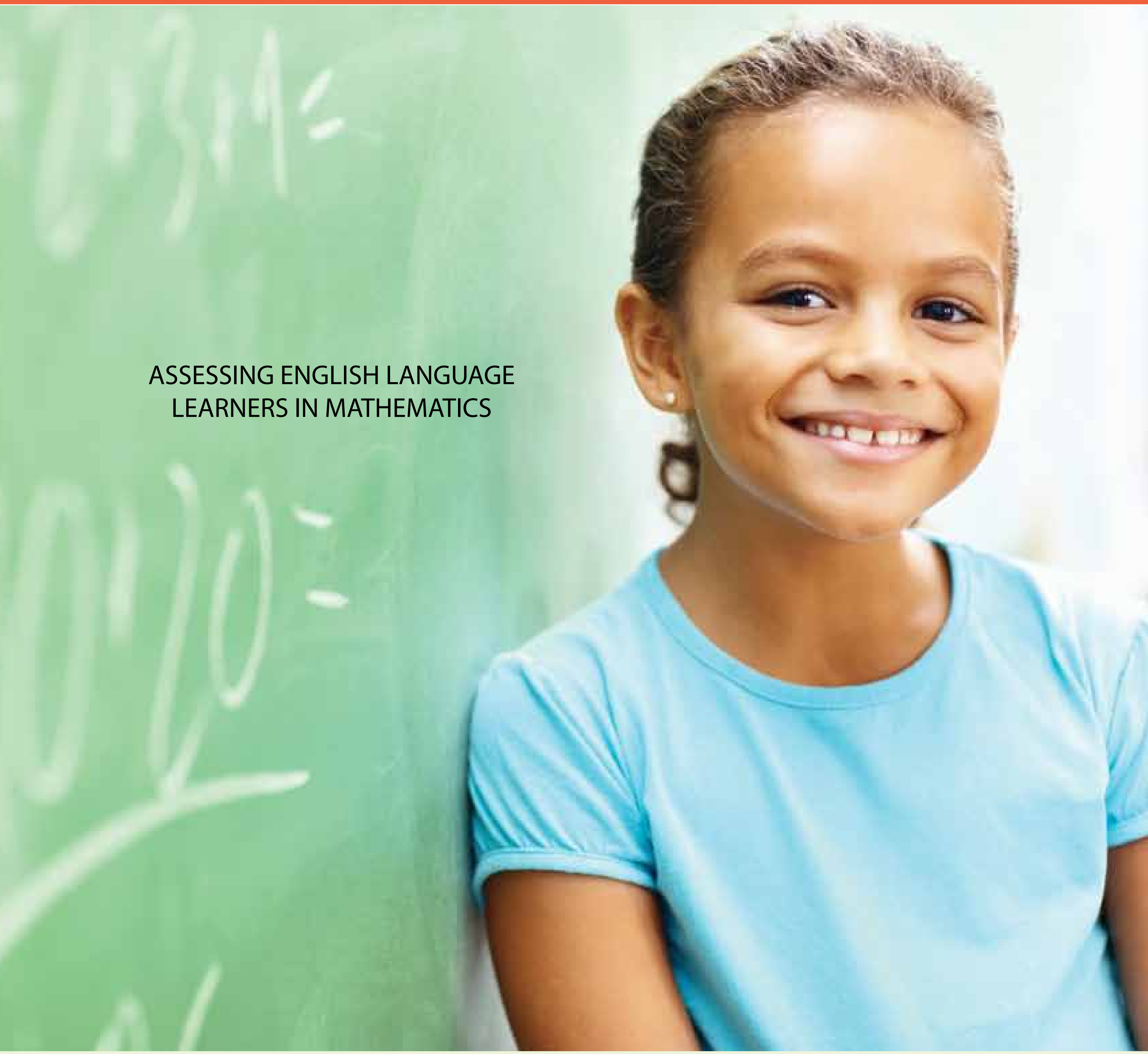
TODOS

VOLUME 2 • MONOGRAPH 2: LINGUISTIC COMPLEXITY IN MATHEMATICS ASSESSMENTS



ASSESSING ENGLISH LANGUAGE LEARNERS IN MATHEMATICS

NATIONAL EDUCATION ASSOCIATION

The National Education Association is the nation's largest professional employee organization, representing 3.2 million elementary and secondary teachers, higher education faculty, education support professionals, school administrators, retired educators, and students preparing to become teachers.

NEA Executive Officers

NEA Executive Staff

TODOS: MATHEMATICS FOR ALL

TODOS Board

TODOS Mission

The mission of *TODOS: Mathematics for ALL* is to advocate for an equitable and high quality mathematics education for all students, in particular Hispanic/Latino students, by increasing the equity awareness of educators and their ability to foster students' proficiency in rigorous and coherent mathematics.

Goals

- To improve educators' knowledge and ability to implement an equitable, rigorous, and coherent mathematics program that incorporates the role that language and culture play in learning mathematics.
- To develop and support leaders who continue to carry out the mission of TODOS.
- To promote the generation and dissemination of knowledge about equitable and high quality mathematics education.
- To inform the public and influence educational policies in ways that enable students to become mathematically proficient.

Published April 2010

Foreword

Edward Silver, Co-Editor • University of Michigan

This is the second research monograph of *TODOS: Mathematics for All*. In the fall of 2006, the TODOS Research and Publications Committee decided to initiate publication of a series of research-based monographs focused on issues of diversity and equity in mathematics education. The Committee decides on a theme for each monograph, issues a call for submissions, reviews the submitted manuscripts, and selects those that appear in each monograph; the final publication and dissemination of the first two monographs have been supported by the National Educational Association (NEA). The first monograph focused broadly on issues related to enhancing the achievement and learning of Hispanic-Latino/a students in mathematics, and this one focuses on mathematics assessment issues, with special attention to those that are pertinent to Hispanic-Latino/a students.

Central to the mission of *TODOS: Mathematics for All* is a belief that the mathematics achievement of Hispanic-Latino/a students in U.S. schools merits continued focused attention from both the educational research and educational practice communities. The papers appearing in the first monograph provoked us to move beyond a conventional view of underachievement in relation to rigid categories of race, class, and language by offering novel theoretical, conceptual, and historical analyses of key issues associated with the mathematics achievement and learning of Hispanic-Latino/a students Papers in that volume also offered some instructional samples to illustrate alternative visions of what might be possible in mathematics classrooms. The papers in this second monograph take another step in the journey toward understanding mathematics teaching and learning in ways that can promote progress for all students, especially Hispanic-Latino/a students.

Why focus on assessment? It is difficult to think of a topic more important to the mathematics education community, nor one that is less well understood, than assessment. In recent years, NCLB-related testing, and the associated pressures on schools and teachers to make *adequate yearly progress* (AYP), have brought externally mandated assessment to the forefront as a factor influencing—some might argue determining—what is taught in the mathematics classroom and how it is taught. The dual requirement that testing performance data be disaggregated by race/ethnicity categories and that AYP be attained within each category has had a complex interaction with the teaching and learning of Hispanic-Latino/a students. On the one hand, the disaggregation requirement has made it less likely that the concerns of these students and their advocates are relegated to the margins as educational decisions are made. On the other hand, the reliance on a single test score, typically derived from an English language assessment under uniform administration conditions with little or no room for accommodation to students who are learners of English as a second language, has created what many see as an onerous burden on many students, teachers, and schools. Although there has long been theoretical and practical attention to the ways in which variations in linguistic content and task format might foster or inhibit the demonstration of mathematical knowledge and proficiency by Hispanic-Latino/a students, the importance of these issues has been intensified under the glare of the NCLB spotlight.

Several of the papers in this monograph offer insights into the subtle, and not-so-subtle, ways in which the language used in mathematics assessment tasks influences the comprehension and performance of Hispanic/Latino/a students in ways that have more to do with linguistic proficiency than with mathematical proficiency. In somewhat different ways, the papers by Martiniello, Lager, and Fernandes et al. all probe the interplay between mathematics assessment items and students who attempt to solve them, uncovering nuances related to language and meaning that challenge simplistic interpretations of test performance that are all too frequent.

Educators, policy makers, politicians, and news media representatives typically take at face value the results of annual tests of mathematics achievement; that is, the tests are seen as providing objective, scientific evidence about the overall achievement of students. They also are often seen as providing convincing evidence of differential attainment of mathematical proficiency by demographic subgroups, especially when the findings conform to the predispositions and expectations of those who seek to use the findings to recommend changes in curriculum, teaching, or educational policies. Unfortunately, the perceptions of consumers of achievement test results often rest on simplistic assumptions about the relationship among various interactive constituent parts of the assessment situation, including the specific mathematical concepts and skills being assessed; the relation between the tested knowledge and the opportunities to learn provided by the curriculum materials and teaching methods employed in particular classrooms and schools; and the extent to which the wording, format, and time allocation of test items influences performance of students with varying degrees of language proficiency. Several of the papers in this volume present analyses of the linguistic entailments of assessment tasks that challenge these simplistic assumptions and reveal nuanced insights and fascinating glimpses into what is (or is not) made available to students in test items and also *how* it is (or is not) made available.

In addition to identifying non-mathematical features that make tasks more complex for Hispanic-Latino/a students, the papers by Martiniello and Fernandes et al. also suggest ways in which one might gain insights into the thinking of students through one-on-one interviews and "think aloud" sessions. These techniques reside at the boundary between testing and instruction. As such, they move us in the direction of formative assessment, which seeks to produce insights that might guide instructional decisions and which is the main focus of the fourth paper in this volume by Kitchen et al.

According to the NCTM Assessment Principle, *"Assessment should support the learning of important mathematics and furnish useful information to both teachers and students"* (NCTM, 2000, p. 22). In fact, it is widely acknowledged that classroom assessment practices are an essential component of quality instruction. Indeed, on the basis of an extensive review of research evidence, Black and William (1998) argued that teachers' ongoing assessments combined with appropriate feedback to students can have large positive effects on student learning—effects that are larger than those associated with many other educational interventions. As Kitchen et al. suggest and illustrate in their paper, high quality formative assessment is a multifaceted concept that includes using a variety of formal and informal assessment approaches and tools: assessing understanding and reasoning as well as factual knowledge; engaging students in assessment of their own work and that of their peers; and, to the extent possible, maintaining an alignment with standards and externally mandated tests (while keeping these tests from becoming the sole or main focus of instruction). If teachers integrate high quality assessment into their daily instructional practice and use the information they obtain to inform instructional decisions and provide feedback and guidance to students, they can derive beneficial consequences both for themselves and for their Hispanic-Latino/a students.

Acknowledgements

Edward Silver and Richard Kitchen, Co-Editors

We would like to thank the many people who contributed to this *TODOS* monograph. First, we want to thank the scholars who submitted papers for possible publication in this second research monograph of *TODOS: Mathematics for All*. Launching a new publication series is never easy, and the success of this venture depends in large measure on the willingness of scholars to contribute their work in this new forum. We hope that others will decide to join in a collective effort to make this monograph series a success by contributing papers to successive volumes of the monograph series.

Second, we would like to thank the other members of the Research and Publications Committee: Marta Civil, Gil Cuevas, Rochelle Gutiérrez, and Carl Lager. Each member of the committee gave selflessly of her/his time, reviewing multiple papers, and providing valuable feedback to the scholars who submitted papers. We also would like to thank the following individuals who served as ad hoc reviewers for the second monograph: Cynthia Anhalt, Laura Burr, Joyce Fischer, Alfinio Flores, Virginia Horak, Mary Marshall, Edgar Romero, Ksenija Simic-Muller, Erin Turner, Barbara Trujillo, and Maura Varley.

Finally, we want to thank Nora Ramirez, *TODOS* president, Miriam Leiva, TODOS past president, and Andrea Prejean, senior policy analyst for the National Educational Association (NEA). Nora has been very supportive of the Research and Publications Committee and the continued success of the TODOS monograph. Miriam was instrumental in launching the TODOS monograph series and continues to be committed to its development as a much needed, high quality research-based publication focused on issues of diversity and equity in mathematics education. As with the inaugural monograph, Andrea and NEA helped with much needed financial resources during difficult economic times to publish and disseminate the second monograph. The continued support of NEA brings instant visibility to the monograph and TODOS, and we very much appreciate the Association's ongoing support and confidence in TODOS. We hope that practitioners and scholars will find the content of this second *TODOS* monograph to be of great benefit in their work to advocate for and advance the mathematics education for not just Hispanic-Latino/a students, but for all students.

Table of Contents

Linguistic Complexity in Mathematics Assessments and the Performance of English-Language Learners

Maria Martiniello • Educational Testing Service

INTRODUCTION

Spanish-speaking students constitute about 80 percent of English-language learners (ELLs) in U.S. public schools (Kindler, 2002). ELLs are the most rapidly growing student group in the country. They also are among the lowest scoring on national and regional assessments of reading and mathematics. Ensuring valid measurements of mathematical skills for these students is a pressing issue in high-stakes educational assessment today. As mandated by the *No Child Left Behind Act*, states must report the achievement of ELLs in their educational accountability systems using assessments that are both reliable and valid (U.S. Department of Education, 2002).

A concern when assessing the mathematics knowledge of ELLs using tests in English is that items with excessive linguistic complexity will reflect their lack of English proficiency rather than their mathematical skills, ultimately making the test not valid and fair for this population of students (August and Hakuta, 1997; National Research Council, 2000, 2002). Experts argue for the need to separate language skills of ELLs from subject area knowledge (Abedi and Lord, 2001; Abedi, Lord, and Plummer, 1997). But they recognize the difficulty in doing so because all assessments administered in English also are measures of English proficiency (AERA-APA-NCME, 1985; August and Hakuta, 1997; National Research Council, 2000).

In an effort to disentangle the influences of language skills and mathematical proficiency, I conducted a validity study of a state fourth-grade mathematics test. This research integrated three sources of evidence: a) analysis of the language in the mathematics items by linguists, language and literacy experts, and mathematics curriculum specialists; b) cognitive interviews with Spanish-speaking Latino ELLs responding to these items; and c) large-scale psychometric analyses of mathematics assessments using differential item functioning (DIF) methods. Used to determine whether test items are fair for assessing the knowledge of students from various groups, DIF identifies the differential performance of an item for members of two groups who have equivalent levels of proficiency on the construct the test is intended to measure (Dorans and Holland, 1993). In this paper, DIF refers to differences in item difficulty for ELLs and non-ELLs with equal mathematics test scores.

The research was guided by the following questions:

1. What are the syntactic and lexical features of items showing large differences in item difficulty for ELLs and non-ELLs of equal mathematics proficiency, i.e., items with large DIF?
2. What comprehension challenges do such features pose to Spanish-speaking Latino ELLs in think-aloud interviews?

To address these questions, expert reviews and textual analysis were conducted examining non-mathematical linguistic complexity features in a fourth-grade mathematics test. Think-aloud interviews with fourth-grade Spanish-speaking ELLs were conducted to investigate the reading comprehension challenges they encountered while solving these items. DIF procedures were implemented to identify those items posing greater difficulty for ELLs than for non-ELLs of comparable mathematics proficiency on the test. Based on the DIF analysis, items showing large DIF favoring non-ELLs over ELLs were selected to undergo thorough review. Textual analysis and think-aloud data confirm that such items exhibit both syntactic and lexical complexity features that challenge ELLs' text comprehension. These items have complex multi-clausal sentences with long noun phrases, unfamiliar vocabulary, and polysemous words.

Research Background and Context

Successful mathematical word problem solving requires competently decoding each of the various semiotic modalities that make up the language of mathematics. First, students must comprehend the item's verbal "language in general" or natural language. The more complex the text is, the more difficult it will be to process, increasing reading time and/or leading to misinterpretations of the problem and, in turn, to incorrect solutions (Mestre, 1988). In addition, students need to know the domain-specific terminology of mathematics, that is, both the specialized vocabulary (i.e., *triangle, coordinates*) and syntactic structures (*greater than*, 10 apples weigh *the same as* 2 melons) that are typical of the mathematical register (Halliday, 1974; Lemke, 2003; Mestre, 1988). Finally, students also need to interpret non-linguistic mathematical symbols and their particular syntax to decode mathematical meaning, such as in equations, as well as make sense of visual displays, diagrams, graphs, and figures.

By definition, ELLs have not yet acquired sufficient mastery of the English language to perform in regular classrooms. Consequently, excessive linguistic complexity is expected to compromise their understanding of mathematical word problems more than non-ELLs' understanding. Research examining the relationship between complexity of natural language in mathematical word problems and DIF for ELLs does in fact confirm this hypothesis. Non-mathematical linguistic complexity functions as a source of *construct-irrelevant*[1] difficulty for students who are not proficient in English, making linguistically complex items more difficult for ELLs than for non-ELLs matched according to their mathematics proficiency. Employing Item Response Theory DIF detection methods, Martiniello (2006a, 2006b, 2007) estimated differences in item difficulty for ELLs and non-ELLs in a fourth-grade mathematics test and showed that it could be predicted from a composite measure of the items' syntactic and lexical complexity. More linguistically complex items in the test tended to show greater DIF favoring non-ELLs over ELLs. In other words, ELLs tended to have a lower probability of answering linguistically complex items correctly than non-ELLs with equivalent mathematics proficiency. At the highest end of the linguistic complexity range, items contained complex grammatical structures that were central for comprehending the item, along with mostly low-frequency, non-mathematical lexical terms whose meaning was central for comprehending the item and could not be derived from context. Conversely, linguistically simple items tended to have lower difficulty

[1] Construct-irrelevant linguistic complexity pertains to the natural language in mathematical word problems and not to their math-related terminology, which is specific to the construct the mathematics test intends to measure.

for ELLs than for non-ELLs with equal mathematics proficiency. These items contained mostly high frequency non-mathematical lexical terms and simple grammatical structures. Some relatively simple items included a "less familiar" term, but its meaning was derivable from context (Martiniello, 2007).

Martiniello (2007) found that the effect of linguistic complexity on DIF is attenuated when items provide certain non-linguistic visual and symbolic representations that help ELLs make meaning of the text. These are schematic rather than pictorial representations. They embody mathematical relationships, either spatial relationships among objects or patterns, or numerical/mathematical relationships through mathematical symbols or algebraic expressions.

Building upon these previous findings, the current study presents a detailed examination of the linguistic features of the items showing large DIF disfavoring ELLs and triangulates this information with think-aloud responses to the mathematics items from Spanish-speaking ELLs.

Methods

The linguistic complexity of the mathematics items was analyzed, and think-aloud protocols were used to gather evidence of comprehension difficulties for Spanish-speaking ELLs. In order to identify those items that posed greater difficulty to ELLs than to non-ELLs with comparable mathematics proficiency, two DIF detection methods were employed. Items flagged for evidence of large DIF disfavoring ELLs were examined; for such items, the degree of DIF, the learning strand they represent, their linguistic complexity, and children's responses to them in the think-aloud interviews are described.

Instrument

The test studied is the English version of the fourth-grade Massachusetts Comprehensive Assessment System (MCAS) mathematics test administered statewide in 2003. This is a standards-based achievement test aligned with the Massachusetts Mathematics Curriculum Framework. It includes five major learning strands: a) number sense and operations; b) patterns, relations, and algebra; c) geometry; d) measurement; and e) data analysis, statistics, and probabilities (Massachusetts Department of Education, 2000, 2003a). The items analyzed included a total of 39 publicly released items, 29 items of multiple choice format and 10 of constructed response (5 of short answer format and 5 open-ended).

MEASURES

Linguistic Complexity

Measures of linguistic complexity were derived from a detailed micro-analysis of the text's syntactic complexity, as well as expert ratings of the items' overall syntactic complexity and overall lexical complexity.

Textual analysis. A coding system was developed to identify elements of complexity in the structural relationships between words, phrases, and sentences in the items (for instance, number of clauses, noun phrases, verbs, and verb phrases); the syntactic function or type of all elements; and the syntactic order of clauses. Two linguists were trained to use this micro-analytic coding manual with items from a different version of the test. The first linguist coded all 39 items, while the second linguist coded 20 percent of the items independently and reviewed the first rater's original coding for the remaining 80 percent. The inter-rater agreement in the micro-analytic coding adjusted for chance agreement was high (Cohen's Kappa coefficient = 0.89). Discrepancies were discussed and when necessary items were recoded.

Vocabulary frequency / familiarity. To assess whether the item vocabulary was likely to be known by the majority of fourth-grade students, words were cross-checked with two word-frequency lists, *A List of 3,000 Words Known by Students in Grade 4* (Chall and Dale, 1995) and the *Living Word Vocabulary* list (LWV) (Dale and O'Rourke, 1979). LWV is a national vocabulary inventory that provides familiarity scores on 44,000 written words (tested in grades 4, 6, 8, 10, 12, 13 and 16) (Dale and O'Rourke, 1979). *A List of 3,000 Words Known by Students in Grade 4*, a subset of the LWV comprising 3,000 words commonly known by 80 percent of fourth graders, is used in the calculation of readability formulas (Chall and Dale, 1995). These word-frequency lists employ samples of students that include ELLs, but are predominantly composed of non-ELLs. Thus, this study used these lists in conjunction with expert judgment of lexical complexity by linguists with expertise on ELLs.

Think-aloud interviews. Think-aloud protocols of the mathematics test were administered to 24 fourth-grade ELLs attending six inner-city Massachusetts public schools. Two of these schools offered dual immersion programs (Spanish and English). The students interviewed were first or second-generation Latin-American immigrants from Colombia, El Salvador, Mexico, Guatemala, Peru, and the Dominican Republic, as well as some students from Puerto Rico. They had between two and four years of schooling in the U.S., came from homes where Spanish was the primary language, and were identified by their teachers as ELLs. The sample was gender balanced. Based on their teachers' ratings, the children interviewed represented a wide range of mathematics proficiency and a relatively wide range of English proficiency. Children who could not read or communicate in English were not interviewed since they would not be able to read the test at all.

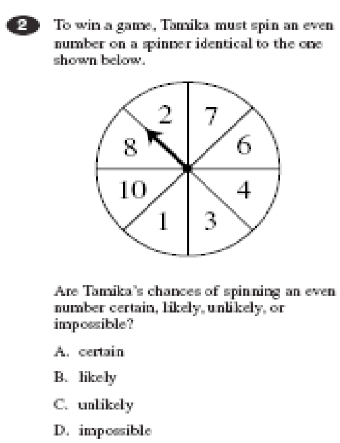
Think-aloud interviews were conducted individually in either one or two sessions, which lasted between 30 and 60 minutes each. They were audiotaped and later transcribed. Interviews were conducted primarily in Spanish, although the children were encouraged to speak in the language they felt more comfortable. The more fluent children often switched between languages during the interviews. The test items were given to the ELLs in English and not in Spanish. Items from the mathematics test were presented to students, who read aloud the item text in English. Attention was paid to the decoding errors, i.e., words students stumbled over, could not pronounce correctly, skipped, or paused at. Children explained what the item was asking them to do. Probe questions assessed whether children could a) understand the text in English; b) rephrase the text in Spanish or English to demonstrate their understanding of its meaning; c) identify which aspects (if any) of the English text they could not understand; d) figure out what the item was requiring them to do, even if they could not understand the text in its entirety; and e) make meaning of the item by relying on non-linguistic visual and symbolic representations. The types of mistakes children made when interpreting linguistically complex text in English were recorded as well as whether they could answer the question correctly based on their understanding of the text.

DIF Detection: Selection of Items for In-Depth Analysis

DIF refers to the differential performance of an item for ELLs and non-ELLs who are matched according to their mathematics proficiency. The proficiency level is most often the total score on the test under study or a purified score of the test (purged of some items suspected of DIF). In this study, a subset of items showing differential difficulty for ELLs and non-ELLs with equivalent mathematics proficiency was identified using the standardization (Dorans and Kulick, 1986), Mantel-Haenszel (Holland and Thayer, 1988), and the item response theory likelihood-ratio tests (Thissen, Steinberg, and Wainer, 1993) methods. These procedures were implemented in two stages using a purification step that purged the matching test score from items flagged for large DIF in the first stage. The sample used for DIF detection was a sub-sample of all the fourth-grade students in Massachusetts who took the test in the spring of 2003. In total, the DIF sample included 68,839 students, 3,179 of whom were ELLs and the rest were non-ELLs. The DIF indices were categorized according to classification guidelines employed by the Educational Testing Service (Dorans and Holland, 1993) and by Zenisky, Hambleton, and Robin (2003).

RESULTS

Two items identified as showing large DIF favoring non-ELLs over ELLs were examined to gauge the possible contribution of non-mathematical linguistic complexity to their unusual difficulty for ELLs. Each item is described in terms of DIF and learning strand in the Massachusetts Mathematics Curriculum Framework, followed by textual analysis of its syntactic and lexical features. In addition, transcripts from think-aloud interviews with fourth-grade Spanish-speaking ELLs responding to the mathematics items are presented to illustrate comprehension errors that these ELLs make when interpreting complex text in these DIF items.



2 To win a game, Tamika must spin an even number on a spinner identical to the one shown below.

Are Tamika's chances of spinning an even number certain, likely, unlikely, or impossible?

A. certain

B. likely

C. unlikely

D. impossible

Item 2. Massachusetts Department of Education, Mathematics, Grade 4 (MCAS, 2003b)

Differential item functioning. One way to represent DIF is to plot for each group the probability of answering the item correctly as a function of mathematical proficiency. For an item with negligible DIF, the two curves will be similar. For an item showing DIF, the curves will be different. As shown in Figure 1, at nearly all levels of mathematical proficiency, a considerably larger proportion of non-ELLs than ELLs is expected to answer item 2 correctly. The odds of answering the item correctly were nearly twice as high for non-ELLs than for ELLs with the same test scores.

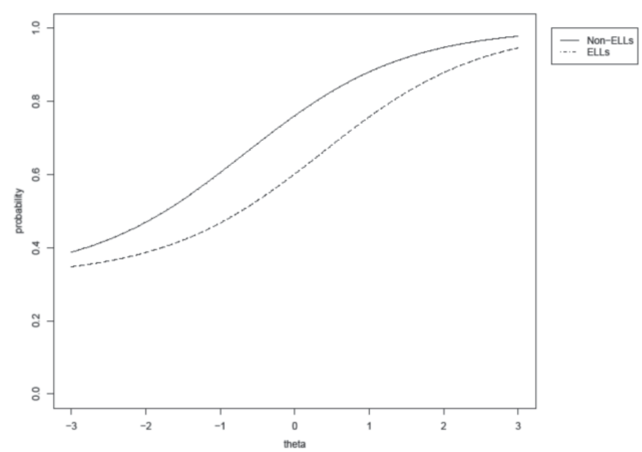


Figure 1. Item Response Theory (IRT) item characteristic curve - Item 2
Source: Martiniello (2007).

Learning strand. Item 2 measures the learning strand *data analysis, statistics, and probabilities*. Responding to this item correctly indicates that students "understand and apply basic concepts of probability," specifically, that they know how to "classify outcomes as certain, likely, unlikely, or impossible by designing and conducting experiments using concrete objects," in this case a spinner (MDOE, 2000, p. 44). Assuming familiarity with spinners and the mathematical concept *even number*, the item requires students to discern that five of the eight numbers on the spinner are even and therefore it is likely that Tamika will spin an even number.

Linguistic complexity. The item stem consists of two long multi-clausal sentences. The first sentence starts with a preposed adverbial clause that includes a non-personal form of a verb *(to win)*. The main clause starts with an uncommon proper noun *(Tamika)* that functions as subject. The verbal phrase displays a complex verb that includes the modal verb must, a noun phrase as a direct object *(an even number)* and a highly complex prepositional phrase with an embedded adjectival clause that includes the past-participle *shown*. The second part of the item stem is a question. This question contains a complex noun phrase with a possessive construction and a prepositional phrase that includes a non-personal form of a verb (progressive *spinning*). The following words are not part of the 3,000-word list known by 80 percent of fourth graders: *even, spinner, identical, likely,* and *unlikely* (Chall and Dale, 1995).

Think aloud/children interviews. Think-aloud interviews revealed that this item was difficult to understand for fourth-grade Spanish-speaking ELLs. Below are transcriptions from two children whose lack of text comprehension led to incorrect solutions to the problem. Here they are explaining their understanding of the English text in the item.

Child 1:

Tamika hizo un juego de éstos	Tamika did one of these games
y tiene que caer en esto…	And it has to fall in this…
[pointing to the spinner]	
¿Cuáles posibilidades hay de que caiga en uno?	What is the likelihood that it would fall in one [the number one slot]?
Maybe puede caer, maybe no puede caer	Maybe it can fall, maybe not
'Likely', es possible	'Likely' means is possible
'Impossible' es que no va a caer	'Impossible' is that it will not fall
'Certain' es que va a caer	'Certain' is that it will fall
'Likely' es posible, tal vez va a caer	'Likely' it is possible, maybe it will fall
'Unlikely' tal vez no va a caer	'Unlikely', maybe it will not fall
[Child pauses to reason his response to the item]	
***Es' unlikely'**, tal vez no va a caer.*	It is 'unlikely', maybe it will not fall

Interv:

¿Cómo supiste que "tenía que caer en uno?"	How did you know that "it had to fall in one"

Child 1:

Porque aquí dice	Because it says so here

Interv:

¿Cómo supiste que "tenía que caer?"	How did you know that "it had to fall?"
[Child points to the spinner]	

Interv:

*¿Cómo supiste que "tenía que caer en **uno**?"* [emphasis]	How did you know that "it had to fall in one? [emphasis]

Child 1:

Uno	One
[Child points to the word **one** in the item stem]	

Of all the content words in the item, this child understood the words *game*, *number*, and *one*, and identified *Tamika* as a proper noun. He did not know what spin, spinner or spinning meant; but he was able to recognize the picture of the spinner. He figured out that Tamika played with the spinner and he must evaluate her chances of getting a particular number.

