的信息是()

- A. 经纬度位置
- B. 山峰的位置
- C. 边境线形状
- D. 首都的位置

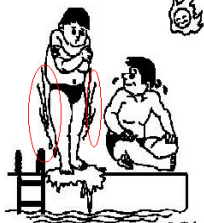


Translation: 8. When answering the question in the following illustration, the useless information in “the map of some neighbor country” is ()

- A. Position of latitude and longitude
- B. Position of mountain peak
- C. Shape of borderline
- D. Position of capital

Title of the left map: the map of some neighbor country

The question the boy in the illustration asks: where does this country border upon China?

2.3.3 Physical condition	Visual device intended to show the state of a person's body, such as freezing or warming up.	
2.3.4 Emotion	Visual device intended to show a strong feeling, such as being sad or angry.	
		
3.1. Text unit	Written textual component in an illustration.	
3.1.1 Non-scientific/mathematical sign	Mark, figure, or symbol conventionally used to represent something not relating to science and/or mathematics.	A punctuation signs such as “” is a Non-scientific/mathematical sign.
3.1.2 Scientific/	Mark, figure, or symbol	N (newton) is coded

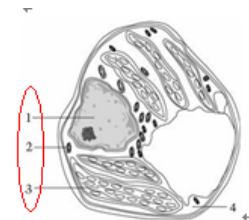
mathematical sign, and notation	conventionally used to represent something in science and/or mathematics.	simultaneously as scientific sign, abbreviation, and letter.
---------------------------------	---	--

		Parentheses () can be a mathematical sign in the context of formula.
3.1.3 Abbreviation	Shortened form of a word.	Do not confuse with acronym (see 3.1.11).

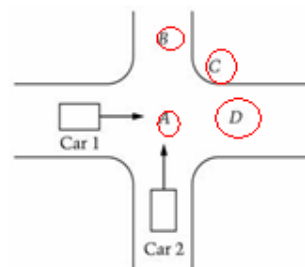
3.1.4 Roman numeral	(Self-explaining).
---------------------	--------------------

I, II, III, IV, V, VI, VII, VIII, IX,..., X, etc.

3.1.5 Arabic number	(Self-explaining).
---------------------	--------------------



3.1.6 Letter	(Self-explaining).
--------------	--------------------



3.1.7 Word/词 (Self-explaining).



3.1.8 Phrase/短语 Group of two or more words without a verb.



Translation: the text in the above illustration (the translation is from left to right, then from top to bottom)

Mental body at the end of the pen, plastic shell, resistance, spring, neon tube, mental tip.

3.1.9 Sentence/句子 Group of words containing a subject and a verb.



3.1.10 Paragraph/段落 Section of a written text consisting of several related sentences that convey meaning on the same topic.

3.1.11 Abbreviation of a set Do not confuse

UNESCO for the United Nations Educational, Scientific, and Cultural

Acronym/缩
略语

of words formed
with the initial
letters or characters
of those words.

with abbreviation
(see 3.1.3).

Organization.

首师大 for 首都师范大学 (Capital Normal University)

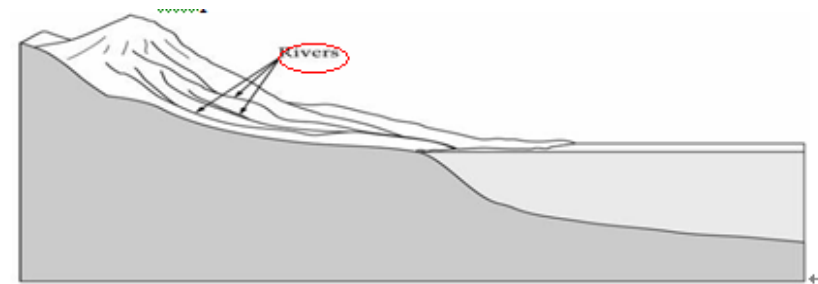
3.2. Text function

Special purpose of
the text used in an
illustration.

3.2.1 Label

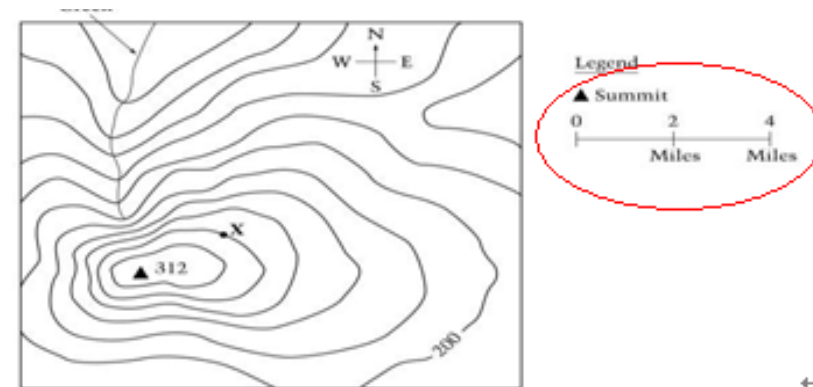
Text of an
illustration intended
to point at a
component in an
illustration.

Label focuses on
the small
components/parts
of the illustration,
while
title/caption/headin
g provides
information on the
whole meaning of
the illustration.



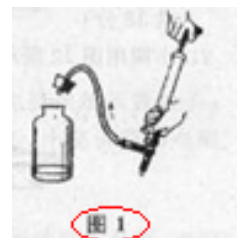
3.2.2 Legend/ code

Text of an
illustration intended
to explain symbol
used in an
illustration,
especially in a map.



3.2.3
Title/caption/
eading

Text of an
illustration intended
to provide a concise
explanation of the
illustration as a
whole.



Translation: Illustration 1

3.2.4Elaborate
/
explain/state/d
escribe

Text of an
illustration intended
to give a detailed
explanation of a
component in the
illustration.

3.2.5Opinions
/
comment/note

Text of an
illustration intended
to express an
opinion or comment.



3.2.6
Instructions

Text of an
illustration intended
to provide direction
and information on
what is expected to
do.

- 3.2.7 Data

Text of an illustration intended to provide data.
- 3.2.8 Text on object

Text that is part of an object (as in a ruler or a scale), included in the illustration with the intent to represent the object accurately but not necessarily with the intent to be read by the viewer.
- 3.3. Text emphasis

Special font style of text in an illustration to accentuate the importance of something.

3.3.1 Capitalization

(Self-explaining).

3.3.2 Bolding

(Self-explaining).

3.3.3 Italicizing

(Self-explaining).
- Tomato Plant Data

Type of Soil	?
W	13
X	6
Y	22
Z	9
- A black and white line drawing showing a hand holding a ruler against a tomato. The hand is positioned on the left, with the thumb and index finger gripping the ruler. The ruler is placed horizontally against the side of the tomato. The drawing is a simple sketch, focusing on the interaction between the hand, the ruler, and the tomato.
- 179

3.3.4 Underlying (Self-explaining).

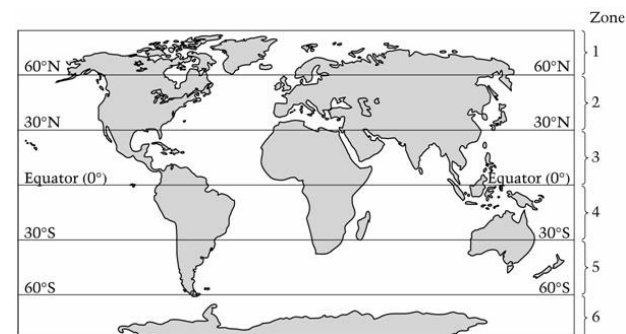
3.3.5 Circling (Self-explaining).

3.4. Text direction Angle in which the text in an illustration is shown.

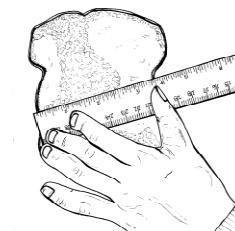
3.4.1 From left to right or vice versa Text in an illustration shown horizontally.



3.4.2 From top to bottom or vice versa Text in an illustration shown vertically.

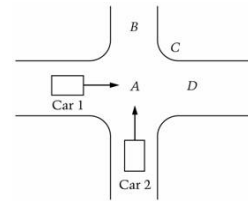


3.4.3 Oblique direction Text in an illustration shown slant.



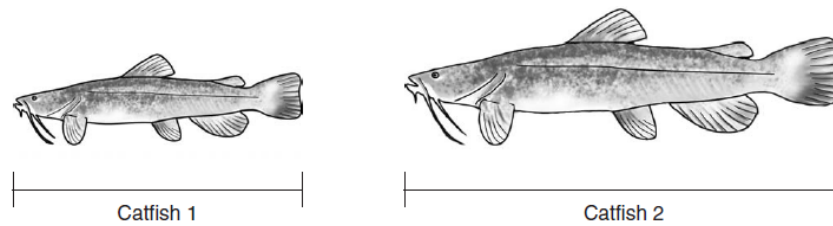
4.1. Variables and constants Level of concreteness or abstraction with which variables and constants are represented.

4.1.1 Cases/facts/events Concrete examples of particular objects, persons, or situations representing different categories or levels of a variable.



Two identical cars travel at 45 miles per hour toward the center of the intersection (point A, as shown above) with equal force. The cars collide at the intersection. If after they collide the cars stick to each other and move together, they will come to rest closest to

1. point A
 2. point B
 3. point C
 4. point D
-



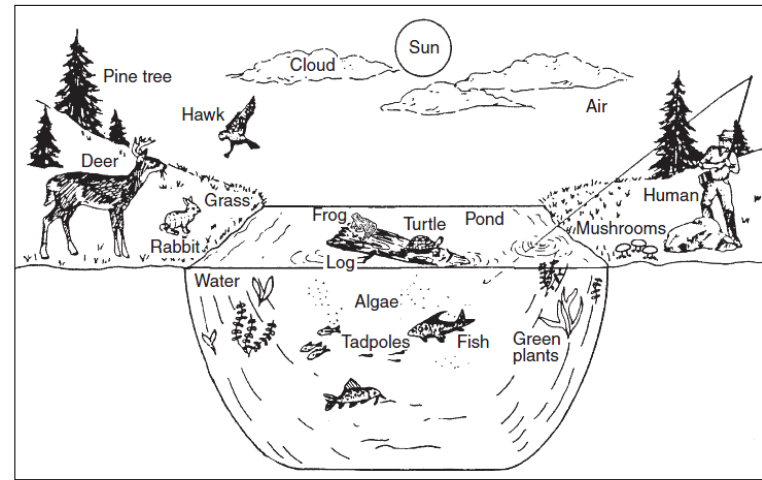
Scale
 $1\text{ cm} = 4\text{ cm}$

- 7 The scale drawings above show two catfish collected from a river. What is the difference in the actual body lengths of these catfish?
- A 3 cm
 - B 9 cm
 - C 12 cm
 - D 24 cm
-

4.1.2
Categories

Concrete representation of generic classes or groups of things in a system.

Base your answers to questions 58 and 59 on the diagram below and on your knowledge of science. The diagram represents an ecosystem.



(Not drawn to scale)

58 Identify *one* producer shown in the diagram. [1]

E.g.: Cats and dogs

Copper and lead

4.1.3
Conditions/treatment/properties/locations

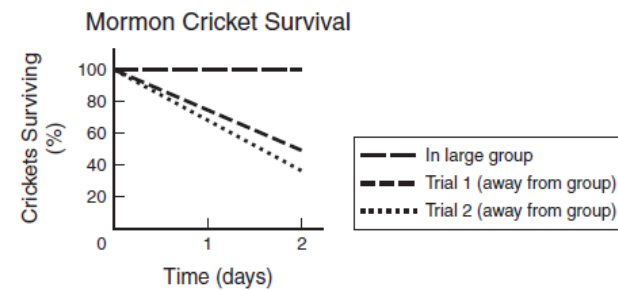
Concrete representation of objects or persons under different conditions that result from experimental manipulation.



4.1.4

4.1.5 Levels Attributes of objects or persons rank-ordered on a continuum of values of variables.

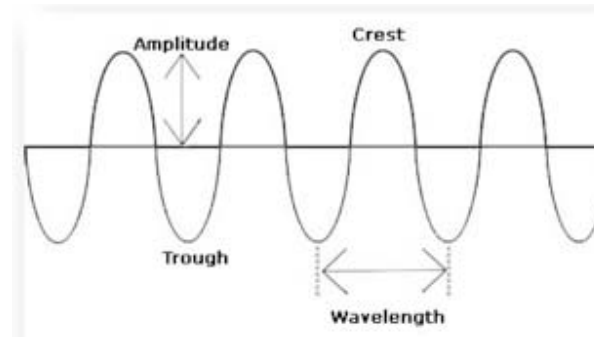
4.1.6 Analogic line Continuous straight or curved line in a Cartesian (x-y axes) coordinate system.



