Function and Form in Research on Language and Mathematics Education

Guillermo Solano-Flores

University of Colorado, Boulder

White paper for the Spencer Foundation

September, 2007

This paper was commissioned by the Spencer Foundation. I wish to thank Dr. Lauren Young for her interest in my research work. Also, I wish to thank Kris Gutierrez, Judit Moschkovich, Khanh Nguyen, Maria Ruiz-Primo, Mary Schleppegrell, and Rich Shavelson for their comments on an earlier version of this paper. The opinions expressed here are not necessarily those of the Spencer Foundation.

Function and Form in Research on Language and Mathematics Education

Guillermo Solano-Flores

University of Colorado, Boulder

Abstract

In this paper, I propose a conceptual framework for examining research that addresses the relationship between language and mathematics education. According to the conceptual framework, this research can be classified based on implicit or explicit views of language—ways in which language is thought of—used by researchers. Four views of language are identified, as a *process,* as *system, as a structure*, and as a *factor*. These views are distinguishable by virtue of differences on seven dimensions: (1) role attributed to language as a key actor in mathematics education, (2) themes or ideas underlying the researchers' actions, (3) units of analysis, (4) language modes most frequently used as sources of data, (5) key concepts, (6) broad areas of research, and (7) theories and disciplines most commonly used. No language view is better than the other; each focuses on different aspects of knowledge. Process and system views can be called, *functional* because they examine the dynamic aspect of language as critical to mathematical communication and the development of mathematical knowledge. Structure and factor views can be called, *formal* because they examine the linguistic features of mathematical problems and the characteristics of linguistic groups. My analysis reveals that functional views are typical of research on teaching (including informal formative assessment) and formal views are typical of research and practice in large-scale testing. Researchers tend to use the perspective of only one view of language in spite of the fact that the four language views are not antithetical. Using views of language in combination can be key to addressing language in mathematics education with a multidisciplinary perspective and may pave the way for addressing teaching and assessment in a coordinated manner. I give some recommendations for future research on teaching and assessment.

Introduction

The symbolic nature of mathematics is intriguing to many people—to the extent that popular conceptions exist about the linguistic nature of mathematics. One conception is that mathematics is a language in its own right, a universal language; another conception is that mathematics is language-free.

A balanced analysis of the linguistic nature of mathematics is provided by Pimm (1987). While it has properties that can be examined from the perspective of linguistics, mathematics is not a language that can be learned in the same way people learn foreign languages. At the same time, natural, ordinary language can be used as a communicative tool for interpreting and constructing meaning in mathematics. Also, as in any discipline and any social activity, there is a register and a set of conventions that are specific to mathematical communication.

Researchers and practitioners in the field of mathematics education deal with language issues one way or another and need to be aware of the multiple aspects of this complex relationship. *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics, 1989) identifies "learning to communicate mathematically" as a major goal for all students. Also, *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics, 2000) recognizes the importance of organizing and consolidating mathematical thinking through communication and the importance of using the "language of mathematics to express mathematical ideas..." (p. 60). *Principles* also has among its goals addressing the instructional needs of students with limited proficiency in the language of instruction. Congruence in research on the teaching and the assessment of mathematics around these or other normative documents is possible only if there is clarity about the multiple ways in which "communication," "language of mathematics," "limited language proficiency," "language of instruction," and other language-related concepts are understood, studied, and measured.

This congruence is especially important in times of standards and accountability. Alignment with standards documents is regarded as evidence of good practice and test validity (see Schoenfeld, 2004; Sloane & Kelly, 2003). Regardless of whether or not we agree with this trend, we cannot help to wonder whether the goals underlying standards documents can be accomplished when teaching and mathematics assessment are guided by different conceptions of language. Also, is language addressed in the same ways by researchers who deal with language in the classroom and those who deal with language in testing? Understanding these differences is critical to both ensuring proper interpretation of results across research areas and improving mathematics teaching and assessment research and practice in relation to language. It also may be helpful in identifying promising areas for future research on language and mathematics and in generating knowledge that can inform practice and policy (see Burkhardt & Schoenfeld, 2003).