In the clause *identical to the one shown below*, the word one is used as a pronoun to refer to the spinner displayed on the page. The child interprets one as referring to the number 1 on the spinner. Recognizing the word one amongst many unknown words, he deduced that the spinner's arrow must fall in the number one slot. He failed to recognize the syntactical function of the word one, used as a pronoun in this sentence, and instead misinterpreted it as the numeral one. Based on this linguistic misinterpretation, he offered a reasonable answer It is *unlikely, maybe it will not fall.*

From his interpretation of the item distracters, it is evident that this child was familiar with the English words *certain*, *likely*, *unlikely*, and *impossible* used to classify the mathematical likelihood of the event *spinning an even* *number*. However, he was not able to identify correctly the event the item referred to, possibly due to the large proportion of unknown words in the sentence and his difficulty understanding the syntactic boundaries of the clauses and the syntactic function of the word one.

Despite having a greater vocabulary, the next child made a similar mistake to the previous one.

Child 2:

Para ganar el juego, Talia...Ta..ka..	To win the game, Talia. . .Ta..ka. . .
no puedo decir ese nombre. . .	I cannot say that name. . .
Tamika necesita hacer así alrededor	Tamika needs to do this around
[spinning gesture]	
para tener el número del 'spinner' 'identical' con el número uno 'shownbelow', igual al 1.	to have the number of the spinner 'identical' to the number one 'shown Below', equal to 1
¿Cuáles son las chances, como la posibilidad que ella tiene ese número?	What are 'the chances', the possibilities that she will get that number?
'certain, likely unlikely, impossible'.	'certain, likely unlikely, impossible'
ella necesita ganar con el ***número 1***	She needs to win with the number 1
Sólo hay un uno [1]	there is only one number 1
'Certain' es seguro que sí,	'certain' is for sure she will
'Likely', maybe ya casi	'likely', maybe almost yes
'Impossible' nunca va a tener	'impossible', she will never get it
Es 'unlikely' *que ella va a tener ese número*	It's 'unlikely' that she will get this number
[pointing to the number 1 (one) in the spinner]	
porque ***sólo hay uno de 1***	because there is only one 1

This was the only student interviewed who knew the word *identical* as meaning *equal*. However, she interpreted the phrase *identical* to *the one shown below* as modifying the noun *number* rather than the noun *spinner*. She resolved the syntactic ambiguity of *identical to the one* [spinner] *shown below as identical to the* [number] *one* [1] *shown below*. Upon completing the item, the student was asked what *even number* meant.

Child 2:

un número con pareja, 2, 4, 6. . .	a number with couple/pair, 2, 4, 6 ...

Instructed to re-read the item one clause at a time, as to scaffold the parsing of the syntactic boundaries, she provided a new answer:

Child 2:

Oh! Es 'likely', porque hay 2, 3, 4, 5; cinco 'even numbers'.	Oh! It is 'likely', because there are 2,3,4,5; five 'even numbers'
[child counts numbers on the spinner]	

In this particular item, the layout of the text does not facilitate the delimitation of the syntactic boundaries for children who may not be fluent readers. The first part of the item stem is divided in three lines as shown below.

Line 1 To win a game, Tamika must spin an even
Line 2 number on a spinner identical to the one
Line 3 shown below

In the transcript above, the child did not perceive the word *even* (end of line 1) as modifying the noun *number* (beginning of line 2), even though she knows the meaning of the phrase *even number*. Likewise, the word *one* (end of line 2) is perceived as separated from *shown below* (line 3). Syntactically, the clause *one shown below* could be rewritten as *one that is shown below*, or *the spinner that is shown below*. The pronoun *one*, referring to spinner, is the subject of the clause shown below. It is possible that the layout favored a perception of the word one as a numeral standing on its own and separated from the next line, instead of a pronoun associated with the past participial clause *shown below* that is in the next line.

In both examples above, the children understood the concept of probability of an outcome using spinners, knew the mathematical meaning of the English words *likely* and *unlikely*, and could correctly classify the likelihood of a particular event occurring. Nonetheless, they could not answer the item correctly because they were unable to understand the text providing them with the information about the event to be classified.

Another important challenge for the ELLs interviewed was their lack of familiarity with the item's vocabulary, in particular some of the distracters. One hundred percent of the children knew the meaning of *impossible*, a Spanish-English cognate that is a high-frequency word in Spanish. In contrast, few children understood the words *identical* (4%) and *certain* (33%); these are also Spanish-English cognates, but, unlike *impossible,* are infrequently encountered in conversation because more colloquial synonyms are available, i.e., *igual*, *equal* and *seguro*, *sure*. About half of the children either ignored or confused the meanings of the words *likely* and *unlikely*, as shown in the transcripts below.

Interv:

¿Qué crees que te están preguntando aquí?	What do you think they are asking you here?

[pointing to the distracters certain, likely, unlikely, impossible]

Child 3:

Si hay de verdad la chance	If there is truly a chance [referring to 'certain']
un chin[2]	a bit [referring to 'likely']
mucho	a lot [referring to 'unlikely']
o imposible, no hay chance	or impossible, there is no chance

While the child above exchanged the meanings of the English words likely and unlikely the next one switched the meanings of certain and likely.

Child 4:

Para jugar, Tamia . . . tiene que. . .	To play, Tamia,. . .has to. . .
el 'spinner'	the 'spinner'
tiene que coger	she has to get
un número que sea 'even'	an 'even' number
que es multiplicado por 2,	that is multiplied by 2,
como 2 por 2 es 4,	like 2 by 2 is 4,
4 es un 'even number'	4 is an 'even number'
entonces tiene que ser el chance, 'certain'?	then, will the chance be certain?
que es como puede pasar pero no pasa seguro,	which is like it can happen but it won't happen for sure
'likely', que puede pasar seguramente	'likely', that can happen for sure
'unlikely', puede pasar pero no estoy tan segura	'unlikely', it can happen but I am not that sure
imposible que nunca va a pasar	impossible, that it will never happen.

Item 8 and Comparison Item 30

Item 8 has the second-largest DIF index disfavoring ELLs in the test. For comparison purposes, item 8 is discussed along with item 30, which measures the same curriculum content.

8 Every Saturday in the fall, Martin has to do 1 inside chore and 1 outside chore. The chores are listed below.

Inside Chores	Outside Chores
vacuum	rake
wash dishes	weed
dust	

How many different combinations of 1 inside chore and 1 outside chore can Martin make?

A. 3

B. 5

C. 6

D. 9

Item 8. Massachusetts Department of Education, Mathematics, Grade 4 (MCAS, 2003b)

[2] Caribbean expression for *un poquito.*

Question 30 is a short-answer question. Write your answer to this question in the box provided in your Student Answer Booklet. Do not write your answer in this test booklet. You may do your figuring in the test booklet.

30 Mr. Garcia gave each of his students a notepad, a pencil, and a ruler on the first day of school. The chart below shows the different colors of the notepads, pencils, and rulers.

Colors of Notepads, Pencils, and Rulers

Notepads	Pencils	Rulers
Yellow	Red	Green
White	Blue	Purple
	Orange	

How many different combinations of 1 notepad, 1 pencil, and 1 ruler can Mr. Garcia make?

Item 30. Massachusetts Department of Education, Mathematics, Grade 4(MCAS, 2003b)

Differential item functioning. Compared with item 30, item 8 is much more difficult for ELLs than for non-ELLs with equivalent mathematics scores. The odds of responding to item 8 correctly for non-ELLs are close to double the odds for ELLs; while for item 30, the odds for non-ELLs are about 1.3 times those for ELLs. Figures 2 and 3 illustrate the discrepancy in the difficulty gaps across groups in these two items. Since both items measure students' ability to count combinations, the discrepancy between their DIF indices cannot be attributed to differences in proficiency in this skill.

Learning strand. Like item 2, items 8 and 30 assess the learning strand *data analysis*, *statistics*, and *probabilities*. They require children to apply basic concepts of probability by counting the number of possible combinations of objects from two/three sets (Massachusetts Department of Education, 2000).

Linguistic complexity. A close examination of the items' linguistic complexity shows that both items' syntactic structures are relatively complex. These two items display particularly lengthy sentences. Item 8 has three sentences with an average of 12 words per sentence. Item 30 also has three sentences with an average of 16 words per sentence. Specifically, the lengthy questions might pose a challenge for ELLs. Experts' reviews of the items' linguistic features ranked these items similarly in their syntactic complexity but rated item 8 as lexically more complex than item 30. In their ratings of lexical complexity, the experts considered the familiarity of vocabulary in the item. The vocabulary in item 8 consists primarily of words related to the home environment (*chores, wash dishes, vacuum*). These English words would tend to be more familiar for non-ELLs than for ELLs. In contrast, the vocabulary in item 30 is mostly related to the school environment (*students, pencil, ruler, colors*). ELLs are likely to have been exposed to these words in their first years of schooling in English. Thus, the school-related words in item 30 are more likely to be familiar for ELLs than the home-related words in item 8.

All the words in item 30 are part of the 3,000-word list commonly known by fourth graders (Chall and Dale, 1995). Two words in item 8 do not appear in this list, *chores* and *vacuum*. However, the word chores appears as known by 67 percent of fourth graders tested nationwide in the LWV list (Dale and O'Rourke, 1979).

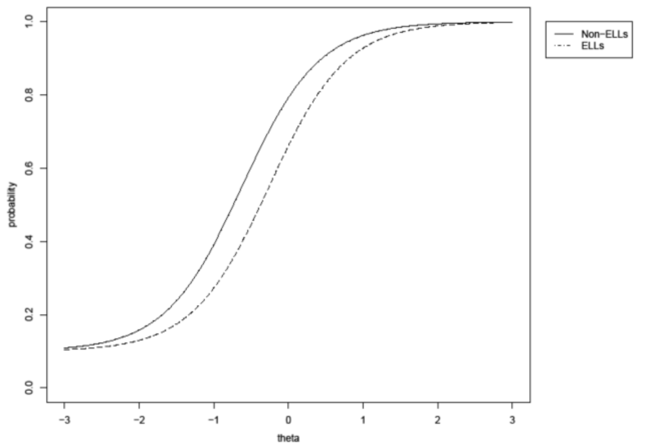


Figure 2. IRT item characteristic curve - Item 8
Source: Martiniello (2007).

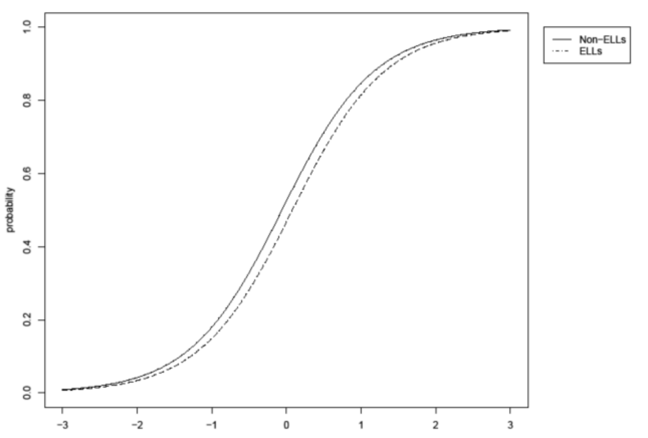


Figure 3. IRT item characteristic curve - Item 30
Source: Martiniello (2007).

Think aloud/ child interviews. In item 30, all the ELLs interviewed understood the meaning of the words *students, notepad, pencil, ruler, day, school,* and *colors*, while only a few had trouble with the past tense of the verb 'to give' (gave) and the adjective *below*. In contrast, most of the ELLs were unacquainted with the word *chores*, the phrases *inside chore* and *outside chore* and with the different words listed in the table in item 8. Only two girls knew what *vacuum*, *wash dishes*, and *dust* meant. None of the ELLs knew the meaning of the words *rake* and *weed*. About 88 percent of the ELLs interviewed were familiar with the mathematical meaning of the word *combination*, a Spanish-English cognate, and understood that they were supposed to create and count combinations of things. However, in item 8, they did not know what to combine, in contrast with item 30. For these ELLs, the proportion of unknown words was larger in item 8 than in item 30, making the reading comprehension of the former more difficult than the latter.

The ELLs interviewed showed a clear understanding of school-related words, whereas words related to home posed bigger challenges. The differential familiarity of ELLs and non-ELLs with the lexical terms in item 8 may explain why this item functions so differently across these groups in contrast to item 30, even though they are measuring the same curricular content.

DISCUSSION

Through textual analyses and children's responses to think-aloud protocols, this study sought to illustrate some of the linguistic characteristics of mathematical word problems that pose disproportionate difficulty for ELLs compared to non-ELLs with equal mathematics proficiency.

Linguistic Features of DIF Items Disfavoring ELLs

Findings indicate that items showing large DIF disfavoring ELLs share some of the following characteristics hindering reading comprehension:

- Item sentences have multi-clausal complex structures with embedded adverbial and relative clauses. They also tend to have long phrases with embedded noun and prepositional phrases. There is reduced syntactic transparency in their text, i.e., lack of clear relationships between the syntactic units. This point is related to the previous two. Embedded clauses along with long noun phrases between syntactic boundaries obscure the syntactic parsing of the sentence. During think-aloud responses to item 2, this lack of transparency prevented some ELLs from interpreting the syntactic boundaries in the text correctly, resulting in a distorted interpretation of the string of words in the item.

- Regarding vocabulary, these items tend to have words that are unfamiliar to most fourth-graders according to the LWV list. Other words more commonly known by English speakers proved to be quite challenging for the ELLs interviewed, e.g., chores, certain. The role of vocabulary knowledge as predictor of reading comprehension for both ELLs and non-ELLs is well established. Since, by definition, ELLs do not have the breadth of English vocabulary that fully English proficient students do, more words will be unknown to them. Sentences in some of these large DIF items have too many English words unknown to ELLs, thus making it very difficult to infer their meaning from context. These results suggest that ELLs might have difficulties with everyday English words usually learned at home (chores, wash dishes) rather than with words they are likely to learn in school (student, pencil, rulers). Most likely, ELLs will know those home-based words in the language they speak at home with their parents. Also, polysemous words, those with multiple meanings, pose additional challenges to reading comprehension. In some DIF items, ELLs could not figure out the syntactic function of a polysemous word in the sentence, misinterpreting the word's appropriate meaning in context, e.g. confusing one as pronoun with one as numeral in item 2.
- Regarding test or text layout, lack of one-to-one correspondence between the syntactic boundaries of clauses and the layout of the text in the printed test may challenge ELLs with poor reading fluency, particularly in complex sentences with multiple clauses and long noun phrases. For instance, in item 2, the visual separation between the words even and number, one and shown below may have hindered ELLs' identification of the grammatical structures. This is consistent with research on the relationship between visual-syntactic text formatting and reading comprehension (Walker, Schloss, Fletcher, Vogel, and Walker, 2005).
- Regarding the role of cognates, the think-aloud interviews suggest that Spanish-speaking ELLs do indeed take advantage of their knowledge of Spanish-English cognates for interpreting the item's meaning. However, it seems that this only applies to high-frequency Spanish words that are thus likely to be familiar to children. In these cases, children were able to recognize the morphological relationship between the English word in the test and the Spanish word they know (eg., impossible-imposible). Other English-Spanish cognates that are less familiar in Spanish and that have more familiar substitutes were difficult for these children (e.g., certain-cierto).

Alternative Sources of Item DIF: Curricular Learning Strands

Linguistic complexity is not the only plausible source of DIF disfavoring ELLs in the items analyzed. Previous studies have identified important association patterns among learning strand, linguistic complexity, and DIF for ELLs in a fourth-grade mathematics test. Compared with the other learning strands in the test, data analysis, statistics, and probabilities items tended to be unusually more difficult for ELLs than for non-ELLs with equivalent mathematics scores (Martiniello, 2006c, 2007). In this study, all the items identified as showing large DIF disfavoring ELLs measure data analysis, statistics, and probabilities. A possible explanation for the differences in difficulty for ELLs and non-ELLs in these items may be the differential teaching and learning of data analysis, statistics, and probabilities subject matter across groups. It is possible that the ELLs tested had been less exposed to the curricular content of this learning strand than were their non-ELL counterparts or, having had the opportunity to learn it, did not master and acquire this knowledge like non-ELLs did for a variety of reasons about which we can only speculate. For example, their low English proficiency may have prevented them from understanding/learning the content during classroom instruction, or they may have had teachers who were less prepared to teach this content. Research shows that teachers of ELL students have less mathematical training and lower certification rates (Gándara and Rumberger, 2003; Mosqueda, 2007).

However, another explanation for the unusual difficulty of this learning strand may lie in its relationship to linguistic complexity. Items measuring data analysis, statistics, and probabilities tend to have text with greater syntactic and semantic complexity than items measuring the rest of the learning strands. Martiniello (2006c) found a significant and positive correlation between this learning strand and a composite measure of linguistic complexity in a fourth-grade mathematics test ($r = 0.455$, $p = 0.007$).

Further studies are needed to examine the interaction between learning strands and linguistic complexity as sources of DIF disfavoring ELLs. There may be important differences among the items measuring data analysis, statistics, and probabilities related to their linguistic complexity and not to subject matter as illustrated by our comparison of items 8 and 30. Since both items measure exactly the same content, differences in the groups' knowledge of the curriculum content/learning strand is not that helpful in explaining their strikingly differing DIF statistics. Item 8's disproportionate difficulty for ELLs relative to item 30 is likely to be related to its unfamiliar vocabulary. Children who speak a language other than English at home are less likely to be exposed to the word chores (and types of chores: rake, weed, dust) than English speakers, who may hear these word when interacting with their parents regarding their household tasks. In contrast, the English words pencils, notebooks, students, and colors are unarguably among the first English words ELLs will learn in the classroom setting. The differential exposure of ELLs and non-ELLs to the lexical terms in item 8 may explain why this item functions so differently across groups in contrast to item 30, even though they are measuring the same curricular content.

What Is Our Inference Based on Item Scores?

Given the impact of linguistic complexity and curriculum content on the differential item performance of ELLs and non-ELL's of equivalent mathematics scores, how should we interpret incorrect answers by ELLs in these DIF items? Ideally, an item should be answered incorrectly only if the student has not mastered the curriculum content measured by the item. However, the evidence from think-aloud interviews shows that our inference based on scores can be distorted by ELLs' unfamiliarity with the English vocabulary and difficulty in parsing complex syntactic structures in the item.

In item 2, a score of zero would mean that students do not know how to classify events as certain, likely, unlikely, and impossible. However, the think-aloud interviews revealed that some ELLs who knew the mathematical meaning of the English words likely and unlikely, and could correctly classify the likelihood of a particular event, could not answer the item correctly (likely) because they were unable to understand the complex clauses providing them with the information about the event to be classified.

Many of the children interviewed did not know one or more of the English words certain, likely, unlikely, and impossible. Based on the Massachusetts mathematics curriculum framework, one could argue that these English words are in fact the mathematical terms the item intends to measure, and that failing to respond correctly to this item for not knowing these words would appropriately reflect children's inability to classify outcomes as certain, likely, or unlikely. However, think-aloud responses indicate that some of these ELLs had a mathematical understanding of a gradient of likelihood for a given event. They could express it in Spanish but had not yet mastered the English vocabulary to label it correctly. Furthermore, some of these children had actually learned to classify the likelihood of events using more familiar English words than those on the item. For instance, some children construed a likelihood continuum ranging from always to never in which certain corresponds to it will always happen, impossible to it will never happen, likely to it will almost always happen, and unlikely to it will almost never happen.

Since the words certain, likely, and unlikely have common meanings in the mathematics classroom and in everyday/social language, it is reasonable to think that children who are fully proficient in English have a greater chance to infer the correct mathematical meanings of these words than ELLs due to greater exposure. For instance, the word certain is listed as a word known by 80 percent of English-speaking fourth graders; but in this group just a few children knew it. This differential familiarity with the item's vocabulary may be contributing to item 2's large DIF. Based on these findings, we may expect the differential performance of ELLs and non-ELLs in this item would decrease if the distracters included more familiar English words. Unfortunately, the results of simplification studies comparing the performance of ELLs in complex and simplified versions of mathematical word problems have been mixed (Abedi et al., 2000; Francis, Rivera, Lesaux, Kieffer, and Rivera, 2006; Moreno, 2006b). More studies using think-aloud protocols are needed to investigate how ELLs understand mathematical word problems of varying linguistic complexity.

IMPLICATIONS

This study has implications for test validation, test development, and mathematics instruction for ELLs. The language of mathematical word problems in large-scale assessments should be carefully scrutinized and their differential functioning for ELLs studied before including them in operational tests. DIF studies routinely done for gender and race groups should now be extended to ELLs. Cognitive lab research could be conducted to learn how students interpret the text of mathematical word problems, since experts' reviews do not always anticipate the actual comprehension challenges children encounter when reading the items.

Item construction for achievement tests often relies on content area specialists who are instructed in long-standing principles of item writing: write clearly, concisely, and avoid irrelevant information (Baranowski, 2006). However, these item writers may not be aware of the specific consequences of excessive linguistic complexity on the performance of ELLs. Further training of item writers and professional test editors is needed to address this void. For instance, item writers could be provided with item performance indicators like DIF and interview transcripts from cognitive lab studies. Abedi (2006) recommends providing them with hands-on exercises in linguistic modification, that is, "simplifying or modifying the language of a text while keeping the content the same" (p. 384).

It is critical that the language simplification is not achieved at the expense of altering the construct/skill to be measured by the item/test. Proponents of linguistic modification do not advocate for language-free mathematics tests for ELLs. Mathematical discourse in classrooms and textbooks combines natural and academic language, mathematical terms, symbols, and graphs. Mathematics assessments should do the same, particularly those designed to assess "mathematics for understanding." Mathematical word problems, like those on the MCAS test, attempt to measure mathematical understanding by providing scenarios and novel situations in which students apply their prior knowledge and establish relationships between mathematical concepts and ideas. We want to know if ELLs can do that. In contrast, a test consisting of computation problems with little or no language interference would provide quite a narrow picture of the mathematical knowledge of ELLs.

Language skills are important for understanding and solving mathematical problems used not only in large-scale assessments, but also in classroom assessments and instruction. However, the important role of language in understanding and communicating mathematical ideas is not generally acknowledged by teachers of ELLs in their teaching practice. Research on teachers' perceptions has found some contradictions in the way teachers view the role of language in the mathematics teaching and assessment of Latino ELLs. Some teachers tend to conceive mathematics instruction for ELLs as relatively free from language. Nonetheless, there are great language demands in the mathematics assessments these teachers use in their classrooms (Bunch, Aguirre, Telléz, Gutiérrez, and Wilson, 2007).

For ELLs, the teaching of mathematics can no longer be conceived as separate from the teaching and learning of language. Teachers of ELLs must provide sustained linguistic scaffolding for ELLs while encouraging the process of mathematical meaning-making (Anhalt, Civil, Horak, Kitchen, and Kondek, 2007). Addressing the educational needs of ELLs requires mathematics teachers to attend to both content instruction and language development support.

ACKNOWLEDGEMENTS

This research was conducted while the author was a member of the Harvard Graduate School of Education and was generously supported by the Spencer Foundation Doctoral Dissertation Fellowship. Any opinions expressed in this article are those of the author and not necessarily of Educational Testing Service. Transcriptions of the think-aloud interviews and analysis were adapted with permission from Maria Martiniello, "Language and the Performance of English-Language Learners in Math Word Problems," Harvard Educational Review, volume 78:2 (Summer 2008), pp. 333-368. For more information, please visit **www.harvardeducationalreview.org** or call 617-495-3432.

REFERENCES

Abedi, J. (2006). Language issues in item development. In S.M. Downing and T.M. Haladyna (eds.), *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Abedi, J., C. Lord, and J. Plummer. (1997). *Final report of language background as a variable in NAEPS mathematics performance* (CSE Technical Report 429). Los Angeles, CA: UCLA Center for the Study of Evaluation / National Center for Research on Evaluation, Standards, and Student Testing.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (AERA-APA-NCME). (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Anhalt, C., M. Civil, V. Horak, R. Kitchen, and L. Kondek. (2007, January). *Preparing teachers to work with English language learning students: Issues of research and practice.* Presentation at the annual conference of the Association of Mathematics Teacher Educators, Irvine, CA.

August, D.E., and K.E.Hakuta. (1997). *Improving schooling for language-minority children: A research agenda.* Washington, DC: National Academy Press.

Baranowski, R. (2006). Item editing and editorial review. In S.M. Downing and T.M. Haladyna (eds.), *Handbook of test development.* Mahwah, NJ: Lawrence Erlbaum Associates.

Bunch, G., J. Aguirre, K. Telléz, R. Gutiérrez, and J. Wilson. (2007, April). *Preservice elementary teachers' views about language and culture in mathematics teaching and learning for Latinos/as.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Chall, J., and E. Dale. (1995). *Readability revisited: The new Dale-Chall readability formula.* Cambridge, MA: Brookline Books.

Dale, E., and J. O'Rourke. (1979). *The living word vocabulary: A national vocabulary inventory study.* Boston: Houghton Mifflin.

Dorans, N.J., and P. Holland. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P.W. Holland and H. Wainer (eds.), *Differential Item Functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dorans, N.J., and E. Kulick. (1986). *Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test.* Journal of Educational Measurement, 23(4), 355-368.

Francis, D., M. Rivera, N. Lesaux, M. Kieffer, and H. Rivera. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for the use of accommodations in large-scale assessments.* Houston: Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston Center on Instruction.

Gándara, P., and R.W. Rumberger. (2003). *The inequitable treatment of English learners in California's public schools.* zMinority Research Institute.

Halliday, M.A.K. (1974). Some aspects of sociolinguistics. In E. Jacobsen (ed.), Interactions between linguistics and mathematical education: Final report of the symposium by UNESCO, CEDO and ICMI, Nairobi, Kenya, September 1–11, 1974 (pp. 64-73). Paris: UNESCO.

Holland, P.W., and D.T. Thayer. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H.I. Braun (eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kindler, A.L. (2002). *Survey of the states' limited English proficient students and available educational programs and services: 2000–2001.* Washington, DC: Office of English Language Acquisition, Language Enhancement and Academic Achievement for Limited English Proficient Students, U.S. Department of Education.

Martiniello, M. (2006a, April). *Linguistic Complexity in Math Tests for Latino English-Language Learners.* Paper presented at the American Educational Research Association, Hispanic Research Issues, San Francisco, 2006.

Martiniello, M. (2006b, April) *Linguistic Complexity and Item-anchored DIF Detection for English Learners in the MCAS Mathematics Test.* Paper presented at the National Council on Measurement in Education, Differential Item Functioning in Special Populations–Paper Session, San Francisco, April 2006.

Martiniello, M. (2006c, October). *Sources of Differential Item Functioning for English learners in word math problems.* Paper presented at the 37th Annual Meeting of the Northeastern Educational Research Association, New York, NY.

Martiniello, M. (2007). *Linguistic complexity and differential item functioning (DIF) for English language learners (ELL) in math word problems.* Dissertation Abstracts International Section A: Humanities and Social Sciences, Vol 68(6-A), 2007. pp. 2422. Publisher: US: ProQuest Information & Learning. ISSN: 0419-4209 (Print)

Massachusetts Department of Education. (2000). *Mathematics curriculum frameworks.* Retrieved September 4, 2004, from www.doe.mass.edu

Massachusetts Department of Education. (2003a). *2002 MCAS technical report.* Retrieved September 9, 2004, from www.doe.mass.edu

Massachusetts Department of Education. (2003b). *Release of Spring 2003 Test Items.* Retrieved September 4, 2004, from http://www.doe.mass.edu/mcas/2003/release/testitems.pdf

Mestre, J. (1988). The role of language comprehension in mathematics and problem solving. In R. Cocking and J. Mestre (eds.), *Linguistic and Cultural Influences on Learning Mathematics* (pp. 201-220). Hillsdale, NJ: Lawrence Erlbaum Associates.

Moreno, R., C. Allred, B. Finch, and M. Pirritano. (2006, April). *Linguistic simplification of math word problems: Does language load affect English language learners' performance, metacognition, and anxiety?* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Mosqueda, E. (2007, April). *English proficiency, academic tracking and the school context: Understanding the mathematics achievement of Latino native and non-native English speakers.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

National Research Council. (2000). *Testing English-language learners in U.S. schools: Report and workshop summary* (K. Hakuta and A. Beatty, eds.). Washington, DC: National Academy Press.

National Research Council. (2002). *Reporting test results for students with disabilities and English-language learners: Summary of a workshop* (J. Anderson-Koenig, ed.). Washington, DC: National Academy Press.

Thissen, D., L. Steinberg, and H. Wainer. (1993). *Detection of Differential Item Functioning using the parameters of item response models.* In P.W. Holland and H. Wainer (eds.), Differential Item Functioning (pp. 67-114). Hillsdale, NJ: Lawrence Erlbaum Associates.

U.S. Department of Education. (2002). *The Elementary and Secondary Education Act Reauthorization (The No Child Left Behind Act).* Retrieved October 10, 2004, from http://www.ed.gov/legislation/ESEA02/

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress. (2007). *NAEP data explorer.* Retrieved December 5, 2007, from http://nces.ed.gov/nationsreportcard/nde/criteria.asp

Walker, S., P. Schloss, C.R. Fletcher, C.A. Vogel, and R.C. Walker. (2005, May). *Visual-syntactic text formatting: A new method to enhance online reading.* Retrieved February 3, 2007, from http://www.readingonline.org/articles/art_index.asp?HREF=r_walker/index.html

Zenisky, A., R.K. Hambleton, and F. Robin. (2003). Detection of Differential Item Functioning in large-scale state assessments: A study evaluating a two-stage approach. *Educational and Psychological Measurement*, 63(1), 51-64

Conserving the Mathematics Construct: Improving Mathematics Item Accessibility for English-Language Learners

Carl A. Lager • University of California, Santa Barbara

In classrooms across the country, hundreds of thousands of mathematics assessment items are annually being created or administered to millions of K–12 English-language learners (ELLs). Whether these items are small-scale or large-scale, multiple choice or constructed response, open, open-middled, or closed (Romagnano, 2006), or formative or summative, they are always couched in language. All items have a level of cognitive demand, most are aligned to a mathematics standard or objective, many are embedded in context, and some include a visual support. Because of the difficulty identifying, understanding, and adequately managing the complex interactions among these item components for all students, but especially ELLs, items can easily become flawed. Daro, Stancavage, Ortega, DeStefano, and Linn (2007) define a flawed item as one that possesses one or more of the following problems:

1. Assumptions that are hidden are unfair to the student.
2. Context is confusing and misleading.
3. Language and/or graphics present unnecessary obstacles to understanding the task.
4. Mathematical errors exist within the task, its response set, or scoring rubric.