Source: Macmillan Publishers Ltd: *Nature*, Vol. 433

44 Flightless Mormon crickets often move in large groups. The graph shows the survival rate of crickets moving in large groups and of some crickets that were moved away from the groups. Which of these inferences about Mormon crickets is most likely accurate?

- F Mormon crickets within a group survive only two days.
- G Mormon crickets away from a group successfully reproduce.
- H Mormon crickets away from a group return to it for protection.
- J Mormon crickets within a group are less likely to be eaten by predators.



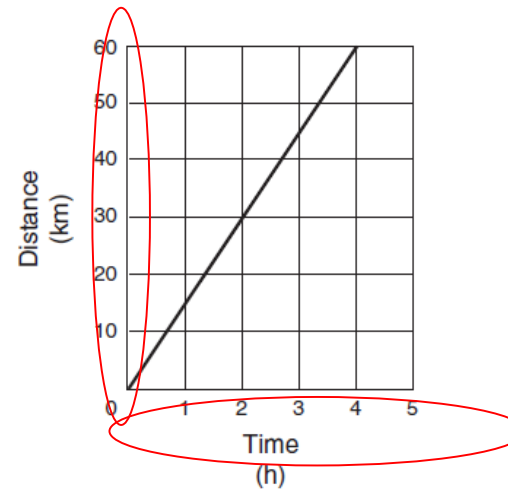
4.1.7 Value Numerical values of variables and/or constants.

Tomato Plant Data

Type of Soil	?
W	13
X	6
Y	22
Z	9

4.1.8 Scale

Values of a variable showing different magnitude values.



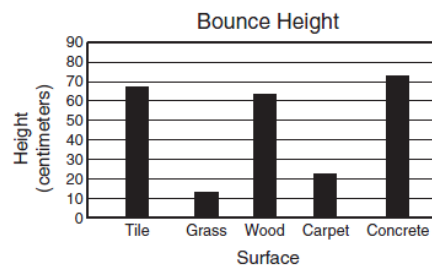
40 The graph shows the movement of a car over time. What is the car's average speed?

- F 10 kilometers per hour
- G 15 kilometers per hour
- H 30 kilometers per hour
- J 60 kilometers per hour

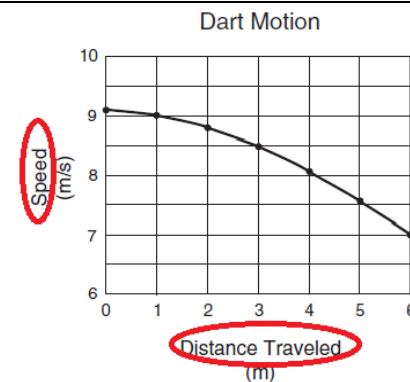
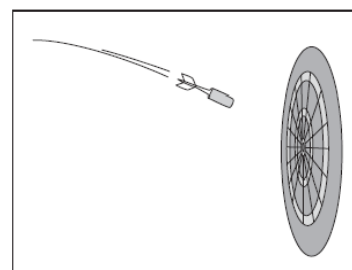
4.1.9 Unit

Quantity of a variable regarded as a standard.

E.g. meter is a unit of length.



4.1.10 Name Name of a variable.

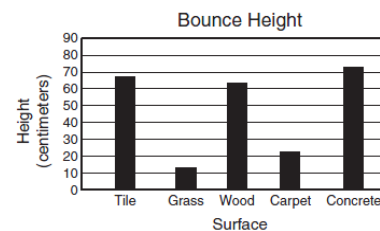


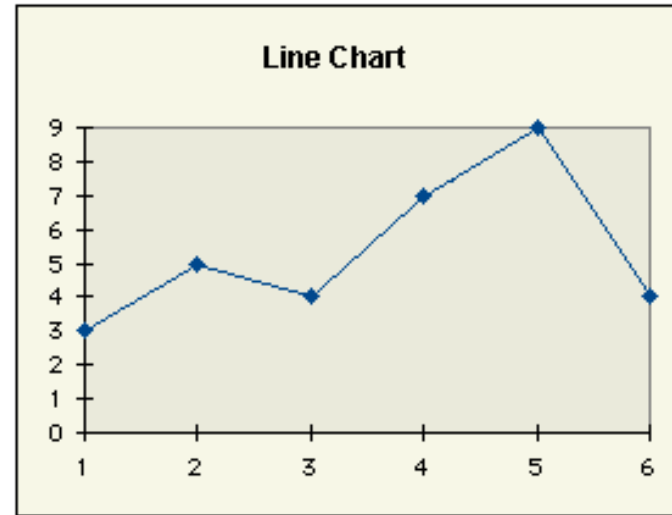
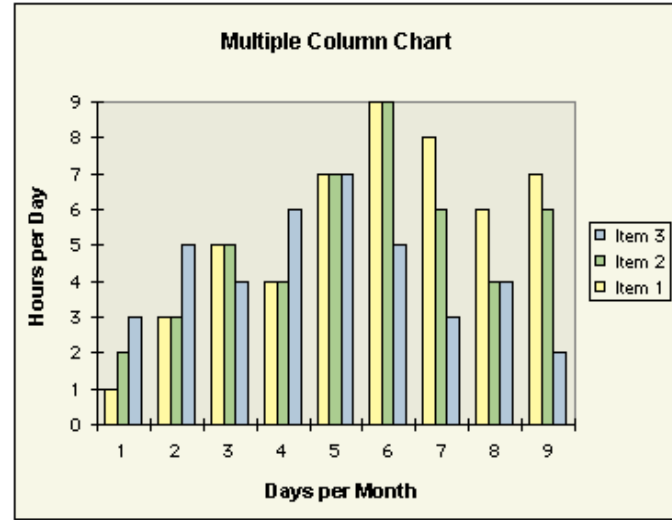
- 31 A foam dart is thrown toward a dartboard. The graph represents its motion. At about what speed is the dart traveling when it hits the dartboard, 7 meters from the starting point?

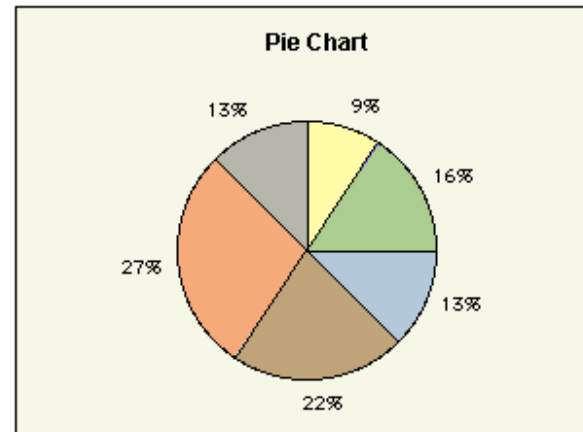
A 6.3 m/s
B 6.9 m/s
C 7.2 m/s
D 7.4 m/s

4.2. Functions Representation of the relationship between variables.

4.2.1 Discrete: Representation of the relationship between variables for a limited number of values.







4.2.2. A
Discrete:
Numeric

Numeric data in an
orthogonal
arrangement.

Tomato Plant Data

Type of Soil	?
W	13
X	6
Y	22
Z	9

4.2.2.B
Discrete:
Textual

Textual data in an
orthogonal
arrangement.

4.2.3 Discrete: Representation of the representation between discrete (1-0) variables as dots (or nodes) and lines (arrows or arcs).

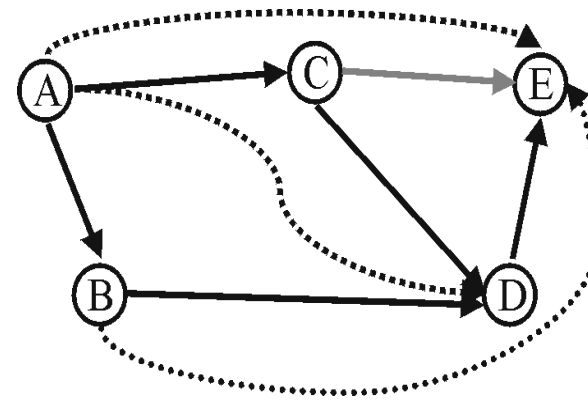
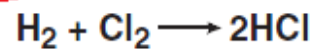


Fig 1. Graphical representation of a graph. Solid arcs like from node A to node B indicate direct connections. The arc between C and E can be circumvented by an indirect path from C to E via D. The dashed arcs indicate that there are indirect paths for which there are no direct paths in the graph.

4.2.4
Formula/equation
Representation of a variable as a formula or as the function between other variables according to a system of notation conventions.

66



Which of these describes the rate of this chemical reaction?

- A an increase in the concentration of HCl and H_2 with time
- B an increase in the concentration of HCl with time
- C an increase in H_2 and Cl_2 with time
- D a decrease in HCl and Cl_2 with time

CSC10369

MATHEMATICS UNIT 27: LICHEN

A result of global warming is that the ice of some glaciers is melting. Twelve years after the ice disappears, tiny plants, called lichen, start to grow on the rocks.

Each lichen grows approximately in the shape of a circle.

The relationship between the diameter of this circle and the age of the lichen can be approximated with the formula.

$$d = 7.0 \times \sqrt{t - 12} \quad \text{for } t \geq 12$$

where d represents the diameter of the lichen in millimetres, and t represents the number of years after the ice has disappeared.

QUESTION 27.1

Using the formula, calculate the diameter of the lichen, 16 years after the ice disappeared.

Show your calculation.

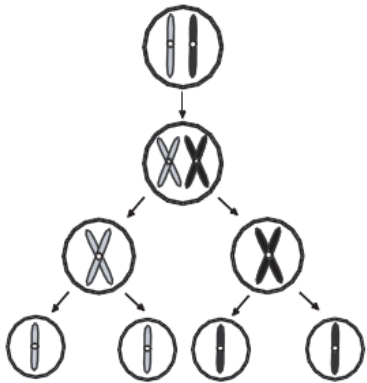
QUESTION 27.2

Ann measured the diameter of some lichen and found it was 35 millimetres.

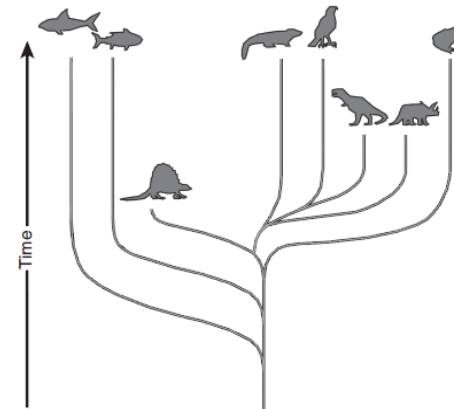
How many years ago did the ice disappear at this spot?

Show your calculation.

4.2.5 Scientific/mathematical symbol	Symbol used to represent a variable or constant commonly used in a scientific and/or mathematical discipline.	Chemical elements (e.g., Au, Hg) are coded as both 3.1.2 Scientific/mathematical sign, and notation, and 4.2.5 Scientific/mathematical symbol.	t =time in physics k =force in physics π =3.1416 g =gravity Δ =increase %=percentage Au=Gold
---	---	--	---

4.3. Structure	Representation of discrete, dichotomous variables.	
4.3.1 Sequential	Single path, linear order of a series of steps or stages in a process or procedure.	Only one path. No branching.  30 People pan for gold by scooping material from a streambed and shaking it in a pan as shown above. The gold nuggets in the pan can easily be separated from the sand and gravel because of their greater density. The material scooped from the streambed is an example of — F a solution G a mixture H an acid J an element
4.3.2 Tree	Branching (hierarchical, breaking down, genealogical) structure of relationship between components.	Only one path between any pair of components. Branching. 

Tree of Some Evolutionary Relationships

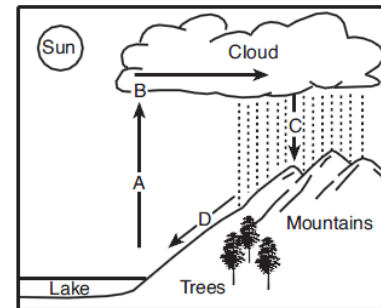


4.3.3 Cycle

Process or procedure in which stages or steps repeat regularly or iteratively in the same sequence.