In this paper, I address the notion that both implicit and explicit conceptions of the phenomena investigated in mathematics education influence which variables researchers regard as relevant, what kind of data they gather, and how they interpret their findings (Schoenfeld, 2006). I offer a conceptual framework for examining research on language and mathematics education. I submit that research on language and mathematics education is guided by views of language—ways in which the phenomenon of language is thought of.

The first section of the paper presents an overview of the conceptual framework. While they are often implicit, views of language guide the reasonings and assumptions researchers use in their investigations. Also, while these language views are not necessarily antithetical or mutually exclusive, researchers tend to adhere mostly to one view in their work.

The second and third sections examine two broad types of views that I identify in educational research in language and mathematics education, functional and formal. Functional views encompass views of language as a process and as a system; formal views encompass views of language as a structure and as a factor. Some examples are provided (although not discussed in detail) with the intent to illustrate the wide variety of investigations that may share each view. These investigations may reflect opposing theoretical perspectives in mathematics education.

In the last section, I give some recommendations and discuss promising areas for research in the field.

Overview of Language Views in Research on Language and Mathematics

Research on language and mathematics education can be characterized according to four language views, as a process, as a system, as a structure, and as factor. None of these language views is better than the others. While they emphasize different aspects of language, they are not antithetical or mutually exclusive.

Table 1 compares these four views of language along seven dimensions: (1) role attributed to language as a key actor in mathematics education, (2) themes or ideas underlying the researchers' actions, (3) units of analysis, (4) language modes most frequently used as sources of data, (5) key concepts, (6) broad areas of research, and (7) theories and disciplines most commonly used. The table is intended to show patterns, rather than clear-cut distinctions between language views. A given investigation may address language in ways that are typical to one view for some dimensions and to another view for other dimensions. However, researchers tend to adhere in their investigations to one view for most of the dimensions, if not all.

Table 1

Views of language as a process and as a system can be referred to as *functional* because they emphasize the dynamic aspect of language in mathematical communication and the development of mathematical knowledge. Views of language as a structure and as a factor can be called, *formal* because they emphasize linguistic features of mathematical problems or types of linguistic groups.

Functional Views

Language as a Process

Views of language as a process are observed in research that examines the role of language in the development of mathematical knowledge. Language can be examined as a reflection of mathematical understanding based on the analysis of verbal reports of students as they solve mathematical problems (e.g., Siegler & Jenkins, 1989), for example, with the purpose of determining how the ways in which children use number words in different contexts reveal aspects of their understanding of the notion of cardinality (see Fuson, 1991). Language as a social process (see Vygotsky, 1936) and culture as a phenomenon that shapes mind (Vigotsky, 1978; Wertsch, 1985) are seen as key in the development of mathematical thinking (see Brenner, 1994; Lampert, 1990). Communication is examined as both a facilitator of learning—*communicating to learn mathematics*—and a learning goal—*learning to communicate mathematically*—(Lampert & Cobb, 2003) in teacher-student and student mathematical conversations (e.g., Brenner, 1998a; Khisty, 1995, Webb, 1991). Also, cultural differences in mathematical reasoning are examined to understand the role of culture and cultural identity in the learning and mental representation of mathematics (Garcia, 1993; Ladson-Billings, 1995; Nasir, 2002; Stigler and Baranes, 1988-1989).

Here I discuss two aspects of language as a process: preserving meaning—how language shapes meaning when it transferred across mathematics and natural language—and negotiating meaning—constructing mathematical knowledge through social interaction.

Preserving meaning. Brown and Yule (1983) propose the existence of two functions of language, transactional (transferring information), and interactional (establishing and maintaining social relationships). The transactional function of language is of special interest in mathematics education because it is concerned with the correspondence between natural language and mathematics.