The National Mathematics Advisory Panel (NMAP, 2008) investigated this issue and concluded that flawed items on National Assessment of Educational Progress (NAEP) and state large-scale assessment tests introduce non-mathematical sources of score variance, such as test takers not understanding the item due to the item's unnecessarily long, complex sentence structure, that affect the meaning and accuracy of the scores. Therefore, a problem solver who generates an incorrect answer to an item that was not fully understood because of its wording or format, but would have correctly solved otherwise, would be demonstrating a worse than expected performance based on actual mathematical competence. Because national and state large-scale assessment items typically are vetted more thoroughly for flaws than local assessment items used in school districts, schools, or classrooms, flawed mathematics items likely exist throughout K–12 education at all levels.

ELLs stand to be disproportionately affected negatively by Flaws 1–3 compared to non-ELLs because ELLs are developing English language proficiency and familiarizing themselves with local and national cultural norms, experiences, and contexts that same-grade non-ELLs are already familiar with (Abedi, Lord, Boscardin, and

Miyoshi, 2000; Rivera and Collum, 2006; Emick and Kopriva, 2007). Some of these meaning-making challenges can come under the umbrella of the mathematical register (Halliday, 1974)—meanings belonging to natural language used in mathematics (Cuevas, 1984). Spanos, Rhodes, Dale, and Crandall (1988) have given specific examples of the register's syntactic, semantic, and pragmatic aspects and others (e.g., Lager, 2006) have added on.

Fillmore and Valadez (1986) argue that to understand mathematics items, connecting these separate language pieces into coherent text is a must. Moschkovich (2000) notes the complexity of that process, explaining that an ELL has to read an item in English first, then translate it into a primary language, then translate that translation into a primary language mathematics register, solve the problem, then either translate the answer back into common English and then into the formal English mathematics register or just go directly back into the English mathematics register. In fact, Khisty (1995) has said that "attention must be given to clarifying confusions caused by both the Spanish and English mathematics register, and to making connections between ways of expressing concepts in both languages" (p. 125). Therefore, flawed items likely contribute to ELLs' consistently documented overall lesser mathematics performance on large-scale mathematics assessments when compared to non-ELLs (U.S. GAO, 2006; NAEP, 2008).

To remedy this systemic problem, the NMAP (2008) recommended that:

> *…test developers be especially sensitive to the presence of these types of flaws in the test development process. To further ameliorate concerns, significant attention should be devoted to the actual design of individual mathematics items and to the evaluation of items for inclusion in an assessment. Careful attention must be paid to exactly what mathematical knowledge is being assessed by a particular item and the extent to which the item is, in fact, focused on that mathematics (p. 60).*

Some guidelines already exist to help test developers better meet ELL needs (e.g., Kopriva, 2008, Solano-Flores, 2003; Abedi and Lord, 2001). However, based on my experiences as a middle and high school mathematics teacher of ELLs, a state's mathematics assessment specialist, a mathematics teacher educator, and a mathematics education researcher, appropriately and consistently applying such guidelines is a scholarship of practice that is not well understood (Kopriva, 2008) or highly valued by many of my mathematics education colleagues. A revision that corrects one fatal flaw often changes a different aspect of the item, such as its mathematical construct or cognitive demand. Therefore, key item components must be identified, considered, and applied together as an integrated whole, in theory and practice.

To raise awareness of this widespread flawed item problem in mathematics assessment and address it for Spanish-speaking ELLs, the Conserving the Mathematics Construct (CMC) theory and framework will be introduced and then applied to the analysis and revision/reconstruction of three mathematics items. The elementary item (Garbage Truck) was developed by a state department of education, the middle school item (Restaurant) was developed by a nationally recognized specialist in the design of assessments for ELLs in pre-K–12 settings (World-Class Instructional Design and Assessment [WIDA], 2007), and the high school item (Medicine) was promoted by the National Council of Teachers of Mathematics (NCTM) (Travis and Collins, 2005). These three items were chosen to show common flaws (and their fixes) across grade spans, content strands, and well-intentioned, but undertrained item development teams/experts. Though no individual mathematics teacher examples are included, current and former classroom mathematics teachers usually are part of the item development and/or vetting processes. Nonetheless, the CMC can be applied to the development and revision of mathematics assessment items at various levels, such as a classroom of 30 students or a state of 3,000,000.

Conservation of the Mathematics Construct

The Conservation of the Mathematical Construct (CMC) theory and framework (Petit and Lager, 2003; Lager and Petit, 2005) was initially created and used to guide the development of over 1,200 large-scale mathematics assessment items for grades 3–8 for three New England states. The goal for the New England Common Assessment Program (NECAP) was to systematically create items that were as accessible to as many students as possible without sacrificing content, cognitive rigor, context, or language. Therefore, the needs of ELLs were considered from the beginning of and throughout the process (e.g., Bielenberg and Wong-Fillmore, 2004; Celedón-Pattichis, 2004; Celedón-Pattichis, 2003; Moschkovich, 2000; Khisty, 1995).

The CMC framework provides material and process guidelines for item developers and reviewers, including classroom teachers, to create and evaluate items that conserve the assessed mathematical construct (content and cognitive demand) when embedding the mathematics in rich contexts (when appropriate), using visual supports (when appropriate), and streamlining the language. Simultaneously attending to content, cognitive demand, context, visuals, and language provides the greatest number of students, especially ELLs, the greatest opportunity to demonstrate their knowledge and skills in relationship to the mathematical construct being assessed. A brief synopsis follows.

Explicitly Aligning the Item to the Mathematical Construct

The content, construct, and cognitive demand of the item should be determined initially. Typically, item developers and reviewers first use local, state, or national mathematics standards to choose or identify the targeted mathematics content of the item. Next, the desired level of cognitive complexity is chosen or identified using work such as Webb's (2002) Depth of Knowledge levels or NAEP's Levels of Mathematical Complexity (2005). For example, moving from least to greatest complexity, Webb's four levels for mathematics can be generally summarized as follows:

Level 1—Recall
Recalling or identifying information; using a procedure; applying an algorithm; single step solutions

Level 2—Skills and Concepts
Showing conceptual understanding; comparing and classifying data, requiring the problem solver to do some basic decision making; multi-step solutions

Level 3—Strategic Thinking
Doing more complex decision making; reasoning; planning; interpreting and using evidence; conjecturing; justifying a solution or decision point when others are possible

Level 4—Extended Thinking
Similar to Level 3 but done over longer periods of time and to greater depth, such as transferring strategic thinking to new situations and/or across content areas.

Timed, large scale assessment items typically are Levels 1 and 2. Small-scale assessment items can be of any of the four levels, with classrooms providing fertile ground for the extended time Level 3 and Level 4 items typically require. Levels 1–3 will be explicated a bit more through the analysis and revision of the three assessment items presented in this article.

Embedding the Item in a Rich Context (when appropriate)

Embedding mathematics items in meaningfully appropriate contexts allows for a natural access to the mathematics, supports student thinking, and shows reality as a source and domain of application, while also motivating greater student engagement (DeLange, 1987). However, contexts that are engaging but unnecessary or inappropriate for the construct assessed, ambiguous (where one mathematical construct is targeted but others can be easily inferred), or imbued with cultural or socioeconomic bias are to be avoided. To increase ELL access, use common and familiar experiences, such as food, geography, school, and business (Winter et al., 2006; Emick, Monroe, Malagon Sprehn, and Kopriva, 2007).

Visual Supports

Kopriva (2008; 2000) has already created a set of ELL best practice guidelines about the use of format and graphic organizers, such as charts, tables, graphics, pictures, and diagrams, to facilitate what is being asked or presented. They fall within seven sets:

1. Use of relevant visuals
2. Use of an effective visual format
3. Use of illustrations to mirror text
4. Use of illustrations to replace text
5. First person visuals
6. Use of visuals to organize events in time
7. Use of visuals to clarify textual meaning

Being simple and to the point, clutter-free, properly labeled, and aligned with the item information are characteristics of the aforementioned sets of practices. These guidelines will be flushed out in greater detail with the analysis and revision/reconstruction of the three items.

Streamlining Language

Simplified Language and Plain Language guidelines already provide important item writing suggestions to improve the language of mathematics assessment items and directions (e.g., Kopriva, 2008, 2000; Abedi and Lord, 2001; Hanson, et al., 1998;). Changing unfamiliar or infrequent words to more familiar words, changing linguistically complex long sentences to shorter sentences, and using present and active voice as much as possible are three actions common to those frameworks. Building upon this previous work, Streamlined Language is the language anticipated to minimize the meaning-making noise between a problem solver and an item, especially for context-rich and language-rich items (Lager, 2004).

Though each problem solver interacts uniquely with an item, Streamlined Language can provide multiple entry points to the item so that all problem solvers, and especially ELLs across English proficiency levels, maximize their opportunities to understand grade-level items. To do this, the complex interactions between construct, context, cognitive demand, visual support, language, and the problem solver are considered together from an ELL's point of view.

Example of Streamlined Language: The following item is typical of the first version of an elementary large-scale NECAP assessment item:

It costs $1.50 to go on the first ride at the Potter Mountain amusement park. Each additional ride costs $0.50. How much does it cost to go on 6 rides?

a) $3.00 b) $3.50 c) $4.00 d) $4.50

At first glance, this item might seem well written for all students, including ELLs, because the sentences are short and there are numerical values and quantities to manipulate. However, upon closer inspection there are several parts that could be improved.

Starting off the first sentence with it is a mistake because that general object pronoun doesn't help the problem solver initially visualize or understand what costs $1.50. *It* actually refers to *the first ride at the Potter Mountain amusement park*, which has needlessly been placed at the back end of the sentence. Therefore, to avoid starting off on the wrong foot, the first sentence should be changed to read, *The first ride at the Potter Mountain amusement park costs $1.50 to go on*. Similarly, the last sentence should be changed to How much do 6 rides cost to go on?

Next, to go on can be difficult for ELLs to interpret because in this context the phrase means to experience the ride's movements, not more common meanings of *to go*, such as to moving in a direction (to go left), at a rate of speed (to go fast), and/or toward a destination (to go home). Though the preposition *on* should cue the student to the difference in this context, the entire phrase *to go on* could be confused with continuation (to go on with your explanation). Therefore, because the cost of the rides is the item's focus, not going on them, *to go on* should be excised.

Further, because *Potter Mountain* is not a nationally known amusement park, like Disneyland, some ELLs may not recognize this two-word phrase as the amusement park's name, despite its capitalization and sentence placement. *Potter Mountain* may unnecessarily confuse students as they try to make sense of an unknown phrase's meaning. Therefore, because the name does not add value to the item, it should be excised. Finally, *at an amusement park* should be moved to the beginning of the first sentence so that the problem solver can immediately contextualize the rest of the item's information.

Now the item reads: *At an amusement park, the first ride costs $1.50. Each additional ride costs $0.50. How much do 6 rides cost?* Yet, looking at the second sentence, *Each additional ride* is likely to be a tricky phrase for many ELLs syntactically and semantically. Syntactically, *additional* modifies *ride* and *each* modifies *additional* ride. Semantically, though the phrase is read left to right, the meaning moves from right to left. Though *additional* and *each additional* are not common conversational phrases for kids, they are part of the mathematics register. Still, *each* can be a challenging pronoun for ELLs (Klingner, 2009). Though *ride* itself is irrelevant, understanding that the additional rides are rides 2, 3, 4…, and that ride 2 costs $0.50, ride 3 costs $0.50, ride 4 costs $0.50, etc., are critical to solving this item. Therefore, one alternative would be to replace the second sentence with: *The second ride costs $0.50, the third ride costs $0.50, the fourth ride costs $0.50, and so on.*

Though the language was streamlined, the item's context, mathematics content, and cognitive demand were unchanged. They were conserved. Access to the item for all problem solvers, but especially ELLs, was likely increased.

Conserving the Mathematics Construct in Action

To show the breadth, depth, and complexity of the possible interactions within an assessment item between mathematics content, cognitive demand, context, visual supports, language, and the problem solver, three mathematics items will be examined and revised using the CMC. The analysis of each item will not be exhaustive, but sufficient to document some of the access-reducing interactions item writers routinely do not take into consideration or address when developing mathematics items for problem solvers, and ELLs in particular. The revision or reconstruction of the item will show one way of addressing the aforementioned interactions. Because increasingly abstract and complex mathematics content are typically more difficult for ELLs to access, the elementary item will be examined first, followed by the middle school and high school items.

Elementary Problem—Garbage Truck

Zehr (2003) shares an elementary school item originally developed for non-ELLs that was revised "successfully" by the Illinois Department of Education to increase its ELL accessibility.

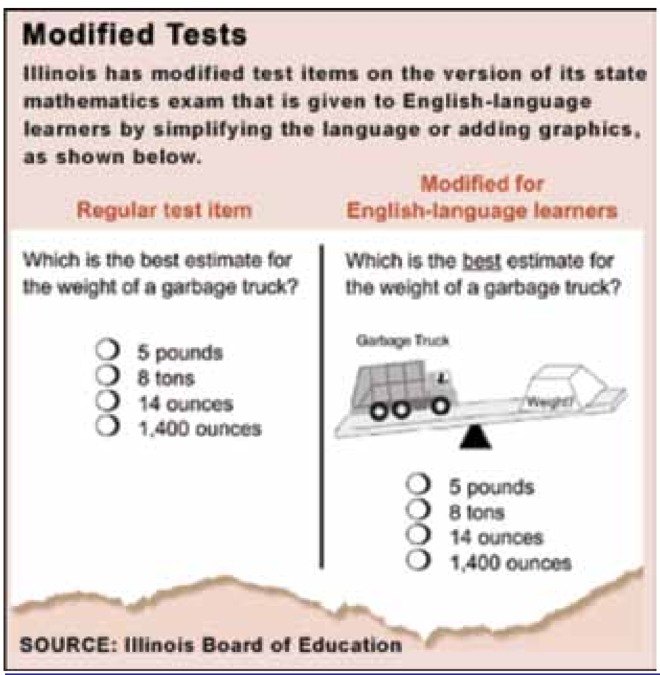


Figure 1. Elementary school item developed for non-ELLs and then modified for ELLs.

Analysis

This closed, selected response item[1] was apparently written to 5th grade Illinois mathematics standard 7B—Estimating measurements and determining acceptable levels of accuracy, with a specific focus on estimating weight using reasonable units (Illinois State Board of Education [ISBEa], 2003). The garbage truck part of the context is good because almost all students have seen one and know its function and general size. The cognitive demand of this task is Level 2 because the student has to compare given data in different units of measure and decide which quantity approximates the weight of a garbage truck. However, there are several problems with the visual representation, context, and language of the task that need to be identified and addressed.

Looking at the picture, only relative, not absolute, weight is clearly being modeled. The unbalanced lever shows that the truck is heavier than the weight. However, because a real garbage truck would never be at one end of such a lever (inappropriate context), the actual weight of the truck comes into question. Since only a toy garbage truck would actually be on such a lever, especially with a weight of roughly equivalent size to the truck on the other side of the lever, the answer logically shifts from 8 tons (real truck) to 14 ounces or 5 pounds (toy truck).

In terms of language, labeling the mass at the other end of the lever *Weight?* likely confuses the construct further because *Weight?* could be interpreted to mean "Is this a weight?" instead of the intended "How much does the truck weigh?" Also, the reader has to infer that *Which* means *Which answer choice of the four answer choices given*. Further, prepositional phrases such as *for the weight, of a garbage truck* are often troublesome for ELLs (Klingner, 2009). In sum, the aforementioned modifications likely **reduced** the item's accessibility for ELLs and therefore their chances of answering it correctly.

[1] An item with a single path to a single solution where the solver selects the correct answer from a list of possible responses (Romagnano, 2006)

Revision

To realign the item to the standard, the picture problems should be addressed first. By replacing the artwork with Figure 2, the garbage truck is now on a street against a backdrop of buildings to model an appropriate real-life context and its relative size. Though the truck may not be drawn exactly to scale, based on perspective (the truck's size and position relative to the buildings), the drawing is accurate enough. In addition, a motorcycle could be placed on the street in front of/behind the garbage truck to tacitly bracket the garbage truck's size between a smaller object (a motorcycle) and a larger object (a building), like in Figure 3.

Figure 2. Revised Garbage Truck artwork.

Figure 3. Alternate Garbage Truck artwork.

In terms of language, the vehicle in Figure 2 is clearly labeled *Garbage Truck*. Though the *Which is the best estimate of the weight of a garbage truck?* query could be kept to explicitly use weight and estimate, words from the standard itself, a better alternative exists. Changing the question to *Approximately, how much does a garbage truck weigh?* uses streamlined language to ask the same query in a more common, accessible way. This option eliminates *Which* and the two prepositional phrases while introducing a query phrase (how much) that signals a quantity is being sought. Also, because *approximately*/aproximadamente is an English/Spanish cognate, that adverb would likely help Spanish-speaking ELLs access the item.

One way to adjust this item for classroom summative or formative assessment use is to start with the revised version, excise the answer choices, and ask students to generate their own responses. This task would require students to select and apply appropriate standard units to measure weight, a 5th grade standard (ISBEa, 2003). The cognitive demand would vary depending on 1) how students were allowed to engage with the task and 2) what they were asked to do. If students could just look up the truck's weight on the Internet and parrot the answer back, then the task would be Level 1. However, if students had to estimate the truck's weight without such assistance and compare and evaluate the reasonableness of their classmates' answers, then the task would be Level 2. Also, because students' answers would likely vary (e.g., 3 tons; 8,000 pounds), their comparison and evaluation would necessitate unit conversions within a measurement system, another 5th grade standard (ISBEa, 2003). However, answers like 4,000 kilograms or 3 cars would necessitate examining standard and non-standards units of measure across measurement systems, which is an 8th grade standard (ISBEb, 2003).

Middle School—Restaurant Problem

The Restaurant problem appears in *Assessing English Language Learners: Bridges from Language Proficiency to Academic Achievement* (Gottlieb, 2006). According to Dr. Margo Gottlieb, a nationally renowned expert in the design of assessments for K–12 ELLs (WIDA, 2007), using a grid within a real world context in the Restaurant problem will help middle school students think mathematically and promote creative uses of mathematics language.

Please do what is asked. Afterward, please consider the mathematical target(s) for this problem. Then reflect on how the interactions of the item's mathematics, context, representation, and language components helped and/or hindered your engagement with the problem. Lastly, please consider how those interactions might affect how ELLs at different English language proficiency levels might engage with the same problem.

A Middle School Math Problem

Here are the floor plans for three different restaurants (A, B, and C). The shaded squares represent a bar; the other squares represent where people eat. You need to think about

- the fractional part or percentage of the restaurant that is the bar
- the total area of the restaurant
- the difference in space between the bar and eating areas

Decide which restaurant you would like to own. Explain, using math language, the reasons why you chose that one.

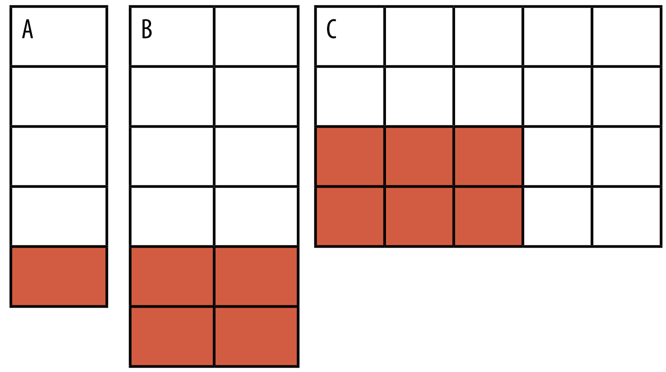


Figure 4. Restaurant floor plans (pp. 68–69).

Though Gottlieb's holistic objectives are important and well-intentioned, she never mentions any specific mathematics standards for this problem. In fact, she expects her readers to assign mathematics objectives to this problem *ex post facto*. However, mathematics item development works oppositely. First a mathematics standard (or standards) to assess is identified, then an item is written to assess those desired standards.

In addition, Gottlieb does not define or identify the kinds of mathematical thinking and creative mathematics language use she expects. Relevant variables such as total area of each restaurant, average number of customers per day at the bar/non-bar areas, average profit per customer at the bar/non-bar areas are not specified either. Also, no exemplary solutions are provided. Therefore, determining the item's intended level of cognitive demand is impossible. In sum, these process and product irregularities render the item, in its current form, useless for standards-based mathematics assessment. However, deconstructing its mathematics content, representation, context, and language can inform its reconstruction into a usable mathematics assessment item.

Analysis

The item's lack of mathematical focus affects how to interpret irregularities within its representations. Four examples follow. First, the individual "squares" in floor plans A and B are rectangles, but not squares. Second, in floor plan A, the bottom two rectangles are not congruent to each other or to the top three rectangles. Third, no scale factor accompanies any of the floor plans. Fourth, floor plans never look like the ones in this problem.

Though nearly all middle school students know what a restaurant is and have experience eating in one, the bar context is inappropriate for teenagers because underage drinking is illegal. Also, students are explicitly asked to compare bar and non-bar areas within and across restaurants, ostensibly, because in a restaurant the bar area usually generates more profit per square foot than the non-bar area (Dr. Vino, 2007). However, expecting a middle school ELL to know this key piece of context-dependent, implicit information about maximizing profit is unreasonable (Daro et al., 2007). More generally, it also is unclear what knowledge students are expected to bring and/or be provided regarding business plans and profit margins for restaurants.

In terms of given instructions, there are several off-target directions. Two examples follow. Asking students to choose a restaurant to own and justify the choice with math language does not guarantee the problem solver will provide the mathematics-based decision or rationale the item writer intended. For example, a student may choose A, because it's easiest to clean. Because "easiest to clean" likely means "floor plan with the least area" to the student, examining part-whole relations, maximizing profits, and including explicit math language like "least," "area," and "percentage" are not part of the response. The accompanying rationale logically supports the choice, based on the criterion important to the student. However, in that kind of response both the mathematics and math language are, at best, implicit.

Second, asking students to think about *the difference in space between the bar and eating areas* for each floor plan is misleading on many levels. First, customers can often eat at the bar within a restaurant, so the dichotomy is pragmatically false. Second, because the bar area appears smaller than the non-bar area in all three floor plans, doing the suggested subtraction as written *(bar–non bar)* will result in the calculation of negative areas. Such nonsensical answers are likely to either get changed to positive answers or get done initially as *non bar–bar* problems because middle schoolers often see the "bigger number" (the minuend) first in subtraction problems. Both adjustments could contribute to ingraining classic subtraction errors that often haunt students throughout algebraic problem solving. Third, to maximize profit, it's not a part-part *(bar–non-bar)* difference that should be examined across floor plans, but a part-whole *(bar–restaurant)* comparison.

Because there are no explicit mathematical or cognitive objectives for this problem, the aforementioned irregularities point to item specifications that seemingly exist only in Gottlieb's mind. Because students are not mind readers, such "spec-in-the-head" (Davidson and Lynch, 2002) expectations will be explicitly examined. First, because gridded floor plans are typically drawn on uniform grids, are students expected to notice and take into account the different sized rectangles within floor plan A and across floor plans A, B, and C? If yes, why use this context to promote investigating how unequal parts within A comprise a whole and how different rectangle units across A, B, and C compare? If no, for what purposes were these discrepancies introduced? In addition, why aren't scale factors provided? Are students expected to create and apply a single scale factor for all three floor plans so that the total areas can be calculated and compared? Are students expected to interpret the lack of a scale factor as a tacit cue to ignore the total area of each restaurant and focus solely on the percentage of each restaurant that is the bar? Though these questions may be good food for thought for an item writing team to discuss, answers should be decided, documented, and shared before the final version of the item is disseminated.

Reconstruction

Though the Restaurant item contains fatal flaws, there are important mathematical ideas embedded in it that could be assessed more effectively for ELLs. Unlike the Garbage Truck revision, the Restaurant reconstruction will result in an almost brand-new item, the Computer Poster. Computer Poster addresses the aforementioned concerns for all students by having clear mathematical objectives, being aligned to specific mathematics standards, and having an explicit level of cognitive demand. Specific for increasing ELL accessibility, Computer Poster employs ELL-friendly contexts and visual supports and is written using streamlined language principles. Because Dr. Gottlieb is based out of the Illinois Resource Center (IRC, 2002), the original task has been reconstructed to meet the following 6th grade Illinois mathematics standards:

- 6A: Demonstrating knowledge and use of numbers and their many representations in a broad range of theoretical and practical settings
- 6B; Investigating, representing, and solving problems using number facts, operations and their properties, algorithms, and relationships
- 6D; Solving problems using comparison of quantities, ratios, proportions, and percents

The reconstructed task:

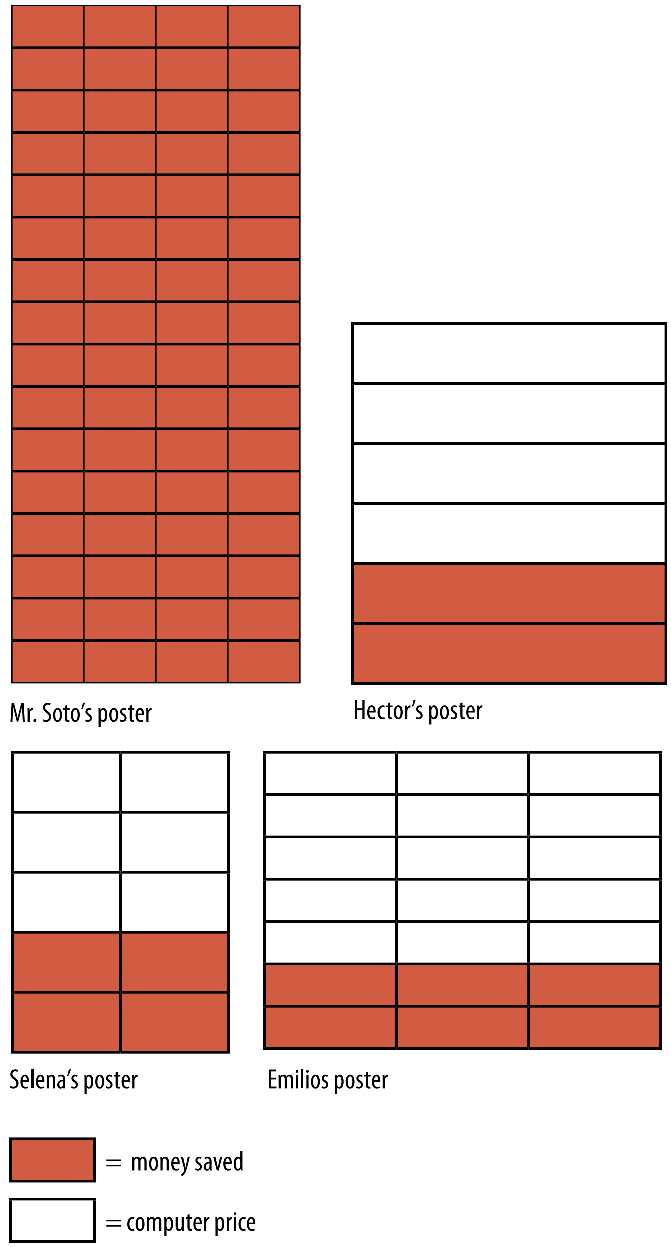


Figure 5. Math poster problem.

Mr. Soto and his students, Hector, Selena, and Emilio, are saving money to buy computers. In Mr. Soto's math class, they made posters comparing their money saved with the computer prices. Mr. Soto's poster shows he saved all the money necessary to buy a computer. Which student is closest to buying a computer?

a) Hector b) Selena c) Emilio d) Impossible to answer

The context, saving money to purchase a computer, is appropriate, relevant, and positive. The mathematics classroom is included because the location is a familiar place for ELLs where such part-whole representations would be made. The task is of Level 2 cognitive demand because conceptual understanding must be shown and data must be compared so that the problem solver makes a decision.

Because all four actors are important, all four are mentioned by name in the problem and in the accompanying representations. A teacher and three of his students were purposefully chosen because a math teacher and his students would make such posters and because the teacher/student dichotomy facilitates the transmission of important life values. By having the teacher also save money for a computer, the teacher is modeling for his students how to save for and invest in their present and future learning. The teacher's poster shows he has successfully reached his goal (another important lesson) and provides a visual example of the money saved/computer price relationship so that the problem solver can make sense of and compare the students' part-whole representations.

In fact, this problem purposefully rewards learners who think conceptually about part-whole relationships (Fosnot and Dolk, 2002) and use the given array models as models for thinking (Fosnot and Dolk, 2002). For example, considering what is constant and what is changing, looking across student posters, the two colored-in rows are constant, but the total number of rows changes. With a common numerator of 2, comparing each student's progress toward saving for a computer (Hector—2/6 of the way there, Selena—2/5, and Emilio—2/7) is easy and fast. Yet, because this problem has multiple entry points, such as seeing the students' saving progress as Hector—2/6 of the way there, Selena—4/10, and Emilio—6/21 and looking for a common denominator to make the comparison, there are many ways to arrive at the same conclusion.

To focus on these part-whole relationships, this problem purposefully did not include any numerical quantities, such as the price of a computer or how much money a student had saved. Notice, also, that whether or not the different computers cost the same amount of money is immaterial. Also, because in this context closest means determining which ratio's value is nearest to 1, not a relative location (the more common meaning), the visual representations purposefully link these two meanings with the colored-in area increasing toward the top of the poster, meaning the money saved/computer price ratio is increasing toward 1.

