CHAPTER 9

What's At Stake in a High-Stakes Math Test? Analysis of Multimodal Challenges for Emergent English Bilingual Learners

Theresa Austin

Introduction

While we live with multimodal literacies in all walks of life, contemporary schooling has traditionally favored only a limited number of modalities, in general verbal and written modalities in the main academic subject areas. With the advance of technologies in learning design, there is increased attention to how the use of multimodality impacts learning in culturally diverse settings. For emergent English bilingual learners, in particular, this multimodal orientation is significant because it holds both potential benefits and obstacles in its implementation. In this chapter, I argue that while multimodal literacies may hold benefits for all learners, these literacies are not culturally neutral and must be examined closely for assumptions that limit their potential use for learners from non-dominant cultural backgrounds. Drawing on tools for multimodal analysis from a critical discourse analysis perspective, assumptions are analyzed in the world of standardized testing through an examination of instructions or guidelines that were designed for math test proctors for the Partnership for

T. Austin (✉)
Department of Teacher Education and Curriculum Studies
University of Massachusetts College of Education, Boston, MA, USA
e-mail: theresaya@gmail.com, taustin@educ.umass.edu, austinty2@gmail.com

F. Pirbhai-Illich et al. (eds.), *Culturally Responsive Pedagogy*,
DOI 10.1007/978-3-319-46328-5_9

Assessment of Readiness for College and Careers (PARCC) math tests. While these instructions were designed so that proctors assure that testing conditions are the same for all learners, I analyzed these to reveal how they provide evidence that the required use of interrelated modalities increases the complexity of the items for emerging second-language learners (L2), often without necessarily revealing what the learner may know about the construct being tested. I point out how the resulting complexity may well impede such a learner from being able to demonstrate what s(he) knows and thereby potentially lead to errors of underestimation and overestimation of the learner's knowledge and skills. In this chapter I provide a brief explanation of the second-language learner and standardized testing movement in the United States, a description of the text and my procedures for analysis. Next I present my results with examples that represent several categories of potential areas of complexity. Furthermore, I include example items from prior test texts that also need to be considered in light of multimodal demands in testing. In closing, I discuss implications and raise questions that we as researchers and educators may take up to disrupt the flow continual disadvantaging of marginalized groups of learners.

English as Additional Language Learners in the United States and Academic Performance on Standardized Tests

By several reports, in some school districts in the United States, students sit for as many as 91 days of testing in a 180-day typical academic school year (Florida). Emergent English bilingual learners, also known as learners of English as an additional language (EAL), may sit for even more. Since the passage of the Bush Administration's No Child Left Behind Act in 2001, and its continued implementation during the Obama administration, these tests increasingly have been relied upon to form part of the US attempt to identify and be accountable for this population's academic progress (Abedi 2002). The makers of these standardized tests go through painstaking effort to produce a defensible measurement instrument that can be used for large-scale assessments and decisions. While biases may exist in these tests, by the time they are released, the tests represent the test makers' best attempt to reduce these. Overtime these tests continue to be monitored for test takers' interactions with items that compromise the test results' ability to render a test takers' "true" score.

However, a continuing significant nagging problem is that these tests consistently produce lower scores for students who are designated as "English Language Learners" (this federal appellation applies to those whose home language is other than English and who enter schools without the level of English required for academic instruction; Hemphill and Vanneman 2011; Mahoney et al. 2009; Short and Fitzsimmons 2007). As early as 1999, the National Council on Measurement in Education in their publication entitled "Standards for Educational and Psychological Testing" stated that "the many psychoeducational tests have been developed for and normed with monolingual, English-speaking children are irrelevant for multilingual learners" (American Educational Research Association, American Psychological Association, & National Council of Measurement in Education [AERA, APA, & NCME] 1999, p. 7). Yet still it is not widely acknowledged that emerging second-language learners are being asked to perform in subject areas' knowledge and skills at the same level as their grade-level peers in ways that do not take into account their developing EAL. Public school districts subject them nonetheless to the demands of these tests, without necessarily preparing teachers adequately to critically interpret or use these scores.

Certainly there are also many external reasons why these learners attain lower scores, such as lack of prior instruction, lack of preparation for testing, and lack of appropriate accommodation in testing conditions (Abedi et al. 2006; Schafer Willner et al. 2008). Nonetheless, there are also test internal issues that merit further analysis. The tests themselves warrant scrutiny of their construct's domains, items, and response formats for their required use of multimodalities in assumed culturally neutral ways. Equally important objects for scrutiny are the instructions for administering the test, the preparation of such administrators and their understanding of accommodations, analysis, and interpretation of learner response patterns, and most importantly their resulting decision making. It can be argued that at every point along the production of a test, we need to be attentive to the cultural assumptions behind the use of modes and the complementarity of their use.

One internal indicator I focus on here is the text that prepares test proctors to administer the test and the subsequent potential effect on the EAL test taker. If we analyze the test guidelines for proctors, we can see how the ideal reader (proctor) must be well prepared multimodally in order to provide sufficient cues aurally, visually, and/or in writing to elicit the second-language test taker's performance.

