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

Chao Wang

&

Guillermo Solano-Flores

University of Colorado at Boulder

Paper presented at the annual meeting of the
American Educational Research Association, April 7-11, 2011.

Note: This study is part of a larger study titled, "Design and Use of Illustrations in Test Items as a Form of Accommodation for English Language Learners in Science Assessment," funded by the National Science Foundation (Award No. DRL 0822362). We are grateful to the funding agency, our colleagues in the project, and the members of our technical advisory board for their support. The opinions expressed are not necessarily those of our colleagues or the funding agency.

Abstract

In this paper, we report on a study that compares state, national, and international assessment programs as to the characteristics and functions of the illustrations used in their science test items. We used our conceptual framework for examining the characteristics of illustrations in science items (Solano-Flores & Wang, 2009, 2011) to code the illustrations of samples of items. We examined the statistical significance of differences in the frequencies of different illustration variables observed in samples of science items from assessments from two countries (China and the U.S,) in four science areas, physics, chemistry, biology, and earth and space science. We observed statistically significant differences between the numbers of features in the illustrations originated in China and the illustrations originated in the U.S. Illustrations from China tended to have more varied and complex characteristics than their U.S. counterparts. We discuss the implications of these findings in the design of science items in assessment projects that involve culturally and linguistically diverse populations in both the U.S. and in the context of international test comparisons.

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

The goal of this study was to contribute to enhanced practice and research concerning the use of illustrations in current large-scale science assessment in the testing of culturally and linguistically diverse populations within the U.S. and in the context of international test comparisons.

While illustrations play a critical role in science education (e.g., in textbooks), their use in the context of assessment is yet to be investigated systematically. Little is known about the ways in which cultural factors shape students' interpretations of illustrations in science assessment or the extent to which illustrations used in test items reflect the characteristics of the culture in which they are generated. We aimed at providing researchers and test developers with empirical evidence on the use of illustrations, which is critical to designing science items that are not culturally biased.

We asked: *How do the features of illustrations used in large-scale science assessments vary across countries and science content areas?*

Conceptual Framework

The theoretical and empirical understanding of visual language is much less advanced than that of spoken and written languages (Freedberg, 1989; Marriott & Meyer, 1998; Moore & Dwyer, 1994; Moriarty, 1996). Research on illustrations in education has focused on the relationship between illustrations and text (e.g., Bloomer, 1960; Miller, 1936, 1938; Whipple, 1953; Zimet, 1966), the influence of illustrations on performance in cognitive tasks (e.g., Bruner, Olver, & Greenfield, 1966; Pressley, 1977; Wendt & Butts, 1962; Wohlwill, 1968), and the characteristics of illustrations according to their intended functions (Fleming, 1966; Svensen, 1993). Unfortunately, this research has neglected the fact that visual representation conventions rooted in culture can lead to misinterpretations of visual language (Boling, Eccarius, Smith, & Frick, 2004; Boling, Smith, Frick, & Eccarius, 2007; Knight, Gunawardena, & Aydin, 2009; Schiffman, 1996).

In the study reported in this paper, we attempted to identify whether different assessment programs differ as to the characteristics of the illustrations used in their items. We examined the features of item illustrations in tests used in the U.S. and in China to identify any cultural differences in the illustrations used in tests.

We used a conceptual framework that we created for characterizing illustration in science tests, and which is described separately in another paper submitted to this conference (Solano-Flores & Wang, 2009, 2011). According to that conceptual framework, an illustration is any device that provides information mainly in non-textual form. We coded the presence of features of illustrations according to over 100 illustration dichotomous variables, which are grouped in five categories that refer to the form, style, and complexity with which information is presented to the examinee: 1) objects and background, 2) metaphorical visual language, 3) text in the illustration, 4) variables, constants, and functions, and 5) illustration-text interaction. Table 1 lists the five categories of illustration variables and provides the full list of variables for the first category, object/background.

Methods

Item sample. Our sample of items was drawn from a corpus of over 800 released multiple-choice and constructed-response science items for Grade 8/9 (Note 1) from the National Assessment of Educational Progress (NAEP), Years 2000, 2005, and 2009, the District of Columbia Comprehensive Assessment System (DCCAS), Year 2009, the New York State Assessment (NYSA) Year 2009, the California Standards Tests (CSTs), Years 2006, 2007, and

2008, the Texas Assessment of Knowledge and Skills (TAKS), Year 2009, China's Middle School Exit Examination (CMSEE; Beijing, Chongqing, Tianjin, Shanghai, Nanjing, Changchun, and Jinan), Years 2009 and 2010. Most of the assessments were generated in most populous cities or regions in the U.S. and China. About 35% of the test items (276) have some kind of illustration.