However, if *closest* is misinterpreted as the greatest colored-in area or the fewest remaining non colored-in rectangular units, then Hector would likely be the logical, but incorrect choice. If *closest* is interpreted as the shortest distance from the top of the colored-in area to the top of the poster, then Emilio would likely be the logical, but incorrect choice. Semi-structured, task-based interviews with middle school ELLs engaging with this item would provide evidence to determine the veracity of these conjectures and the value of these item distracters.

Overall, the visual representation followed Kopriva's (2008) ELL guidelines. The posters were purposefully of different sizes, divided into different, equal-sized units, and labeled to match the four actors. The posters were ordered to match the order of the four actors in the text, which was consistent throughout. Together, the posters and legend defined and clarified textual meaning. In fact, for the legend, the *money saved* key was purposefully placed directly over the *computer price* key, without the fraction bar, to visually hint at the part-whole relationship that needs to be calculated for each student's poster. Altogether, the visual representation was well-planned, well-organized, and clutter-free.

Still, there are many ways to adjust and extend this reconstructed item. Three follow. One is to start with the reconstructed item, excise the answer choices, and ask problem solvers to generate and justify their own response. A second would be to start with the reconstructed item and ask problem solvers who saved the most money and to justify their response. A third would be to start with the reconstructed item, excise the answer choices, and ask problem solvers if the students could together purchase one computer. The second and third allow problem solvers to determine multiple solutions using different interpretations of the item and document sources of ambiguity in the item. According to the Northwest Regional Education Laboratory (NRWEL) mathematics problem-solving scoring guide (NRWEL, 2000), such competencies would show proficient and exemplary insights into the deeper structure of the problem.

High School—Medicine Problem Please solve the following assessment exemplar from *Mathematics assessment sampler, grades 9–12: Items aligned with NCTM's Principles and Standards for School Mathematics* (Travis and Collins, 2005) before continuing.

> If a certain medicine is absorbed by your body at a rate so that 1/3 of the original amount is left after 8 hours and if your doctor gives you 10 grams today and does not want more than 10 grams to accumulate in your system, how much medicine should she give you tomorrow at the same time? (58)

Now, please consider the mathematical target(s) for this problem. Next, reflect on how the interactions of the mathematics, cognitive demand, context, and language components of the item helped and/or hindered your engagement with the problem. Lastly, consider how those interactions might affect how secondary ELLs of different levels of English language proficiency levels might engage with the same problem. Then proceed.

Analysis

According to *NCTM's Mathematics Assessment Sampler: Grades 9–12* (2005), the problem was written to the following grade 9–12 standards (NCTM, 2000):

1. Represent and analyze mathematical situations and structures using algebraic symbols
   a. Use symbolic algebra to represent and explain mathematics relationships
   b. Judge the meaning, utility, and reasonableness of the results of symbolic manipulation, including those carried out by technology
2. Use mathematical models to represent and understand quantitative relationships
   a. Use symbolic expressions to represent relationships from various contexts
   b. Draw reasonable conclusions about a situation being modeled

The item is of Level 2 cognitive demand because the problem solver must show conceptual understanding and take a multi-step path to reach the solution. The context is appropriate and relevant. However, there are many unnecessary syntactic, semantic, and pragmatic language mathematics register challenges with its written form that likely restrict access for many problem solvers, especially ELLs.

First, in terms of syntax, the "if and if-then" construction of this one-sentence question is unnecessarily long and complex. Here are the five (a–e) components of this construction:

> a—a certain medicine is absorbed by your body at a rate
>
> b—1/3 of the original amount is left after 8 hours
>
> c—your doctor gives you 10 grams today
>
> d—does not want more than 10 grams to accumulate in your system
>
> e—how much medicine should she give you tomorrow at the same time?

The *if a so that b* and *if c and d* constructions, followed by the e interrelationship, are data connections that are not explicit or typically taught by secondary mathematics teachers to their ELLs. *"So that"* meaning *b* modifying *a*, not *a* causing *b*, and the comma between *system* and *how* meaning "then" are two examples of common logical connectives that ELLs need to recognize and understand to access and solve the item.

Second, though starting off with the most abstract information (the absorption rate) is logical from a mathematical perspective, since the student is expected to iteratively apply that rate to data given later in the problem, from an English language development perspective placing the *if a so that b* clause first likely hinders reading comprehension. Semantically, that clause does not identify the medicine "I've" absorbed, how much medicine "my" body started with initially, or how long "I've" had such medicine in my body already. Therefore, constructing a chronologically correct mental movie for the situation (a common reading comprehension strategy (Strong et al., 2002; Vacca and Vacca, 2005) to prompt a viable concept image (Tall and Vinner, 1981) of the absorption is likely to be more difficult than need be.

Third, specific non-mathematics vocabulary words in this problem will likely unnecessarily cause meaning making difficulties for ELLs. For instance, the problem solver must recognize that *your body*, *you*, and *your system* are not three distinct entities, but synonyms in this item. This connection will likely not be obvious, in part because system is a non-descriptive, abstract noun that has other, more common meanings, but refers to the body as a functional unit in this context. In addition, not only does *original amount* mean "initial amount," but the initial amount is not specified when the phrase is introduced. Therefore, though 10 grams is mentioned later, an ELL must infer that the *original amount* of medicine must be the 10 grams the doctor provides. This connection is not obvious. Finally, *left* can be challenging because it means "remains" in this context, not direction or motion.

Fourth, there are words that ELLs sometimes do not understand or recognize as unimportant to solving the problem, yet these words slow them down or stop them from doing so. For example, in this context, *certain* refers to some unknown, but particular kind of medicine instead of the state of being sure. Though the name and type of medicine are irrelevant in this context, ELLs may not recognize this point. They may not know that they can substitute any name/type of medicine for *certain* and continue on, so long as the given data drive their problem solving and not their personal experience with the name/type of medicine they choose. In fact, thinking about some unknown, but particular kind of medicine might be as difficult as considering how a variable represents an unknown, but specific quantity. Further, even if an ELL does recognize or constructs the correct meaning of *certain*, this non-descriptive, abstract adjective is not likely to conjure a sharp mental image for an ELL, much less similar mental images across ELLs, whereas a common, concrete adjective, such as "liquid," likely would.

Fifth, there are specific verb tenses that beginning ELLs are likely to find difficult. The passive voice, *is absorbed* in this problem, is a difficult language construct because the object being acted upon is the focus of the action not the actor doing the action. Though, from a mathematics perspective, item writers intentionally use passive voice to draw attention to relevant information (the absorption rate) and away from unimportant information (whose body is doing the absorbing), from an English language development perspective this syntactic structure does not help ELLs mentally generate a chronological frame-by-frame sequence of the action, like a noun-verb-object structure would (e.g., "You absorb medicine.").

Sixth, the problem solver is required to make many inferences from the text that are not obvious, yet influence determining the solution. For example, from the data given, the absorption rate is unclear. If the absorption rate is linear, then *all* the medicine ingested would be absorbed in half a day (2/3 medicine is absorbed in 8 hours, so 3/3 medicine is absorbed in 12 hours). However, if the rate is exponential, then 1/27 of the medicine is left unabsorbed after one day (1/3 of the medicine ingested being unabsorbed after 8 hours, 1/9 after 16 hours, and 1/27 after 24 hours). This distinction is not trivial mathematically or pragmatically because some absorption rates (like in real estate) are linear. In addition, the meaning of *accumulate* is unclear because the substance(s) being accumulated is (are) not specified. Within the body, is the accumulated medicine the medicine that is absorbed only, unabsorbed only, or absorbed and unabsorbed (meaning both kinds)? Lastly, though the doctor gives "me" 10g of medicine today and makes clear that her preferred upper bound for the accumulated medicine in my body is 10g inclusive, the problem does not state explicitly how much medicine I actually ingested, how much I had yesterday, or how much I should have in my system tomorrow.

In fact, to generate the "correct" answer of 9.63g[2] the problem solver must infer that:

1. Yesterday, he had 0 g of accumulated medicine in his body.
2. He ingests all 10g of medicine as soon as it is given to him.
3. His body absorbs the medicine exponentially.
4. His body accumulates unabsorbed medicine only.
5. He must be given a specific amount of medicine tomorrow so that when he ingests all of it exactly 24 hours after the first ingestion, he will have exactly 10g of unabsorbed medicine in his body at that moment.

These required inferences raise many questions. Are these kinds of inferences requisite parts of successful mathematical modeling and problem solving? If so, is expecting problem solvers, and especially ELLs, to make them reasonable? If a problem solver makes reasonable, but different inferences than what the item writer and classroom teacher intended and successfully solves a related, but different problem, should the problem solver receive partial or full credit for her/his performance? How aware are item writers and classroom teachers of the embedded inferences they are requiring problem solvers to make? Should there be a continuum of inferences upon which item developers and classroom teachers draw? How much information should an item writer and classroom teacher lay out explicitly and how much should a problem solver be expected to infer? Should ELL inference expectations be further delineated by levels of English language proficiency? Should mathematics teachers be teaching students how to recognize their own inferences and decide which ones seem most reasonable for particular problems? These questions merit further investigation.

Revision

Keeping all of the aforementioned interactions in mind, while keeping the same mathematical foci and level of cognitive demand, I created the following version of the task:

> *María was healthy on Sunday. However, Monday she wakes up sick. So, Monday at 7:00 a.m., she swallows 10 grams of liquid medicine. For María to feel better, her body must absorb the medicine. The chart below shows María's medicine absorption rate.*

[2] After 8 hours, 1/3(10g) or 10/3g of medicine remain in the body. After 16 hours, 1/3(10/3g) or 10/9g of medicine remain in the body. After 24 hours, 1/3(10/9g) or 10/27g of medicine remain in the body. Therefore, the doctor gives (10 - 10/27) grams of medicine = 260/27 grams or 9.63 grams.

Medicine in María's Body

Day	Time	Absorbed	Unabsorbed (not absorbed)
Day	7:00 a.m.	0	1
Day	1:00 p.m.	2/3	1/3
Day	7:00 p.m.	8/9	1/9
Day			
Day			
Day			

Figure 6. Absorption rate chart.

> *For example, Monday at 1:00 p.m., María's body has absorbed 2/3 of the medicine and not absorbed 1/3 of the medicine. How much medicine must María swallow Tuesday at 7:00 a.m. to have exactly 10 grams of unabsorbed medicine in her body?*

The one long, complex sentence in the original version was replaced in the revised version by seven shorter sentences to help ELLs better chunk, sequence, and connect the information. Sentences 1–2 introduce María and her health. Sentence 3 states her ameliorating action, transitioning the focus from María to the medicine itself. Sentences 4–6 explain how she absorbs the medicine. Sentence 7 requires the student to apply information from sentences 1–6 to solve the problem.

At the phrase and word levels, many changes and insertions were made. Because *you* was too personal in the original, *María* is now the protagonist. The *doctor* was removed to limit the number of actors in the problem to one. The *Sunday/Monday* and *healthy/sick* dichotomies were inserted to demarcate reasons for not having medicine in your system on Sunday but for taking it initially on Monday. The *she swallows 10 grams of liquid medicine phrase* was added to explicitly state how much medicine was initially taken and because liquid medicine is common, tangible, and easy to visualize. Also, *liquid* and *medicine* are both cognates for their Spanish equivalents, *líquido* and *medicina*, thus increasing Spanish-speaking ELLs' item access. In contrast, *certain, system, accumulated, left*, and the passive voice have been excised from the original item because of their potential to limit ELL accessibility.

For María to feel better, her body must absorb the medicine was included to rationalize the importance of medicine absorption, help define *absorb* in context, and scaffold the introduction of *absorption rate* in the next sentence. Though *absorb* is not a high frequency word in mathematics, because it is part of the science register the concept and word would likely have been taught already in a science course. *Rate*, however, is a high frequency concept and word in mathematics, and therefore a semantic part of the mathematics register.

Further, because *unabsorbed* is likely to be an unfamiliar concept and word, *un* and *not* are intentionally underlined throughout Version 2.0 to help ELLs see that *unabsorbed* and *not absorbed* are synonymous as well as integral to correctly solving this problem. To avoid any confusion with the rest of the text, *un* and *not* are the item's only underlined character strings. Also, though difficult for some beginning and intermediate ELLs to recognize and/or understand from the context itself, *exactly* ensures no more and no less than 10 grams of medicine Tuesday at 7 a.m. Without *exactly*, a student could give any response greater than 9.88 grams and still be correct because, technically, if a student answered "12 grams," for example, María would have 10 grams of unabsorbed medicine in her body. Therefore, the original problem's condition is conserved.

In addition, changes were made to the data themselves and their presentation. Though the original problem consisted of three 8-hour absorption cycles, the revised version of the task consists of four 6-hour absorption cycles to facilitate the use of specific data to define the exponential absorption rate without saying "exponential." Also, the given data allow the problem solver to generate the next two cycle iterations to algebraically determine the solution, like in the original version. However, shortening the period within the same 24-hour interval increases the final answer to 9.88g[3].

The chart was introduced to organize and share enough data so that the problem solver can clearly see the absorption rate and to reduce item wordiness. The chart's structure makes clear that absorbed medicine and unabsorbed medicine are distinct, related entities that are simultaneously inside María's body. The *For example,…* sentence exists to help students explicitly connect the absorption rate to the amount of medicine swallowed, which *…at a rate so that 1/3 of the original amount is left…* did in the original version. Also, unlike the original item, the problem solver can use the information from the 2/3, 1/3 example to go up a row on the chart to determine that when María swallows medicine, 100 percent of it is initially unabsorbed. Then the student can look across and down rows 1–3 to calculate the absorption rate and apply it to the next two absorption cycles. Having high school ELLs do this revision and subsequently participate in task-based interviews would provide evidence to examine how and to what extent the students' thinking corresponds to the assessed mathematics standards (Kulm, Wilson, and Kitchen, 2005).

In fact, the chart's three blank rows were included to visually hint that the medicine absorption continues past Monday at 7:00 p.m. Though *María's medicine absorption rate* should lexically clarify that key point, past research has shown that ELLs sometimes give primacy to the visual representation in similar iterative situations (Lager, 2006). Two extra rows would have matched the exact number of absorption cycles needed to go from Monday at 7 p.m. to Tuesday at 7 a.m., thereby possibly giving away what students must determine on their own. With three blank rows, problem solvers who choose to "fill in the blanks" must reconcile going one iteration past what the problem asks.

However, because many students are accustomed to using all given information to solve a problem, there are some potential, but illustrative drawbacks to the three blank rows, such as the "answer row" trap. Based on related prior research with middle school ELLs (Lager, 2006), two "answer row" traps will be explicated with hypothetical students A and B. If Student A believes that the last row must be the "answer row," he may shorten the absorption cycle to 4 hours (12 hours/3 blank rows) to make sure the last row is Tuesday at 7 a.m. Such work would evidence a rate deviation from the given data to accommodate the chart and the posed question. Student B, however, may keep the 6-hour absorption cycles, but respond to Tuesday at 1:00 p.m., the extra iteration, instead. Though B's fidelity to the rate concept would appear stronger than A's, both A and B would be giving more credence to what they believe the chart is saying than from the accompanying text itself. When problem solvers struggle to make meaning of an item's text, their dependence on and confidence in visual cues tends to increase, regardless of their English language proficiency (e.g., Lager, 2006).

Lastly, though the item's word count has been increased to 84, the chart breaks up the text into two

---

[2] After 6 hours, 1/3(10g) or 10/3g of medicine remain in María's body. After 12 hours, 1/3(10/3g) or 10/9g of medicine remain in her body. After 18 hours, 1/3(10/9g) or 10/27g of medicine remain in her body. After 24 hours, 1/3(10/27g) or 10/81g of medicine remain in her body. Therefore, the doctor gives María (10 - 10/81) grams of medicine = 800/81 grams or 9.88 grams.

42-word sections. Aesthetically, the chart provides visual balance between the sections. Effectively, two 42-word sections often seem less intimidating than one 84-word section to an ELL from a reading comprehension perspective.

CONCLUSION

Understanding the unique challenges English-language learners encounter on content tests and finding equitable ways to assess their content understanding (Durán, 2008; Solano-Flores, 2008) are significant theoretical and pragmatic concerns, especially for the mathematics education community. Because assessing an ELL's mathematical knowledge in English is likely to be significantly influenced by the student's English language proficiency (Menken, 2008), item developers and reviewers must ensure that mathematics assessment items are written in ways to provide ELLs unfettered access to understanding them. In addition, these items must not compromise the mathematical construct being assessed, the item's cognitive demand, or be language-free. The Conserving the Mathematical Construct (CMC) process can be used to train mathematics item writers and evaluators who have little or no training in task and item design, to create new items or analyze and revise/reconstruct previously written items.

As seen across the three analyzed and revised/reconstructed items in this article, each item is written and presented differently. There is no single best format or structure for an assessment item, much less all items. However, the integrated principles discussed here, considered as an integrated whole, can be used for item development across grades and content strands.

Also, there is no single path to creating, revising, or reconstructing an item or any formulaic way to generate mathematics items for all ELLs. Just as mathematics teaching (Lampert, 2003) and learning (Fosnot and Dolk, 2002) are messy, non-linear processes filled with numerous questions, concerns, and decision points, so is item development. Go to **www.huertodemanzanas.com**, click on the CMC tab, then click the CMC Schematic hyperlink to get an overview of the entire iterative, CMC process in detail.

Finally, invested stakeholders need to work together to do this important access and equity work for ELL mathematics assessment development. As seen in this article, no single development group or individual has all the answers. Mathematics-focused specialists usually consider the mathematics primary, and the context, language, and format of the item secondary (if at all); vice versa for ELL specialists. For optimum mathematics assessment development for all students, and especially ELLs, both kinds of specialists need to equally value each other's expertise. Mathematics and ELL specialists from national testing companies, state departments of education, universities, and school districts should work together with mathematics teachers and ELLs themselves to professionally develop each other and build and share items. Doing research on those collaborative products and processes and disseminating the findings would help mathematics educators and our ELLs move forward.

REFERENCES

Abedi, J., C. Lord, C.K. Boscardin, and J. Miyoshi. (2000). *The effects of accommodations on the assessment of LEP students in NAEP* (CRESST Tech. Rep. No. 537). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

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

Bielenberg, B. and L. Wong-Fillmore. (2004). The English they need for the test. Educational Leadership, 62, (4), 45-49.

Celedón-Pattichis, S. (2004). Research findings involving English-language learners and implications for mathematics teachers. In M. F. Chappell, J. Choppin, and J. Salls (eds.), *Empowering the beginning teacher of mathematics: High school* (p. 45-46). Reston, VA: National Council of Teachers of Mathematics.

Celedón-Pattichis, S. (2003). Constructing meaning: Think-aloud protocols of ELLs on English and Spanish word problems. *Educators for Urban Minorities*, 2(2), 74-90.

Daro, P., F. Stancavage, M. Ortega, L. DeStefano, and R. Linn. (2007). *Validity study of the NAEP mathematics assessment: Grades 4 and 8.* (Chapters 2 and 3). Washington, DC: American Institutes for Research. Retrieved on January 4, 2009 from http://www.air.org/publications/documents/NAEP_Math_Validity_Study.pdf.

Davidson, F. and B.K. Lynch. (2002). *Testcraft: A Teacher's Guide to Writing and Using Language Test Specifications.* New Haven, CT: Yale University Press.

DeLange, J. (1987). *Mathematics Insights and Meaning.* Utrecht, The Netherlands: University of Utrecht.

Dr. Vino (2007). *Are restaurants pushing wine for profit?* http://www.drvino.com/2007/04/02/restaurant-profits-in-the-wine-cellar/. Retrieved on December 20, 2008.

Durán. R. (2008). Assessing English-Language Learners' Achievement. *Review of Research in Education* 32, 292-327.

Emick, J. and R.J. Kopriva. (2007). *The validity of large-scale assessment scores for ELLs under optimal testing conditions: Does validity vary by language proficiency?* Presentation at the American Educational Research Association annual meeting. Chicago, IL, April.

Emick, J., R. Monroe, M. Malagon Sprehn, and R.J. Kopriva. (2007). Culture, education, and testing in the US: A novel response. Article submitted for publication.

Fillmore, L.W. and C.M. Valadez. (1986). Teaching bilingual learners. In M.C. Wittrock (ed.), *Handbook of research on teaching* (3rd ed.). New York, NY: Macmillan, 648-668.

Fosnot, C.T. and M. Dolk. (2002). *Young mathematicians at work: Constructing fractions, decimals, and percents.* Portsmouth, NJ: Heinemann.

Gottlieb, M. (2006). *Assessing English language learners: Bridges from language proficiency to academic achievement.* Thousand Oaks, CA: Corwin Press.

Halliday, M.A.K. (1974). Some aspects of sociolinguistics. In E. Jacobsen (ed.), *Interactions between linguistics and mathematical education: Final report of the symposium by UNESCO, CEDO and ICMI, Nairobi, Kenya, September 1–11, 1974* (pp. 64-73). Paris:UNESCO.

Hanson, M.R., J.R. Hayes, K. Schriver, LeMahieu, and P.J. Brown. (1998). *A Plain language approach to the revision of test items.* Paper presented at American Educational Research Association annual meeting. San Diego, CA.

Illinois Resource Center (IRC). (2002). http://www.thecenterweb.org/irc/. Retrieved on May 25, 2009.

Illinois State Board of Education. (ISBEa). (2003). Illinois Learning Standards—Mathematics—Classroom Assessments and Performance Descriptors—Printable Descriptors Grades 1–5. http://www.isbe.state.il.us/ILS/math/capd.htm. Retrieved May 25, 2009.

Illinois State Board of Education. (ISBEb). (2003). Illinois Learning Standards—Mathematics—Classroom Assessments and Performance Descriptors – Printable Descriptors Grades 6–12. http://www.isbe.state.il.us/ILS/math/capd.htm. Retrieved May 25, 2009.

Khisty, L.L. (1995). Making inequality: Issues of language and meanings in mathematics teaching with Hispanic students. In W. Secada, E. Fennema, and L. Adajian (eds.), *New directions for equity in mathematics education*. Cambridge, United Kingdom: Cambridge University Press, 279-297.

Klingner, J. (2009, March 9). Teaching English language learners to read [Web log message]. Retrieved from http://cecblog.typepad.com/rti/2009/03/teaching-english-language-learners-to-read.html.

Kopriva, R.J. (2008). *Improving Testing for English Language Learners.* New York, NY: Routledge.

Kopriva, R.J. (2000). *Ensuring accuracy in testing for English language learners: A Practical guide for assessment.* Council of Chief State School Officers, SCASS-LEP Consortium. Washington, DC.

Kulm, G., L.D. Wilson, and R. Kitchen. (2005). Alignment of content and effectiveness of mathematics assessment items. *Educational Assessment*, 10(4), 173 -196.

Lager, C.A. (2004). *Streamlined Language.* Unpublished manuscript.

Lager, C.A. and M. Petit. (2005). Professional development: Universal Design and Conserving the Mathematical Construct in J. Burrill and A.M. Collins (eds.) *Mathematics assessment sampler, grades 6–8: Items aligned with NCTM's Principles and Standards for School Mathematics.* 238-241. Reston, VA: NCTM.

Lager, C.A. (2006). Types of mathematics-reading interactions that unnecessarily hinder algebra learning and assessment. *Reading Psychology*, 27(2-3), 165-204.

Lampert, M. (2003). *Teaching problems and the problems of teaching.* Yale University Press.

Menken, K. (2008). *English learners left behind: Standardized testing as language policy.* Clevendon, UK: Multilingual Matters.

Moschkovich, J.N. (2000) Learning mathematics in two languages: Moving from obstacles to resources. In W. Secada (ed.), *Changing the Faces of Mathematics* (Vol. 1): Perspectives on multiculturalism and gender equity. Reston, VA: NCTM. (pp. 85-93).

National Assessment of Education Progress (NAEP). (2008). NAEP Data Explorer. http://nces.ed.gov/nationsreportcard/nde/criteria.asp. Retrieved May 14, 2008.

National Assessment of Educational Progress. (2005). Mathematical complexity. http://nces.ed.gov/nationsreportcard/mathematics/complexity2005.asp. Retrieved on January 12, 2009.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and Standards for School Mathematics.* Reston, VA: NCTM.

National Council of Teachers of Mathematics. (2005). B. Travis and A. M. Collins (eds.) *Mathematics assessment sampler, grades 9–12: Items aligned with NCTM's Principles and Standards for School Mathematics.* Reston, VA: NCTM.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* Washington, DC: U.S. Department of Education. Retrieved December 30, 2008, from http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf.

Petit, M. and C.A. Lager. (2003). *Conserving the Mathematical Construct.* Unpublished manuscript.

Rivera, C. and E. Collum. (2006). *A national review of state assessment policy and practice for English language learners.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Romagnano, L. (2006). *Mathematics assessment literacy: Concepts and terms in large-scale assessment.* Reston, VA: NCTM.

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

Solano-Flores, G. (2003). The multidimensionality of test review and test design: A conceptual framework for addressing linguistic and cultural diversity in testing. Paper presented at the 10th Biennial Conference of the European Association for Research on Learning and Instruction, Padova, Italy, August 26–30.

Spanos, G., N. Rhodes, T. Dale, and J. Crandall. (1988). Linguistic features of mathematical problem solving. In R. Cocking and J. Mestre (eds.), *Linguistic and cultural influences on learning mathematics*, 221-240.

Strong, R.W., M.J. Perini, H.F. Silver, and G.M. Tuculescu. (2002). *Reading for academic success: Powerful strategies for struggling average and advanced readers, grades 7–12.* Thousand Oaks, CA: Corwin Press, Inc.

Tall D.O. and S. Vinner. (1981). Concept image and concept definition in mathematics, with special reference to limits and continuity. *Educational Studies in Mathematics*, 12 151-169.

U.S. Government Accountability Office. (2006). *No Child Left Behind Act: Assistance from education could help states better measure progress of students with limited English proficiency (GAO-06-815).* Washington, DC: Author. Retrieved May 29, 2009, from http://www.gao.gov/new.items/d06815.pdf.

Vacca, R.T. and J.L. Vacca. (2005). *Content area reading: Literacy and learning across the curriculum.* Boston: Pearson.

Webb, N. (2002). Depth-of-Knowledge Levels for Four Content Areas—Mathematics. http://www.paec.org/delta/mathhighquality/Levels.pdf. Retrieved on January 10, 2009.

World-Class Instructional Design and Assessment. WIDA. (2007). http://www.wida.us/aboutus/staffBios/Admin/margogottlieb.aspx. Retrieved on May 25, 2009.

Winter, P., R.J. Kopriva, C.S. Chen, andJ. Emick. (2006). Exploring individual and item factors that affect assessment validity for diverse learners: Results from a large-scale cognitive lab. *Learning and Individual Differences*, 16, 267-276.

Zehr, M.A. (2003). States developing tests for English-Learners. *Education Week*, 23(13), 13 - 15.

Going beyond multiple choice: Probing Mexican-American students' thinking and communicating about assessment items in measurement

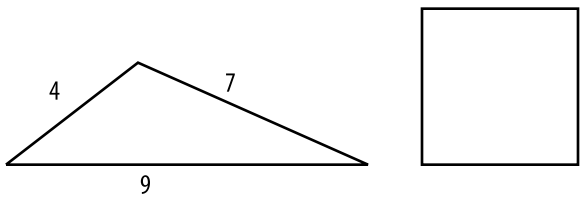
Anthony Fernandes • The University of North Carolina Charlotte
Cynthia O. Anhalt • The University of Arizona
Marta Civil • The University of Arizona

Large-scale summative assessments like the National Assessment of Educational Progress (NAEP) are used to certify what students know and can do in a given content area. These assessments contrast with formative assessments that act as a bridge between teaching and learning and provide the teacher with crucial information that can be fed back into teaching (Wiliam, 2007). Even though summative and formative assessments are employed for different purposes, we will illustrate how NAEP items can be used to extract information about English-language learner (ELL) students, which can then be used for formative purposes. The purpose of this study is to illustrate that items from NAEP could be used to get a better understanding of issues of language and mathematics that affect Latino/a[1] students, who comprise the majority of ELLs. In particular, the goals of this study are to a) gain an understanding of how a group of 15 Mexican-American students approached selected NAEP measurement items and; b) uncover some of the challenges that these items presented to this group of students.

Focus on measurement

We focus here on one of the five NAEP content strands: measurement. Lubienski (2003) noted that this content area had the largest achievement gaps between White and Hispanic scores at the eighth-grade (in NAEP 2000). We conjectured that the large gaps in this content area would allow us both to identify certain test items that seemed very challenging for Latino/a students and to probe their understanding on these items to uncover their conceptions and ways of thinking. In this article, we focus on what we learned from interviews with 15 Mexican-American students—drawn from grades 4 through 6 in schools serving working class communities—as they solved two NAEP measurement tasks. These items were the Perimeter problem, which was classified as a *hard* problem for fourth-grade students on the 1996 NAEP (NAEP items are classified as hard, medium, or easy at each grade level), and the Area Comparison problem, which was classified as a 'hard' problem for fourth-, eighth- and twelfth-grade students in the 1996 NAEP **(see Fig. 1).** We chose these

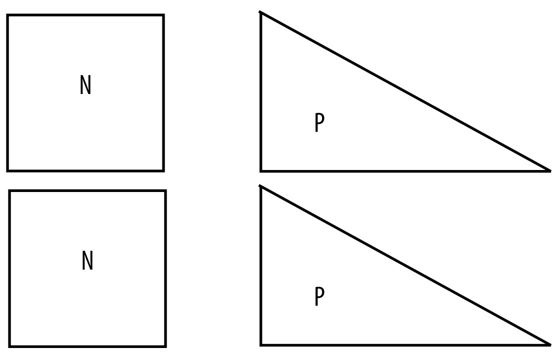
The Perimeter problem



If both the square and the triangle above have the same perimeter, what is the length of the side of the square?

a. 4
b. 5
c. 6
d. 7

The Area Comparison problem



(Cut outs of N and P were provided with the base of P being twice the side of the square)

Bob, Carmen, and Tyler were comparing the areas of N and P.