I focus on a critical analysis of two areas of standardized testing: the instructions given to those who administer the audio segments of the tests, and then focus on multimodal test items themselves. These analyses are undertaken on a segment of the PARCC math tests, as they apply to EAL learners. In essence, the guidelines themselves are a text that demands the test administrator to be an "ideal performer" who must consistently and accurately solicit expectations for test takers to perform multimodally. We could say that test guidelines genre itself then is "doubly" multimodal for both the administrator and test taker as each is expected to produce results multimodally. Thus we need to review the level of complementarity between how the test construct is elicited multimodally and how the test taker is expected to perform in a corresponding manner.

By taking this approach, I make visible what is at stake in a subset of the multimodal literacies that are demanded of second-language learners in high-stakes testing. There appears to be an assumption by the test makers that the general multimodalities embedded in the tests are part of decodable information that can be recognized by the test taker. However, these multimodalities need to be deconstructed for EAL learners whose prior cultural and linguistic repertoires may not have been considered, and whose abilities within subject domains could be underrepresented without such a focus. I argue that the results of these analyses demonstrate the need to understand how multimodal items and aural instructions affect the ability of EAL learners to demonstrate what they know and can do.

Drawing on critical multimodal analyses (Kress and Van Leeuwen 2006), this chapter makes visible the multimodal assumptions in these instructions about visual–verbal synergy as well as aural–visual synergy. I argue for similar analyses to be conducted both of the prompts, and response formats of the items included in all high-stakes assessment systems. The current study points out several areas that need significant attention to help learners negotiate unfamiliar high-stakes multimodal texts.

Positioning the Researcher–Teacher Education in the Context of the Global Spread of High-Stakes Standardized Testing

I am a multilingual teacher educator and second-language researcher who has taken up the call to build a critical testing literacy (Austin and Ites 2012; Guillerme 2007; Shohamy 2014). Currently, while I face pressure

to teach to the areas of teacher testing that will prepare aspiring teachers for professional licensure, there is even greater need to prepare all teachers to understand and navigate the obstacles in the path of second-language learners and students whose home language varieties differ from the dominant language. One of the biggest obstacles in second-language learners' academic progress is standardized testing that, in general, does not account for what these learners have accomplished, particularly in language arts, math, and science, but also in their family and community life. At the present time, second-language learners in the US K-12 public schools are also being held to standards that have been set in international standardized testing, Trends in International Mathematics and Science Study (TIMMS), as well as nationally set standards such as those set by the Common Core Curriculum through its assessment, World Instruction Assessment (WIDA).

One such standard is set by the PARCC. As of 2014, despite criticisms of these standards, over 40 of the 50 states in the United States have adopted Common Core Standards. This partnership has developed Model Content Frameworks, which have aligned standards, instruction, and assessment from the above-mentioned Common Core Initiative. Their alignment includes curriculum, materials, and PARCC assessments to fortify these discourses within public education, making them "norms" that govern instruction and monitoring of progress under new accountability regulations already in place in the No Child Left Behind (2001) and Race to the Top (2010) US federal initiatives. Increased deep thinking and high expectation for all learners form part of the promise. The use of this test will affect second-language learners' future and begs for further analysis of the demands placed on these learners. How can progress of second-language learners from diverse educational backgrounds and levels of English be fairly evaluated in their understanding of content and in progress toward becoming English users?

Toward Culturally Responsive Tests – Recognizing Multimodal Challenges

Understanding Kress and van Leeuwen's (1998) account of multimodality that "[l]anguage always has to be realized through and comes in the company of, other semiotic modes" (p. 186) helps me explain how standardized tests are designed to make use of more than mere print-based text

to elicit test taker responses. Royce (2007) points out that "[a] multimodal text (e.g., page or screen-based) is a text where the modes utilized 'work together' in various ways to produce comprehensible meanings – there is a synergy in their combined meanings" (p. 374). This can be seen in the relations of these modalities to shaping ideational meanings, interpersonal meanings, context, and coherent textual meanings.

The study of multimodal literacies recognizes the complementarity of modes in the production of meaning in texts. When two or more modes co-occur, they constitute new potential meanings that need to be interpreted and often reproduced by learners to demonstrate their understanding. In order to make sense, learners draw on their background knowledge and past experiences which are always culturally situated. For example, at the interpersonal level, beginning L2 learners may not be culturally prepared to interact in the roles designed by the test makers. Thus the L2 learners' verbal or written response may suffer from coherence or cohesion on the textual level because they may not be familiar with a visual prompt to meaningfully draw on its represented information. Particularly in a testing context, the culturally diverse learner would then be cut off from meaning-making resources, which could penalize them if they are at beginning stages of language development. In this way, the interaction of visual and language modes in the test items need to be examined for their level of complementarity and for how L2 learners use these resource options. With this analysis we can see if, at the ideational level, the combined modalities offer L2 test takers a sufficient range of information for conceptualizing and responding to the targeted item. At a minimum with the results of the analysis a test's affordances for these learners could be augmented to create a more culturally responsive test.