Any given element is connected with itself. No branching. Only one path between any two given components.

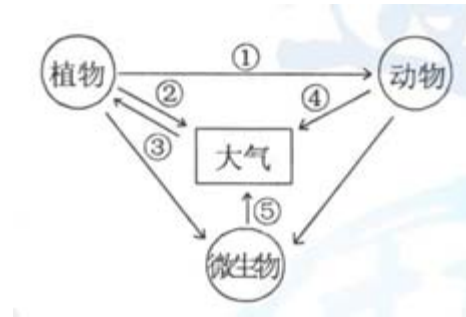
Water Cycle



4.3.4 Network Multi-path interaction among different components in a complex system.

Multiple paths between at least one pair of components.

Food webs are coded as network.



Translation: the text in the above illustration (the translation is from left to right, then from top to bottom)

Plant Animal Atmosphere Microorganism

4.3.5 Composition Components of a whole.



4.3.6 Inclusion Representation of a relationship of components in which some components are part of or include other components.

21. 下列有关细胞核中染色体、DNA 和基因三者之间的包含关系的表述中，正确的是

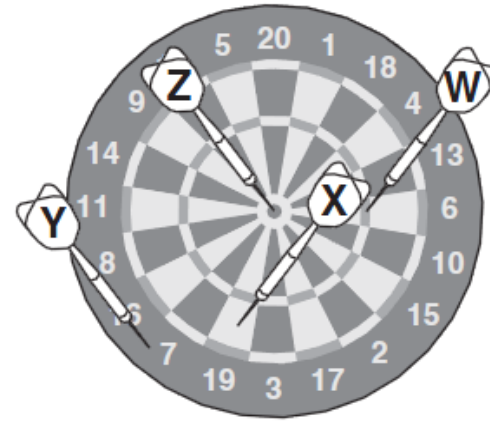


Translation: 21. Among the demonstration of the inclusive relationships between chromosome, DNA, and gene, the correct one is

A Gene DNA Chromosome B DNA Gene Chromosome
C Gene Chromosome DNA DDNA Chromosome Gene

5.1. Location Position of an illustration relevant to the text of the test item.

5.1.1 Above stem/prompt Illustration located above stem (in a multiple-choice item) or prompt (in a constructed-response item).



- 5 If the dartboard above is used to model an atom, which dart indicates where the protons and neutrons are located?
- A Dart W
 - B Dart X
 - C Dart Y
 - D Dart Z

5.1.2 Between stem/prompt and options/response format

Illustration located between the stem/prompt and the options/ response format.

Stem includes item question.

19. 蔬菜中的维生素，对青少年的生长发育非常重要。下图是某同学由于不合理的饮食所导致的结果，其主要原因是



Translation: 19. Vitamins in vegetables are very important to the healthy growth of youngsters. The following illustration shows the results due to an inadequate diet. The major reason that causes this is

Translation for the words in the illustration:

I don't like to eat these things. Why do my gums bleed?

A. Picky eating habits, lack of vitamin C B. Picky eating habits, lack of vitamin D

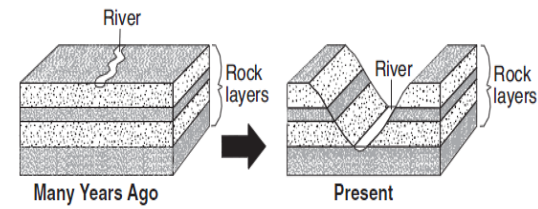
C. Picky eating habits, lack of vitamin A D. Picky eating habits, lack of vitamin B₁

5.1.3

Embedded in
stem/prompt

Illustration
embedded in a stem
or prompt.

5 The diagram below shows an area of land that changed after many years.



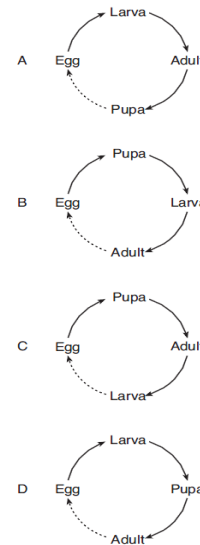
Which process changed the shape of the rock layers over time?

- A condensation
- B evaporation
- C erosion
- D magnetism

5.1.4
Embedded in
options/respon
se format

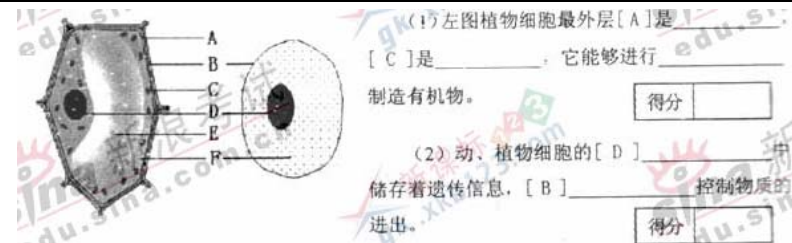
Illustration
embedded in options
or response format.

28 Which diagram shows the order of the stages in the butterfly's life cycle?



5.1.5 On the
left of the text

Illustration located
on the left of the text
of item.



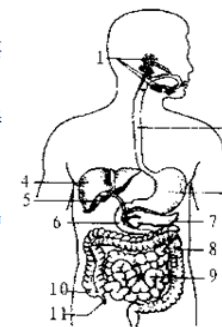
Translation: (1) The outmost part of the plant cell in the left illustration [A] is _____. [C] is _____, which is able to _____ to produce organic matter.

(2) [D] in animal and plant cell is _____, which stores genetic information. B is _____, which controls the entrance and exit of substances.

5.1.6 On the right of the text

Illustration located on the right of the text of item.

1. 右图为人体的某个系统的结构示意图，请据图分析：
 - (1) 该结构图表示人体的_____系统，此系统的主要功能是_____。
 - (2) 图中“3”表示的器官是_____；“4”表示的器官是_____。
 - (3) 除了右图所示的系统能参与维持内环境的物质稳定以外，还有其他一些系统也参与维持内环境的物质稳定。请写出其中的两种系统：_____系统、_____系统。
2. 右图为缩手反射的反射弧模式图，请分析并回答：



Translation: 1. The right illustration is a schematic representation of a system in human body, please analyze based on the illustration:

- (1) This illustration shows _____system of human body. The major function of this system is _____.
- (2) 3 in the illustration refers to the organ of _____; 4 refers to the organ of _____.
- (3) Besides the system shown in the illustration, which is able to maintain the stableness of substance within the body environment, there are other systems which also contribute to maintain the stableness of substance within the body environment. Please write the names of two such systems: _____ system, _____ system.

5.1.7

5.1.8

5.1.9 Under the text

Illustration located under the text of item.

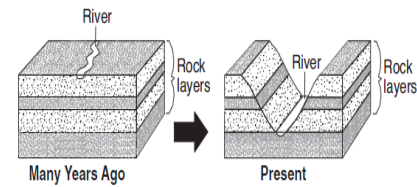
5.2. Reference

Mention of the

to illustration existence of an illustration in the text of the test item.

5.2.1 Explicit The text of the item mentions the illustration explicitly.

5 The diagram below shows an area of land that changed after many years.



Which process changed the shape of the rock layers over time?

- A condensation
 - B evaporation
 - C erosion
 - D magnetism
-

5.2.2 Not
stated

The text of the item
does not mention the
illustration.

Vanessa is watching her friends play soccer. She sees Breanna and Julie kick the soccer ball at the same time in opposite directions.



Then Vanessa sees the ball rolling to the right. Which of these best describes why the ball started rolling to the right?

- F** The girls exerted the same amount of force on the ball.
- G** The force due to gravity caused the ball to roll to the right.
- H** One girl exerted more force on the ball than the other girl.
- J** An upward force by the ground caused the ball to roll to the right.

Content: G8 Science
Item Number: 22
Item ID: 01172894
Item Type: Multiple Choice

Answer Key: H
Max. score points: 1
Standard: 8.7.3
Reporting Category: Forces/Density and Buoyancy

5.3. Stated
actions to
perform with
the illustration

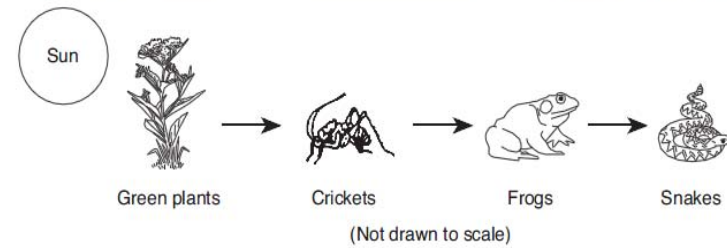
Explicit guidance for
examinees about
actions relating to
the use of an
illustration in the test
item.

5.3.1 Observe
or examine

Text of the test item
asks examinees to
look at or examine
the illustration.

Code illustration
5.3.1 if the item
contains key word
such as “refer to”
and “according
to”.

Base your answers to questions 41 and 42 on the diagram of a food chain below.



41 Identify a predator in this food chain. [1]

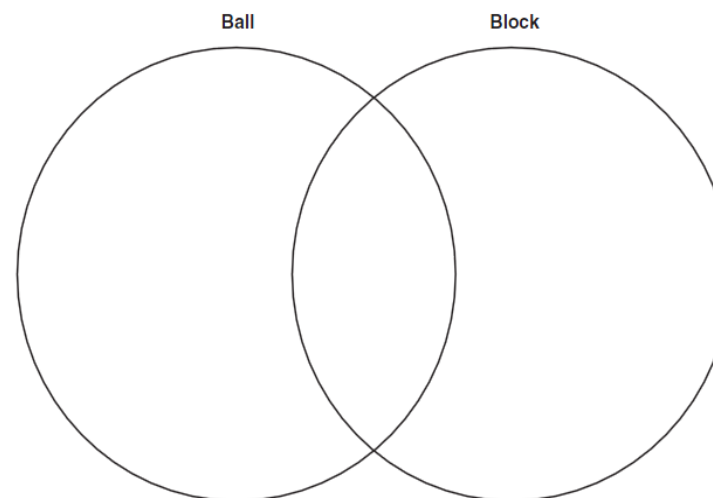
5.3.2 Write, draw or mark on illustration provided

Text of the test item asks examinees to write, draw, or mark on illustration provided.

31 Some properties of a ball and a block are listed below.

Ball	Block
rubber	red
red	300 g
sphere	rough
300 g	cube
smooth	wood

Complete the Venn diagram below to compare and contrast the ball and block. Use all of the properties listed above. [1]



5.3.3 Generate an illustration

Text of the test item asks examinees to draw an original illustration.

5.3.4 No actions stated

Text of the test item requests no actions from examinees to

perform with the illustration.

5.4. Commonality Attribution of illustration relating to the independence of the test item.

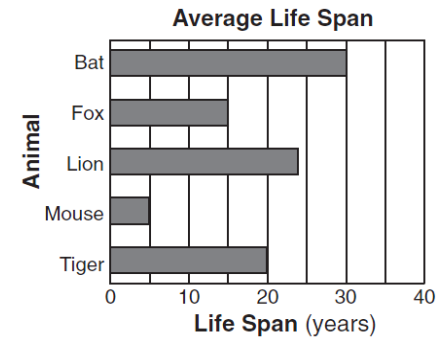
5.4.1 Part of a stand-alone item Illustration depicted specifically for one independent test item.



- 7 The picture shows a type of gazelle that lives in Africa. It eats leaves, twigs, flowers, and fruits. Which characteristic most helps this gazelle get food?
- A Long neck
 - B Tan fur
 - C Curled horns
 - D Large ears

5.4.2 Same illustration for several items Illustration depicted for a bundle of related test items.

Base your answers to questions 31 and 32 on the bar graph below and on your knowledge of science. The bar graph shows the average life span of five different animals.



31 Which animal has the *shortest* life span? [1] _____

32 How much longer is the life span of a bat than the life span of a fox? [1]
_____ years

5.4.3

Illustration for one item within a series of related items

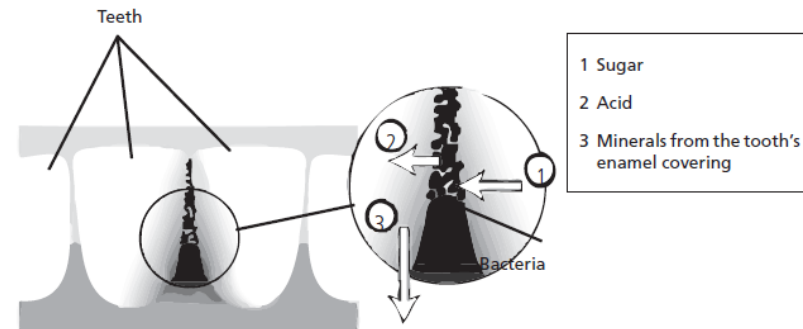
Illustration depicted specifically for one item which belongs to a bundle of related test items.