Some scientific concepts may be difficult to learn because the words used to refer to them have different meanings in everyday life (e.g., Meyerson, Ford, Jones, & Ward, 1991). In their attempts to make sense of word mathematics problems, students may substitute words (e.g., *for every—for each; through—in*) when they rephrase them (Mitchell, 2002). Due to the overlap of meaning in a discipline and in the natural language, students need to be familiar with the meaning of words at a level of understanding that goes beyond the knowledge passed on by definitions (Nagy, 1988).

The notion of register is critical to examining how the characteristics of natural language shape students' interpretations of mathematical representations. *Register* is a term that refers to the fact that written or spoken language varies across situations and activities (Halliday, 1978) and, more specifically, to the ways in which certain things and concepts (e.g., *integer, subtract*) are referred to by a community (e.g., the community of mathematicians or the community of mathematics educators) as a result of a social process that involves specialization in a content area, certain contexts, and certain specific activities.

As with science, writing in mathematics is distinguished from ordinary writing by virtue of a high frequency of features such as nominalization, impersonal style, passive voice, lexical density, and the use of interlocking definitions, among many others (Halliday, 1993; Morgan, 1998). Altogether, these features make it difficult for students to understand technical writing. For example, *glass crack growth rate* may be more difficult to understand than *how quickly cracks in glass grow* (Halliday, 1993, p. 79), which conveys the same meaning and has a more familiar style. However, it could be argued that, since abstraction is essential to mathematical reasoning (see Sfard, 2000), being knowledgeable in mathematics necessarily involves understanding the language of mathematics (Greeno, 1989) and being able to talk as mathematicians do (see Lave and Wenger, 1991).

Discourse and stylistic properties of writing shape the reader's perception of the extent to which meaning is preserved. Suppose that students in a class are asked to state why, given certain premises, t is a positive number. Here are two hypothetical students' responses:

Student 1: *As shown in [5],* $t > x+m$. *Therefore,* $t > 0$.

Student 2: *I saw in Equation 5 that* $t > x+m$. *That is why I think t is positive.*

While the two responses show the same conclusion and provide the same kind of justification, Student 1's response could be judged as reflecting a deeper understanding of mathematics than Student 2's response, simply because its style is the same as the style typically used in mathematics textbooks.

In discussing how writing reflects student's knowledge of mathematics, Morgan (1998) observes that features of text such as "the presence of algebra" (e.g., the use of letters to denote variable names), "abstracness" (e.g., the absence of references to persons, the use of present tense), "correct" terminology, and the absence of evidence of process (e.g., the absence of a step-by-step descriptions of the reasoning used in solving a problem) may influence teachers' judgments of their students' mathematical skills. Thus, mathematical understanding and the use of conventions of the investigation report genre may be difficult to dissociate.

At the core of the debate around the use of mathematical and informal language is the tension between mathematizating thinking and making mathematics meaningful as forms of mathematical enculturation—the entry into the mathematical community through interaction with others (see Schoenfeld, 1992). Some (e.g., Sfard, 2000; Sfard & Cole, 2002) argue that an excessive emphasis on real-life mathematics and real-life context in the teaching of mathematics takes its essence away from the discipline—the ability to deal with abstract ideas and symbols. Others (e.g., Brenner, 1998b; Ladson-Billings, 1995) argue in favor of making mathematics meaningful by connecting students' everyday life experiences to school curriculum.

Negotiating meaning. Findings from research on the role of social interaction in the classroom (e.g. Leung, 2005) underscore the importance of allowing students to construct meaning through language beyond the simple use of mathematical vocabulary. Unfortunately, normative documents may overemphasize "correct vocabulary" and formal language and dismiss the importance of natural language, thus limiting the linguistic resources students can use to construct mathematical knowledge.