Preparing Test Administrators – Invisible/Inaudible Multimodal Issues in Aural Testing Prompts for Math

Aside from the test construction itself, preparing people to properly administer an oral exam is a critical to obtaining an accurate sample representing what learners understand and can do in the subject matter being tested. Rigorous training attempts to ensure that standardized conditions for administering a test do not vary significantly because variations in administering the oral sections have potential to alter the test takers' responses. However

standardized conditions are often in reality a greater challenge than is generally recognized, particularly in terms of meeting optimal conditions for second-language learners. It is here that the variance caused in actual conditions can adversely affect second-language learners who are at the most vulnerable and earliest stages of developing their comprehension of multimodal test items. The PARCC Math Assessment is leading the current wave of standardized testing in the United States. It is promoted as rigorously aligned with the Common Core Standards. At the time of writing, 13 states and the District of Columbia have joined together to support the development and use of these assessments. For research purposes, I obtained a copy of the manual for preparing testers to administer the PARCC Math Assessment Audio Guidelines Version 3.0 (PARCC 2014). This 84-page-long document is one of many documents describing the rigorous processes undertaken to develop these assessments that are available for public scrutiny. Due to the need to maintain test security these most likely are not the actual or most current guidelines, but they do provide a window to examine potential multimodal issues in administering such tests.

This particular document provides instructions on how test administrators are expected to provide oral instructions as the test takers interact with test items. The web-published version includes a Change History Log, indicating that it has undergone at least three revisions across four dates, in which items have been revised or deleted by particular authors. The guidelines include instructions for the test administrator to describe the following items: visuals and the symbols, numbers and expressions/equations/operations, and diagrams/figures and keys that are included in the tests. The guidelines (PARCC 2014) also provide a classification of the embedded codes used to describe items for text speech. The codings reveal the three levels of all items with visual elements, for example: "[1] is not construct-relevant and can be eliminated"; "[2] is construct-relevant and can be represented using accompanying textual description"; and "[3] is construct-relevant and can be represented using accompanying textual description together with a tactile representation or physical manipulative" (p. 7). Explicit instructions for the test administrator's reading of each section of items included in the guidelines are necessary because the aural "reading" is needed to guide the test taker and must not inadvertently provide cues to the answer being sought nor vary greatly in the details of the instructions. More importantly, notice that each instruction exists because there is variance in orally rendering each item. These guidelines

provide an example of the item, then provide instructions, and finally present an application of how to read each.

I selected this document precisely because it provides an opportunity to analyze how training of a test administrator is alerted to the multimodal challenges to represent the tasks, but moreover to point out how these might impact L2 learners listening in these tasks and working with the items. Since the test taker needs to use the audio section to focus attention on the item in order to understand the question, the relationship between this prompt and the item ideally should guide the test taker to perform the anticipated processes in order to produce the response that best represents the test taker's knowledge and skill.

Since I theorize that second-language learners at beginning stages of English development and new to learning math concepts will predictably perform differently from the second-language learners at beginning stages of English development and who are already familiar with and understand the math concepts being tested, I envision the former population as I conduct this analysis of these guidelines to identify items that could potentially cause them confusion or misinterpretation of the oral instructions and written language and symbols used together in math items. In reviewing this document I used three features of these test items: ideational, interpersonal, and textual. Ideational are meanings that are

> concerned with the identification of participants (who, or what is involved in any activity), the activity (the processes in terms of what action is taking place, events, state, types of behavior), the circumstances (where, who with, by what means the activities are taking place), and the attributes (the qualities and characteristics of the participants). (Royce 2007, p. 375)

The interpersonal level of a test consists of how the designers address their viewers/readers, express degrees of involvement, and exert degrees of power relations through forms of address (questions, commands, statements, etc.) and attitude (necessary/unnecessary, possible/impossible, true or false, etc.). The textual level is how the designers make use of combined modalities to produce coherent meanings. In examining these three levels of the test items described in the guidelines for proctors, I found several types of potentially problematic issues for L2 learners (interpersonal) at the beginning stages of aural comprehension.

A Critical Analysis of Instructions for Math Test Proctors

In the following paragraphs, I discuss two categories: (1) orally representing written math symbols and (2) polyphonic items/repetition of cues/coordination of modes and provide corresponding examples.

Imagine listening to an oral passage on a math test that uses symbols and having to select a correct answer afterward. In the guidelines (PARCC 2014), the administrators are challenged to orally perform this task that requires the second-language learner to see a symbol in a number of items but listen to the word that is being represented. This task is made more difficult if abbreviations are used. On the ideational level, the test item's use of abbreviation would expect the second-language learner to recognize and understand the symbol representing abbreviation as well as the word corresponding to the abbreviation. However, they are repetitions of each other only if the test taker knows/recognizes the abbreviation visually and understands the corresponding orally rendered item. The use of abbreviations in math is one category that merits attention because it is integral to showing and representing math knowledge. In the guidelines (PARCC 2014), two examples of how items included in the visual and symbols section are displayed for use in the test below:

Abbreviations (ft., km)
Example 1
3ft.