SCIENCE UNIT 20: TOOTH DECAY

Bacteria that live in our mouths cause dental caries (tooth decay). Caries have been a problem since the 1700s when sugar became available from the expanding sugar cane industry.

Today, we know a lot about caries. For example:

- Bacteria that cause caries feed on sugar.
- The sugar is transformed to acid.
- Acid damages the surface of teeth.
- Brushing teeth helps to prevent caries.



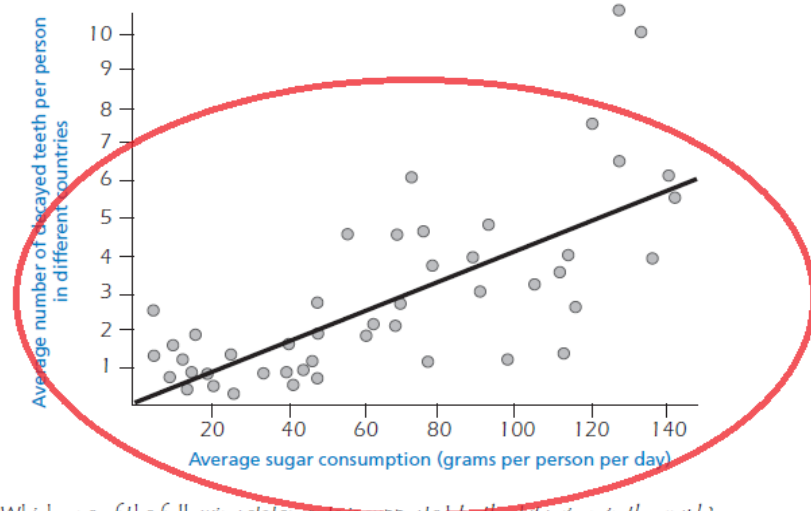
QUESTION 20.1

What is the role of bacteria in dental caries?

- E. Bacteria produce enamel.
- F. Bacteria produce sugar.
- G. Bacteria produce minerals.
- H. Bacteria produce acid.

QUESTION 20.2

The following graph shows the consumption of sugar and the amount of caries in different countries. Each country is represented by a dot in the graph.



Which one of the following statements is supported by the data given in the graph?

- A. In some countries, people brush their teeth more frequently than in other countries.
- B. The more sugar people eat, the more likely they are to get caries.
- C. In recent years, the rate of caries has increased in many countries.
- D. In recent years, the consumption of sugar has increased in many countries.

* Illustrations from

Arizona Department of Education. (2009). *Arizona's instrument to measure standards AIMS Science, 2009 sample test for Grade 8.*

Retrieved October 28, 2009, from <http://www.ade.state.az.us/standards/aims/sampletests/Gr8-AIMSSampletestscience.pdf>

California Department of Education. (2009). California Standards Test, 2006-2008 released test questions for grade 8 science.

Retrieved September 10, 2010, from <http://www.cde.ca.gov/ta/tg/sr/css05rtq.asp>

California Department of Education. (2009). California Standards Test, 2006-2008 released test questions for grade 10 science.

Retrieved September 10, 2010, from <http://www.cde.ca.gov/ta/tg/sr/css05rtq.asp>

China's Middle School Exit Examination. (2008). Released test questions for middle school exit examinations in physics, life science,

and earth science. Retrieved March 18, 2011, from <http://shiti.edu.sina.com.cn/shiti/index.php>

China's Middle School Exit Examination. (2009). Released test questions for middle school exit examinations in physics, life science,

and earth science. Retrieved March 18, 2011, from <http://shiti.edu.sina.com.cn/shiti/index.php>

China's Middle School Exit Examination. (2010). Released test questions for middle school exit examinations in physics, life science,

and earth science. Retrieved March 18, 2011, from <http://shiti.edu.sina.com.cn/shiti/index.php>

DC office of the State Superintendent of Education. (2009). *District of Columbia Comprehensive Assessment System, 2009 released*

items fifth grade science. Retrieved March 18, 2011, from

<http://www.nclb.osse.dc.gov/2010items/Science%20Grade%205%20ALL%20Released%20Items.htm>

Georgia Department of Education. (2004). Georgia End-Of-Course Tests Physical Science, released test booklet. Retrieved September 10, 2010, from http://www.doe.k12.ga.us/ci_testing.aspx?PageReq=CI_TESTING_EOCT&SubPageReq=GUIDES

National Assessment of Educational Progress. (2005). *NAEP 2005 Science Assessment Public Released Items for Grade 8*. Retrieved September 18, 2009, from <http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=science>

National Assessment of Educational Progress. (2009). *NAEP 2009 Science Assessment Public Released Items for Grade 8*. Retrieved March 18, 2011, from <http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=science>

Texas Education Agency. (2004). *Texas Assessment of Knowledge and Skills, 2004 released science test for Grade 5*. Retrieved September 10, 2010, from <http://www.tea.state.tx.us/student.assessment/released-tests/>

Texas Education Agency. (2009). *Texas Assessment of Knowledge and Skills, 2009 released science test for Grade 8*. Retrieved September 10, 2010, from <http://www.tea.state.tx.us/student.assessment/released-tests/>

The University of the State of New York. (2009). *Grade 4 elementary-level science test*. Retrieved March 18, 2011, <http://www.nysedregents.org/Grade4/Science/home.html>

The University of the State of New York. (2010). *Grade 4 elementary-level science test*. Retrieved March 18, 2011, <http://www.nysedregents.org/Grade4/Science/home.html>

The University of the State of New York. (2010). *Grade 8 intermediate-level science test*. Retrieved March 18, 2011, <http://www.nysedregents.org/Grade4/Science/home.html>

Trends in International Mathematics and Science Study. (2007). *TIMSS 2007 Science Assessment Public Released Items for Grade 8.*

Retrieved March 18, 2011, from <http://timss.bc.edu/TIMSS2007/items.html>

<http://bestclipartblog.com/clipart-pics/men-clipart-1.png>

<http://images.clipartof.com/small/22891-Clipart-Illustration-Of-A-White-Man-Wearing-A-Tie-Shaking-Hands-With-Another-Upon-Agreement-Of-A-Business-Deal.jpg>

<http://love.catchsmile.com/wp-content/uploads/2011/06/Angry-10.jpg>

http://pica.nipic.com/2007-11-02/2007112115038369_2.jpg

<http://rigsamarole.files.wordpress.com/2009/08/snoring.jpg?w=500>

http://www.clker.com/cliparts/b/1/7/9/11949849671589982655male_symbol_dan_gerhards_01.svg.med.png

<http://www.ebi.ac.uk/microarray/Research/networks/DisruptionNetwork/graph.gif>

http://www.intropsych.com/ch02_human_nervous_system/02homunc.jpg

http://www.sapdesignguild.org/goodies/diagram_guidelines/DIAGRAMS/LiniendgmTH.jpg

http://www.sapdesignguild.org/goodies/diagram_guidelines/DIAGRAMS/MFSaeulDgn.GIF

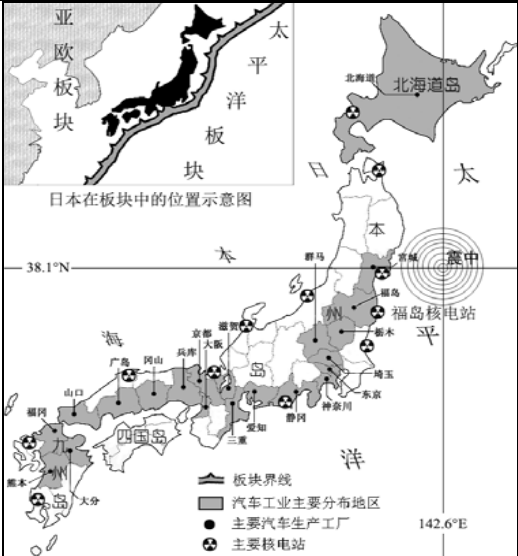
http://www.sapdesignguild.org/goodies/diagram_guidelines/DIAGRAMS/TortendgmTH.jpg

http://www.un.org/Depts/dhl/maplib/images_maplib/unflag.gif

APPENDIX D

Nine illustrated items identified with the most frequent outstanding IDVs (Translation of the text of Chinese items in italics)

Chinese items in original format	English translation of Chinese items
Item 1 BJ-ES-2011-6 6. 2011 年 3 月 11 日, 日本东北宫城县以东海域发生里氏 9 级特大地震并引发海啸, 福岛核电站受到严重影响并发生了泄漏。读图 8、图 9 和表 2, 回答下列问题。 (1) 3·11 特大地震发生的海域是_____洋, 震中的地理坐标为_____、_____。 (2) 日本境内火山、地震频发, 主要是位于_____板块和_____板块交界处所致。 (3) 福岛核电站位于日本的_____。(选择填空) A. 本州岛 B. 九州岛 C. 四国岛 D. 北海道岛 (4) 此次强震的危害巨大, 其主要原因是_____。(选择填空) A. 震级高、震源深、破坏力大 B. 未能在震前发布预报和警报 C. 主震后余震频繁且等级较高 D. 强震及诱发灾害的共同影响	Item 1 BJ-ES-2011-6 6. <i>March 11, 2011, the Richter 9 devastating earthquake and tsunami occurred in the east sea of the Miyagi Prefecture, northeastern Japan. The Fukushima nuclear power plant has been seriously affected, and begun to leak. Read Figure 8, 9, and Table 2, and answer the following questions.</i> (1) <i>March 11 devastating earthquake occurred in _____ ocean, the geographic coordinates of the epicenter are _____ and _____.</i> (2) <i>In Japan volcanoes, earthquakes occur frequently, mainly because it is located between _____ plate and _____ plate at the junction.</i> (3) <i>The Fukushima nuclear power plant is located in Japan's _____. (Select to fill in the blank).</i> A. <i>Honshu Island</i> B. <i>Island of Kyushu</i> C. <i>Shikoku Island</i> D. <i>Hokkaido Island</i> (4) <i>The damage caused by this earthquake is tremendous, mainly because _____. (Select to fill in the blank)</i>



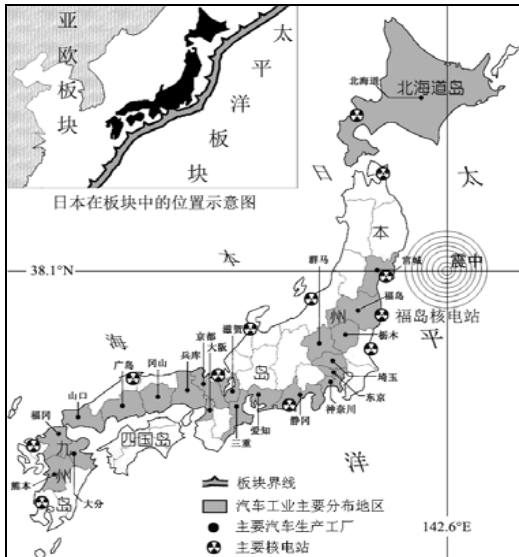
(5) 地震发生时，下列做法不正确的是_____。（选择填空）

- A. 躲避在高楼、大烟囱、立交桥之下 B. 迅速切断电源和煤气、天然气
- C. 远离易燃、易爆、有毒物品储藏区 D. 躲避在室内支撑结构较多的空间部位

(6) 从图中可以看出，日本的汽车工业分布不均，主要集中分布在_____。

表 2 日本主要产品产量占世界总产量的比重

- A. High magnitude, focal depth, the damage is great
- B. Forecasts and warnings failed to be released before the earthquake
- C. Frequent and higher grade aftershocks after the main shock
- D. The combined effect of the earthquake and induced disasters



(5) When an earthquake occurs, the following is incorrect _____.(Select to fill in the blank)

- A. Hide in tall buildings, large chimney, under the overpass
- B. Quickly cut off the power, gas, and natural gas
- C. Away from storage area which contains flammable, explosive, or toxic substances

主要 产品	轿 车	摩托车	船 舶	车 床	合成 纤维	录像机	照相机
产量 比重	43. 1%	78. 1%	72. 6%	57. 2%	48. 7%	77. 3%	87. 0%

(7) 根据表 2 推测, 受此次超强地震的影响, 短期内国际市场最有可能出现的波动现象是_____。(选择填空)

- A. 含碘食盐、蔬菜价格上涨 B. 录像机、照相机等配件的价格上涨
- C. 船舶、铁矿石的价格下跌 D. 汽车轮胎的原料——橡胶价格下跌

震后, 日本的一些漫画家通过绘制漫画鼓励民众渡过难关。如《七龙珠》和《阿拉蕾》的作者鸟山明, 让书中人物“悟空”和“阿拉蕾”一起为灾民加油打气, 鼓励人民坚强。

D. Hide in the interior support structure which has more space

(6) From the figure, the uneven distribution of the Japanese automobile industry mainly concentrated in _____.