They each conclude the following:

Bob:	N and P have the same area;
Carmen:	The area of N is larger;
Tyler:	The area of P is larger.

Who was correct? Use pictures and words to explain why.

Figure 1: The NAEP problems

interview questions since they were the ones that yielded big differences in the percent correct when comparing White and Hispanic scores in NAEP. In the case of the Perimeter problem, the percent correct of fourth-grade White students was 29 percent compared to 13 percent of Hispanic students at the same grade. For the Area Comparison problem the percent correct for fourth-grade students was 7 percent for White students compared to 1 percent for Hispanic students. Further, at the eighth-grade the percent correct was 32 percent for White students solving the Area Comparison problem compared to 20 percent of Hispanic students. Note that by choosing these questions we are not trying to explain why the Hispanic student population underperformed, especially since our sample is too small, but we conjectured that these questions have the potential to elicit interesting thinking from these students that could inform teaching. Because NAEP is administered only to students in the fourth-, eighth- and twelfth-grades, we assumed that a hard problem for fourth-grade students in the NAEP sample could be considered less difficult for fifth- or sixth-grade students. These considerations guided our design of the set of NAEP items that were used to interview students at a particular grade.

Lubienski (2003) treated the gaps in performance of Whites and Hispanics on the Area Comparison problem as illustrative of gaps seen on other NAEP multistep problems. She conjectured that the gaps revealed a lack of opportunity for Hispanic students to solve multistep problems, and she suggested that teachers could act to redress this limitation by providing these students with the appropriate opportunities. Two specific suggestions she offered were to avoid leaving measurement until the end of the year, when it might be eliminated due to lack of time, and to tie measurement to other content areas like algebra, geometry, data analysis, and number sense.

Strutchens, Martin, and Kenney (2003) analyzed student performance in the NAEP measurement strand and concluded that, despite overall gains from 1990 to 2000 at the fourth-, eighth- and twelfth-grade levels, only a small number of students performed at the high end of the scale. The researchers looked closely at the concepts of length, area, and volume and suggested that examining successful student strategies could inform possible routes for instruction. In examining the concept of area, they used the solution strategies of successful students in the Area Comparison problem to conclude that the students had different levels of understanding of area. Students with a conceptual understanding of area could discover the relationships between the sides of the triangle and square and solve the problem flexibly without recourse to the use of area formulas. On the other hand, they also conjectured that the manipulation of the shapes may have been more efficient for the students who found it difficult to remember the formulas. One of their recommendations for instruction was that the students' first experience with the concept of area should include activities that involve area comparison of shapes without the presence of specific numerical measurements.

THEORETICAL PERSPECTIVES

This study is part of the research agenda of the Center for the Mathematics Education of Latinos/as (CEMELA), which aims to understand the interplay of mathematics, language, and culture among Latino/a1 students. Our perspective is essentially a combination of a sociocultural perspective and a cognitive perspective (Brenner, 1998; Civil, 2006; Cobb and Yackel, 1996). Our student interviews were cognitively based, and our analysis of these interviews was guided by a sociocultural perspective of mathematics cognition and language (Moschkovich, 2002, 2007b) with an emphasis on students' communication *about*, and *in* mathematics (Brenner, 1994).

Communication was central to our study in that we were investigating students' use of language as they interpreted the tasks and explained their thinking, even though this is not the intention of NAEP tasks; in some cases, our students were bilingual (English and Spanish), but more proficient in one of their two languages. Moschkovich (2002) points out that communication is multifaceted involving gestures, expressions, drawings, and objects as resources to simultaneously communicate mathematical ideas. These resources are especially crucial for students who may be less proficient in English, but are being educated in the U.S. in an all English instruction classroom, even though the students may have been able to use Spanish as a resource. A student's mathematical competence became more visible with the use of a sociocultural perspective that allowed for multiple resources from the situation. For example, students could manipulate the shapes that were provided and point to places where they wanted the interviewer to focus as they explained their thinking. In doing this, the students could convey their point, even in the absence of precise mathematical terminology.

Brenner's (1994) *Communication Framework for Mathematics* distinguishes among three kinds of mathematics-related discourse, namely: (a) communication *about* mathematics, which entails the need for description of problem solving processes and their own thoughts about these processes; (b) mathematical communication *in* mathematics, which entails using the language and symbols of mathematical conventions; and (c) communication *with* mathematics, which refers to the uses of mathematics that empower students by enabling them to deal with meaningful problems. Brenner emphasizes that all three kinds of mathematical communication are needed in the classroom for developing useful mathematical understanding. The third kind, communication *with* mathematics, was not part of our coding of the data since our focus was on NAEP assessment items, which do not match well Brenner's notion of meaningful problems. Thus, our approach focused on the students' use of resources as they communicated *about* and *in* mathematics.

METHODS

We conducted and videotaped task-based interviews (Goldin, 2000), both in English and Spanish, with 24 Latino/a students in grades 4 through 8. The students were attending elementary and middle schools in predominantly working class neighborhoods. In this article, we focus on the 15 students, in grades 4 through 6, who were interviewed in English (the others were interviewed in Spanish) and who worked on the Perimeter and Area Comparison problems in their set. We had 6 different measurement problems, and each interview used 4 of the problems, contingent on the grade level of the student being interviewed. The Perimeter problem and the Area Comparison problem were the only two problems that were common to the students interviewed in grades 4 through 6. Each of the authors conducted some of the interviews; sometimes one of the authors conducted the interview, while another author was videotaping and would occasionally also ask questions; at other times, one author conducted the interview and a person other than an author operated the camera.

The interviews of these 15 students were videotaped and conducted in English and each interview lasted approximately 30 minutes. The students first solved the problems independently as a simulation of an actual assessment setting. After the students gave an initial response, they were asked to explain their thinking. We then asked probing questions based on their responses and our interview script. In some cases, the students' interactions with the researcher prompted them to revise their initial solution. Our interview scripts for the problems were focused on questions that would help externalize the students' thinking (Goldin, 2000).

During the interviews, it was pertinent for us to consider the language demands that these NAEP assessment items would make on the ELL students. Campbell, Adams, and Davis (2007) examined the cognitive demands that two realistic-type problems in a high-stakes assessment presented to ELL students and illustrated a complex interaction between the culture, language, and mathematics that took place during the solution process. Because a limited amount of information can be stored in working memory, students need to recall pertinent information from long-term memory, store this temporarily in working memory, and process the problem information at the same time. As students try various strategies, they also are keeping in mind the various pieces of information provided in the problem. If the combined pieces of information are more than the capacity of the working memory, cognitive overload is said to occur.

In the case of ELL students, Campbell, Adams, and Davis (2007) pointed out that there were extra cognitive demands placed on the students as they navigated the language in the task and that cognitive overload was more likely to occur as they tried to juggle the mathematics and language required for the problem. We were aware of this as we interviewed the students and ensured that our questioning procedures served to address the issue of cognitive load by focusing the students' attention on smaller pieces of the task. Details of our probing questions for each problem are outlined below.

The Perimeter problem

The students were asked to solve the problem independently and reminded that they would be asked questions to explain their thinking. Once the students had solved or attempted the problem, we began with questions such as, “Can you explain your thinking to me?” Even if the student gave a correct explanation, we would still ask clarifying questions about their calculations: (i) “Why did you divide 20 by 5?” (ii) “Why did you multiply 5 x 4?” (iii) “What do you know about a square?” If the student provided an incorrect explanation, particular attention was paid to understanding the student’s method, and questions were directed to achieve this goal. In some cases, the student was asked to read the problem or we read the problem to them. Most of the probing questions took the following form: (i) “Do you know what the word perimeter means?” (ii) “What does this phrase mean, ’If the square and triangle above have the same perimeter”’?” and (iii) “Do you think that these two figures, the triangle and the square, have the same perimeter?” Once the students could answer these questions, but still could not work out the problem, the script focused the their attention on trying to get the length of the square: (i) “What is the perimeter of the square” (ii) “What do you know about a square?” or more pointedly, (iii) “What do you know about the sides of a square?”

The Area Comparison problem

The students were asked to read the question **(see Fig. 1)**, and the physical cut-outs of the two squares and triangles were provided. Once again the students were allowed to independently think about the problem, give a solution, and then asked about their solution method.

In this question, there were at least two distinct ways of solving the problem, and we probed the students on both methods. In the first method, successful students could superimpose the triangle and square, cut-off the extra portion on the triangle, and rearrange the two pieces to form the square. In the second method, the student could rearrange all four pieces to form two rectangles of the same area and then conclude that the area of the triangle and the square were the same. If the students solved the problem one way, they were also probed about the other method. If the students could not provide an answer to the Area Comparison problem initially, we asked questions to check their understanding of the problem. We first ensured that the students understood what was being asked in the question before asking them to explain their thinking of the mathematics. In some cases, the students confused the area and perimeter and tried to compare the perimeter of the shapes instead of the area. So we asked probing questions to clarify these concepts with the students. Again, our goal was to learn what possible obstacles the students would encounter that might prevent them from correctly solving the problem.

DATA ANALYSIS

We initially watched the videotapes individually of the students solving the two problems, and then we watched and discussed key segments together to reach agreement on interpretation of interactions. We summarized the students’ interactions, based on the agreed upon interpretations, in a table using key words and short descriptions. (See Table 1 for a summary of one student’s work on both problems.) The table helped organize the data and the colored highlights allowed for a holistic view that helped us isolate themes that we saw in the video clips across problems and students. Our first task was to categorize the solutions as being correct or incorrect based on the students’ explanations. In doing this, we ignored careless mistakes (e.g. choosing an incorrect answer in the multiple choice question only if the student had verbally given a correct explanation). A solution was considered incorrect if the student had the correct answer but could not back it up with a correct explanation. For example, a student who

Table 1

Name	Perimeter problem	Area Comparison problem
Student Name (sixth-grade)	Independent solution: Incorrect	Independent solution: Correct
	Does not seem to understand the question. Knows that the perimeter is the outside of the shape when asked. Makes a box around the triangle when asked to find its perimeter. Says that his teacher mentioned that they have to make a box to find the perimeter. He is not sure if it was the perimeter or the area that was referred to in the teacher's method. Realizes that he was doing the area: "That's why I messed up". Hesitant about using the lengths of the sides of the triangle to get the perimeter: "Do I add 'em?" Wants to divide 20 by 3 to get the perimeter since there are 3 numbers shown on the triangle. Again thinks that the 20 he got is the area. Interviewer reframes the problem in the context of fencing for an enclosure and the student immediately says 20 ft. On rereading the problem, he notices that he was reading "what is the length of each square?" instead of "what is the length of each side of the square?" Groans about the error and writes 5 on the sides of the square. Reflects that he got the perimeter and area mixed up.	Takes one of each shape and bends the triangle around the square to conclude that the areas are the same. Uses the "cut" and 'paste" argument to convince the interviewer. Uses a lot of gestures to make the argument. On being asked for another way, he just folds differently and uses the previous argument. Again puts a triangle and a square together and thinks that the areas are the same because by looking he can see that the small part jutting out can be put on the square. Interviewer reminds him that it was his first way and asks him if there was another way that he could convince her that used two triangles and two squares. Places the squares and the triangles together and says that they form two rectangles of equal area. Then looks at the rectangles with two triangles and says that if we cut that down the middle then it would form one square like the original. Interviewer pushes him to explain why one triangle would have the same area as one square from this argument. After going back to his first method on a few occasions he says "This is half of a rectangle (square) and this will be half of this rectangle (points to the triangle). Further convinces by saying, "Let's say that the areas of the rectangles are a 100, so when you cut 'em in half, this will be 50 and this will be 50."

Area and Perimeter confusion
Language issues/understanding the question
Connecting representations, connections in general
Recall and memorization
Visual approx
Mathematical communication

used a visual approach and could "see" that one side of the triangle was about the same length as that of the square in the Perimeter problem was considered incorrect because the student did not explain his thinking. A student who found the perimeter of the triangle by adding the three side lengths and then divided the sum by four to find the length of a side of the square was considered to have produced a correct solution. After classifying each solution attempt as correct or incorrect, we analyzed the videos closely for issues of language and communication.

We made a note of instances when the language in the problem was not clear to a student. For example, one cue was when students read the problem slowly and did so multiple times to understand it. In other cases, students requested guidance in understanding the problem. These instances were coded as *linguistic complexity*. In coding for *linguistic complexity*, we focused attention on the students' comprehension of the problem. We want to point out that distinguishing between issues related to language and those related to mathematics was not always straightforward, underscoring how intertwined these two areas are. Thus, in some cases, what we coded as linguistic complexity may have had elements of mathematical complexity.

Students' communication of their thinking was noted in the table—coded either as *area and perimeter* or *visual*—as we tracked patterns across students and across problems. For example, the code *area and perimeter* pointed to instances when the students confused either one or both of these concepts in their solutions and later in their communications of their solutions. The *visual* code focused on the students' use of grids and visual approximations of lengths in their explanations of their thought processes. These codes informed us about the students' communication *about* mathematics (Brenner, 1994).

In tracking the students' communications *in* mathematics, we paid attention to the use of definitions and technical terminology (e.g., perimeter, area, square, triangle, height, width), their conceptions of geometric figures, and their translations between representations; and we noted these instances in the table. We also looked at the instances when the students participated in a mathematical discussion. We associated the code *discourse* if the students made sense of the mathematical arguments and statements that were being discussed and tried to communicate their ideas. Another code, *connections*, captured the fluidity of translation between representations (physical, verbal and symbolic) displayed by the students. For example, if they could move fluidly from verbal explanations using physical cutouts (in the Area Comparison problem) to diagrams and written explanations on their papers.

Once the data were coded, we individually cycled between the video clips and the codes to ensure that our coding was accurate. Further, all the researchers were in agreement over the data analysis and the themes that emerged. Our description of the students' methods and communication was built from the above analysis. The coding provided us a holistic view of the data and the emerging themes. At this point we transcribed portions of the video clips that demonstrated students' communication in and about mathematics and some that exemplified linguistic complexities associated with the problems. In the next section we address general strategies used by the students, highlight issues of linguistic complexity, and discuss students' communication about mathematics and communication in mathematics.

RESULTS

We first give a brief overview of the students' independent solutions for the two problems, focusing primarily on the difficulties students had with these problems. Though we attend to students' difficulties, our goal is not to catalogue what students could not do, but to extract from these students' approaches to the problems information that might inform assessment and instruction (see Fernandes, Anhalt, and Civil, in press). After discussing the students' performance on the problems we discuss our findings in terms of linguistic complexity, communication about mathematics, and communication in mathematics.

About the Perimeter problem

Nine of the 15 students incorrectly solved the Perimeter problem. All of these students displayed, at some point in the interview, that they did not have a clear understanding of the question. Among these nine students, seven relied on what we refer to as a visual approach. The visual approach consisted of either "seeing" that the side of the triangle with length 4 was about the same as the side of the square or creating an arbitrary grid to estimate the perimeter by counting the number of squares on the border. The other two of the nine students did not choose an answer and said that they could not understand the question. In one of these cases, providing a context to the problem actually helped the student to solve it. When asked about the meaning of perimeter, six of the nine students could explain the meaning of perimeter, one student could not, and two had a wrong concept of perimeter (e.g., one student described the perimeter as the number of sides of a polygon).

Three students among the nine had difficulty connecting the two pieces of information in the problem, namely that the perimeter of the square was 20 and that a square had sides of equal length. These students put arbitrary numbers that added up to 20 as the sides of the square when asked about the length of its side. On the other hand, when queried about the properties of a square, some of them mentioned that the sides had to be of equal length. Two sixth-grade students confused area and perimeter. These students tried to enclose the triangle in a rectangle and attempted to compare the area of this rectangle to the area of the square. One of the students mentioned that he was recalling a procedure taught by his teacher and was not sure if it was for the area or the perimeter. He eventually realized that he had the perimeter and area mixed up.

The six students who arrived at the correct solution appeared to be comfortable with both conceptual and procedural understanding needed to answer this question. They knew the concept of perimeter and how to apply it to the problem and also were proficient with the arithmetic operations required in the task.

About the Area Comparison problem

Seven of the 15 students were incorrect. Five of the students used visual differences to draw conclusions. One of the five students mentioned that the square had a larger area since it had four sides as opposed to three of the triangle. A second student relied on grids and counted the squares, a third student conjectured that the triangle had a larger area since there was a part of the triangle that was sticking out after placing the shapes on top of each other. A fourth student assumed that the triangle was larger since the rectangle formed by the two triangles had sides that were larger to those of the square. The fifth student concluded that the square was larger since there was a part of the square that did not overlap with the triangle when they were superimposed. The confusion between perimeter and area accounted for the incorrect response of a sixth-grade student who moved the square around the triangle and compared the perimeters. Finally, one sixth-grade student reasoned that the area of the triangle had to be less than that of the square since there was a factor of half in the formula (1/2 x base x height as opposed to base x height).

Eight of the 15 students successfully solved this problem. In most cases, students compared one triangle and one square and used the area preservation property to work through the problem as shown below **(Fig. 2)**. The arrow indicated a "cut" and "paste" operation, in which part of the triangle was cut and arranged in the dark area. This operation helped the students draw the conclusion that the areas of the triangle and the square were the same. Four of the eight students came up with a method that used the pair of triangles and squares. These students

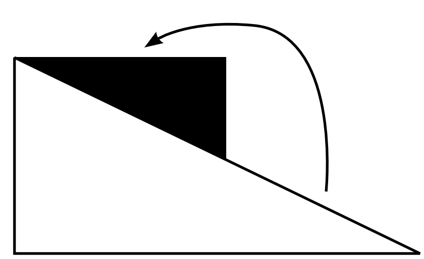


Figure 2: The "cut" and "paste" strategy in the Area Comparison problem

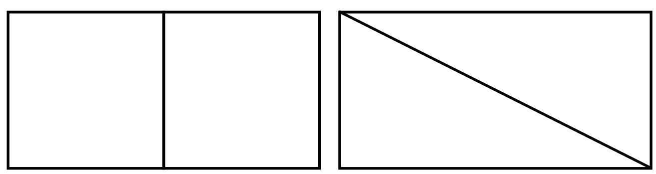


Figure 3: Strategy involving all four cut-outs in the Area Comparison problem

observed that by placing the two triangles together and the two squares together, the two newly formed rectangles were equal in area, and each triangle and square represented half the area of the respective rectangles **(Fig. 3)**.

We now discuss areas of linguistic complexity, communication in, and communication about mathematics as observed across problems and students in our data analysis.

Linguistic Complexity

We observed two forms of linguistic complexity, one with regard to reading and comprehending the test items and the other with regard to the written answers. Note that the students were expected to write an explanation only for the Area Comparison problem (as per NAEP instructions). The first form of linguistic complexity that we observed was in phrasing of the Perimeter problem. The problem read, "If both the square and the triangle above have the same perimeter, what is the length of each side of the square?" One of the fourth-grade students interpreted the "if" statement as "they do not have the same perimeter." When probed, she said "but they do not because it says IF (emphasis added)." This child was interpreting the "if" statement as a negation statement, therefore, the square and the triangle could not possibly have the same perimeter. Her facial expression indicated that she was faced with conflict and was not able to engage with the mathematics as the problem intended.

Interviewer: Okay, now read the first part of the problem again, what does it say?

Student 1: If both the square and the triangle above have the same perimeter, what is the length of each side of the square?

Interviewer: So does that triangle and that square, do they have the same perimeter? (pause)

Student 1: No.

Interviewer: And what does the problem say? The first part.

Student 1: If both the square and the triangle above have the same perimeter ...

Interviewer: Okay, stop. What does that mean? What are they telling you?

Student 1: That if both of them have the same perimeter, what is the length of each side?

Interviewer: So, are they supposed to have the same perimeter?

Student 1: Well, yeah.

Interviewer: According to the question, are they telling you that they have the same perimeter? [Student 1 shakes head 'no']

Interviewer: No, how is it? Read that part again.

Student 1: If both the square and the triangle above have the same perimeter, . . .

Interviewer: Okay, stop right there, if both triangle and the square have the same perimeter, are they telling you that they have the same perimeter?

Student 1: [shakes head 'no']

Interviewer: No, how come they're not?

Student 1: Because they're saying, "IF both."

Interviewer: What does that mean?

Student 1: That means that. . .like. . .if. . .if both of them have the same perimeter.

Interviewer: So are they saying, "If both of them, meaning they don't have it, but if they did," that's how you interpret the question that they're asking?

Student 1: [nods 'yes']

By the student's interpretation of the language used in this problem, it was difficult to assess her mathematical understanding of two shapes having the same perimeter. In this case, the interviewer decided to drop the word "if" and rephrased the question to, "The square and the triangle above have the same perimeter," and then asked the student to proceed with the problem. However, this same student was unable to solve the problem as she continued because she made an assumption that the square had "four sides" and not four equal sides. Thus, in her discussion of her solution, she noted that the sides of the square were 1,3,7,9. Our overall conclusion was that this student's difficulties with this problem involved both linguistic aspects and mathematical concepts. We contrast this episode to our next example of a sixth-grade student who solved the same problem.

The sixth-grade student first added the sides of the triangle to get 20 and then mentioned that she did not understand the question. After the interviewer explained the question to her, she was able to figure out the length of the sides of the square mentally.

[First she adds the lengths of the sides of the triangle and works that out to be 20.]

Student 2: I don't understand it.

Interviewer: Okay, do you want me to read it or do you want to read it aloud?

[The student reads the question.]

Interviewer: Do you know what the word perimeter means?

Student 2: The outside of umm [points to shapes]

Interviewer: Okay, and so they give you this triangle and give you the measurements of the sides of this triangle and they give you a square and they don't give you the sides. So, they are telling you that "If the square and the triangle have the same (emphasis) perimeter what is the length of the sides of this square?"

Student 2: Five!

Interviewer: How'd you do that?

Student 2: This together is 20 [points to the triangle and the work she did before] and this has 4 sides [pointing to the square] and 5 times 4 is 20.

In this case it seems that the student had not seen the relevance of the word "same" in "same perimeter." By emphasizing the word "same", the interviewer assisted the student in comprehending the problem, and after that she could successfully solve the problem. We wonder if the student's difficulty in the problem was not noticing the word "same" or again an issue with the "if-then" statement.

Overall, the students found the Perimeter problem more linguistically challenging than the Area Comparison problem. For this second problem, the main linguistic complexity occurred when the students were asked to give a written explanation for their work. Among the eight students who correctly solved the problem, three students did not provide any explanation for their work. We did not insist that they provide a written solution if they expressed an inability or reluctance to do so. Out of the other five correct responses, only one could have been interpreted as a correct solution if no verbal communication had taken place with the student. The written work of the remaining four students would be difficult to interpret without interacting with them. These students were able to explain their work orally but had difficulties providing written explanations. Although this may be the case for many students, we argue that this is more likely with ELL students who, in communicating orally, were able to use (and did use) gestures and were able to interact with the interviewer about the mathematics. Moschkovich (2007a) also found that ELL students were better at orally expressing their ideas but were not mathematically precise in writing. This is an important issue to consider for teaching in assigning and grading tasks submitted by ELL students.

In another example, a fifth-grade student provided a figure and a written explanation for the Area Comparison problem (**Fig. 4**), but the explanation would be difficult to interpret without interacting with the student. The student writes "If N has 4 units but is a different shape than P, both of them might have different units. But when I cut out N into the shape of P, having still 4 units, P would have 4 units also." The student meant that the square and the triangle were different shapes and he assumed that the square was 4 units and conjectured that the triangle "might have different units" (a different area). But then he observed that by placing the triangle on top of the square, a part of the square could be cut and rearranged into the triangle. We confirmed this latter aspect through interactions with the student. His written explanation did not cover all the thinking that he displayed with the researcher and it could have been misinterpreted. Only one fifth-grade student provided a complete explanation in words and a diagram to go along with her verbal interactions with the researcher **(Fig. 5)**. She wrote, "I put the square on top of the other shape. Since [I] saw that one little space was left I put that space in the square and saw that N and P were the same area."

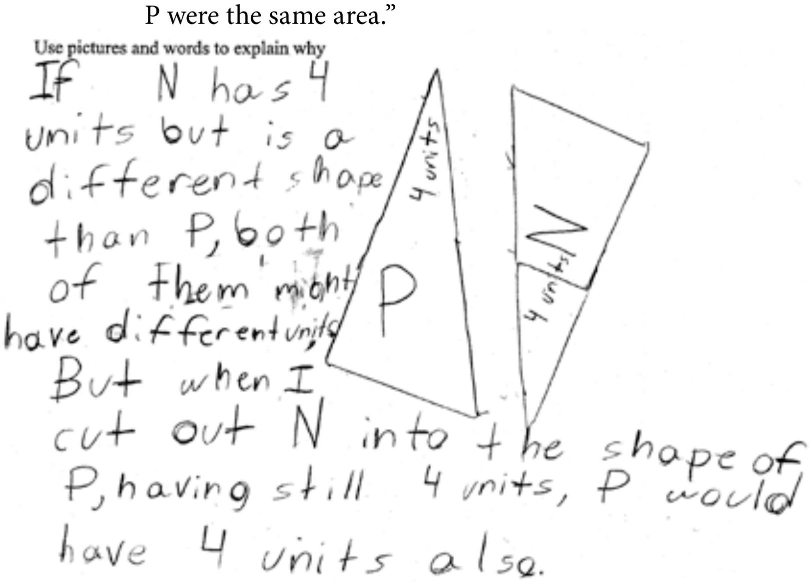


Figure 4: An example of an unclear written solution of the Area Comparison problem

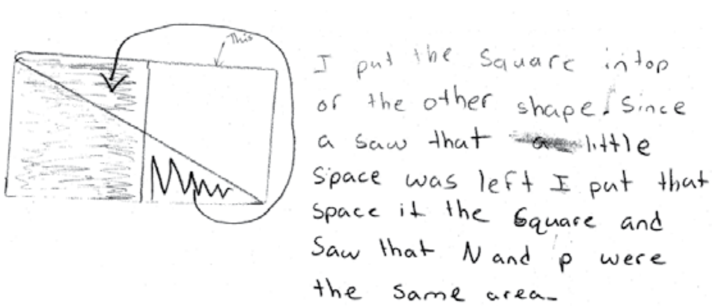


Figure 5: An example of an clear written solution of the Area Comparison problem

Communication *about* Mathematics

According to Brenner's *Communication Framework* (1994), communication about mathematics entailed the need for the students to describe the problem-solving processes and their own thoughts about these processes. The majority of the students in the study were able to describe the process when asked to explain their thinking, even if their reasoning was not complete. There were a few cases in which the students guessed the answer and could not provide a reason. Most of the students were able to convey their ideas either on their own or with some probing.

In communicating their solutions, some students confused area and perimeter conceptually in their explanations. For example, in the Area Comparison problem, a fifth-grade student concluded that the area of the triangle was larger than the square. On being asked to explain her thinking, she rotated the square cut-out around the triangular cut-out **(Fig. 6)**, thus comparing perimeters instead of areas. On being asked about area and perimeter, the student said that the perimeter "was the outside of the shape" and the area was "the inside of the shape." However, she was confused when trying to use these ideas in solving the problem.

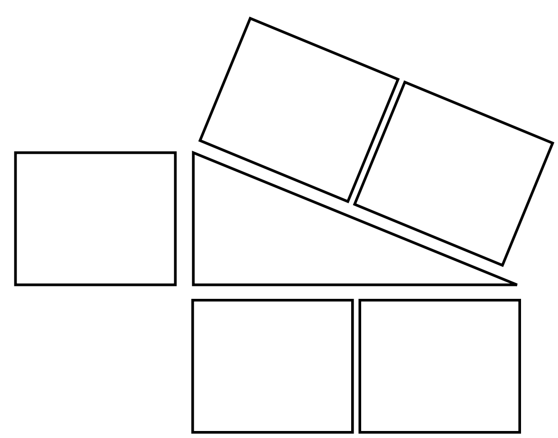


Figure 6: An example of the confusion between area and perimeter

Some students relied on visual cues and drawings as their key sources of information and in explaining their work. A few students requested a ruler for finding the lengths of the sides for some of the figures in the Area Comparison problem since no measurements were provided. On the Perimeter problem, a fourth-grade student drew the 5 x 5 grid **(Fig. 7)** in the square and concluded that the length of the side had to be 5, which was the correct answer. This student stated that he did not need to know the other conditions given in the problem. During probing, the researcher asked the student about drawing a 10 x 10 grid and if the side of the square was now 10? This caused a conflict in the student's thinking that he was later able to resolve with more guidance from the researcher, but this left us wondering about this student's understanding of the concepts involved.