Example 2
What is the correct abbreviation for kilometer?
A: kl
B: K
C: km
D: klm

Audio Guideline
Present abbreviations by speaking the whole word the abbreviation represents.

If the item measures the ability to identify the meaning of the abbreviation, then read the abbreviation letter by letter.

If speaking the abbreviation violates the construct being measured, then read letter by letter.

If the item has measurements that are all uppercase or lowercase, then it is not necessary to reference the cases.

Application of Audio Guideline
Example 1
Three feet

Example 2
What is the correct abbreviation for kilometer?
A: kl
B: K
C: km
D: klm
(PARCC 2014, pp. 14–15)

As you can see here, the guidelines (PARCC 2014) specify when the test proctor must read the text of the item as a whole word or letter-by-letter or when the texts vary visually in use of upper/lowercase. However, this particular example overlooks the role of punctuation as a visual cue. Note that in the first example, feet is abbreviated as "ft." Yet in subsequent abbreviations punctuation is not evident in any of the choices for "correct abbreviation." While punctuation here may be considered a minor visual cue that can be corrected through subsequent editing of the multiple choice responses, attention to such details is important. In other abbreviations, such as in measurement where the marks ″, ′, and cm^2 indicate meaningful measurements of inches, feet, and square centimeters, should these be missing all learners could be affected. Other patterns of aural/visual text issues that involve symbols appear in this version of the guidelines (PARCC 2014) may be even more problematic to second-language learners who are expected to listen to the prompt and map what they are hearing to the symbols. One item belonging to this category is illustrated later.

A second category of potentially problematic items are those that have multiple oral renditions for the same written symbol. An illustrative example consists of those items with a negative number (–x) versus the symbol representing the operation of subtraction symbol (–). However, the oral reading of parenthesis in math varies more dramatically across test items of probability and multiplication. In probability items, the test proctor is instructed to read the parenthesis as "*of*", for example: "P(orange) = $\frac{1}{6}$" is

rendered as "P of orange is one sixth" (PARCC 2014, p. 31). The word probability is abbreviated to the notation only mentioning the letter "p." In contrast in the multiplication items, the tester is asked to

> Read the multiplication symbol as "times" when it appears in a math item.
>
> When a number, symbol, or another set of parentheses appears before a set of parentheses, read the number or symbol as is and "open parenthesis" before what is within the parentheses. When multiple sets of parentheses appear consecutively, read as "open parenthesis and closed parenthesis."
>
> If there are two variables or a variable and a number consecutively, do not read "times" to represent implied multiplication. (PARCC 2014, p. 32)

Expressions/Equations/Operations

Multiplication
Example 1
$3 \times 5 = X$

Example 2
$xy + 4x = 10$

Example 3
$(3 + x)(y - 2)$
(PARCC 2014, p. 32)

These are only two examples where items using multiple oral readings of the same symbol may cause confusion for emerging learners of English. There are also potential problems when visuals and written text are supposed to be used in a complementary manner to respond to a test item.

A Critical Analysis of Visuals and Language in Multimodal Math Test Items: Cultural Issues in Interpreting Multimodal Math Test Items – Visuals and Language Use

When language tests are designed to elicit "receptive skills" such as reading and listening comprehension, they inevitably require cultural knowledge of context to make sense. Reading and listening call on the test takers to use

their knowledge of sociocultural expectations for the use of language in relation to its nonlinguistic context and knowledge of how utterances and functions of these utterances are organized to create coherent communication. Specifically test takers must attend to a stimulus provided by a text or illustration, interpret meanings that they expect that test makers want, and register appropriately their responses in formats that are provided. For example, a typical listening comprehension test item may require a test taker to listen to a short narrative, and understand main points and details before answering by reading to select an appropriate multiple choice answer before accurately bubbling-in the corresponding answer on a response protocol or computerized form. For newcomers to such standardized testing, becoming accustomed to juggling so many performances requiring attention becomes a taxing short-term memory task aside from knowing how to solve the problem and answer within the requirements of the mode – narrative, visual, and/or orally. For those in the early stages of second-language development who know how to resolve the item but do not have skill in providing the required short answer narrative genre and appropriate accompanying visual may well be misidentified as not knowledgeable by the requirement to perform multimodally. If asked to respond orally, they may be able to explain their processes and their visual in their stronger language. However if asked in a language they are still in the beginning stages, their explanation may underrepresent their math knowledge, herein raising the questions about where the items are actually measuring the targeted constructs' math knowledge or math literacy development (ability to use their second language in math literate ways).

Moreover when test makers construct standardized subject matter tests, second-language learners are required to draw on visual, aural, and oral and written modes to interpret communication, leaving language almost as neutral or taken for granted. Yet the learner's coherent interpreting and producing of meaning depends upon multiple representations simultaneously in a cultural context. For example, problems occur for second-language learners when too much or too little information is conveyed by a single modality. Trumbull and Solano-Flores (2011a) point out that the language and cultural demands placed on learners becoming bilingual through English as a second language may obscure what they can actually demonstrate in subject matter. Here multimodality of an assessment item is also a factor that can distract students attention rather than focus their perception and support understanding of the text and the task it presents (see Fig. 9.1).