Table 2 Japan's major products accounted for a proportion of world production

Main product	Car	Motor cycle	Ship	Lathe	Synthetic fibers	VCR	Camera
Proportion	43. 1 %	78. 1 %	72. 6 %	57. 2 %	48. 7%	77. 3 %	87. 0%

(7) According to Table 2, affected by this devastating earthquake, the most possible fluctuations as in the phenomenon of _____ will appear in the international market in the short term (Select to fill in the blank)

- A. Increase in the prices of iodized salt and vegetables*
- B. Increase in the prices of VCRs, cameras and other accessories*
- C. Decrease in the prices of ships, and iron ore*
- D. Decrease in the prices of raw materials - rubber for car tires*

After the earthquake, some Japanese cartoonist encouraged people to ride out the storm by drawing cartoons. For example, Akira Toriyama, the author of "Dragon Ball" and "Ala Lei", let the characters "Sun Wukong," and "Ala Lei" cheer together for the victims, and encourage people to be strong.



(8) 从鸟山明的漫画作品可以看出，日本的文化_____。（选择填空）

- A. 放弃了自身传统内容
- B. 不再拥有传承与创新
- C. 不受西方文化的影响
- D. 深受中国文化的影响



(8) From Akira Toriyama's manga, we can see that Japanese culture _____. (Select to fill in the blank)

- A. abandon their own traditional content
- B. no longer has the heritage and innovation
- C. free from the influence of western culture
- D. influenced by Chinese culture

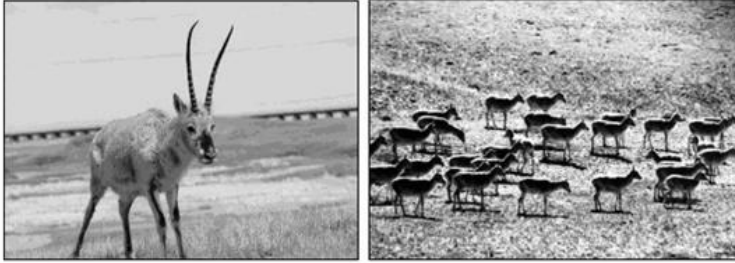
Item 2 BJ-LS-2007-33

33. 热门话题你参与:

Item 2 BJ-LS-2007-33

33. Please join the discussion of hot topics:

青藏铁路线上的美丽风景

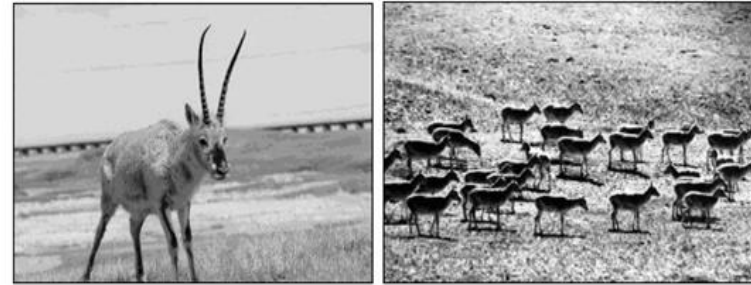


A 一只藏羚羊在青藏铁路清水河特大桥附近觅食 B 一群藏羚羊在青藏铁路沿线的楚玛尔河一带迁徙

你知道藏羚羊吗？藏羚羊是青藏高原特有的保护动物(如图-A)，它体表被毛，浑身是宝，具有胎生、哺乳的特点。它身手敏捷，是田径好手，北京 2008 奥运吉祥物五个可爱的福娃中，“迎迎”的原型就是一只机敏灵活、驰骋如飞的藏羚羊。

自 2001 年青藏铁路开工建设以来，铁路设计、施工的相关单位高度重视可可西里生态环境的保护工作，为保护区藏羚羊的迁徙(如图-B)专门设置了动物通道，并在施工过程中采取了严禁惊扰藏羚羊等管理措施。经过两年的适应

Beautiful scenery along the
Qinghai-Tibet railway line
青藏铁路线上的美丽风景



A 一只藏羚羊在青藏铁路清水河特大桥附近觅食 B 一群藏羚羊在青藏铁路沿线的楚玛尔河一带迁徙

A A Tibetan antelope is
looking for food near the
gigantic bridge of Qingshui
River along the Qinghai-Tibet
railway

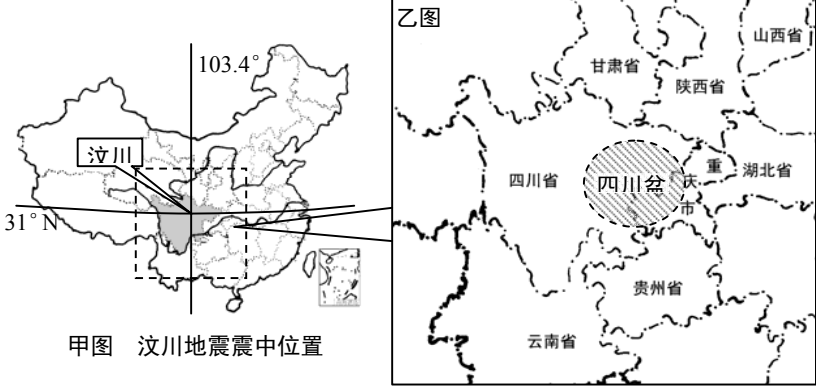
B A group of Tibetan
antelopes is migrating nearby
Chumaer River along the
Qinghai-Tibet railway

Do you know Tibetan antelope? Tibetan antelope is protected animal unique to the Qinghai-Tibet Plateau (see Figure A). It is covered with wool, viviparous and lactational. It is agile like skillful athletic players. Among the five mascots for Beijing 2008 Olympic Games, the prototype of “Yingying” is such an agile Tibetan antelope.

Since the Qinghai-Tibet Railway started construction in 2001, the relevant departments on railway design and construction attach great importance to Hoh Xil eco-environmental protection. They set up a special animal channel to protect the migration of Tibetan

期，可可西里藏羚羊已完全适应了青藏铁路。从 2006 年 5 月中旬开始，已有 300 多只藏羚羊安全、顺利地穿越了铁路。据介绍，青藏铁路正式通车后，游客们将可以在列车内或铁路沿线设立的观景站台上，观赏到藏羚羊、野牦牛、藏野驴等高原珍稀物种。 (1) 通过分析以上资料，你认为藏羚羊在分类上属于脊椎动物中的_____类动物。 (2) 羚羊的迁徙是一种动物行为，动物的行为可分为先天性行为和_____行为。 (3) 藏铁路的建设者们为保护区藏羚羊的迁徙采取了哪些措施？（至少回答出一点） (4) 藏铁路的建设者们保护藏羚羊等珍稀生物的做法给你的启示是：	<i>antelopes in the protected areas. In the process of construction, they also adopted management procedures which prohibited any actions disturbing the Tibetan antelopes. After two years of the adaptation period, Hoh Xil Tibetan antelope has been fully adapted to the existence of Qinghai-Tibet Railway. Beginning in mid-May 2006, more than 300 Tibetan antelope safely and smoothly have come through the railway. According to reports, after the official opening of the Qinghai-Tibet Railway, the visitors will be able to watch plateau rare species of Tibetan antelope, wild yak, Tibetan wild ass in train or viewing platform along the railway.</i> <i>(1) Based on the analysis of the above material, you think Tibetan antelope belongs to _____animal in the classification of animals as the vertebrate.</i> <i>(2) The migration of Tibetan antelope is an animal behavior. Animal behavior can be divided into congenital behavior and _____behavior.</i> <i>(3) What are the measures taken by the builders of the Qinghai-Tibet railway to protect the Tibetan antelope migration of the District? (Answer at least one point)</i> <i>(4) What you learn from the practice of the builders of the Qinghai-Tibet railway to protect the Tibetan antelope and other rare creatures is:</i>
Item 3 SH-ES-2008-2 二、2008 年 5 月 12 日 14 时 28 分，四川汶川县发生 8.0 级地震，除了造成四川省严重受灾，甘肃、陕西、重庆、云南、山西、贵州、湖北等七个省市也不同程度受灾。读图	Item 3 SH-ES-2008-2 <i>In 2008, at 14:28 on May 12, an 8.0 earthquake occurred in Wenchuan County in Sichuan. In addition to causing a serious disaster in Sichuan Province, seven provinces and municipalities</i>

回答：（10分）

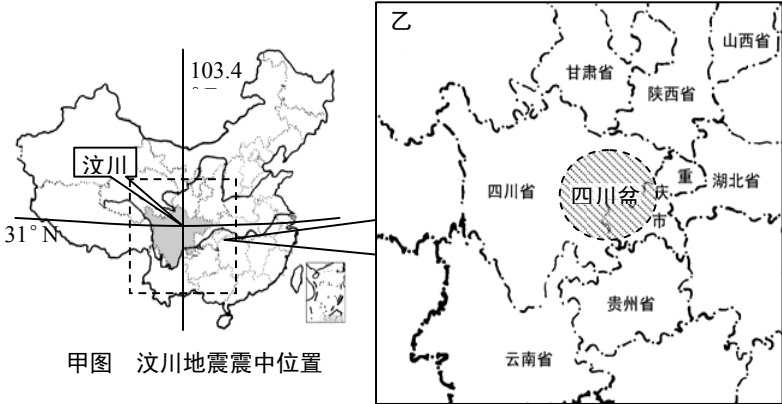


1. 汶川地震震中的纬度是_____。
2. 甲乙两图中，比例尺较大，内容较详细的是图。
3. 汶川县位于_____盆地西北边缘，地处我国第_____级阶梯和第_____级阶梯交界处，由于地形复杂，给救灾工作带来很大困难。
4. 在上述受灾的省区中，没有与四川省相邻的是山西省和_____省。

Item 4 SH-ES-2008-4

四、2008年5月6日至10日，国家主席胡锦涛对日本进行了为期五天的友好访问。读图文资料回答：（10分）

including Gansu, Shaanxi, Chongqing, Yunnan, Shanxi, Guizhou, Hubei were also affected to varying degrees. Read maps and answer questions: (10 points)



1. The latitude of earthquake epicenter is_____.
2. Between these two maps, the one with larger scale and more detailed information is _____.
3. Wenchuan County is located in the northwestern edge of the _____ basin, at the junction of the _____ ladder and the second level of the ladder. The complex terrain caused tremendous difficulties to the relief work.
4. Of the above affected provinces, Shanxi Province and _____ Province are not adjacent to the Sichuan Province.