The students who were successful in explaining their solutions and reasoning were able to make connections during their discussions about mathematics. Two successful students, on the Area Comparison problem, reasoned that the area of the triangle and the square were the same by assigning numbers to the unknown lengths and areas in a way that accurately reflected the relationships. One of these students (in sixth-grade) assumed that the rectangles

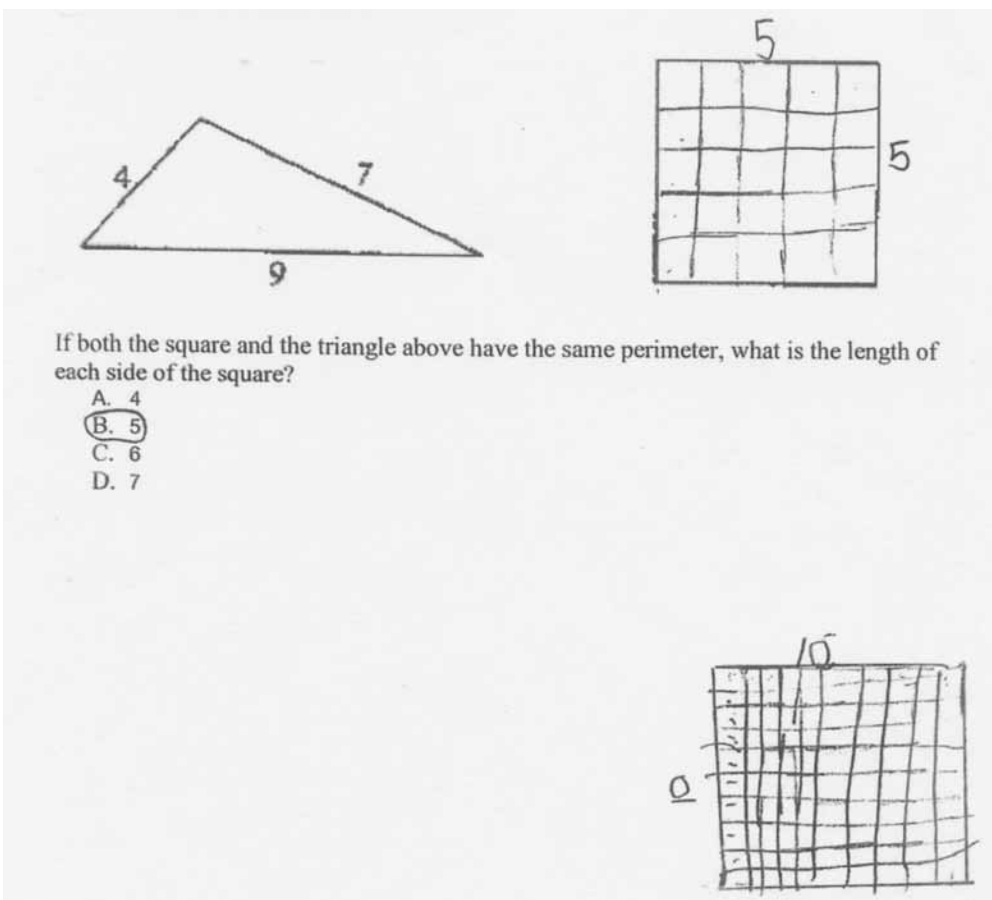


Figure 7: An example of the use of visual cues in the Perimeter problem

formed by the two triangles and the two squares had an area of 100 and so one triangle and one square each had an area of 50.

> **Interviewer:**. . .I am going to leave here what you had a second ago in front of you, [Interviewer arranges the squares and the triangles to form two rectangles near each other] and you just told me that this [rectangle formed from the two squares] and this [rectangle formed from the two triangles] are the same . . . area you mean or the . . .
>
> **Student 4:** [interrupts] yeah, area ... the same area.
>
> **Interviewer:** Okay, same area, same area [points to rectangles]. How can you use these to tell me how the area of one of them [points to square] compares to the area of one of them [points to the triangle]?
>
> **Student 4:** 'Cause this is half of a rectangle [points to the square] and this will be half [points to the triangle] of this rectangle [points to the rectangle formed from the two triangles]
>
> **Interviewer:** Ah, huh.
>
> **Student 4:** Like, say this rectangle ... they both . . . the area is a 100, and if you cut them in half, this will be 50 [holds up the square] and this will be 50 [holds up the triangle]

The use of the 100 and the 50 by the student is especially significant when compared to the explanations of the other students who did not make progress in the Area Comparison problem when asked to use all the four shapes. The unsuccessful students usually concluded that the two rectangles formed with the two squares and the two triangles had equal area but were unable to use this to conclude that the individual triangle and square had equal areas. It was this latter point that the sixth-grade student communicated so elegantly through his use of the numbers 100 and 50.

Communication *in* Mathematics

Communication in mathematics (Brenner, 1994) referred to students' proper use of language and symbols of mathematical conventions. Students who were successful with the Perimeter and Area Comparison problems displayed competent communication in mathematics. These students could calculate the perimeterand work out the length of the side of the square, either using division or addition strategies. Further, in the Area Comparison problem, the students understood the concept of area preservation and were adept at showing that the square and the triangle had the same area by cutting and rearranging either a portion of the square or the triangle.

We found that the students who were successful in solving these problems were fluid in their translations between representations. For example, a successful sixth-grade student simultaneously represented his manipulations of the concrete shapes in the Area Comparison problem with the equation 2P=2N ("P" represented the triangle and "N" represented the square). He reasoned that if 2P=2N, then half of each rectangle is the triangle P and the square N, therefore, P=N and the areas were equal. Here is the dialogue after the student had written 2P= 2N:

> **Student 5:** Two Ps equals two Ns.
>
> Interviewer: Okay, can you tell me . . . I mean this is a very nice . . . you know algebraic expression . . . can you tell me what it means? I like . . .
>
> **Student 5:** Okay. Two of these which are Ps [holds up triangle] equals two of the Ns which are squares [holds up the square] because as you can see, you put these together . . . like that [the triangles], and if you put these together they makes a rectangle, and if you put these on top then it makes the . . . the squares [repeats making the two triangles] put the squares on top, and it's the same size.
>
> **Interviewer:** Okay . . . same size okay, so that's nice, so 2P=2N, so how can you use that information if you can to tell me about the area of one N and one P?
>
> **Student 5:** They are the same.
>
> **Interviewer:** And why?
>
> **Student 5:** Because . . . because it's like if you are saying . . . like ,if you were to cut this in half [the rectangle], if you cut this off right here [puts the triangle and the square together and then indicates a cutting motion with his fingers], you could put it right here [points to where the extra part of the triangle would fit into the square], and if you were to do that, then you could do it on this side, too [holds up the other triangle and square], this one . . . which are the same [brings the things that he is holding in his hands together] so they are the same size, the square . . . well N and P are the same area.

In this case it is interesting to note that the student only wrote "2P=2N" and "P=N" on his paper, and this did not capture the innovative thinking and verbal explanation of his solution. The student used the concept of the conservation of area to explain "P=N" rather than divide both sides of "2P=2N" by 2. At the time, this surprised us somewhat since the interviewer's thinking was that from 2P = 2N, the student would perhaps divide both sides by 2 and conclude P = N, but in looking at all the data from all the students, we noticed that this type of algebraic thinking did not occur.

There were some students who did not use precise language and terminology in explaining their work. These students used a lot of referents and had to be constantly probed by the researcher to get a clear meaning of their work. For example, below is the dialogue of a sixth-grade student (an advanced ELL), in which the interviewer constantly asked her to clarify her meaning of "length" as she discussed her work on the Perimeter problem.

[Works on the Perimeter problem on her own and chooses 5 as her solution]

Interviewer: Okay, can you tell me how you came up with 5? [pause]

Student 6: Well, umm [clears throat], since the … length of the square, this one [points to the triangle instead] is the same …

Interviewer: The length of the square is the same as what?

Student 6: This one [points to the square].

Interviewer: This is the square [points to the square] right?

Student 6: uh-huh.

Interviewer: … what do you mean by the length of the square is the same? The same as ... as …

Student 6: No this is the same [points to the triangle] as this [points to the square]

Interviewer: Okay, when you say this, you mean the shape?

Student 6: The length.

Interviewer: The length … the length of what? [pause] I mean there is something there that is the same … yes … but I am not really sure I'm understanding what you are saying.

In this discussion, the interviewer probed for more clarifications about what the student referred to in the figure. By saying that "this [pointing to the triangle] and this [pointing to the square] are the same," it is not clear if the student was referring to the lengths of the sides, perimeter, or area. Here the student did not make use of precise vocabulary that was required to communicate her ideas.

Proper mathematical communication was displayed in the interactions with a sixth-grade student who successfully solved the Perimeter problem. The student was very clear in the process he used to arrive at the solution and could justify his steps when asked by the researcher. Further, the student knew the mathematical terms that were part of the problem such as perimeter, square, etc. He also was proficient in translating between the various representations of diagrams, verbal, and symbolic representations. This was representative of the conversations between other successful students and the researcher. Here this sixth-grade student was being queried about his method in solving the Perimeter problem.

[Thinks about the problem.]

Interviewer: Okay, do you want to explain your thinking?

Student 7: Ahh, I added, I added, uhh 4, 7, and 9. I got 20 and then I divided 20 by 4 and I got (hesitates), I got ... oh I messed up [erases his choice of 4]

Interviewer: Oh, you are changing your answer?

Student 7: Yeah, I messed up on this one.

Interviewer: Why did you change it? What happened?

Student 7: Because it's 20, the, the [indicates all around the square] the (unclear) whole perimeter of the square is 20, so I know, umm ... there are four, umm … divided by 20 equals 5. So, it's 5 for each side.

Interviewer: Okay, tell me about the math sentence that you just mentioned … umm … you added this and got 20 [points to the triangle] and then tell me what you did over here?

Student 7: I, I, umm, divided 20 by 4.

Interviewer: Oh, okay, and 20 divided by 4 …

Student 7: Equals 5.

Interviewer: Equals 5. Okay, I see what you did. Initially, when you got the 4, what were you thinking? Did you just . . .

Student 7: Divided by 5 instead of 4 . . . I messed up

Interviewer: Oh, okay ... and how did you know to divide by 4?

Student 7: Well, because I know that 4 times 5 equals 20

Interviewer: Okay, . . . umm, how did you know to take the 20 and divide it by 4?

Student 7: Because there's 4 sides.

Interviewer: Oh, what do you know about the sides of a square?

Student 7: It has 4 . . . 4 equal sides.

Interviewer: Ahh . . . equal, okay, I see your thinking.

The student showed a good understanding of participating in mathematical communication. He knew how to go about solving the problem and explains the process to the researcher. He also was able to recognize and rectify his error independent of the interviewer. The student understood the concept of perimeter and made proper use of the mathematical terminology.

DISCUSSION

Through this study we sought to a) gain an understanding of how a group of 15 Mexican-American students approached selected NAEP measurement items and b) uncover some of the challenges that these items presented to this group of students. In this section, we address these goals and include teaching implications that emerged from our analysis and some consideration for assessments. Examining the overall results from the interviews, the linguistic complexity was a challenge that seemed to be prominent with this group of Latino/a students, especially in the case of the Perimeter problem. These students' struggle with the hypothetical assumption involved with the 'if' in the statement of the Perimeter problem. This is similar to the observation of Fillmore (2007) who described similar struggles of the students with the word 'suppose' in the sentence, "For example, suppose you are randomly choosing marbles one after another, …" (p. 340). Fillmore pointed out that in some cases these students were categorized as English 'proficient' and yet struggled with the academic discourse that was needed. We agree that hypothetical assumptions such as "if-then" and "suppose" may be difficult for native speakers of English, and we argue that these may even be more difficult for ELLs or, at the very least, should be an area for teachers and assessment developers to give serious consideration.

Compared to the Perimeter problem, the Area Comparison problem was more accessible to the students, and there were fewer issues of linguistic complexity observed in the students' understanding of the question. This could be attributed to the lower complexity in the language used and the presence of cut-outs. These shapes helped mediate the students' thinking and communication with the researcher as the students could indicate their thinking processes by manipulating the shapes. Further, the cut-outs allowed for informal methods that may have connected better to what they did in their regular classroom. Even though we have focused on the results of two problems in this article, the bigger study also indicates that the ELL students whom we interviewed were more comfortable with problems that included visuals or concrete objects. In the case of the Area Comparison problem, it is interesting to note that a bigger percentage of our students were correct compared to the results on the NAEP. A probable reason for this difference could be that in the interview setting we accepted a verbal explanation of their reasoning and this verbal exchange was linguistically less challenging for these students.

Our findings show that most students were able to communicate about and in mathematics. In some cases, the students provided an incorrect answer, but they could explain their thought processes. Although we saw a number of cases where the students could verbally express their thinking, it was much more difficult to convey their thinking in writing. This was the case of the Area Comparison problem where the task asked for a written explanation. Our findings underscore the need to provide students with multiple opportunities to express their thinking in a pencil and paper format and to reinforce their learning of academic English. Though this is necessary for all students, it may be particularly important for students, such as those in our study, who are juggling two languages and may be at varying levels of English proficiency.

We conjecture that the linguistic difficulties that our students faced in understanding the questions influenced their approaches to the problem. If students did not understand the question completely, they relied on other resources like visual cues, concrete objects, numbers, and partial information from the question. They also relied on their memory of activities that they did in the classroom, such as the grids that seemed to bring about some confusion between area and perimeter. Although these strategies may also be observed in non-ELL students, we posit that ELL students, especially in English-only classrooms, may take a longer time to understand all the subtleties in a teaching approach of a mathematical concept given that they are grappling with understanding the language and learning the concept at the same time (also expressed by Chamot and O' Malley, 1994).

Since we also have classroom observation data for many of these students, we can confirm that, in fact, several of the teachers often used manipulatives in their teaching of mathematics. Likewise, some of these teachers used a grid approach to finding the area of rectangles (the area model), and this representation seemed to have a powerful influence on some of the students in our study. Our interviews showed several students using grids in the Perimeter problem. Although students in general were able to relate area to the "inside of the shape" and perimeter to the "outside of the shape," their understanding of these concepts was not clear in some cases. For example, some students constructed a grid on their figure and counted the squares to work out the area and the perimeter. In this strategy there was a potential for the students to either count the wrong set of squares, and thereby confuse the concepts of area and perimeter, or they could confuse the units when measuring the perimeter (since they are counting squares on the boundary of the shape and could think that the perimeter was 10 square units instead of 10 units). The students needed more experience in making a transition from working with the grids to working without them and the links between the grids, linear measurements, and concepts of area and perimeter. Kamii and Kysh (2006) also have discussed potential points of confusion when discrete quantities like unit squares are used to help students understand the concept of area, a continuous quantity. Successful students had made the transition from using the grids to operating solely with the lengths.

Some Possible Implications for Classroom Instruction

This study shows the potential of certain NAEP items as tools to elicit thinking from all students but to collect especially vital information in the case of Latino/a students. This information can be fed back into the planning of further instruction, which is the key step in formative assessments. The Perimeter problem presented several linguistic and mathematical challenges (e.g., the "if-then" statement; the segment on "same perimeter"; the misleading visual appearance of the two shapes). Knowing the kinds of difficulties that students in our study faced with the problem could inform teachers to provide more experiences to address these directly. For example, a teacher could engage with the students on equivalent (from the mathematical point of view) rewritings of problems having an "if-then" statement, but switching from "if the square and the triangle above have the same perimeter" to "the square and the triangle have the same perimeter."

The use of cut-outs in the Area Comparison problem enabled students to communicate their thinking by means of words and gestures while manipulating the shapes; this wider definition of communication that includes multiple resources that students can use to express their thinking is particularly significant with students who are working with more than one language (Moschkovich, 2002). This also is the case for the use of multiple representations; however, as our study shows, it is necessary to pay close attention to how students are interpreting these representations (e.g., the use of grids to find the area of a shape) and what kinds of mathematical connections they are making across representations.

The teacher also can build on some of the strategies used by students. For example, the teacher could first allow the students to assign concrete numbers for the Area Comparison problem if it helps them understand the underlying relationships. Later in a class discussion, the teacher could ask the students to formulate a general statement that would hold regardless of the particular numbers that they chose. Further, by paying attention to the strategies provided by the students, the teacher could link the exclusive use of visual cues or the partial use of information from the problem to probable linguistic difficulties that the students may have and provide the appropriate support.

CONCLUSION

This study gives a closer look at Latino/a students' thinking, which is often masked by multiple-choice test items. For example, in our study, the linguistic complexity of the question interfered with the students' mathematical thinking, and the nature of this interference was only clear after interacting with the students. Further, paper-and-pencil tests that rely solely on the written work of the students (even in the case of constructed response tasks) also may fail to reveal important facets of student thinking. We found that our students' written responses in the case of the Area Comparison problem were difficult to understand and had to be amplified through verbal elaborations. In our interactions with students, we observed that it was important to consider the students' use of language resources to explain their thinking, and thus it is important to broaden our conception of competency in mathematics. The use of resources also has been discussed by Moschkovich (2002, 2007) and Radford, Bardini, and Sabena (2007), and they are in agreement that it is not possible to capture the students' mathematical thinking only by examining their written work.

We conjecture that interviewing students with NAEP mathematical tasks that show large gaps in performances of different groups by ethnicities could be used to draw out interesting thinking from the students. This information could improve instruction and has the potential for promoting equity for all students in terms of reaching more students of diverse linguistic backgrounds. The interviewing process and the newly gained perspectives have heightened our understanding of student thinking, especially in the case of ELL and multilingual students. Improvement in instruction is more likely to happen with increased teacher understanding of students' thinking of the mathematics through dialogue that engages students in communication *about* and *in* mathematics.

NOTES

(1) We use the term *Latinos/as* to refer to the student population in the U.S. whose origins are of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture regardless of race as defined by *The Oxford Encyclopedia of Latinos and Latinas in the United*

States, 4 vls, Oxford University Press 2006. In our local context, most Latino/a students are of Mexican origin. Because NAEP reports use the term *Hispanic* to refer to this population, that term is used in this paper when we refer to data from NAEP reports.

ACKNOWLEDGEMENT

CEMELA is funded by the National Science Foundation under grant No. ESI 0424983. The views expressed here are those of the authors and do not necessarily reflect the views of the funding agency.

REFERENCES

Brenner, M.E. (1998). Adding cognition to the formula for culturally relevant instruction in mathematics. *Anthropology & Education Quarterly*, 29, 214-244.

Brenner, M.E. (1994). A communication framework for mathematics: Exemplary instruction for culturally and linguistically diverse students. In B. Mc Leod (ed.), *Language and learning: Educating linguistically diverse students*. New York, NY: State University of New York Press.

Campbell, A.E., V.M. Adams, and G.E. Davis. (2007). Cognitive demands and second-language learners: A framework for analyzing mathematics instructional contexts. *Mathematical Thinking and Learning*, 9, 3-30.

Charmot, A.U., and J.M. O' Malley. (1994). The CALLA *Handbook: Implementing the cognitive academic language learning approach*. New York, NY: Longman.

Civil, M. (2006). Working towards equity in mathematics education: A focus on learners, teachers, and parents. In S. Alatorre, J.L. Cortina, M. Sáiz, and A. Méndez (eds.), *Proceedings of the twenty-eighth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 30-50). Mérida, Mexico: Universidad Pedagógica Nacional.

Cobb, P., and E. Yackel. (1996). Constructivist, emergent, and sociocultural perspectives in the context of developmental research. *Educational Psychologist,* 31, 175-190.

Fernandes, A., C.O. Anhalt, and M. Civil, 2009. With an eye toward equity: Assessing Latino students through mathematical interviews. *Teaching Children Mathematics.*

Fillmore, L.W. (2007). English learners and mathematics learning: Language issues to consider. In A.H. Schoenfeld (ed.), *Assessing mathematical proficiency* (pp. 334-344). New York, NY: Cambridge University Press.

Goldin, G.A. (2000). A scientific perspective on structured, task-based interviews in mathematics education research. In A.E. Kelly and R.A. Lesh (eds.), *Handbook of research design in mathematics and science education* (pp. 517-546). Mahwah, NJ: Lawrence Erlbaum Associates.

Kamii, C., and J. Kysh. (2006). The difficulty of "length x width": Is a square the unit of measurement? Journal of *Mathematical Behavior*, 25, 105-115.

Lubienski, S.T. (2003). Is our teaching measuring up? Race-, SES-, and gender-related gaps in measurement achievement. In D.H. Clements and G. Bright (eds.), *Learning and teaching measurement* (pp. 282-292). Reston, VA: National Council of Teachers of Mathematics.

Moschkovich, J. (2007a). Beyond words to mathematical content: *Assessing English learners in the mathematics classroom*. In A.H. Schoenfeld (ed.), Assessing mathematical proficiency (pp. 345-352). New York, NY: Cambridge University Press.

Moschkovich, J. (2007b). Bilingual mathematics learners: How views of language, bilingual learners, and mathematical communication affect instruction. In N. Nasir and P. Cobb (eds.), *Improving access to mathematics: Diversity and equity in the classroom*. New York, NY: Teachers College Press.

Moschkovich, J. (2002). A situated and sociocutural perspective on bilingual mathematics learners. *Mathematical Thinking and Learning*, 4, 189-212.

Radford, L., C. Bardini, and C. Sabena. (2007). Perceiving the general: The multisemiotic dimension of students' algebraic activity. *Journal for Research in Mathematics Education*, 38, 507-530.

Strutchens, M.E., W.G. Martin, and P.A. Kenney. (2003). What students know about measurement: Perspectives from the National Assessment of Educational Progress. In D.H. Clements and G. Bright (eds.), *Learning and teaching measurement* (pp. 195-207). Reston, VA: National Council of Teachers of Mathematics.

Wiliam, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F.K. Lester (ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 2, pp. 1051-1098). Charlotte, NC: Information Age Publishing.

Cultivating a Culturally Affirming and Empowering Learning Environment for Latino/a Youth through Formative Assessment

Richard S. Kitchen • The University of New Mexico

Laura Burr • The University of New Mexico

Libni B. Castellón • The University of New Mexico

INTRODUCTION

Non-Latino teachers' characteristic lack of knowledge of the Spanish language and dismissive attitude toward Mexican culture makes them unlikely to be familiar with the cultural definition of educación. Thus, when teachers deny their [Latino/a] students the opportunity to engage in reciprocal relationships, they simultaneously invalidate the definition of education that most of these young people embrace. And, since that definition is thoroughly grounded in Mexican culture, its rejection constitutes a dismissal of their culture as well. (Valenzuela, 1999, p. 23)

In this excerpt from her book, *Subtractive school: U.S.-Mexican youth and the politics of caring*, Angela Valenzuela describes what happens when teachers do not understand the importance ascribed by the family in Mexican culture to educate youth nor sufficiently respect the inherent dignity of the individual. To be *bien educada/o* (well-educated), one knows "how to live in the world as a caring, responsible, well-mannered, and respectful human being" (Valenzuela, 1999, p. 23). Unfortunately, the research literature is endemic with examples of how teachers and administrators seldom form the sorts of meaningful relationships with youth valued within Mexican culture (e.g., see, Suárez-Orozco, 1989; Valenzuela, 1999). Even more disturbing, the attitudes of school administrators and teachers toward Latino/a parents and students are often negative and result in low expectations and a lack of academic rigor (McKown and Weinstein, 2008). Not surprisingly, "real learning [for Latino/a youth] is difficult to sustain in an atmosphere rife with mistrust" (Valenzuela, 1999, p. 5).

For years, researchers have documented how low academic expectations so often have been the norm for racial/ethnic minorities such as Latinos/as[5] and students living in poverty (Ferguson, 1998; Grant, 1989; Khisty, 1995; Knapp and Woolverton, 1995; Winfield, 1986; Zeichner, 1996). Latino/a students often attend crowded schools in poor neighborhoods and, in most of cases, those schools do not have adequate resources to attend to students' needs (Borjian, 2008; Fry, 2005; Lockwood and

[5] We use the term "Latino/a" to denote a person of Latin-American or Spanish-speaking descent <http://en.wikipedia.org/wiki/Latino>. We often use "Latino/a" and Mexican interchangeably, though not all Latinos/as are of Mexican descent.

Secada, 1999). In their review of social class and schooling, Knapp and Woolverton (1995) claimed that controlled forms of instruction teach students living in poverty that little is expected from them except compliance to a rigid classroom environment. Similarly, studies have documented how educators of Latino/a students often make the memorization of math facts, algorithms, vocabulary, and procedures the focal point of their instruction, rather than teaching students using complex, challenging problems (Flores, 2007; Moschkovich, 2007).

Immigrant Latino/a students experience additional challenges in U.S. schools. They often enter U.S. schools performing below their English speaking peers in core academic subjects such as mathematics and their academic progress usually is measured with inadequate tools that do not accurately represent their learning (Abedi and Gándara, 2006; Abedi and Lord, 2001). Research also has documented how immigrant Latino/a youth often are not mentored in ways that could assist them to have more success in school or to be better represented in honors-level courses (e.g., see Romo and Falbo, 1996; Olsen, 1997). Lastly, instead of viewing the language resources that immigrant Latino/a youth bring as "Spanish dominant" or as potential bilinguals, they are generally pigeonholed as "limited English proficient" (Valenzuela, 1999). When immigrant Latino/a students speak with an accent, use English words incorrectly, or speak in Spanish as a means to express themselves, educators, peers, and community members may assume they lack the capacity to perform well in mathematics (Gutiérrez, 2007; Moll and Ruiz, 2002; Moschkovich, 2007).

"Deficit perspectives" such as these attribute lower levels of academic achievement to specific ethnic/racial groups based upon characteristics such as lack of fluency in English, life experiences that do not parallel those of the dominant society, or low family income (Khisty, 1995; Lubienski, 2007). Instead of looking at students and their communities through a deficit lens, they can be viewed as having funds of knowledge such as knowing one language and learning another, having experiences that are richly grounded in their culture, and having extensive mathematics experiences in their daily lives (Moll and Ruiz, 2002). If educators build on the attributes students possess and treat them as mathematically competent, there is greater potential for increased academic success and an enhanced mathematical identity (Empson, 2003; Turner, Celedón-Pattichis, and Marshall, 2008).

To counteract deficit views of immigrant Latino/a students, researchers have documented how their prior knowledge, language, and culture must be integrated into instruction and assessment tasks (Abedi and Gándara, 2006; Abedi and Lord, 2001; Lockwood and Secada, 1999). Bilingual students benefit from the use of their home language and other feedback techniques; they benefit from the translation and explanation of key words and sentences, by reflecting on their own thinking, assessing their own errors, and having teachers and peers revoice their explanations (Abedi and Gándara, 2006; Borjian, 2008; Lockwood and Secada, 1999). Instruction of immigrant Latino/a students should be supportive of students using resources such as gestures, concrete objects such as drawings, and the use of their first language to communicate their mathematical thinking (Moschkovich, 2002).

In the study reported here, immigrant Mexican students often used gestures and mathematical representations to explain complex mathematical ideas. Commencing in spring 2008, we began conducting intensive one-on-one interviews with four bilingual immigrant Mexican students to document the progression of their mathematical thinking through a series of rational number tasks. Using a formative assessment format that we refer to as the "interactive interview protocol," two interviewers created a learning environment in which the four participating students had multiple opportunities to solve and then refine their solutions to the tasks. The following overarching research question guided our research:

> *In what ways can an interactive interview assessment protocol as a formative assessment tool support the demonstration of mathematical knowledge of sixth-grade, bilingual students?*

We set out to understand whether the interactive interview protocol fostered equitable and accessible bilingual learning opportunities for the participants. We quickly learned that the research protocol supported the development of positive relationships and interactions among the participating students and researchers. Furthermore, through the use of the protocol, the researchers prioritized respecting the students' thinking and actively sought out their mathematical ideas. Subsequently, students responded in very positive, innovative, and empowered ways to mathematics. Before describing the study and our findings in more detail, an introduction to alternative assessment formats in mathematics is first provided. We then proceed to offer the theoretical framework used to interpret the mathematical thinking of the participating students and the context within which this thinking took place. After delineating the methodology used in this investigation, we will summarize and analyze the research findings. We conclude with a final discussion of the pertinence of this study to the research literature.

An introduction to alternative assessment formats in mathematics

For almost two decades, researchers and policymakers have been advocating for revisions in assessment practices to bring about changes in instruction based on how children learn (e.g., Kulm, 1994; O'Day and Smith, 1993). Often coined as "alternative assessment formats," new approaches to assessment promote higher order thinking among students, elicit a range of student responses, and require students to communicate their thinking (Wiggins, 1993). Alternative assessment formats align with mathematic education reforms (National Council of Teachers of Mathematics, 1989, 1995, 2000; National Science Foundation, 1996) in which the primary goal is for students to develop mathematical understanding by making connections, communicating, representing, and problem solving (Hiebert and Carpenter, 1992; Hiebert, Carpenter, Fennema, Fuson, Human, Murray, Olivier, and Wearne, 1996).

At a time when large-scale assessments such as No Child Left Behind mandated tests (NCLB, 2001) dominate the landscape, alternative assessment formats for use at the classroom-level are receiving little attention. This is troubling primarily because of the limitations of "the test" to demonstrate the depth of students' mathematical reasoning, particularly for students who may speak a language other than English as their first language (Abedi and Lord, 2001).

Classroom assessments are used to inform teachers, students, and parents about student knowledge and understanding of mathematical concepts, processes, and skills (Wiggins, 1993). There are two categories of classroom assessments: summative and formative. Summative assessments are most often done individually but can be done in dyads or groups (Fuchs, Fuchs, Karns, Hamlett, Katzaroff, and Dutka, 1998). Summative assessment formats focus on what students know at a given time (Guskey and Bailey, 2001). Formative assessments differ from summative assessment in that the focus is not just on summarizing students' learning, but on using student learning data to inform instruction. After examining 250 research studies on classroom assessments, Black and Wiliam (2001) found that when teachers focus on formative assessment, student achievement gains are among the largest ever reported for educational interventions. Formative assessments at the level of the classroom can include any of the following: classroom observation, inquiry, group work, whole class discussions, peer assessment, written work, individual interviews, student self-assessment, and portfolio assessment (Gearhart and Saxe, 2004; Stiggins, 2001).

For the most part, assessment formats utilized in classrooms do not yet reflect the paradigm change that places a premium on students making sense of mathematics. Shepard (2000) and Pegg (2003) call for changes in assessment to reflect a change in practice in which students actively make meaning of mathematical concepts by building on their previous knowledge and experiences and making connections to previous knowledge and new understandings. Our goal in designing the interactive interview protocol was to create a formative assessment format that would promote mathematical sense making for bilingual Mexican immigrants. A fundamental belief of our research team is that promoting classroom-level assessment practices that dynamically validate students' ideas and thinking is a social justice issue, particularly for Latino/a students who have historically been marginalized in the mathematics classroom.