Raiker (2002) investigated the extent to which the characteristics of the spoken language used in mathematical conversations influence the teaching of mathematical concepts and examined the possibility that the *National Numeracy Strategy* (NNS)—Great Britain's official document intended to provide teachers with mathematics standards, course structure, and class activities—may overemphasize "the correct use of mathematical vocabulary." Through discourse analysis of classroom interactions, Raiker was able to observe that both teachers and students tend to ascribe different meanings to technical terms. Raiker also observed that teachers' use of some terms included in *Mathematical Vocabulary*—a document that supplements NNS and which contains terms that its authors believed can facilitate the learning of certain concepts—appeared to hamper, rather than facilitate, the learning of the targeted concepts.

Barwell (2005) notes that, while ambiguity is a common occurrence in mathematical conversations and can be a valuable resource in the teaching of mathematics (see also Rowland, 2000), strategies used by NNS appear to be based on the premise that mathematical language is always precise. He cites statements from *Mathematical Vocabulary* intended to provide teachers with guidance on the use of mathematical language in the classroom:

- "children need support to move on from 'informal' to 'technical' language in mathematics, and from hearing and speaking new vocabulary to reading and writing;
- "teachers should ascertain the extent of children's mathematical vocabulary and the depth of their understanding." (cited by Barwell, 2005, p. 120).

Thus, the guidelines appear to imply that academic language occurs only in the reading and writing modes, which contradicts current thinking in functional linguistics that certain forms of oral discourse are highly academic (Schleppegrell, 2004).

An important issue in research that examines mathematical conversations is the need for appropriate conceptual frameworks for characterizing social interaction in the classroom. Lampert and Cobb (2003) note that the study of the relation between communication in the classroom and student achievement has been based on loose definitions of communication. What counts as student-student interaction or group discussion may be different across studies. Also, features of discourse measured tend to be too generic and may not address aspects of communication that are specific to mathematics.

An example of a conceptual framework for examining classroom interactions comes from the field of formative assessment. Formative assessment has been characterized as the set of assessment activities intended to support learning—assessment *for* learning—as opposed to those intended to appraise learning—assessment *of* learning (Black, 1993; William 1999a; 1999b). Ruiz-Primo and Furtak (2006; 2007) propose a model for examining informal formative assessment in the classroom—the set of unplanned, unstructured forms of assessment that take place in classroom conversations. They characterize classroom interactions as cycles comprising four steps: the teacher elicits a question; the students respond; the teacher recognizes critical information from the students' responses; and the teacher uses that information to support student learning. These cycles are not a prescribed teaching formula. Rather, they occur naturally in classroom conversation and may be initiated at any of its four steps. Ruiz-Primo and Furtak's results show that informal formative assessment is linked to successful student learning. Students who performed well in assessments embedded in instruction tended to have teachers who completed these cycles more frequently.

Language as a System

Views of language as a system are observed in research that addresses the confluence of languages and language varieties in the construction of mathematical language. *System* refers to the fact that different forms of language (e.g., world languages, the dialects of a given language, everyday language) are governed by rules and conventions. It also denotes choice in an individual's use of a language or a dialect according to social contextual factors (see Coulmas, 2005; Fishman, 1965). *Linguistic diversity* refers to different languages (e.g,, English, Haitian-Creole), different dialects of a given language (e.g., Standard English, African American Vernacular English), different forms of a language (e.g., informal language, formal language, academic language), different levels of proficiency in a given language, the condition of being bilingual, and each of the languages of a bilingual individual (e.g., the first language, the second language).

Here I discuss language, dialect, and the condition of being bilingual as instances of linguistic diversity. I also discuss three aspects of language in research that addresses language as a system: code-switching in problem solving, the tension between languages, and the influence of language in the interpretation of mathematical problems.