Item 4 SH-ES-2008-4

May 6 to 10, 2008, President Hu Jintao was in Japan for a five-day goodwill visit. Read the following figure and table to answer

日期	行程	
5月6日	抵达东京羽田机场	
5月7日	赴东京首相官邸与日本首相会谈	
5月8日	赴东京早稻田大学演讲	
5月9日	参观川崎环保城 参观横滨中华街	
5月10日	访问古都奈良 由大阪关西机场登机回国	

- 日本是东亚岛国，此次访日行程都位于____岛上，该岛东临__洋。
 - 日本海运发达，此次访日城市中大阪和____都是世界著名大港。
 - 日本工业高度集中，主要分布在____沿岸地带，原因是该地带……（ ）
- A. 资源丰富 B. 运输便利 C. 人口稀少 D. 气候宜人

questions: (10 points)

Date	Itinerary
May 6	Arrive at Tokyo's Haneda airport
May 7	Meet with Japanese Prime Minister at his residence in Tokyo
May 8	Give a speech at Waseda University in Tokyo
May 9	Visit the Kawasaki green city Visit Yokohama
May 10	Visit the ancient capital Nara Return from Osaka's Kansai Airport

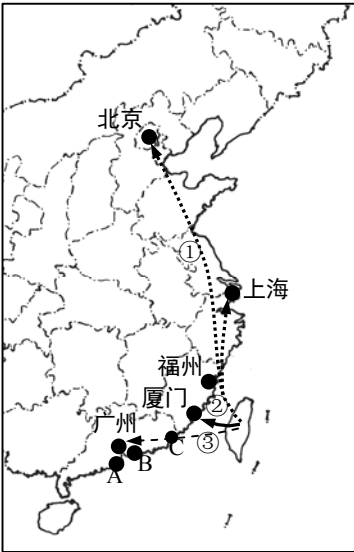


- Japan is island nation in East Asia. This visit to Japan are located on _____ island, this island is in the east of _____ Ocean.
 - Japanese ocean shipping has been well developed. Among the Japanese cities visited in this trip, Osaka and _____ are world famous seaports.
 - Japanese industry is highly concentrated, mainly in the _____ coastal zone, because this zone..... ()
- A. Resource-rich B. Convenient transport
C. Sparsely populated D. Pleasant climate

Item 5 SH-ES-2008-5

五、据报道，为实现台湾岛和大陆之间的“三通”（通邮、通航、通商），两岸商讨今年7月实现周末包机直航，不再绕行香港。读图回答：
(10分)

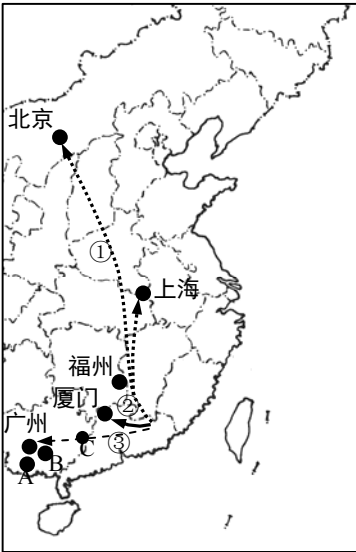
1. 右图标有字母的城市中，表示香港的是_____。
2. 飞机从台湾起飞，走航线①抵达北京，此航线飞行方向大致是自_____向_____。
3. 走航线②抵达厦门，该城市所在省级行政区的简称是_____。
4. 走航线③进入_____三角洲，该地区农民充分利用当地自然条件，创造了独具特色的“生态农业”——_____农业。



Item 5 SH-ES-2008-5

According to reports, in order to realize the Three Linkages (direct postal, transportation, and trade) between the island of Taiwan and the mainland, the two sides negotiated that direct flights (without connecting in Hong Kong) in weekends be permitted in July this year. Read the map and give answers: (10 points)

1. Among the cities labeled with letters in the right map, the letter refers to Hong Kong is _____.
2. Flights taking off from Taiwan, which follow skyway ①, arrive at Beijing. The direction of the skyway is approximately _____.
3. Flights taking off from Taiwan, which follow skyway ②, arrive at Xiamen. The shortened name for the province, where Xiamen is located, is _____.
4. Flights taking off from Taiwan, which follow skyway ③, enter _____ Delta, where peasants efficiently make good use of local natural conditions, and create “Eco-friendly Farming” — _____ Farming.



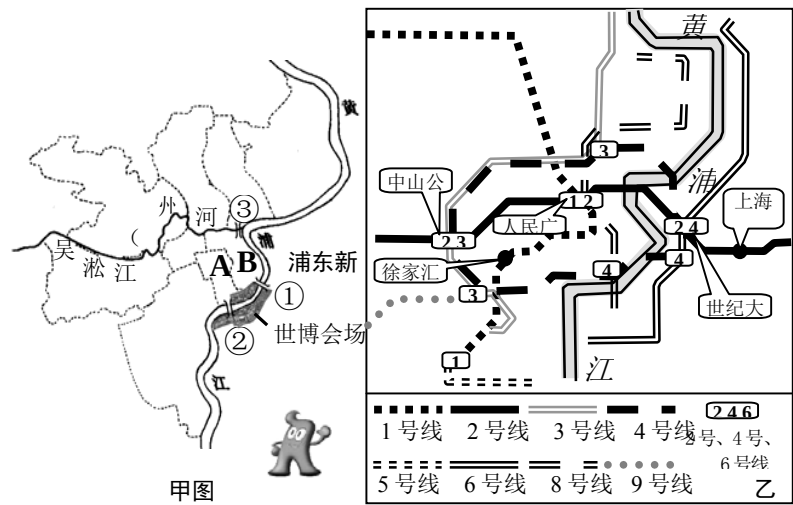
Item 6 SH-ES-2008-10

十、2010 年即将召开的上海世博会推动了上海的城市交通

Item 6 SH-ES-2008-10

The upcoming 2010 Shanghai World Expo has promoted the

建设。下列甲乙两图分别是“上海中心城区图”和“上海轨道交通运营示意图”。读图回答：（10分）

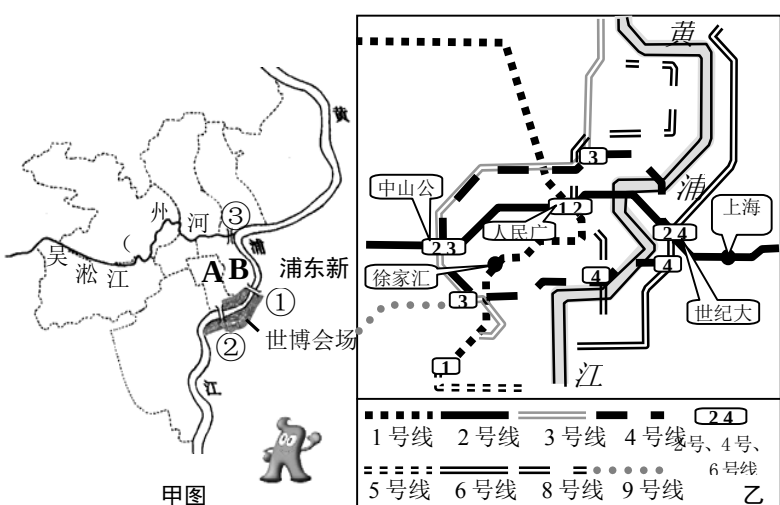


1. 甲图中字母 A 和 B 分别表示中心城区中的卢湾区和.....（ ）

- A. 黄浦区 B. 杨浦区 C. 长宁区

2. 上海世博会场地主要位于南浦大桥和卢浦大桥之间，沿着 A、B 两区和浦东新区的黄浦江两岸布局，图中卢浦大桥的数码是_____。

construction of Shanghai's urban transportation. The following Map Jia and Yi are respectively "Shanghai downtown map "and" Shanghai rail transit operations schematic diagram". Read the maps and answer questions: (10 points)



1. In Map Jia, the letters A and B respectively represent the Luwan District and () in the central urban area

- A. Huangpu District B. Yangpu District
C. Changning District D. Xuhui District

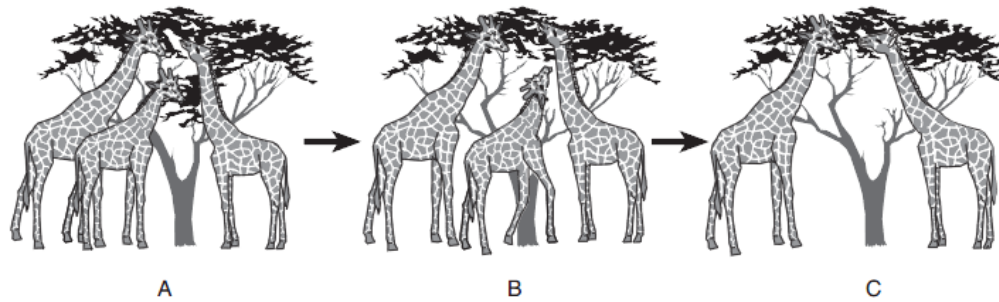
2. The site of Shanghai World Expo is located between Nanpu Bridge and Lupu Bridge. Along the A, B two districts and Pudong New Area, the number for the Lupu Bridge is _____.

3. Number ③ in Map Jia represents an already hundred old

3. 图中数码③是已有“百岁”高龄目前正在整修的外白渡桥，该桥跨_____两岸。 4. 目前上海已有八条轨道交通建成营运，其中完全位于黄浦江以东的是……（ ） A. 1 号线 B. 2 号线 C. 6 号线 5. 各轨道交通线换乘便利，其中 2、3、4 号线均可换乘的站点是……（ ） A. 徐家汇站 B. 中山公园站 C. 人民广场站 6. 轨道交通为上海市民出行带来极大的便利。写出从徐家汇至上海科技馆乘坐轨道交通最便捷的线路和换乘站名称。 徐家汇 → _____ 号线 → _____ 站 ^{换乘} → _____ 号线 → 上海科技馆 7. 甲图右下角以汉字“人”为核心创意的世博会吉祥物叫_____，他的头发像翻卷的海浪，点明了该吉祥物出生地濒临_____海的区域特征。	<i>White Bridge currently being refurbished. This bridge comes cross the two sides of _____.</i> <i>4. At present more than eight-track lines are able to operate, among which _____ lie entirely east of the Huangpu River. ()</i> A. line 1 B. line 2 C. line 3 D. line 9 <i>5. The transfer at the rail transits is convenient. Either Line 2,3, or 4 can transfer to..... ()</i> A. Xujiahui Station B. Zhongshan Park Station C. People's Square station D. Century Avenue Station <i>6. Rail transport brings great convenience for the people of Shanghai. Write the most convenient rail transit line and the name of the transfer station from Xujiahui to the Shanghai Science and Technology Museum.</i> Xujiahui → Line _____ → _____ Station ^{Transfer} → Line _____ → Shanghai Science and Technology Museum <i>7. At the lower right corner of Map Jia, there is the Expo mascot _____, created based on the Chinese character for "person". His hair is like the ocean waves, which highlights the regional features of _____ Ocean, the birthplace of the mascot.</i>
---	---

Item 7 NY-LS-2009-36

36 The diagram below shows a population of adult giraffes over time. Letters *A*, *B*, and *C* represent three time periods.



Bernstein et al (1998), Concepts and Challenges in Life Science
Teachers Edition (3rd), Globe Fearon, Inc., p. 389 (adapted)

Which process does this diagram best represent?

- (1) ecological succession
- (2) genetic engineering
- (3) natural selection
- (4) asexual reproduction

Item 8 TX-ES-2006-28

Before



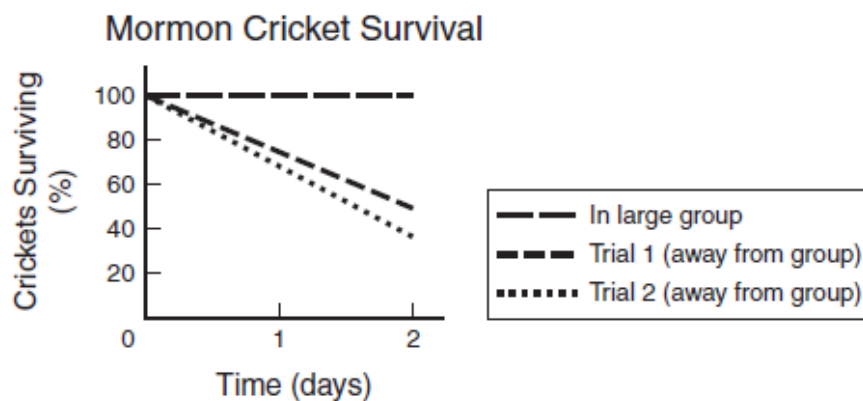
After



© AP Photo/Jim Cole (both photos)

- 28 In spring 2003 a natural rock outcropping in New Hampshire called the Old Man of the Mountain collapsed. Which of the following most likely loosened the rock and caused it to fall?
- F Heat turning sedimentary rock into metamorphic rock
 - G Water freezing and thawing inside cracks in the rock
 - H Volcanic activity producing pressure at the rock's base
 - J Oxygen reacting with iron on the surface of the rock

Item 9 TX-LS-2009-44



Source: Macmillan Publishers Ltd: *Nature*, Vol. 433

- 44 Flightless Mormon crickets often move in large groups. The graph shows the survival rate of crickets moving in large groups and of some crickets that were moved away from the groups. Which of these inferences about Mormon crickets is most likely accurate?

- F Mormon crickets within a group survive only two days.
- G Mormon crickets away from a group successfully reproduce.
- H Mormon crickets away from a group return to it for protection.
- J Mormon crickets within a group are less likely to be eaten by predators.