THEORETICAL FRAMEWORK

The theoretical framework on which our data analysis is based draws upon two areas: (a) the notion of "teaching for diversity" and (b) social-constructivism. Our goal is to infuse notions of equity and justice into well-established theoretical frameworks that are concerned with student learning and the context of that learning. In so doing, we hope to push theoretical constructs that may not necessarily challenge taken-for-granted macro-level educational structures and practices that may be detrimental to Latino/a students.

Teaching for Diversity. For progressive educators, a potential role of the mathematics education reform movement is to promote more egalitarian and democratic societies in which all students, not just a select few, have the opportunity to develop mathematical literacy (Kitchen, 2005). While much has been written about the need to implement standards-based curriculum and instruction in mathematics classrooms, in the mathematics education community, little emphasis has been placedon preparing teachers of mathematics to implicitly and explicitly incorporate socially, culturally, and politically equitable instructional strategies in their classrooms, that is, "teach for diversity" (Rodriguez and Kitchen, 2005). Ultimately, teaching for diversity entails teachers of mathematics teaching in more culturally responsive, gender-inclusive, and socially relevant ways (Rodriguez and Kitchen, 2005).

There is a developing body of inquiry into the social, cultural, and political context of the teaching and learning of mathematics (see Atweh, Forgasz, and Nebres, 2001; Gutstein, 2003; Kitchen, 2005; Martin, 2000; Roy and Rousseau, 2005; Secada, 1995; Tate, 1995). Research and teaching in mathematics education that takes seriously the social, cultural, and political context of learning examines how tracking affects learning, whether diverse students have equitable opportunities to learn challenging mathematics, and how race and class play out in the classroom. Other studies show how teachers use mathematics as a means to build critical consciousness in students (Frankenstein, 1995; Gutstein, 2003; Kitchen and Lear, 2000; Ladson-Billings, 1995; Tate, 1995). Some scholars have employed a multidisciplinary framework to investigate the interaction between mathematics and students' linguistic and cultural practices (see Adler, 1998; Brenner, 1998; Civil and Andrade, 2002; Gutiérrez, 2002; Khisty, 1997; Lipka, 1994; Moschkovich, 1999). Still other scholars have applied a social reconstructionist orientation in their teaching to prepare prospective teachers to incorporate equitable and socially just instructional strategies in their classrooms (see Dunn, 2005; Leonard and Dantley, 2005). Such an orientation works to link and potentially challenge specific instructional practices when reflected on relative to social and political considerations.

The significance of these studies is that they redefine traditional notions of "effective pedagogy" (Roy and Kitchen, 2005). Effective teaching is viewed as more than engaging students in constructivist-based mathematics activities. Specifically, teaching for diversity promotes the development of students' cultural identity, empowerment, and social justice. These ideals are beyond that of the equity vision put forth in the Principles and Standards for School Mathematics [PSSM] document (NCTM, 2000), which largely supports learning dominant, albeit reform-based, mathematics (Gutiérrez, 2002; Rodriguez and Kitchen, 2005) with little attention given to issues of culture and social criticism.

Social-Constructivism. The emergent social-constructivist paradigm borrows from cognitive, constructivist, and sociocultural theories (Shepard, 2000). Within the cognitive psychology paradigm, scholars seek to understand an individual's learning in term of internal cognitive structures and processes (Cobb, 2007). The learning of mathematics is viewed as an active process of mental construction and sense making. Within this paradigm, frameworks have been developed to locate students' thinking within specific mathematical domains such as multiplicative reasoning (e.g., Confrey and Smith, 1995). A potential pitfall of domain-specific cognitive frameworks is that they may not take into account cultural and social issues such as the cultural practices of the communities in which the learner lives; nor are issues of equity and access necessarily considered.

In the sociocultural perspective, learning is developed through socially supported interactions. We will borrow a central idea from Vygotsky's work (1979) that learning and child development and brought about from the beginning through communication. "Instruction and development do not meet for the first time at school age; rather, they are in fact connected with each other from the very first day of a child's life" (Vygotsky, 1956, cited in Lerman, 2001, p. 5). From this perspective, cognition is inherently social and learning is viewed as an element of a system of cultural practices (Cobb, 2007). Vygotsky advocated that we not only look at mental activity but at situated practices and that the process must be studied, not just the outcome of activities (Forman, 2003). Thus, sociocultural theory provides a means to explain the complex relationship between social context and learning.

Nevertheless, a shortcoming of Vygotsky's work is the lack of analysis of how individual agency can transform these contexts (Rodriguez, 2005). As students participate in mathematical learning communities, they build on their previous experiences and knowledge to achieve a more advanced understanding of challenging mathematical concepts. They also may begin to ask critical questions such as: "Why should I bother to solve this problem? For whom am I solving this problem? Whose mathematics is this, anyway?" In other words, students are not simply participants in pre-existing cultural practices, they also are active participants in transforming systems of cultural practices.

Sociocultural theory also has been criticized for its lack of usefulness at the classroom level. According to Paul Cobb (2007), "Sociocultural theory provides only limited guidance because the classroom processes on which design [experiment] focuses are emergent phenomena rather than already-established practices into which students are inducted" (2007, p. 24). Nevertheless, scholars have pointed to the potential contribution of sociocultural theory that centers on the notion of a community of practice (Lave and Wenger, 1991; Stein, Silver, and Smith, 1998; Franke and Kazemi, 2001). This research has provided insights into how teachers' instructional practices are influenced by institutional constraints such as the availability of teaching resources and instructional support provided to them (Cobb, McClain, Lamberg, and Dean, 2003).

Merging Teaching for Diversity with Social-Constructivism. In the social-constructivist paradigm, classroom expectations and social norms are examined to understand how important dispositions, such as students' willingness to persist in trying to solve difficult problems are developed (Shepard, 2000). The notion of teaching for diversity contests general references to students and takes seriously how race, ethnicity, gender, social economic status, sexual orientation, etc., may affect opportunities and access students have and how this influences the development of student disposition. Teaching for diversity brings issues of cultural and linguistic diversity and equity to the forefront in all considerations having to do with classroom learning and also with the very structures of schools and schooling.

For us, overlaying teaching for diversity with social-constructivism inspires a commitment to equitable and just educational opportunities for all learners in which each student's ways of thinking is honored. In addition, in this emerging paradigm, it is vital to critically analyze the social context of learning such as the obstacles that could hinder learning (e.g., poorly trained teachers). In the research project described here, we investigated opportunities afforded bilingual Mexican immigrant students in a learning context afforded through the use of a formative assessment format employed by two caring adults.

METHODOLOGY

In spring 2008, videotape data were collected of four students as they estimated, calculated, and explained their solutions to tasks involving fractions, mixed numbers, percents, and proportional reasoning; first on a pre-assessment administered prior to instruction of a unit on rational numbers from a "reform" mathematics curriculum, then on a post-assessment administered after two to four weeks of instruction. The four bilingual Mexican immigrant students were in a sixth-grade class taught by the first author. A four-stage interactive interview protocol was designed and used throughout the pre- and post-assessments of these students. The interactive interviews were conducted with students on an individual basis by a team of two researchers (the second and third authors). One interviewer is English dominant (Interviewer A) and the other is Spanish dominant (Interviewer B). Throughout the interviews, students were given the option of explaining their work in English or Spanish.

The series of rational number tasks were designed by the researchers using similar type tasks to those in the mathematics textbook in use, but were situated in settings to which the students could relate. Names of actual students in the class also were included in the tasks. Task scenarios included students' purchasing grapes from a local market, riding bicycles in their neighborhood, or helping parents purchase gasoline to go on a trip to visit a student's grandmother.

The design of the interactive interview protocol built upon and was an extension of an interview protocol used by the first author in a previous study (see Kitchen and Wilson, 2004; Kulm, Wilson, and Kitchen, 2005). The unique feature of the interactive interview protocol is that it offers students multiple opportunities in differing learning contexts to problem solve. In addition to students expressing solutions in writing, the research protocol includes stages in which students are encouraged to think out loud. The interactive interview protocol consists of four stages.

In the estimation stage of the interactive interview protocol (stage 1), each student was presented with a mathematics task written in English and given the option of having the task translated into Spanish. After being read the task by Interviewer A, the student was asked to estimate a solution without having the benefit of utilizing any tools (e.g., ruler, paper and pencil, calculator, etc.). The student also was not permitted to write down ideas while approximating. Throughout this initial stage, Interviewer A and Interviewer B could ask clarifying questions based on the student's response.

In the second stage, the student went to a separate room where, working independently, she/he developed written solutions to all the tasks for which she/he had developed estimates for previously. During the explanation stage (stage 3), the student was asked by Interviewer A to explain his/her reasoning to solve each task. The student was encouraged to write on a dry-erase board to demonstrate his/her mathematical thinking. Interviewer A also asked clarifying questions, revoiced the student's explanations, and/or referenced aspects of the student's work.

In the phone simulation stage, (stage 4), the student had one last chance to modify the task solution based upon feedback previously received and any new insights.

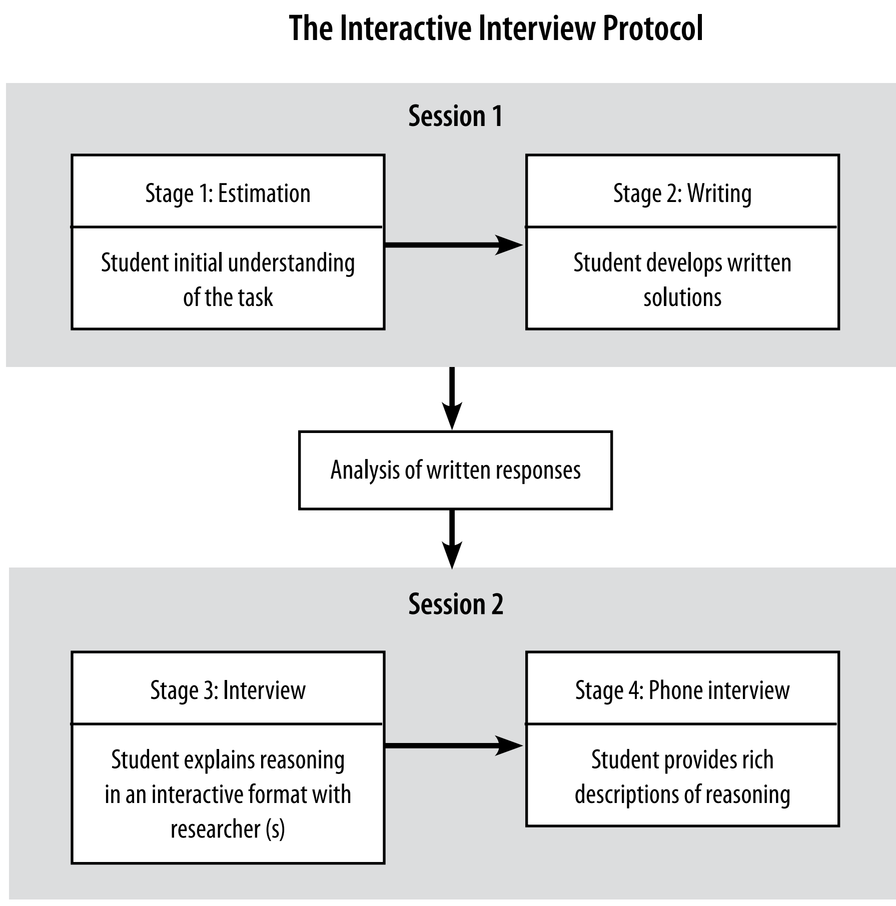


Figure 7: Stages of Interactive Interview Protocol

During this stage, the student was asked to explain her/his mathematical thinking for a task by Interviewer B in a simulated telephone interview. Interviewer B was selected to conduct this interview since Spanish is her first language. We wanted each student to have the choice to discuss her/his mathematical thinking on the phone in either Spanish or English. While the interviewer in stage 4 could not see what a student wrote, the student often used the dry-erase board as a means to recall the process used to solve the task previously or used the board to develop a new solution. Interviewer B could not view any of the student's written work during this stage since the goal was to motivate the student to have to provide rich descriptions of her/his mathematical reasoning to solve the task. Similar to stage 3, Interviewer B could ask clarifying questions, re-voice explanations, and/or reference written solutions completed by the student during stage 2.

During stages 3 and 4, students were allowed to review and reference their written solutions produced during stage 2. Interestingly, students often modified their earlier solutions as they interacted with the interviewers during stages 3 and 4. Throughout, interviewers communicated the expectation that students should thoroughly explain how they obtained answers during each stage of the process. On occasion, interviewers asked questions to assist students to clarify their thinking and encouraged students to persist with problem-solving strategies. It also was not uncommon for the interviewers to provide a "scaffolded" or mini-lesson to assist students make connections, justify a generalization, expound further on their reasoning, or even abandon a non-productive problem-solving strategy. Students were videotaped during stages 1, 3, and 4. Transcripts were created for each of the videotaped sessions.

The researchers met weekly to watch the videotapes and review the transcripts. The data subsets were analyzed using interpretive methods (Erickson, 1986; Maxwell, 2005). Each data subset was read as a whole, followed by a period of open coding to allow for the emergence of themes. An iterative process of coding, memo writing, focused coding, and integrative memo writing followed (Emerson, Fretz, and Shaw, 1995). Creation of the codes went through multiple revisions, as the data were repeatedly read to check the consistency of themes. This process continued until either no new categories were developed or consistency was achieved. After a set of themes were obtained from the dataset, we searched for commonalities and differences in our data subsets. We also sought both confirming and disconfirming evidence by searching for supportive and non-supportive evidence (Erickson, 1986).

This method of analysis coupled with the integrated theoretical framework used enabled an analysis of student understanding of mathematical concepts in the context of a highly relational learning environment engendered by the two interviewers. When we examined students' thinking, we explored how students used traditional mathematical algorithms, produced representations such as graphs, interpreted number lines, applied reasoning, formed mathematical connections, and produced generalizations. We also analyzed the nature of students' interactions with the interviewers, their responses to instructional scaffolding provided and their willingness to persevere in their attempts to problem solve. All our analyses were done with an eye on equity issues and access, such as whether participating students felt their ideas were valued and whether they had opportunities to freely express their ideas without fear of reprisal.

The school and student participants

The four students who participated in this study attended a small, progressive faith-based middle school in a large city in the southwest. At the time the study was undertaken, they were in sixth-grade. The participating students attended a school in which all the school's students lived in poverty and 94 percent are people of color. The school had recently opened in the fall of 2007 and was designed based upon findings from a study that offered insights into how to structure an urban school to productively serve poor and diverse student populations (Kitchen, DePree, Celedón-Pattichis, and Brinkerhoff, 2007). For example, the school had an extended school day to provide significant support for student learning that included a mandatory tutoring session at the end of the academic day. All students were integrated into the regular academic trajectory, which was a college preparatory track. Parents and people living both within and outside the school's community served as volunteers to support the school's high academic expectations.

Developing students who could think critically and read and write academic Spanish also goals as reflected in the school's vision statement: "Teachers will be provided extensive support to develop an inquiry-based curriculum in which experiential learning is central to expand students' critical thinking skills. The core academic subjects as well as Spanish will be the focus of the academic program. All students will learn how to 'critically read the world' ..." Moreover, validating students' cultural backgrounds was a goal of the school from the outset. The mission statement explicitly stated that the school would be "a culturally relevant and affirming school."

A foundational principle embedded in the school's mission was that each student is endowed by God with unique and specific talents. Teachers at the school were expected to explore and to develop each student's full potential. To accomplish such an ambitious goal, the school had adopted the stance that everyone at the school, including students, were responsible for supporting each other to learn and grow emotionally and spiritually to the greatest extent possible. In such an environment, competition was minimized, grades were seldom given, and the holistic health of the school community was of utmost importance.

Seven sixth-graders returned the needed student consent forms with parental approvals to participate in this study. Of these students, we asked four students to participate in the study, being sure to select students so as to achieve optimal diversity in representation across gender and achievement in mathematics. Throughout, we often describe the participants as "bilingual Mexican immigrant" or simply as Mexican students since all are bilingual in Spanish and English and had at least one parent who had migrated from Mexico to the U.S. More importantly, though, we identify the participants as Mexican because they referred to themselves as Mexican. Though some students in this study were born in the U.S., all identified strongly with Mexico and Mexican culture and all spoke Spanish in their homes.

In the year prior to when this study was undertaken, all four research participants had attended a local elementary school with a strong dual language program and spoke English at a "Very Good User level as described by the International English Language Testing System (IELTS)" (Baker, 2006, p. 29). While students were not tested using the IELTS, this rating is based on teacher observation. Two students, Veronica and Zenia were high achievers in mathematics, while Marisol and Andres were considered by their teacher to be performing at an average level in mathematics.

RESEARCH FINDINGS

In this study, we found that the interactive interview protocol provided the means for an in-depth understanding of participating bilingual immigrant Mexican students' mathematical knowledge, reasoning, and procedural ability. Furthermore, the research protocol provided students with the means to ask questions, to be creative, to test and revise their hypotheses, and to explore mathematical concepts deeply (Kulm, 1994). Throughout, it was not unusual for students to develop and connect mathematical ideas as they solved problems. Students not only spoke in English and Spanish to express their thinking, but also communicated their ideas non-verbally, using gestures, diagrams, and mathematical representations. In general, we found that the interactive interview protocol allowed students to explore mathematical concepts deeply without fear of reprisals when they made errors.

We have organized the research findings to highlight three themes that provide insight into how an alternative assessment format can foster a culturally affirming approach to the teaching and learning of Latinos/as. Specifically, the findings are arranged to describe how the interactive interview protocol provided the means to develop student agency through problem solving, support and encourage mathematical innovation, and cultivate a shared sense of purpose in mathematics. First, though, we provide an example of Zenia working through the four stages of the interactive interview protocol as a means of holistically demonstrating how one student's thinking progressed and developed as she solved a problem.

Four Stages of the Interactive Interview Protocol

We begin by illustrating the four stages of the interactive interview protocol. In this problem, Zenia is trying to solve the following task: Andres drove his bike 39 1/6 m and Ned drove his bike 28 5/9 m. How many more meters did Andres drive than Ned?

In stage 1, Zenia was asked to estimate the answer for the problem described above. After much thought, Zenia said the answer would be less than 10 but then changed her mind deciding that it would be a little less than 11 instead. To arrive at the solution, she explained that 39 – 28 = 11 but the answer would be less than 11 because 5/9, which is close to half is bigger than 1/6, and when you subtract, it would be a negative number.

In stage 2 (when students used pencil and paper to problem solve), Zenia invoked the traditional algorithm in her written response to change mixed numbers to improper fractions, found equivalent fractions that had a common denominator, and then subtracted the two fractions as follows:

39 1/6 = 234/6 x 9/9 = 2106/54

28 5/9 = 252/9 x 6/6 = 1512/54

2106/54 - 1512/54 = 594/54 = 11

As you may have noticed, Zenia failed to add the numerators of the fractional portions of the mixed numbers to the numerators of the improper fractions.

During stage 3 (interactive interview based on student's written response in stage 2), Zenia started to demonstrate how she changed the mixed numbers to improper fractions and found that she had made an error. She noted that she needed to add the numerators of the fractional portions of the mixed numbers after multiplying each denominator by the whole number. She modified the calculations she had performed on paper during stage 2 on a small white board as follows:

39 1/6 = 235/6 x 9/9 = 2115/54

28 5/9 = 257/9 x 6/6 = 1542/54

2115/54 - 1542/54 = 573/54

At this point, Zenia said she would change her answer to a mixed number by dividing 573 by 54. To do this, she showed how she had applied the division algorithm with the aid of a white board. At first, Zenia determined the answer to be 1.6 because she neglected to write a 0 after 1 in the quotient. When the interviewer asked her to approximate how many times 54 went into 573, Zenia responded, "About 10." She then recognized and corrected her mistake and derived the quotient 10.6 applying the traditional long division algorithm as follows:

```
       10.6
54 | 573.0
      54
        33
         0
       330
       324
         6
```

Note here that Zenia's reliance on the division algorithm resulted in her obtaining an approximation with the decimal solution instead of deriving the more precise solution of 10 33/54 or 10 11/18.

The interviewer continued by asking Zenia if she could have used a number instead of 54 as a common denominator and Zenia suggested 36. After being prompted to look for a lower common denominator, she identified 18. She used this common denominator to solve the problem in another way in stage 4.

During stage 4 (simulated telephone interview), the interviewer asked Zenia if there was another way to solve the problem after she had explained the identical problem solution used in stage 3. At this point, she reverted to a strategy alluded to during stage 1 in which she worked with the whole number and fractions independently to estimate a solution. First, Zenia subtracted the whole numbers 39 and 28 and got 11. She then converted the fractional portions of the given mixed numbers to 3/18 and 10/18. Zenia subtracted the fractions deriving a solution of -7/18. After some thought, she determined the answer to the task to be 10 11/18 meters. When asked how she arrived at her solution, Zenia said she needed to take 1 from the 11 which was 18/18 and then subtract the 7/18 leaving 11/18.

This vignette revealed some of Zenia's mathematical knowledge, provided insights into her mathematical values, and demonstrated her openness to explore alternative problem solving strategies. After being asked to consider alternative solutions to the bicycle task at the conclusion of stage 3, Zenia needed minimal prompting during stage 4 to pursue a solution strategy that she had hinted at when estimating a solution to the task during stage 1. During this final stage, Zenia developed an efficient strategy to solve the bicycle problem when she converted 1 whole to 18/18 and then derived a solution without getting lost in complex calculations.

If Zenia had just been asked to solve this task using only paper and pencil, the robust mathematical knowledge that she demonstrated during the interview would not have been revealed. Interestingly, she abandoned her estimation strategy (operate on the whole numbers and fractions independently to arrive at a solution) during stages 2 and 3 and initially during stage 4. Only through some prompting from an interviewer did Zenia return to this strategy, which she was able to formalize and correctly apply to obtain the solution. We hypothesize that the reason that Zenia moved away from her estimation strategy to the more traditional algorithmic approach to solve the problem was because she attached high status to this strategy. Most likely, because of the formal instruction received in the past, Zenia had developed the sense that deriving common denominators was the approach that was highly valued by her teachers and her textbook authors.

On a summative assessment (e.g., quiz), Zenia would have most likely received a relatively low score because she made some calculation errors as she attempted to solve the task using learned algorithms. Though she did not immediately derive a correct solution, Zenia established that there is much she did know. For instance, she demonstrated knowledge of common denominators, the relationship of mixed numbers to improper fractions, how to convert an improper fraction into a decimal and the meaning of negative numbers. Viewing Zenia's work through an equity lens highlights what she knew, whereas viewing her work through the deficit lens would highlight what she did not know (Moll and Ruiz, 2002, Khisty, 1995). The interactive interview protocol provided the means to respect Zenia's mathematical thinking, her mathematical values, and how she adapted her thinking to derive a correct solution.

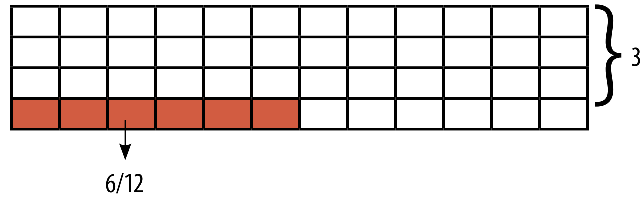
Alternative Assessment as a Means To Encourage Student Agency

In the following vignette, a description is provided of how Marisol solved the following problem:

Veronica has $\boxed{5\frac{1}{4}}$ pounds of grapes. She gave $\boxed{2\frac{2}{3}}$ pounds to Marisol. How many pounds of grapes does Veronica have left?

Initially, Marisol subtracted the whole numbers, she then converted the fractional portion of the mixed numbers 1/4 and 2/3, to 3/12 and 8/12, respectively. She indicated that she was not sure how to compute 3/12 - 8/12, because she did not believe it to be possible to take eight from three. So, Marisol decided to subtract three from eight and got six due to a computational error. She then inserted a negative sign in front of the number, arriving at -6/12.

The interviewer asked Marisol how she could combine three and the fractional part of the solution, -6/12. She indicated that the six (the numerator of the fraction) could borrow something from the three. To represent the three wholes, she drew three rectangles and divided each one into 12 pieces. She then drew an extra rectangle to represent 6/12:



Marisol spent several minutes thinking about what to do next and decided to leave the answer as $\boxed{3\frac{6}{12}}$. She stated, "I do not like negatives."

During the telephone simulation (stage 4), Marisol explained the same procedure to solve the problem just described, but realized that she made a mistake when subtracting the fractions and decided to change her solution to $\boxed{3\frac{5}{12}}$. The interviewer questioned her about the meaning of the negative fraction within the mixed number and asked if the actual solution is less than or greater than three. Marisol was able to recognize that this value would be less than three. In the end, Marisol did not combine the whole number and the negative fraction.

Though Marisol did not arrive at what may be considered a legitimate solution, she clearly demonstrated correct mathematical thinking. Working in a learning environment in which her ideas were actively sought out and respected, Marisol created a strategy to solve the problem. This strategy appeared to be of her own making, that is, she was not applying a learned algorithm to derive a solution. Precisely because she felt encouraged to explore her ideas without the threat of negative consequences, Marisol took the chance of deriving a solution in her own way. The learning environment that had been created for Marisol paved the way for her to experiment and pursue a problem solving strategy that made sense to her. Without a doubt, she would have been penalized for arriving at a solution of $\boxed{3\frac{6}{12}}$ on a paper/pencil examination, and would mostly likely have been penalized for the solution $\boxed{3\frac{5}{12}}$.

Alternative Assessment as a Means To Support Creative Innovation

In the following two vignettes, Andres is asked to find the difference of two fractions and to demonstrate his thinking with the aid of a number line. These vignettes are offered to demonstrate how through the use of the interactive interview protocol, one student developed innovative strategies to problem solve that would not have been developed on a paper/pencil assessment.

During the pre-assessment, Andres was asked to compute 1/4 – 1/2 and to show how he could use the number line to arrive at an answer. Without a lot of thought, he gave three different answers to the question, but when asked to explain his reasoning, he realized that his solutions were incorrect. Ultimately, Andres decided to access his understanding of money to determine a solution to this task.

To derive his solution, Andres established a relationship between the fractional and money representations; a quarter for 1/4 and 2 quarters for 1/2 (explaining that 1/2 is equivalent to 2/4). He then proceeded to explain his thinking with the use of the quarters. Andres represented the numbers with a stack of two negative quarters and another stack of one positive quarter, respectively. He simulated how one of the negative quarters would cancel out the positive quarter resulting in an answer of one negative quarter. To conclude, he translated his solution into a fraction.

Interviewer: Okay, your problem is to take one fourth minus one half.

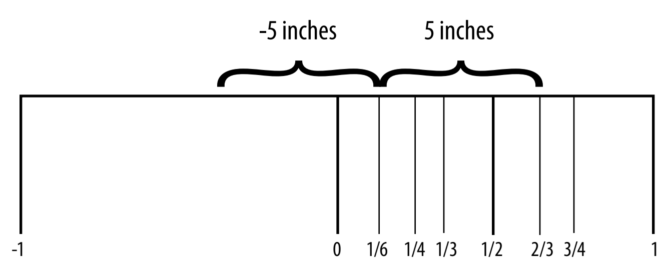
Andres: I think it's minus one fourth because, once again you have the two quarters and a quarter over here (With hand movements, he simulates each one of the movements).

Interviewer: Okay.

Andres: And one of the quarters you subtract minus this quarter and it just gives you nothing. And it's like you have twenty-five cents here and you have minus zero here, and you subtract it, and it's going to give you like minus twenty-five. And, so then twenty-five as a fraction is one-fourth, and so it's minus one-fourth.

Based on how Andres solved this first task, we wondered how he would solve a similar task if the fractions could not be so easily modeled with coins. So, in the post assessment, Andres was given the problem 1/6 - 2/3 = .

Andres first tried to divide the units on the number line into fourths and thirds and labeled them. He then indicated where 1/6 would be on the number line. At this point, he focused on the distance between 1/6 and 2/3 as representing the difference between the two numbers. Andres verbalized that because the problem involved subtracting a larger number (2/3) from a smaller number (1/6), he believed that he could find the distance between 2/3 and 1/6, and then start at 1/6, derive a solution by going the same distance in the opposite direction on the number line. Andres used 5 inches as a sample distance and talked about moving 5 inches to the left to find the answer. He said, "I think I'll probably go like minus 2/3 or something like that."



At this point, the interviewer began scaffolding, building on concepts Andres knew to assist him in connecting them to ideas that he appeared to understand with some fragility. She asked, "So, my question is, if you start at 1/6 and go this far (pointing to the space between 0 and 2/3) in the other direction, are you going to be at a negative two thirds, because you're starting at 1/6?"

Andres responded by holding his index finger and thumb apart the distance between 1/6 and 2/3 as depicted on the number line and rotated his hand to indicate going the same distance but in the opposite direction. The following dialogue then took place:

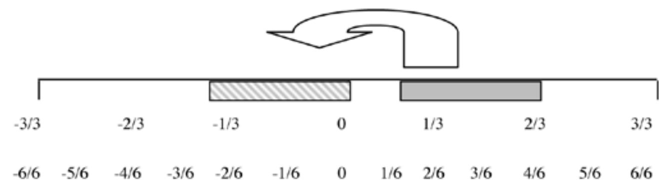
Interviewer: And that makes sense. Since this, if you're going two thirds, that is your distance here and you're at 1/6 and you're going the same distance, two thirds here. If you went that same distance and started at zero, where would you end up?

Andres: I'll probably end up at 2/3.

Interviewer: Right, yeah. So I think it's a really good strategy, but what about the one sixth?