You are going to make jello for 12 people.

Heat the water for 3 minutes

Ingredients	Price per unit
fruit	$2.00 total
2 cups of hot water	from the sink
2 cans of juice	$1.50 per can
2 packages of jello	$1.00 per package

With the information above, find the total price of the ingredients. You can use pictures, numbers, and/or words.

Explain how you used the information above to find the total price of the ingredients.

Fig. 9.1 The language demands of mathematics assessments. Taken from E. Trumbell and G. Solano-Flores (2011b)

In this figure, interpreting meaning can be problematic – as it may produce an additional source of cultural bias because this item requires at a minimum two steps in cognitive tasks. One is to discern the relevant information in understanding what information counts in this "hide-and-seek" problem. Information such as "Heat the water for three minutes" and the number of people for the recipe required the second-language learner using valuable time for reading but were irrelevant for the task. In addition, the first question asks the price but the figure only lists costs for three items. The cost for heating water is assumed to be not important to include. It provides a linguistic element that must be understood in relation to the graph and illustrations that is "per" in order to answer correctly. In addition, the prompt makes use of the phrase "Find the total price," which is also repeated in the label listed in the graphic as "total." Potentially this repetition creates a confusing need to find the word "total" in contrast to the word "sum."

The second task has been created to understand the test takers' processes to "calculate the sum." While this explanation can be done multimodally by drawing each ingredient and labeling the costs, it can also be accomplished through the use of numbers and a formula. Neither is excluded but one may be preferred over another. The learner must discern which ways are more highly valued than others.

What does this indicate? In the standardized testing field, rather than illustrating or clarifying a relationship between text and visual, this test item uses multimodality as a distractor and, moreover, treats this as normal. Distractors are common in multiple choice testing formats when only one answer is correct, and the other options are distractors. Therefore, in a typical math word problem, extraneous information may be communicated in a text that the learner must disregard to answer correctly. In this use of a visual there is a deliberate effort to distract, which reflects a reasoning process that is valued in the field of constructing multiple choice standardized testing. For a learner not prepared to deal with visuals meant to mislead and the need to discern this fact and thereby ignore that visual data, this can be confounding to say the least. In essence, this poses a type of hidden cultural bias that is part of a wider field of cultural production and use of multimodality that is dangerous for a learner prepared to display math knowledge and unprepared to deal with this "tricky" visual.

How many of these types of items cause even second-language learners who are "good" at math to be evaluated as "weak" at math? Some would argue that test-taking skills might be needed to prepare learners for the expectations in these cultural traps. While this may help, many test-taking

skill lessons focus on vocabulary skill building (Robison 2010), largely ignoring cultural expectations regarding the relation between the visual and text modalities. In other words, they presume multimodal interpretation is shared in standardized testing. Others argue for removing these types of items as they are construct irrelevant, meaning they elicit responses that are not relevant to the underlying knowledge/performance that the item is attempting to test. Since all items will have cultural bias to some degree, it would be important to monitor which visuals and text combinations present the most difficulty to second-language learners of diverse linguistic and cultural backgrounds.

Another issue occurs when the text accompanying the visual is too limited to indicate the focus of the tested item. Ostensibly this strategy in test item construction is to lessen the linguistic load of the math problem for the second-language learner, an effort that has many advocates (Abedi et al. 2001). Yet, this tact also presents problems when viewed from the perspective of questioning how this type of multimodal representation affects the learner's ability to interpret what is expected. Many have critiqued the cultural assumptions that standardized test creators fail to consider (Emihovich 1994; Trumbull and Solano-Flores 2011b) but few have identified the multimodal nature in test items that has potential to cause confusion and thereby weakens the reliability of the questioning format's ability to assess the test taker's abilities to use mathematical thinking. Note the figures in the following item (Fig. 9.2) and the sparsity of procedural text, an interpersonal feature that could be improved.

The issues raised in this item (Fig. 9.2) include procedural language that is absent. In addition, the logic assumes that a second-language learner knows this, as this knowledge is required to answer the item. However, the test maker relies on the learner interpreting the figures without this guidance. While many test reviewers do attempt to examine tests for bias in responses to items and in terms of specific language used, the level of appropriateness of the language for guiding the intermodal (verbal–visual) interpretation of the figure, thus the test item, needs to be also scrutinized. In the instructions for test administrators, this intermodality link is not made explicit. In fact, scrutiny of such items often escapes bias reviews as visuals are taken to be explicit cues. Scrutiny of the relation between the ideational affordances maybe a case of taken for grantedness in assuming all logic is culture free (O'Connor 2006). Hence, again the conditions are not clear for gaining an accurate measure of the second-language learners response to the mathematical construct being tested.