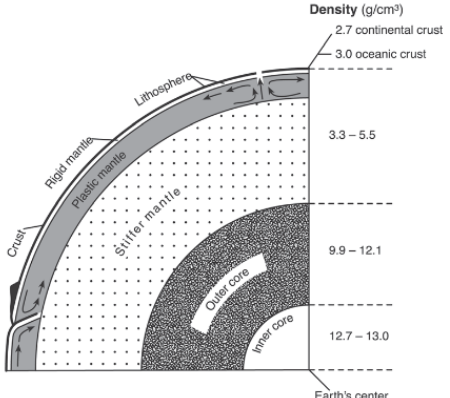
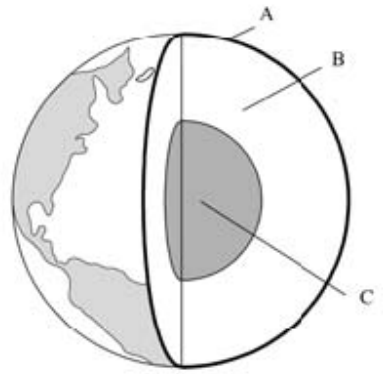
APPENDIX E

Example of Flagged Illustrated Items across Sub-Content Areas in Earth Science, Life Science, and Physical Science from China Four Municipalities, U.S. Four States, and TIMSS

(Translation of the text of Chinese items in italics)

Table E1

Example of Flagged Illustrated Items Across Sub-Content Areas in Earth Science From China Four Municipalities, U.S. Four States, and TIMSS.

Sub-content area	Assessment system		
	China	U.S.	TIMSS
Earth layers		NY-2009-13	TIMSS-1999-B01
		Base your answers to questions 13 and 14 on the cross section below and on your knowledge of science. The cross section compares the densities of different Earth layers. 	The picture shows the three main layers of the Earth.  Where is the hottest?

13 Which Earth layer is most dense?

- (1) plastic mantle (3) outer core
(2) stiffer mantle (4) inner core

A. Layer A

B. Layer B

C. Layer C

D. All three layers are the same temperature.

14 Convection currents, which may be the driving force for the movement of lithospheric plates, are mostly found in Earth's

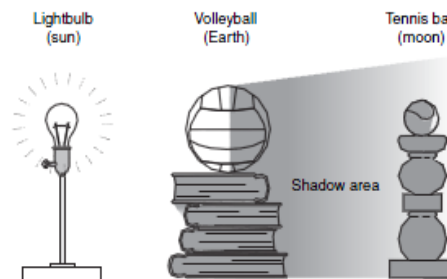
(1) crust (3) outer core

(2) plastic mantle (4) inner core

TX-2006-22

TIMSS-2003-S032532

Earth in the
solar
system and
universe



Draw the position of the Moon on the diagram below to show what is meant by an eclipse of the Sun.



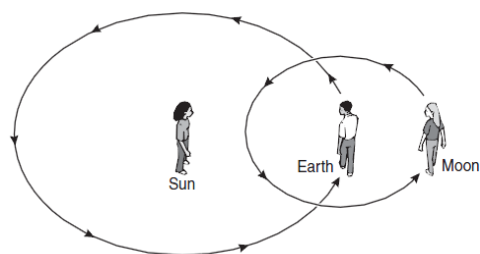
22. The model above is set up to show how a lunar eclipse occurs. What is the greatest limitation of this model?

F The lightbulb is standing straight up instead of tilted on an axis.

G Comparative sizes and distances are inaccurate.

-
- H** The shadow is being cast in the wrong direction.
- J** The heat released is much less than that released by the sun.

TX-2009-37



37. Three students use their bodies to show how the sun, the moon, and Earth are aligned during the phases of the moon. What is one limitation of this model?

- A** It cannot show the relative motion of the three objects.
- B** It cannot be safely used to show gaseous objects such as the sun.
- C** It cannot show how the sun's light affects the moon's appearance.
- D** It cannot be used to show the direction of Earth's revolution.

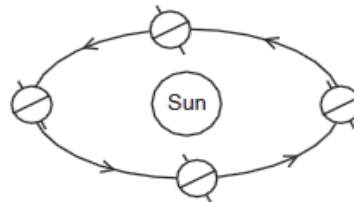
Earth's
rotation

NY-2010-29

TIMSS-2007-S032663

and
revolution

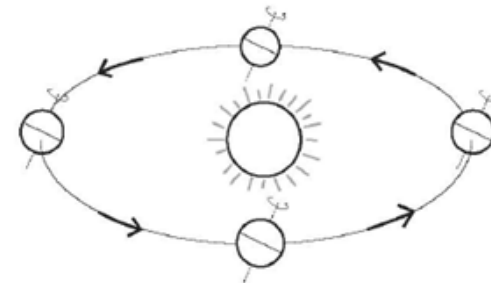
29 The diagram below shows Earth at four locations in its orbit around the Sun.



(Not drawn to scale)

Which motion do the arrows in the diagram represent?

- (1) Earth's rotation
- (2) the Sun's rotation
- (3) Earth's revolution
- (4) the Sun's revolution



The diagram above shows the Earth's path around the Sun and the tilt of Earth's axis. Which of the following patterns on Earth is caused by the tilt of Earth's axis?

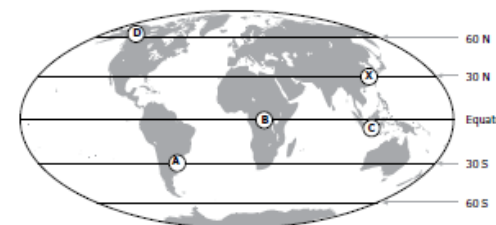
- A seasons
- B day and night
- C years
- D time zones

NY-2008-81

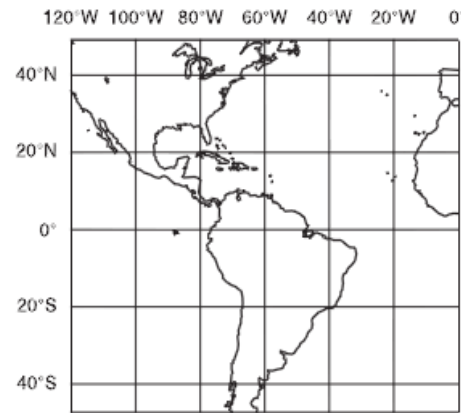
Place an **X** on the map below to indicate a location at 20° S 60° W.

World map
with lines
of latitude

TIMSS-2003-S032652



The diagram above shows a map of the world with the lines of latitude marked. Which of the following places marked on the map is most likely to have an



average yearly temperature similar to location X?
 a location A
 b location B
 c location C
 d location D

TJ-ES-2010-15

近来全球地震多发，依据资料回答 15—20 题 (*Recently, earthquakes occur frequently in the entire earth. Answer question 15 to 20 according to the following materials*). □

Earth processes

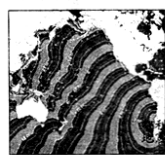


图 1
Figure 1

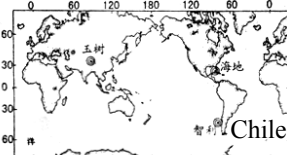
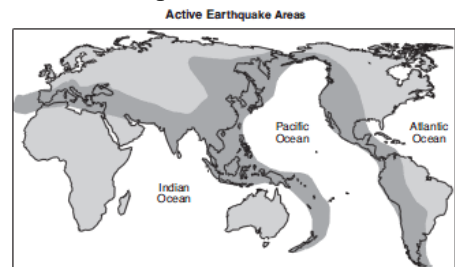


图 2
Figure 2

15. 根据图 1 判断，智利地震引

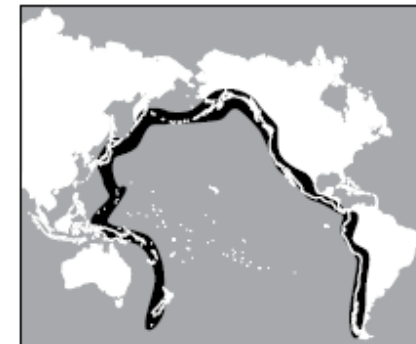
NY-2009-51

Base your answers to questions 51 and 52 on the map below and on your knowledge of science. The darker shading on the map shows the most active earthquake areas.



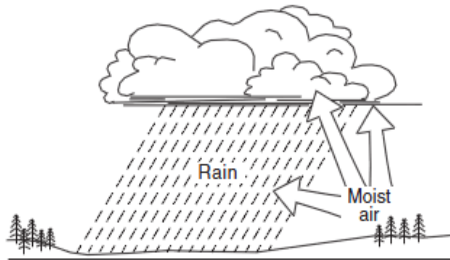
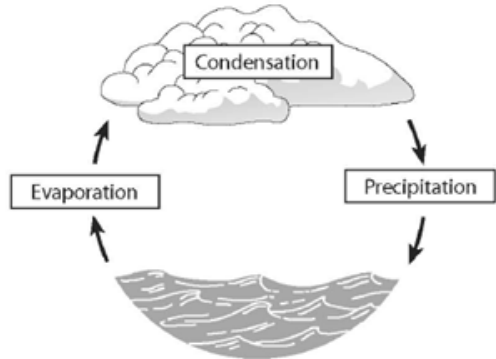
51 The most active earthquake areas are associated with the boundaries

TIMSS-2003-S032656



The diagram above shows the Pacific Ring of Fire. Earthquakes and volcanic activity occur along the Ring of Fire. Which of the following best explains why?
 a It is located at the boundaries of tectonic plates.

发海啸波及的主要海洋是 () <i>According to Figure 1, the major ocean affected by the tsunami caused by Chile earthquake is</i> A. 太平洋 (<i>Pacific Ocean</i>) B. 印度洋 (<i>Indian Ocean</i>) C. 北冰洋 (<i>Arctic Ocean</i>) D. 大西洋 (<i>Atlantic Ocean</i>)	of lithospheric plates. Explain what happens to the lithospheric plates at these boundaries that causes an earthquake. [1] _____ _____ _____ 52 Identify one geologic event, other than an earthquake, that may also occur in the darker shaded areas on the map. [1] _____	b It is located at the boundary of deep and shallow water. c It is located where the major ocean currents meet. d It is located where ocean temperature is the highest.
Weathering	TX-2006-28 Before  After  28 In spring 2003 a natural rock outcropping in New Hampshire called the Old Man of the Mountain collapsed. Which of the following	TIMSS-2003-S012013 The pictures show two different mountains. The mountains in Picture A are rough and jagged. The mountains in Picture B are smooth and rounded.  Picture A Picture B Which statement about these mountains is probably true? a The mountains in Picture A are older. b The mountains in Picture B are older. c The mountains are about the same age but were formed in different ways.

	most likely loosened the rock and caused it to fall? F Heat turning sedimentary rock into metamorphic rock G Water freezing and thawing inside cracks in the rock H Volcanic activity producing pressure at the rock's base J Oxygen reacting with iron on the surface of the rock	d The mountains are about the same age but are in different hemispheres.
	NY-2008-23 23 The diagram below shows a material being cycled between the living and nonliving environments.  Which material is being cycled? (1) carbon dioxide (3) oxygen	TIMSS-2007-S022294 The diagram below shows Earth's water cycle.  What is the source of energy for the water cycle?

Earth's
water cycle

(2) nitrogen

(4) water

A The Moon

B The Sun

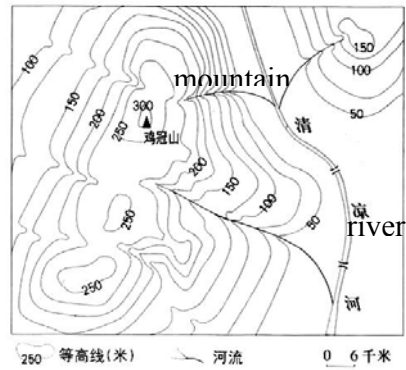
C The tides

D The wind

CQ-2007-6/7

TX-2006-29

TIMSS-1999-D03



Reading
contour
maps

读右地形图回答 6—7 题 (*Read the contour map on the right to answer question 6 and 7*):

6. 图示范围主要地形是 (*the major landform shown in the map is*)

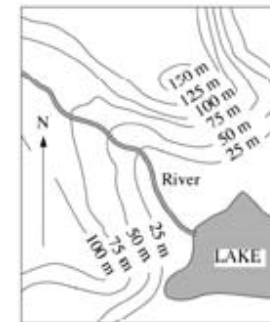
- A. 山地 (*country*)
- B. 丘陵 (*hill*)
- C. 盆地 (*basin*)
- D. 高原 (*tableland*)

7. 清凉河的大致流向是 (*The approximate direction of the river*



29. Which of the following is the most reasonable topographic map of the area shown in the picture?

On the diagram, hills and valleys are shown by means of contour lines. Each contour line indicates that all points on the line have the same elevation above sea level.