Andres proceeded by putting -1/6 on the number line and a discussion on equivalent fractions occurred as he explored a way to express the distance on the number line between 1/6 and -1/6. He determined that it would 2/6. After further discussion of where the answer should probably be located on the number line, Andres said he thought it would be smaller than 2/3 and bigger than 1/2 (on the negative portion of the number line). The interviewer gave an example of common denominators and how they can be useful when solving a problem. Andres decided to construct a new number line after being given the hint, "How about if you use thirds on your number line." He successfully put 0, 1/3, 2/3 and 3/3 on the number line and then, after placing 1/6 on the number line, he was unsure where to put 2/6. He was asked if 2/6 was bigger or smaller than 1/3 at which he responded, "I think it is the same because you multiply it (1/3) by 2/2 and get 2/6." He then successfully placed 3/6, 4/6, 5/6 and 6/6 on the number line.

After the interviewer prompted, "So, back to our problem," Andres said, "Minus one sixth . . . I think minus three sixths is the answer." He then explained that, if he took the distance between 1/6 and 2/3, moved it on the number line to the zero, the distance would go from 0 to 3/6. He then rotated it to cover the distance from 0 to -3/6.



During the phone interview (stage 4), Andres was able to explain clearly how he solved the problem and synthesized all the procedures with the following explanation. He had already divided the number line into sixths and realized that 2/3 is equal to 4/6.

Andres: I tried to subtract that [1/6-2/3] and, well, I thought that if I had zero through four-sixths, and if I subtracted that the answer would just be minus four-sixths, but since I have one-sixth through four-sixths, I'm going to try to subtract that. And the space that I have between zero and one-sixth is going to be the same space that I have from three over six to four over six. Okay? So, then I'm going to, like, move the chart, so now I'm just going to make it from zero to three-sixths 'cause now it's going to be the same distance.

In this vignette, Andres demonstrated creativity when he first used quarters as a means to assist him to find the difference of 1/4 and 1/2. He used "tools" to assist him to find the difference of the two fractions (Forman, 2003). When it became obvious that the tools provided the means for him to derive this difference during the pre-assessment, we modified the task during the post-assessment to get a more accurate sense of what he understood about subtracting two fractions with a negative difference. Once again, after being given significant scaffolding support, Andres demonstrated mathematical ingenuity as he pursued an exhaustive process to derive the correct solution. After solving the task, Andres spoke about how much he liked figuring things out and how, when possible, he essentially avoided using algorithms he had been taught.

While reviewing the videos of Andres solving this task, it was quite evident that he enjoyed his social interactions with the interviewers and he was clearly motivated to impress them with his mathematical thinking. This observation points to the strength of the highly interactive nature of the research protocol and how it promoted positive social relationships between the interviewers and participating students. Andres was motivated by his social interactions with the interviewers, while also being encouraged to persist. As opposed to growing weary or frustrated while solving the second task, Andres persevered as he relentlessly pursued a solution to the task.

Andres' persistence and innovative ideas were his way of responding to a task that he wanted to solve. This ambition was not something his teacher had observed often as a trait possessed by Andres. Interestingly, Andres had been selected to participate in the study primarily because he had been identified by his teacher as "average," as a student who had not particularly distinguished himself in mathematics. Yet, when given support and encouragement in a highly relational learning environment in which he was the focus, Andres demonstrated great mathematical creativity as he endured to solve a task that was challenging for and interesting to him.

Alternative Assessment as a Means To Develop a Shared Sense of Purpose

Another student, Veronica solved the following problem:

Veronica has $\boxed{5\frac{1}{4}}$ pounds of grapes. She gave $\boxed{2\frac{2}{3}}$ pounds to Marisol. How many pounds of grapes does Veronica have left?

During the explanation stage (stage 3), Veronica described how she initially subtracted the whole numbers, then she converted the fractional portion of the mixed numbers 1/4 and 2/3, to 3/12 and 8/12 respectively. She explained that five minus two would be three but since 3/12 is less than 8/12, she argued that the three needed to be reduced by one, leaving two as the whole number part of the answer. Veronica then added the fractions and got 11/12 and specified her final solution to be $\boxed{2\frac{11}{12}}$.

The interviewer proceeded by asking why she had added the fractions. Veronica's answer was unclear, so the interviewer asked her to write down 3/12 - 8/12 and said "Okay, so we have a subtraction problem." At this point, it became clear that the effort to solve the problem became a collaboration between the student and the interviewer. Veronica openly solicited the interviewer's ideas to solve the problem after reaching a point where she was not sure how to proceed with the task.

After receiving some instruction from the interviewer, Veronica subtracted the fractions and derived $\boxed{-\frac{5}{12}}$ as the difference of the two fractions. Veronica then decided to change her answer to $\boxed{-3\frac{5}{12}}$. The interviewer asked her to represent $\boxed{-3\frac{5}{12}}$ in an alternative way and Veronica wrote $\boxed{2\frac{11}{12}}$. The interviewer then requested an explanation of how she got that value. Veronica decided to convert three to the fractional representation of 36/12. She then subtracted 5/12 from 36/12 and got 31/12. Veronica then converted this number to $\boxed{2\frac{11}{12}}$. Further questioning about the procedure to convert improper fractions into mixed numbers led Veronica to revise her solution to be $\boxed{2\frac{7}{12}}$.

While detailing her solution to the task during the telephone simulation interview (stage 4), Veronica made it clear in her explanation of her solution that it had been developed in collaboration with the first interviewer. This was evidenced by her use of the pronoun "we." She used the pronoun "we" to describe the same procedures she used to obtain the correct solution during the previous stage, after the interviewer had actively collaborated with her to modify her initial solution. For example, Veronica started one explanation, "What we did was 5 minus 2 is equal to 3 and …"

In this vignette, Veronica developed a solution in partnership with the interviewers. As she invited the interviewer to collaboratively make meaning of the given problem, a shared sense of purpose was developed to support one another to arrive at a solution. The processes intrinsic to the interactive interview protocol contributed to the formation of a productive collaboration that resulted in Veronica successfully solving the task. This, in turn, led her to derive a sense of herself as a capable mathematics student.

Interestingly, Veronica did have knowledge of an algorithm that could be used to derive a solution (i.e., convert three to 36/12 and then subtract 5/12). She struggled to rectify the solution she derived using this approach with her initial result of $\boxed{2\frac{11}{12}}$. Thus, this vignette provides information that could be used to inform mathematical instruction for Veronica. Specifically, Veronica needed assistance to become a more reflective problem solver who regularly looked back at her solutions to be sure that they were well-reasoned and made sense.

DISCUSSION AND IMPLICATIONS

Similar to other studies that have examined whether summative assessments in mathematics disadvantage students who are learning English as a second language, (Abedi and Lord, 2001; Lampert and Cobb, 2003; Morgan and Watson, 2002), we found that participating students' written responses gave a very limited snapshot of their mathematical reasoning and communication. The interactive interview protocol proved to be a valuable formative assessment format to "gather evidence of [a student's] knowledge and then infer what the student knows" (Wilson and Keeney, 2003, p. 53). Our understanding of the participating students' mathematical knowledge is far greater than what we could have obtained solely through traditional means. In contrast to large scale, high-stakes tests that only afford insight on specific tasks on which a student fails, a formative assessment tool such as the research protocol provides the means for in-depth individual diagnosis (Noddings, 2004). Furthermore, the research protocol adds to the research literature by showing the effectiveness of a tangible tool that can serve as a means to provide individualized, cognitively demanding mathematics instruction for bilingual, Mexican immigrant students.

The interactive interview protocol proved quite effective at uncovering the four participating students' mathematical reasoning and inspired multiple opportunities for the researchers to engage in mathematical discourse with students, encourage students, provide scaffolding when students struggled to connect ideas, and to teach mathematical ideas that students could use to problem solve. Opportunities to solve tasks accompanied by the interviewers' high expectations, resulted in students expressing knowledge and demonstrating skills and abilities needed to problem solve throughout the various stages of the research protocol. When students' thinking was challenged by an interviewer, their knowledge and expressive abilities grew as they verbalized the processes they were exploring to arrive at a problem solution.

In the introductory excerpt, Valenzuela describes teachers who "deny their [Latino/a] students the opportunity to engage in reciprocal relationships" (1999, p. 23) as essentially repressing their Latino/a students as cultural beings. In this study, we learned of the potential of the interactive interview protocol to support the development of trusting and affirming relationships that ultimately led to mathematically inspiring Latino/a youth. Through the fruition of these positive relations, the research protocol provided a means to position these students as competent problem solvers (Empson, 2002; Forman, 2003; Turner, Celedón-Pattichis, and Marshall, 2008) and to support bilingual learners using the resources that they bring to the learning process (Moschkovich, 2002). Because of the interactive interview protocol's usefulness, we argue that this alternative assessment format has great potential for use in classrooms that serve majority Latino/a populations.

In retrospect, the research protocol served as a means to cultivate a culturally affirming and empowering learning environment for the participating bilingual, immigrant Mexican students. The social-constructivist framework provided us with the means to analyze students' thinking, while continually reminding us of the constraints of summative assessment formats such as traditional paper/pencil tests. After reflecting upon our insights about students' learning (e.g., innovative, willing to take risks, etc.), a compelling image emerged of the participating students and of their mathematical competencies and potential. The result was a revelation of sorts of what is possible if formative assessment practices are coupled with attention to the cultures and social identities of Latino/a youth.

A question that emerges from this study is what specific features of the interactive interview protocol, those that were embedded or aspects of the research protocol that emerged, might account for the research findings? For instance, was the most vital feature of the research protocol the repeated, layered opportunities students were provided to continually consider and re-consider their mathematical ideas? Perhaps, giving students' multiple opportunities to verbalize their mathematical ideas is the most noteworthy characteristic of the interactive interview protocol. On the other hand, maybe the most important salient attribute of the research protocol has more to do with the caring adult attention that the bilingual Mexican immigrants received throughout. This study was not designed to reveal which qualities of the research protocol are the most significant. Further research is needed to examine whether all the explicit and implicit features of the interactive interview protocol are necessary to produce results similar to those found in this study.

Accurately assessing what a student knows is very difficult and the techniques that are most effective are the most time intensive. A great deal of time was spent in conducting the interviews and analyzing the results, time most educators would not have for assessment. We plan to continue this line of research to develop efficient applications of the interactive interview protocol that will be of use to classroom teachers. Taking as a given the enormous challenges of teaching mathematics for understanding, particularly in schools that serve largely poor, immigrant students (Kitchen, 2003), we believe that it is not enough to simply present a viable assessment format without attending to the classroom-level and school-level structures that must be transformed.

Research has often documented structural constraints that must be overcome to provide increased educational opportunities for students of color and the poor (e.g., Kitchen, DePree, Celedón-Pattichis, and Brinkerhoff, 2007; Noguera, 2003). For example, secondary teachers generally teach large classes making it almost impossible for them to develop healthy professional relationships with more than a handful of their students. Thus, at schools that serve large numbers of students, such as comprehensive public high schools, structural constraints do not allow teachers to develop, nor do they support teachers in developing, culturally affirming relations and teaching practices that serve Latino/a youth well. This may partially explain why schools become structured in ways that "fracture students' cultural and ethnic identities, creating social, linguistic, and cultural divisions among the students and between the students and the staff" (Valenzuela, 1999, p. 5)

We posit that all educational practices, particularly taken-for-granted ones such as having only one teacher in the classroom and contributing structural constraints, be evaluated with regards to their capacity to develop each student's full potential. At a time when so much national attention is being placed on improving tests scores in mathematics, school districts are investing heavily in short-cycle assessments and diagnostic software to try and determine areas in which students need mathematically remediation. The irony is that while much is being invested in testing, hardware, and software to assess students' mathematical wherewithal, professional development that focuses on preparing teachers to effectively use formative assessment formats continues to lag (Desimone, Smith, and Ueno, 2006).

We also question why so many districts and schools are willing to devote significant resources to assessing students and remediating with expensive hardware and software, but less willing to hire more qualified teachers, which would allow more adults to be available to work directly with students. To restructure school in ways that affirm and support students burgeoning cultural and ethnic identities, more well-trained teachers need to be in the classroom so that students have opportunities to develop the sorts of positive relationships that will affect their learning for the better. Furthermore, having more than one teacher in the classroom would make it much more feasible to engage in resource intensive classroom practices such as employing the interactive interview protocol with individual or small groups of students. Instead of spending huge amounts of money on testing and highly technical approaches to education that just intensify the fractured nature of schooling, more humane approaches to instruction and assessment need to be implemented that have the potential to positively influence students' school experiences in general, and Latino/a students' experiences in particular.

Finally, we find it problematic that so many high-stakes decisions such as how to place students in mathematics classes are being made throughout the U.S. based upon limited assessment data (e.g., a student's result on a short-cycle assessment). Above all, we are concerned that decisions on how to place students in mathematics that are based on limited assessment data have a particularly adverse effect on students who speak English as a second language. In this study, if we had only examined participating students' written responses, we would have had a very limited snapshot of their mathematical reasoning and communication.

In summary, this study provides evidence that the interactive interview protocol has tremendous potential to significantly improve the teaching and learning of mathematics, particularly for bilingual Mexican immigrants. Our work with the research protocol demonstrated that when the participating students were given significant opportunities to develop close relationships with a knowledgeable teacher, they were inspired and responded enthusiastically to mathematics. By utilizing assessment formats that promote students using the bilingual resources they bring to the process, teachers can simultaneously determine what their bilingual students understand while positioning students as competent and enhancing their identities as mathematical problem solvers. Pursuing more humane assessment practices must also become a priority given the high-stakes nature of traditional assessments that have proven to be particularly harmful for students who speak English as a second language. Given the potential of culturally affirmative assessment formats to transform schooling practices for Latinos/as, more resources should be devoted to implement these approaches through supporting classroom teachers with the required training and human capital support.

This research was supported by the National Science Foundation, under grant ESI-0424983, awarded to the Center for the Mathematics Education of Latino/as (CEMELA). The views expressed here are those of the authors and do not necessarily reflect the views of the funding agency.

REFERENCES

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

Abedi, J. and P. Gándara. (2006). Performance on English Language Learners as a Subgroup in Large-Scale Assessment: Interaction of Research and Policy. *Educational measurement: Issues and practices*, 25 (4), 36-46.

Adler, J. (1998). A language for teaching dilemmas: Unlocking the complex multilingual secondary mathematics classroom. For the Learning of Mathematics, 18, 24-33.

Atweh, W., H. Forgasz and B. Nebres. (2001). Sociocultural research on mathematics education: An international perspective. Mahwah, NJ: Lawrence Erlbaum Associates.

Baker, C. (2006). *Foundations of bilingual education and bilingualism.* Clevedon, U.K.: Multilingual Matters Ltd.

Black, P., and D. Wiliam. (2001). Inside the black box: Raising standards through classroom assessment. Retrieved July 29, 2008, from http:ngfl.northumberland.gov.uk/keystage3ictstrategy/Assessment/blackbox.pdf.

Borjian, A. (2008). A new approach in meeting the needs of Latinos in American schools: Students' views on attending a redesigned small high school. *The International Journal of Learning*, 15 (2), 277-285.

Brenner, M.E. (1998). Adding cognition to the formula for culturally relevant instruction in mathematics. *Anthropology and Education Quarterly*, 29(2), 213-244.

Civil, M., and R. Andrade. (2002). *Transitions between home and school mathematics: Rays of hope amidst the passing clouds.* In G. de Abreu, A.J. Bishop, and N.C. Presmeg (eds.), *Transitions between contexts of mathematical practices* (pp. 149-69). Dordrecht: Kluwer.

Civil, M. (2007). Building on community knowledge: An avenue to equity in mathematics education. In N.S. Nasir and P. Cobb (eds.), *Improving access to mathematics: Diversity and equity in the classroom* (pp. 105-117). New York, NY: Teachers College Press.

Cobb, P. (2007). Putting philosophy to work: Coping with multiple theoretical perspectives. In F.K. Lester, Jr. (ed.), *Second handbook of research on mathematics teaching and learning* (2nd ed., pp. 3-38). Charlotte, NC: Information Age Publishing.

Cobb, P., K. McClain, T. Lamberg, and C. Dean. (2003). Situating teachers' instructional practices in the institutional setting of the school and school district. *Educational Researcher*, 32(6), 13-24.

Confrey, J., and E. Smith. (1995). Splitting, covariation, and their role in the development of exponential functions. *Journal for Research in Mathematics Education*, 26, 66-86.

Desimone, L.M., T.M. Smith. and K. Ueno. (2006). Are teachers who need sustained, content-focused professional development getting it? An administrator's dilemma. *Education Administration Quarterly*, 42(2), 179-215.

Dunn, T.K. (2005). Engaging prospective teachers in critical reflection: Facilitating a disposition to teach mathematics for diversity. In A.J. Rodriguez and R.S. Kitchen (eds.), *Preparing mathematics and science teachers for diverse classrooms: Promising strategies for transformative pedagogy* (pp. 143-60). Mahwah, NJ: Lawrence Erlbaum Associates.

Emerson, R.M., R.I. Fretz, and L.L. Shaw. (1995). *Writing ethnographic fieldnotes.* Chicago: The University of Chicago.

Empson, S.B. (2003). Low-performing students and teaching fractions for understanding: An interactional analysis. *Journal for Research in Mathematics Education*, 34(4), 305-343.

Erickson, F. (1986). Qualitative methods in research on teaching. In M.C. Wittrock (ed.), *Handbook of research on teaching* (3rd ed., pp. 119-161). New York, NY: Macmillan.

Ferguson, R.F. (1998). Teachers' perceptions and expectations and the Black-White test score gap. In C. Jencks and M. Phillips (eds.), *The Black-White test score gap*, (pp. 273-317). Washington, DC: The Brookings Institution.

Flores, A. (2007). The opportunity gap. *TODOS Research Monograph*, 1, 1-18.

Forman, E.A. (2003). A sociocultural approach to mathematics reform; Speaking, inscribing, and doing mathematics within communities of practice. In J. Kilpatrick, W.G. Martin, and D. Schifter (eds.), *A research companion to principles and standards for school mathematics* (pp. 333-352).

Franke, M.L. and E. Kazemi. (2001). Teaching as learning within a community of practice: Characterizing generative growth. In T. Wood, B.C. Nelson, and J. Warfield (eds.), *Beyond classical pedagogy in elementary mathematics: The nature of facilitative teaching* (pp. 47-74). Mahwah, NJ: Lawrence Erlbaum Associates.

Frankenstein, M. (1995). Equity in mathematics education: Class in the world outside the class. In W.G. Secada, E. Fennema, and L.B. Adajian (eds.), *New directions for equity in mathematics education* (pp. 165-190). New York, NY: Cambridge University Press.

Fuchs, L.S., D. Fuchs, K. Karns, C. Hamlett, M. Katzaroff, and S. Dutka. (1998). Comparisons among individual and cooperative performance assessments and other measures of mathematics competence. *The Elementary School Journal*, 99 (1), 23-51.

Gearhart, M. G.B. Saxe. (2004). When teachers know what students know: Integrating mathematics assessment. *Theory Into Practice*, 43(4), 304-313.

Grant, C.A. (1989). Equity, equality, teachers and classroom life. In W. G. Secada (ed.), *Equity in education* (pp. 89-102). London: Falmer Press.

Guskey, T.R., and J.M. Bailey. (2001). *Developing grading and reporting systems for student learning.* Thousand Oaks, CA: Corwin Press.

Gutiérrez, R. (2002). Beyond essentialism: The complexity of language in teaching mathematics to Latina/o students. *American Educational Research Journal*, 39, 1047-1088.

Gutiérrez, R. (2007). (Re)defining equity: The importance of a critical perspective. In N.S. Nasir and P. Cobb (eds.), *Improving access to mathematics: Diversity and equity in the classroom* (pp. 37-60). New York, NY: Teachers College Press.

Gutstein, E. (2003). Teaching and learning mathematics for social justice in an urban, Latino school. *Journal for Research in Mathematics Education*, 34(1), 37-73.

Hiebert, J., and T.P. Carpenter. (1992). Learning and teaching with understanding. In D. Grouws (ed.), *The handbook of research on mathematics teaching and learning* (pp. 65-97). New York, NY: Macmillan.

Hiebert, J., T.P. Carpenter, E. Fennema, K. Fuson, P. Human, H. Murray, A. Olivier, and D. Wearne. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher*, 25(4), 12-21.

Khisty, L.L. (1997). Making mathematics accessible to Latino students: Rethinking instructional practice. In J. Trentacosta and M. Kenney (eds.), *Multicultural and gender equity in the mathematics classroom: The gift of diversity*, 97th Yearbook of the National Council of Teachers of Mathematics. Reston, VA: National Council of Teachers of Mathematics.

Khisty, L.L. (1995). Making inequality: Issues of language and meaning in mathematics teaching with Hispanic students. In W.G. Secada, E. Fennema, and L.B. Adajian (eds.), *New directions for equity in mathematics education* (pp. 279-297). New York, NY: Cambridge University Press.

Kitchen, R.S. (2005). Making equity and multiculturalism explicit to transform the culture of mathematics education. In A.J. Rodriguez and R.S. Kitchen (eds.), *Preparing mathematics and science teachers for diverse classrooms: Promising strategies for transformative pedagogy* (pp. 33-60). Mahwah, NJ: Lawrence Erlbaum Associates.

Kitchen, R.S. (2003). Getting real about mathematics education reform in high poverty communities. *For the Learning of Mathematics*, 23(3).

Kitchen, R.S. and J.M. Lear. (2000). Mathematizing Barbie: Using measurement as a means for girls to analyze their sense of body image. In W. Secada (ed.), *Changing the faces of mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Kitchen, R.S. and L.D. Wilson. (2004). Lessons learned from students about assessment and instruction. *Teaching Children Mathematics*, 10(8), 394-399.

Kitchen, R.S., J. DePree, S. Celedón-Pattichis, and J. Brinkerhoff. (2007). *Mathematics education at highly effective schools that serve the poor: Strategies for change.* Mahwah, NJ: Lawrence Erlbaum Associates.

Knapp, M.S., and S. Woolverton. (1995). Social class and schooling. In J. Banks and C. Banks (eds.), *Handbook of research on multicultural education* (pp. 548-569). New York, NY: Macmillan.

Kulm, G. (1994). *Mathematics assessment: What works in the classroom.* San Francisco, CA: Jossey-Bass.

Kulm, G., L.D. Wilson, and R.S. Kitchen. (2005). Alignment of content and effectiveness of mathematics assessment items. *Educational Assessment Journal*, 10(4), 333-356. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Lampert, M. and P. Cobb. (2003). Communication and language. In J. Kilpatrick, W.G. Marin and D. Schifter (eds.), *A research companion to principles and standards for school mathematics* (pp. 237-249). Reston, VA: National Council of Teachers of Mathematics.

Lave, J., and E. Wenger. (1991). *Situated learning: Legitimate peripheral participation.* NY: Cambridge University Press.

Leonard, J., and S.J. Dantley. (2005). Breaking through the ice: Dealing with issues of diversity in mathematics and science education courses. In A.J. Rodriguez and R.S. Kitchen (eds.), *Preparing mathematics and science teachers for diverse classrooms: Promising strategies for transformative pedagogy* (pp. 87-118). Mahwah, NJ: Lawrence Erlbaum Associates.

Lerman, S. (2001). A cultural/discursive psychology for mathematics teaching and learning. In W. Atweh, H. Forgasz, and B. Nebres (eds.), *Sociocultural research on mathematics education: An international perspective* (pp. 3-17). Mahwah, NJ: Lawrence Erlbaum Associates.

Lipka, J. (1994). Culturally negotiated schooling: Toward a Yup'ik mathematics. *Journal of American Indian Education*, 33(3), 14-30.

Lockwood, A., and W. Secada. (1999). *Transforming education for Hispanic youth: Exemplary practices, programs, and schools.* Washington, DC: National Clearinghouse for Bilingual Education.

Lubienski, S. (2007). Research, reform and equity in U.S. mathematics education. In N.S. Nasir and P. Cobb (eds.), *Improving access to mathematics: Diversity and equity in the classroom* (pp. 10-23). New York, NY: Teachers College Press.

Martin, D.B. (2000). *Mathematics success and failure among African-American youth: The roles of sociohistorical context, community forces, school influence, and individual agency.* Mahwah, NJ: Lawrence Erlbaum Associates.

Maxwell, J. (2005). *Qualitative research design: An interactive approach.* Thousand Oaks, CA: Sage Publications Inc.

McKown, C. and R. Weinstein. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology*, 46, 235-261.

Moll, L.C., and R. Ruiz. (2002). The schooling of Latino children. In M.M. Suárez-Orozco and M.M. Páez, (eds.), *Latinos: Remaking America*, (pp. 362-374). Berkley, CA: University of California Press.

Morgan, C., and A. Watson. (2002). The interpretive nature of teacher's assessment of students in mathematics: Issues for equity. *Journal for Research in Mathematics Education*, 33(2), 78-110.

Moschkovich, J.N. (1999). Supporting the participation of English language learners in mathematical discussions. *For the Learning of Mathematics*, 19(1), 11-19.

Moschkovich, J. (2007). Bilingual mathematics learners: How views of language, bilingual learners and mathematical communication effect instruction. In N.S. Nasir and P. Cobb (eds.), *Improving access to mathematics: Diversity and equity in the classroom* (pp. 89-104). New York, NY: Teachers College Press.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1995). *Assessment standards for teaching mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Science Foundation. (1996). *Indicators of science and mathematics education 1995.* Arlington, VA: Author.

No Child Left Behind Act of 2001. (2001). Retrieved on January 2, 2009, at http://www.ed.gov/policy/elsec/leg/esea02/index.html.

Noddings, N. (2004). High stakes testing: Why? Theory and Research in Education, 2(3), (pp. 263-269).

Noguera, P. (2003). *City schools and the American dream: Reclaiming the promise of public education.* New York, NY: Teachers College Press.

O'Day, J.A., and M.S. Smith. (1993). Systemic school reform and educational opportunity. In Susan H. Fuhrman (ed.), *Designing coherent educational policy: Improving the system* (pp. 250-311). San Francisco, CA: Jossey-Bass.

Olsen, L. (1997). *Made in America: Immigrant students in our public schools.* New York, NY: The New Press.

Pegg, J. (2003). Assessment in mathematics: A developmental approach. In J.M. Royer (ed.), *Mathematical cognition* (pp. 227-259). Greenwich, CT: Information Age Publishing.

Rodriguez, A.J. (2005). Using sociotransformative constructivism to respond to teachers' resistance to ideological and pedagogical change. In A.J. Rodriguez and R.S. Kitchen (eds.), *Preparing mathematics and science teachers for diverse classrooms: Promising strategies for transformative pedagogy* (pp. 17-32). Mahwah, NJ: Lawrence Erlbaum Associates.

Rodriguez, A.J. and R.S. Kitchen. (eds.) (2005). *Preparing Mathematics and Science Teachers for Diverse Classrooms: Promising Strategies for Transformative Pedagogy.* Mahwah, NJ: Lawrence Erlbaum Associates.

Romo, H.D., and T. Falbo. (1996). *Latino high school graduation: Defying the odds.* Austin, TX: University of Texas Press.

Roy, F., and R. Kitchen. (2005). *Teachers' conceptions about mathematics and diversity in two urban districts: Results from an effective schools study.* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Roy, F., and C. Rousseau. (2005). Student thinking as a context for high expectations. *For the Learning of Mathematics*, 25(2), 16-23.

Secada, W.G. (1995). Social and critical dimensions for equity in mathematics education. In W. Secada, E. Fennema, and L.B. Adajian (eds.), *New Directions for Equity in Mathematics Education*, (pp. 146-164). New York, NY: Cambridge University Press.

Sheets, R.H. (1995). From remedial to gifted: Effects of culturally centered pedagogy. *Theory Into Practice*, 34(3), 186-193.

Shepard, L.A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.

Stein, M.K., E.A. Silver, and M.S. Smith. (1998). Mathematics reform and teacher development: A community of practice perspective. In J.G. Greeno and S.V. Goldman (eds.), *Thinking practices in mathematics and science learning* (pp. 17-52). Mahwah, NJ: Lawrence Erlbaum Associates.

Stiggins, R.J. (2001). *Student-involved classroom assessment.* Upper Saddle River, NJ: Merrill Prentice Hall.

Suárez-Orozco, M.M. (1989). Psychosocial aspects of achievement motivation among recent Hispanic immigrants. In H.T. Trueba, G. Spindler, and L. Spindler (eds.), *What do anthropologists say about dropouts?* New York, NY: The Falmer Press.

Tate, W.F. (1995). Economics, equity, and the national mathematics assessment: Are we creating a national tollroad? In W.G. Secada, E. Fennema, and L.B. Adajian (eds.), *New directions for equity in mathematics education* (pp. 191-208). New York, NY: Cambridge University Press.

Turner, E., S. Celedón-Pattichis, and M. Marshall, (2008). Cultural and linguistic resources to promote problem solving and mathematical discourse among Hispanic kindergarten students. *TODOS Research Monograph, 1*, 19-40.

Valenzuela, A. (1999). *Subtractive schooling: U.S.-Mexican youth and the politics of caring.* Albany, NY: State University of New York Press.

Vygotsky, L. (1979). Consciousness as a problem in the psychology of behaviour. *Soviet Psychology*, 17, 5-35. (Original work published in 1924)

Wiggins, G.P. (1993). *Assessing student performance: Exploring the purpose and limits of testing.* San Francisco, CA: Jossey-Bass.

Wilson, L.D., and P.A. Kenney. (2003). Classroom and large scale assessment. In J. Kilpatrick, W.G. Martin, and D. Schifter (eds.), *A research companion to principles and standards for school mathematics* (pp. 53-67). Reston, VA: National Council of Teachers of Mathematics.

Winfield, L.F. (1986). Teacher expectations and low-achieving students. *The Urban Review*, 18(4), 253-268.

Zeichner, K.M. (1996). Educating teachers to close the achievement gap: Issues of pedagogy, knowledge, and teacher preparation. In B. Williams (ed.), *Closing the achievement gap: A vision for changing beliefs and practices* (pp. 56-76). Alexandria, VA: Association for Supervision and Curriculum Development.

NOTES:

NOTES:

TODOS

nea
NATIONAL EDUCATION ASSOCIATION
nea.org
Great Public Schools for Every Student

4314 | 0410