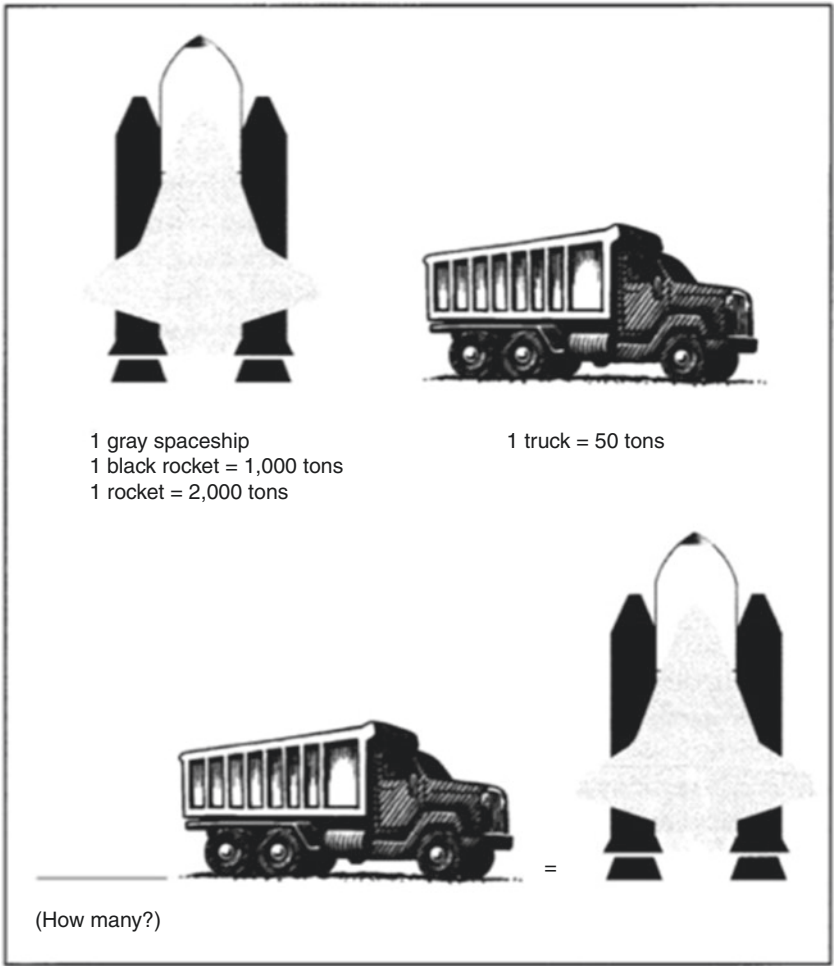

Fig. 9.2 Sample of a formative assessment (Mihai 2010)

Some test designers would argue that assumptions about shared interpretation have to be made. Others would argue that problematic assumptions need to be identified in a dynamic fashion by providing students support to help complete the task. This would allow testers a way to determine which prompts help test takers display their knowledge.

Thus by examining the expectations of the test items using the ideational, interpersonal, and textual meanings that are assumed, these elements of cultural bias are made visible. Such an analysis can be applied to other aspects of standardized tests to uncover problematic validity issues in the test domains, items, and response formats.

Conclusion

Discourses concerning the "new basics" underlie and reinforce global standards that are implemented through standardized testing in mathematics and reading such as the Programme for International Student Assessment which is a triennial international survey that takes place across national borders conducted under the auspices of the Organization for Economic Co-operation and Development. Concurrently, the spread of English as a language of wider communication during this time of globalization continues to fuel the demands for testing regimes such as the Test of English as a Foreign Language (TOEFL), Graduate Record Examination (GRE), S.A.T™, and so on. The reading of both discourses of new basics as well as the actual spread of culturally influenced multimodal tests create unusually demanding material conditions that are particularly problematic for underserved students and those who seek to educate and assess them in culturally sustainable ways. Paris and Alim (2014) define culturally sustaining pedagogies' goal as one that "seeks to perpetuate and foster – *to sustain* – linguistic, literate, and cultural pluralism as part of the democratic project of schooling and as a needed response to demographic and social change" (p. 88). Culturally sustaining pedagogies ask us to not simply value our current communities' assets, rather also to be concerned about assessing the critical skills, knowledges, and ways of being needed for success in evolving meaningful participation and contribution both in the present and future. Past research has alerted us to the major role language plays in testing (Abedi and Lord 2001; Garcia et al. 2010; Fairbairn and Fox 2009) but attention to developing learners critical multimodal understandings of tests must be developed as well.