In which direction does the river flow?

- A. Northeast
- B. Southeast
- C. Northwest
- D. Southwest
- E. It is not possible to tell from the map.

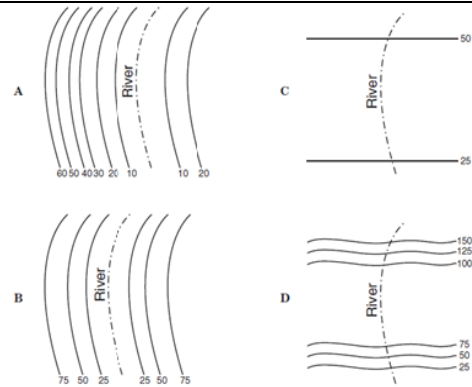
is)

A. 自北向南流 (from north to south)

B. 自南向北流 (from south to north)

C. 自东北向西南流 (from northeast to southwest)

D. 自西往东流 (from west to east)



TJ-2010-4

4. 下面的等高线地形图所对应的地形景观素描图是 ()

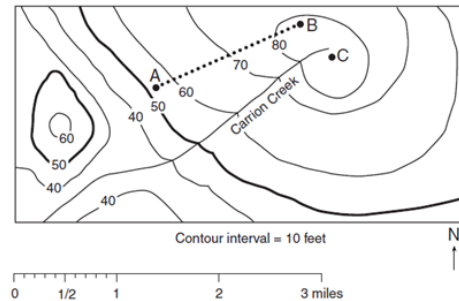
Which of the following is the corresponding picture of landform to the topographic map ()



A. 甲 B. 乙 C. 丙 D. 丁

NY-2007-40/41

Base your answers to question 40 and 41 on the topographic map below, which shows the elevation of land in feet above sea level. Points A, B, and C are locations on the map.



40 A camper walked from point A to

point *B* by taking a path shown by the dotted line. What is the approximate distance the camper walked?

- (1) 1.5 miles (3) 3.0 miles
(2) 2.5 miles (4) 3.5 miles

41 What is a possible elevation of point *C*?

- (1) 75 feet (3) 95 feet
(2) 85 feet (4) 105 feet
-

Table E2

Example of Flagged Illustrated Items Across Sub-Content Areas in Life Science From China Four Municipalities, U.S. Four States, and TIMSS.

Sub-content area	Assessment system		
	China	U.S.	TIMSS
	TJ-2009-22		TIMSS-2003-S032202
	22. 下图为某生态系统的食物网简图，请据图回答（6分）		
Food web	(The following illustration is a simplified food web for some ecosystem. Please answer questions according to the illustration 6 points)		
			The diagram above shows a community consisting of mice, snakes and wheat plants. What would happen to this community if people killed the snakes? D. All three layers are the same temperature.

(1) 图中所示的食物网中, 共有_____条食物链。

(2) 请写出该食物网中最长的一条食物链

(3) 此食物网中, 鹰与兔的关系是_____。

(1) In the food web shown in the illustration, altogether _____(number of) food chains.

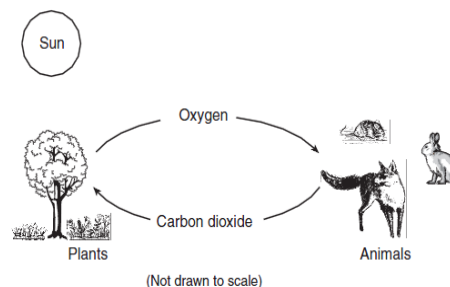
(2) Please write down the longest food chain_____.

(3) In this food web, the relationship between hawk and rabbit is_____.

NY-2008-56

TIMSS-2007-S022115

56 The diagram below represents gas exchange between several different organisms.

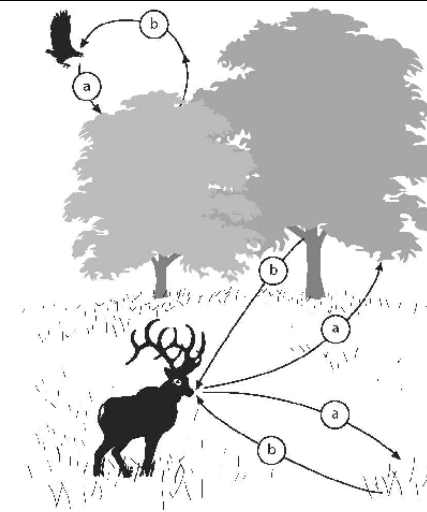


The diagram below shows an example of interdependence among organisms. During the day the organisms either use up or give off (a) or (b) as shown by the arrows.

Gas
exchange

The animals in the diagram are

dependent on the plants for oxygen.
Identify *one* other way in which
animals are dependent on plants. [1]



Choose the right answer for (a) and (b) from the alternatives given.

- (A) (a) is carbon dioxide and (b) is nitrogen.
- (B) (a) is oxygen and (b) is carbon dioxide.
- (C) (a) is carbon dioxide and (b) is water vapor.
- (D) (a) is carbon dioxide and (b) is oxygen.

SH-2008-21

TIMSS-2007-S042006

Human
body

1. 右图为人体的某个系统的结构示意图，请据图分析 (*The right illustration shows a system in human body, please analyze according to this illustration*)

(1) 该结构图表示人体的系统，此系统的主要功能

是_____。

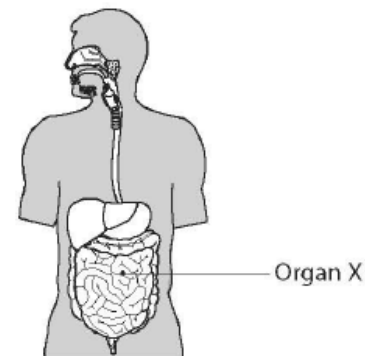
(1) *The main function of the system shown in this illustration is _____.*

(2) 图中“3”表示的器官是_____；“4”表示的器官是_____。

(2) *Number 3 in the illustration refers to the organ _____; 4 refers to the organ _____.*

(3) 除了右图所示的系统能参与维持内环境的物质稳定以外，还有其他一些系统也参与维持内环境的物质稳定。请写出其中的两种系统：_____系统、_____系统。

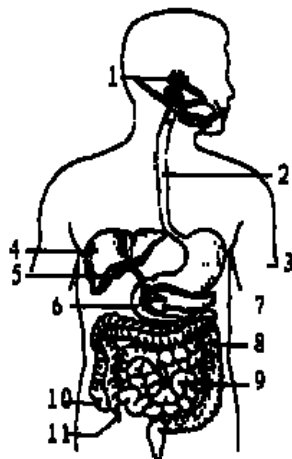
(3) *Besides the system shown in this illustration which helps maintain the balance of inner circulation, there are other systems having the same function. Please write the names of two such systems: _____system,*



What is organ X?

- (A) liver
- (B) stomach
- (C) small intestine
- (D) large intestine

_____system.



TJ-2009-21

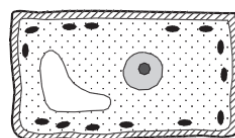
21. 下图是一个植物细胞模式图，请据图回答（8分）

21. The following illustration shows a plant cell, please answer questions according to this illustration (8 points)

Plant cell

NY-2008-52

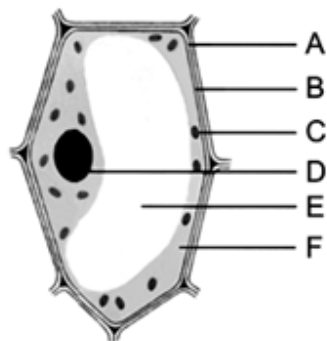
Base your answers to questions 52 and 53 on the diagram below and on your knowledge of science. The diagram shows a typical plant cell and some of its parts.



Plant cell

Key	
	Cell membrane
	Cell wall
	Large vacuole
	Nucleus
	Chloroplast

52 Which cell part directs the



- (1) 上图植物细胞的最外层[A]是_____。
- (2) 图中紧贴[A]内侧的[B]_____能够控制物质出入细胞。
- (3) 图中的[D]_____是遗传信息库。
- (4) 西瓜甘甜可口，主要是因为[E]_____内的细胞液中含有大量糖分。

(1) The outmost part of the plant cell in the left illustration [A] is _____.

(2) [B] next to [A] is _____, which controls the entrance and exit of substances.

(3) [D] in the illustration stores genetic information.

(4) Watermelons usually sweet, mainly because [E] _____, inside which cell sap contains a large amount of sugar.

activities of the cell? [1]

53 Identify two cell parts that indicate this diagram represents a plant cell and *not* an animal cell. [1]

- (1) _____
- (2) _____

Table E3

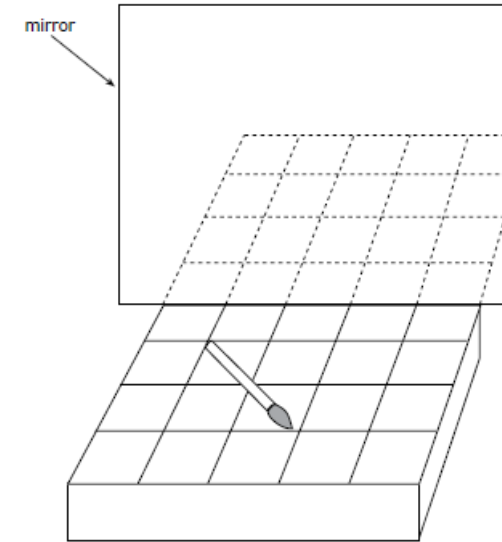
Example of Flagged Illustrated Items Across Sub-Content Areas in Physical Science From China Four Municipalities, U.S. Four States, and TIMSS.

Sub-content area	Assessment system		
	China	U.S.	TIMSS
Reflection of light in a mirror	BJ-2010-30(1/2)		TIMSS-2003-S022279
	30. 小明利用如图 16 所示的实验器材探究平面镜成像的特别。请你解答以下问题： 30. Xiaoming uses the experimental materials shown in Figure 16 to investigate the characteristics of mirror image. Please solve the following questions: （1）该实验采用薄透明平板玻璃作为平面镜，是为了能确定_____，能比较_____。 (1) This experiment uses thin transparent plate glass as a plane mirror, which is for ensuring_____, and being able to compare_____. （2）该实验选择两个相同的蜡烛 A 和 B，是为了观察蜡烛 B 是否与蜡烛 A 所成的像_____。		The picture shows a paint brush that is lying on a shelf in front of a mirror. Draw a picture of the paint brush as you would see it in the mirror. Use the patterns of lines on the shelf to help you.

(2) This experiment selects two identical candles A and B, which for observing if Candle B and the image formed by Candle A



图 16



SH-2010-6

6. 质量相同的甲、乙两小车同时同地做匀速直线运动，它们运动的 $s-t$ 图像如图 2 所示。由图像可知，甲小车 ()

Motion and speed

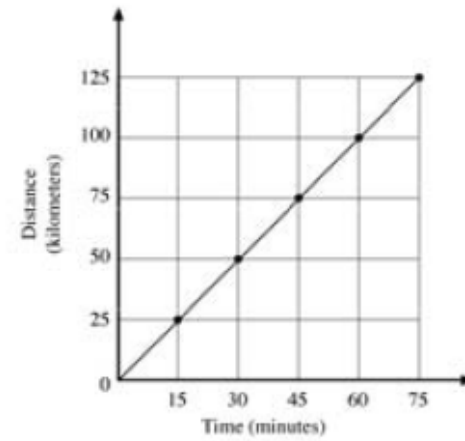
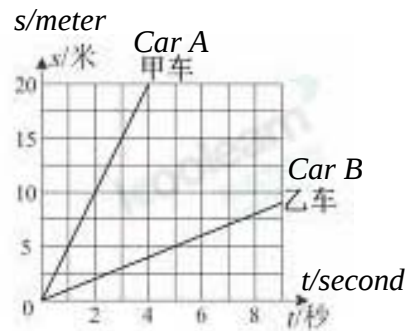
6. Car A and Car B have the same mass, and simultaneously do uniform linear motion. Figure 2 shows the $s-t$ image of their motion. Based on the image, Car A ()

A 具有的惯性大
B 所受的合力大

TIMSS-1999-P01

The graph shows the progress made by a car traveling along a straight road.

- C 具有的动能大
 D 具有的势能大
A with larger inertia
B with bigger composition of forces
C with bigger kinetic energy
D with bigger potential energy



- What is the speed of the car?
 A. 25 kilometers per hour
 B. 50 kilometers per hour
 C. 75 kilometers per hour
 D. 100 kilometers per hour