Table 1. Excerpts from the c*onceptual framework for analyzing illustrations in science test items (Solano-Flores & Wang, 2011). The list of illustration dichotomous variables (in italics) is shown only for the Objects and Background category*.

Objects and Background variables: Style, complexity, and concreteness-abstraction with which objects and background are represented.

- Image concreteness: *photograph, realistic line drawing, schematic, silhouette, cartoon, logo, icon, emblem, metonymy, symbol, reference, entity*
- Background: *with background, without background*
- Zooming: *zoom-in, zoom-out, zero zooming (naked eye)*
- View: *external, internal*
- Dimension: *three dimensional, two dimensional*
- Relative position of objects: relevant, irrelevant
- Relative scale of objects: *proportionate, disproportionate*
- Color: black & white, multicolor, gray scale
- Constituents: subject, object, action, background

Metaphorical visual language variables: Ways of representing space, time, dynamics, manifestations of energy, and persons' states.

Text variables: Form and complexity of textual information.

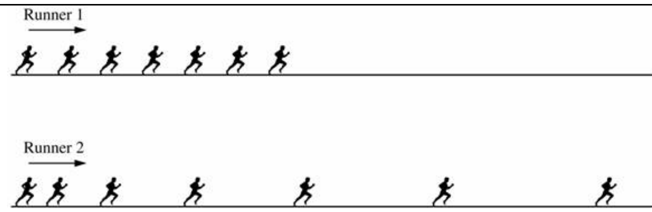
Variable/Function variables: Form, complexity, and concreteness-abstraction with which variables, functions, and constants are represented.

Illustration-text Interaction variables: 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.

From this corpus of items, we drew a sample of items using a multistage cluster sampling method—in which items are randomly selected from each cluster (Freedman, Pisani, & Purves, 2007)—that allowed us to attain a balanced design with equal sizes (10 items) from each of four science content areas—Physics, Chemistry, Earth Science, and Life Science for each country.

Coding

Unlike other approaches for examining images (e.g., Evans, Watson, & Willows, 1987; Hunter, Crismore, & Peason, 1987), our coding approach is not based on characterizing images simply by giving them general labels such as “chart,” “diagram,” or “graph”, whose meanings lend themselves to different interpretations and which are not sensitive to complexity.



5. The picture above shows the positions of two runners at <u>one-second intervals</u> as they move from left to right. For each runner, indicate whether the runner's speed seems to be constant, increasing, or decreasing. Explain how you can tell this from the pictures.

____
____
____

Coding of the Item’s Illustration: Illustration Variables Identified

Object/Background variables:
- *Silhouette*: The illustration shows only the shape of the object, with no other details
- *No background*: The object (a man running at different point in time) is self-standing, with no other objects as background.
- *Zero zooming*: The object is shown in a way in which would be seen with the naked eye.
- *Natural view*: No cross-sections or virtual windows to the interior or the object are shown

Metaphorical visual language variables:
- *Time*: Time is shown as a succession of events (the same man at different points in time)

Variable/Function variables:
- *Case*: A variable is shown as cases (Runner 1 and Runner 2)
- *Structure*: A variable (time) is shown as structure, from left to right

Text variables:
- *Label*: “Runner 1” and “Runner 2”

Illustration-Item Interaction variables:
- *Location of illustration*: Above the text of the item
- *Reference to illustration*: Explicit (“The picture above...”)
- *Actions examinees are expected to perform with the illustration*: Observe or examine the illustration

Figure 1. Example of a science item on comparing two runners’ speeds in 2005 NAEP for Grade 8 students.

In our coding approach, the illustrations of the selected items (Table 3) were coded dichotomously (1-0), by recording the presence or absence of each of the variables from the seven categories listed in Table 1. Figure 1 provides an example of coding for a physics item.

Data Analysis

We examined the 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. A series of two-way ANOVAs were performed to examine the statistical significance of any differences in the frequencies of illustration variables observed by science subject area (e.g., physical sciences), and item cultural origin.

Results

Table 2 shows the mean total numbers of different features observed in items from the two countries. The mean is higher for the illustrations from items from China than items from the U.S.

Table 2
Mean on Total Number of Features Across Country of Origin. Standard Deviations in Italics.

Items' Country of Origin	Number of Features
U.S. (n=40)	16.68 *3.83*
China (n=40)	18.73 *4.05*
Total (n=80)	17.70 *4.05*

The histograms in Figure 2 show the frequencies of numbers of different features between items generated in the U.S. and items generated in China. As the shape of the distributions show, Chinese illustrations tend to have a wider range of features than illustrations from the U.S. Illustrations from China tend to have more varied and complex characteristics than their U.S. counterparts.