This chapter highlights the institutionalized practices of standardized testing that not only neglect culturally and linguistically diverse learners but also jeopardize their access to higher education and to becoming productive contributors in their communities. Throughout the examples shown in these guidelines for proctors as well as in the item design, there are assumptions about multimodality being understood and produced in

the context of testing, which highlights culturally situated performances required to demonstrate certain knowledge. While the particular items may have been addressed by the time this chapter is published, most likely the multimodal issues will not have been adequately addressed. I have shown how the use of multimodalities in math test items requires a much closer scrutiny of the test makers' cultural assumptions about logical correspondences between modalities. Considering this, if the ideal reader (proctor) must be so well prepared to administer the test, how much more must be done to prepare the EAL learner to sit for these types of items? Because the test constructs are represented multimodally, test takers who are second-language learners need to be better prepared for these expected performances to interpret and produce their responses using multimodality. Furthermore, if standardized testing will always need to examine this aspect, why not disrupt this orientation to envision other means to capture the student's processes for answering an item rather than just right or wrong responses. Bavali et al. (2011) support dynamic assessment by gathering this type of information as it would be more productive in assessing both the item and the child's zone of proximal development. In essence this accumulated information would be useful to

> modify learners' performance level in order to enable the mediators not only to understand individual learners' current level of abilities but to predict (assess) their unassisted potential future abilities based on their present performance in assisted (instructed) completion task settings. (p. 896)

Inevitably such information would reveal the test's cultural assumptions and the learners' level of knowledge used to respond to the item.

Given the spread of high-stakes testing, from the very initial test development stage in the testing world these items merit piloting to provide test makers of a better understanding of how these multimodalities affect second-language learners' meaning making in specific subject areas. In a globalized world, why not begin with linguistic and cultural diversity as a starting point for each context of testing? How do the text and visuals impact students display of what they know and how well they can use this knowledge? For example, in one study on the use of diagrams in quadratic functions, how learners created their diagrams provided additional information to researchers about the processes evidencing their mathematical thinking (Lobato et al. 2014). Under what contexts are particular concepts represented better in multimodal forms over written texts? With the

increasing presence of computerized testing, these questions will become even more significant. Under what contexts are there concepts that require less written text and more detail in the graphics in the prompts in order to assess second-language learners?

In classrooms, teachers can play an important role in ascertaining and developing second-language learners production and comprehension of multimodal texts, even beyond performances in the genres of testing. If teachers can take up the perspective of assessment as "inquiry" into the modalities needed (language and literacy) in different text types, second-language learners can be socialized into ways of critically interpreting and producing texts. The "tricky" parts of tests can be unpacked and scrutinized as a social practice. Using multimodal analysis as part of the pedagogy helps students understand forms of representing knowledge and even better how to make use of these to create knowledge through using multimodal communication skills, collaboration, problem solving, and creative thinking.

Furthermore, researchers collaborating with teachers can help document these learner engagements in activities that learners production and comprehension of multimodal texts are evidenced. Consequently, these activities can then be used for generating situated evidence that captures second-language learners development and progress in much more valid and reliable ways than bubble-in assessments. The performances of all those involved, the test designers, the testing proctors, teachers, and the students together produce the results we end up with. To hold only teachers and students accountable would be ethically misguided in an era of globalization.

References

Abedi, J., Hofstetter, C., Baker, E., & Lord, C. (2001). *NAEP mathematics performance and test accommodations: Interactions with student language background*. Technical Report No. 536. Retrieved from National Center for Research on Evaluation Standards, & Student Testing. Website http://www.cse.ucla.edu/products/Reports/TR536.pdf.

Abedi, J. (2002). Standardized achievement tests and English Language Learners: Psychometrics issues. *Educational Assessment*, *8*(3), 231–257.

Abedi, J., Courtney, M., Leon, S., Kao, J., & Azzam, T. (2006). *English Language Learners and math achievement: A study of opportunity to learn and language accommodation*. Technical Report No. 702. Los Angeles, CA: CRESST/UCLA. Retrieved from ERIC. Website http://eric.ed.gov/?id=ED495848.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*(3), 219–234.

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education (AERA, APA, & NCME). (1999). *Standards for educational and psychological testing*. (2nd ed.). Washington, DC: American Education Research Association (AERA).

Austin, T., & Ites, E. (2012). Testing/assessing literacies: How do educators build useful knowledge about second language development? *Languages and Linguistics*, 29(30), 181–210.

Bavali, M., Yamini, M., & Sadighi, F. (2011). Dynamic assessment in perspective: Demarcating dynamic and non-dynamic boundaries. *Journal of Language Teaching and Research*, *2*(4), 895–902. doi:10.4304/jltr.2.4.895-902.

Emihovich, C. (1994). The language of testing: An ethnographic-sociolinguistic perspective on standardized tests. In K. Holland, D. Bloome, & J. Solsken (Eds.), *Alternative perspectives in assessing children's language and literacy* (pp. 33–54). Norwood, NJ: Ablex.

Fairbairn, S. B., & Fox, J. (2009). Inclusive achievement testing for linguistically and culturally diverse test takers: Essential considerations for test developers and decision makers. *National Council on Measurement in Education. Educational Measurement: Issues and Practice*, *28*(1), 10–24. doi:10.1111/j.1745-3992.2009.01133.x.

Garcia, E., Lawton, K., & Diniz De Figuereido, E. H. (2010). *Assessment of young English language learners in Arizona: Questioning the validity of the state measures of English proficiency*. Los Angeles, CA: The Civil Rights Project, University of California.