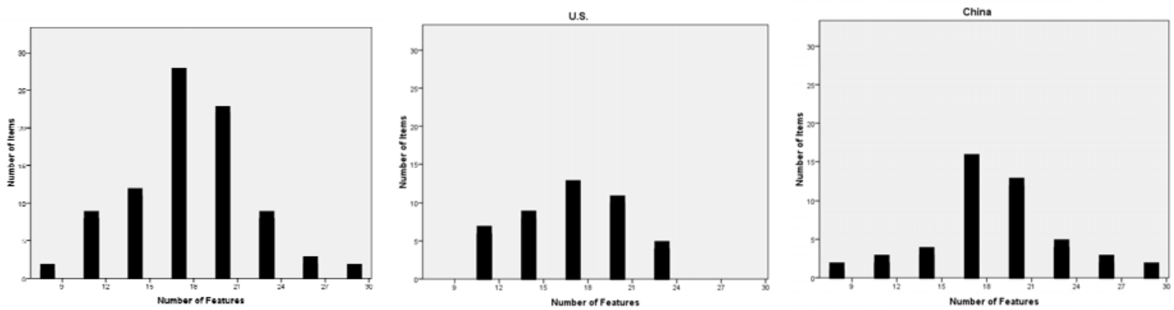


Figure 2. Different number of features shown in different number of items generated correspondingly in the U.S.-China combined, the U.S, and China.

A two-way ANOVA revealed statistically significant mean differences on the total number of illustration features only for the main effect of country of origin ($p=0.02$), although with a small effect size ($\eta^2=0.018$). This small effect size may be an artifact of including in the computation 22 variables titled, “other,” for types of features not considered in the coding sheet of each of the subcategories in our coding system. Since the “other” variables were rarely coded 1, their zero values may have masked the proportion of the differences between countries in the numbers of different features. Further analysis will be conducted to address this possibility.

Summary and Conclusions

In this exploratory study, we examined the numbers of different features in testing illustrations used in large-scale tests form the U.S. and from China on four science content areas. We observed a statistically significant differences between the numbers of features in the illustrations originated in China and the illustrations originated in the U.S. .

Our results should be regarded as preliminary. Additional analyses will be performed to examine whether and how the countries differ on the number of different item illustration features for each of the five illustration categories in our coding system. The findings from the full set of analyses will contribute to informing the process of assessment design in international test comparison programs. For now, the differences between countries in the shapes of the frequency distributions and the numbers of features observed in the illustrations suggest important cultural differences in the complexity of illustrations used in science tests.

Notes

Note 1. While in the U.S. large-scale science tests on all science content areas are administered in 8th grade, in China large-scale tests are administered in 8th grade for Earth and Space Science and Life Science, and in 9th graders for Physics and Chemistry.

References

Bloomer, R. H. (1960). Children's Preferences and Responses as Related to Styles and Themes of Illustration. *The Elementary School Journal, 60*(6), 334-340.

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). *Graphical devices in instructional illustrations: Designers' intentions and viewers' interpretations*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Bruner, J. S., Olver, R. R., & Greenfield, P. M. (1966). *Studies in cognitive growth.* New York: Wiley.

Carletta, J. (1996) Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics, 22*(2), pp. 249–254.

Cohen, J. (1968). "Weighed kappa: Nominal scale agreement with provision for scaled disagreement or partial credit". *Psychological Bulletin 70*: 213–220.

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.

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.

Freedberg, D. (1989). *The power of images: Studies in the history and theory of response.* Chicago and London: The University of Chicago Press.

Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics* (4th ed.). New York & London: W. W. Norton.

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.

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.

Marriott K., & Meyer, B. (1998). *Visual language theory.* New York: Springer-Verlag.

Miller, W. A. (1936). The Picture Choices of Primary-Grade Children. *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.

Moore, D. M. & Dwyer, F. M. (1994). *Visual literacy: A spectrum of visual learning.* Englewood Cliffs, New Jersey: Educational Technology Publications.

Moriarty, S. E. (1996). Mapping the visual communication field. In R.Griffin, D. Beauchamp, J. Hunter, & C. Schiffman (Eds.), *Eyes on the future: Converging images, ideas and instruction* (pp. 379-386). The International Visual Literacy Association.

National Assessment of Educational Progress. (n.d.). *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

Pressley, M. (1977). Imagery and Children's Learning: Putting the Picture in Developmental Perspective. *Review of Educational Research, 47*(4), 585-622.

Schiffman, H.F. (1996). *Linguistic culture and language policy.* London, UK: Routledge.

Solano-Flores, G. (2009). *Conceptual Framework for Classifying Illustrations in Multiple Choice Test Items.* University of Colorado at Boulder, October, 20, 2009.

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 accepted for presentation 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.

Wendt, P. R., & Butts, G. K. (1962). Audiovisual Materials. *Review of Educational Research, 32*(2), 141-155.

Whipple, G. (1953). Appraisal of the Interest Appeal of Illustrations. *The Elementary School Journal, 53*(5), 262-269.

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.