Guilherme, M. (2007). English as a global language and education for cosmopolitan citizenship. *Language and Intercultural Communication*, *7*(1), 72–90.

Hemphill, F. C., & Vanneman A. (2011). *Achievement gap: How Hispanic and White students in public schools perform in mathematics and reading on the national assessment of educational progress*. National Center for Educational Statistics. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/studies/2011459.pdf website.

Kress, G., & Van Leeuwen T. (1998). Front pages: (The critical) analysis of newspaper layout. In A. Bell & P. Garrett (Eds.), *Approaches to media discourse* (pp. 186–219). Oxford, UK: Blackwell.

Kress, G., & Van Leeuwen T. (2006). *Reading images: The grammar of visual design*. (2nd ed.). London: Routledge.

Lobato, J., Hohensee C., & Diamond J. M. (2014). What can we learn by comparing students' diagram-construction processes with the mathematical conceptions inferred from their explanations with completed diagrams? *Mathematics Education Research Journal*, *26*(3), 607–634.

Mahoney, K., Haladyna, T., & MacSwan, J. (2009). The need for multiple measures in reclassification decisions: A validity study of the Stanford English Language Proficiency Test. In T. G. Wiley, J. S. Lee, & R. Rumberger (Eds.), *The education of language minority immigrants in the United States* (pp. 240–262). Bristol, UK: Multilingual Matters.

Mihai, F.M. (2010). *Assessing English language learners in the content areas: A research into practice guide for educator*. Ann Arbor, MI: The University of Michigan Press.

O'Connor, M. C. (2006). The implicit discourse genres of standardized testing: What verbal analogy items require of test takers. In J. Cook Gumperz (Ed.), *The social construction of literacy* (2nd ed.., pp. 264–287). Cambridge, UK: Cambridge University Press.

Paris, J., & Alim, H. S. (2014). What are we seeking to sustain through culturally sustaining pedagogy? A loving critique forward. *Harvard Educational Review*, *84*(1), 85–100.

Partnership for Assessment of Readiness For College and Careers (PARCC). (2014). PARCC Accessibility Features and Accommodations Manual. Appendix J: PARCC Mathematics Audio Guidelines Version 3.0. Retrieved from http://www.jcboe.org/boe2015/images/PARCCCCSS/PARCC%20appendix-j-math-audio.pdf.

Robison, S. (2010). Teaching math to English learners-myths and methods. In E. N. Whelan Ariza (Ed.), *Not for ESOL teachers. What every classroom teacher needs to know about the linguistically, culturally, and ethnically diverse student* (pp. 127–139). Boston, MA: Pearson, Allyn & Bacon.

Royce, T. (2007). Multimodal communicative competence in second language contexts. In T. Royce & W. Bowcher (Eds.), *New directions in the analysis of multimodal discourse* (pp. 361–390). Mahwah, NJ: Lawrence Erlbaum Associates.

Schafer Willner, L., Rivera, C., & Acosta, B. (2008). *Descriptive study of state assessment policies for accommodating English language learners*. Oregon Department of Education. Retrieved from website www.ode.state.or.us/wma/teachlearn/commoncore/descriptivestudyell.pdf.

Shohamy, E. 2014). *The power of tests: A critical perspective on the uses of language tests*. ([Kindle version]. Retrieved from Amazon.com.

Short, D. J., & Fitzsimmons, S. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners – A report to Carnegie Corporation of New York*. Retrieved from Carnegie Corporation of New York website: https://www.carnegie.org/media/filer_public/bd/d8/bdd80ac7-fb48-4b97-b082-df8c49320acb/ccny_report_2007_double.pdf.

Trumbull, E., & Solano-Flores, G. (2011a). The role of language in assessment. In M. R. Basterra, E. Trumbull, & G. Solano-Flores (Eds.), *Cultural validity in assessment: Addressing linguistic and cultural diversity* (pp. 22–45). New York: Routledge.

Trumbull, E., & Solano-Flores, G. (2011b). Addressing the language demands of mathematics assessments: Using a language framework and field-based research findings. In M. R. Basterra, E. Trumbull, & G. Solano-Flores (Eds.), *Cultural validity in assessment: Addressing linguistic and cultural diversity* (pp. 218–253). New York: Routledge.

Theresa Austin As a professor of language, literacy, and culture at the University of Massachusetts College of Education: Department of Teacher Education and Curriculum Studies, Theresa teaches courses that explore the impact that becoming multilingual has on individuals and their community, teachers, and institution administrators. This exploration leads to curricular planning as well as policies that are designed to be more culturally nurturing and sustaining across race, class, and gender. In particular, as a critical ethnographer of L2 language and literacy contexts, Theresa collaborates with teachers, L2 learners, and administrators to examine how assessments can better function to increase learning. Theresa also volunteers in numerous local, national, and international projects to increase culturally and linguistically diverse population's access to meaningful education that furthers realization of their aspirations. Her publications can be found in journals such as *Modern Language Journal, Languages and Linguistics, Journal of Latinos and Education* and in numerous book chapters.