Language, dialect, and bilingualism. English, Spanish, Náhuatl, Swahili, and any other language are rule-governed systems, each consisting of a unique set of arbitrary conventions of sounds, symbols, and a unique set of rules for combining those sounds and symbols in ways that allow communication among its users (see Wardhaugh, 1978; Fasold & Connor-Linton, 2006). The word, *arbitrary* stresses the fact that no language is more natural than any others. Calling *tree* a tree in English is as natural as calling it with another word in any other language. Likewise, no language is *better* than others, as each language develops in a way that meets the communicative needs of their users (see Nettle & Romaine, 2002).

Dialects are also rule-governed systems, varieties of a same given language that are, within broad limits, mutually intelligible, and which can be distinguished from one another by virtue of such features as pronunciation, grammar, vocabulary, discourse conventions, and the use of certain sets of idiomatic expressions and colloquialisms (see Crystal, 1997). While the term, *dialect* is sometimes used to characterize a variety of a language as corrupted, everybody speaks dialects (Preston, 1993). All dialects of the same language have comparable levels of sophistication and complexity (Farr & Ball, 1999). While some dialects may be more prestigious than others, no dialect is better, as system, than others (Corson, 2001). The most prestigious variety of a language is frequently referred to as the standard dialect of that language, such as *Standard English* (Wardhaugh, 2002).

The notion of system also can be applied to examine bilingualism. A bilingual individual can be thought of as someone who has a language system comprising two languages, the native language and the second language (Bialystok, 2001). The condition of being bilingual is not the addition of two separate languages; rather, those two languages make the bilingual person's language system (see Grosjean, 1989). While political discourse sometimes characterizes bilingualism as a deficit (see Crawford, 2000), no scientific evidence exists that supports such conception (Baker, 2006). Indeed, there are some cognitive advantages that result from being bilingual, including an increased flexibility in the performance of certain cognitive tasks and an increased metalinguistic awareness (Bialystok, 2002).

"Bilingual" is a term that reflects a wide range of degrees of proficiency that an individual may have in two languages. This proficiency may vary considerably not only across that person's two languages but also across language modes (e.g., listening, speaking, reading, and writing) within each language. Thus, individuals who are not proficient in the language of instruction (e.g., English language learners) can be viewed as bilingual students, even if their bilingualism is incipient (see Valdés & Figueroa, 1984).

Code-switching in problem solving. Moschkovich's (2006) study of code-switching in mathematical conversations is a good example of research that addresses language choice (see also Moschkovich, 2002). Code-switching is a term from the field of sociolinguistics which refers to the alternate use of two languages or two dialects during conversation. Current thinking in sociolinguistics and bilingual development holds that, rather than a deficiency, code-switching is a complex skill that involves sophisticated knowledge of the syntactical structures of an individual's two languages (see Bialystok, 2001; Poplack, 1980).

Moschkovich examined the transcript of a conversation between two Grade 9 bilingual, native Spanish speaking Chicana students engaged in solving a problem connecting a linear equation and its graph (Figure 1).

Figure 1

An analysis of the transcript revealed that code-switching allowed the students to build arguments efficiently. For example, in explaining why the line's steepness should be lower one of the students, Marcela, said:

"Porque fíjate, digamos que este es el suelo.

[Because look, let's say that this is the ground.]

Entonces, si se acerca más, pues es menos steep.

[Then, if it gets closer, then it's less steep.]

. . . . 'cause see this one [referring to the line y = x] . . . is . . .

está entre el medio de la x y de la y. Right?

[is between the x and the y]"

Marcela builds an argument in which the register learned from formal instruction in English is used to refer to a mathematical concept ("steep") and Spanish—Marcela's native language—is used to provide illustrations and emphasize some parts of the argument. A constructive mathematical discussion takes place because the students use their language system without restrictions. A "view of everyday language as unscientific and as contrasted with the precision and specificity of scientific terminology does not do justice to how human beings use language to think and to learn" (Warren, Ballenger, Ogonowski, Rosebery, & Hudicourt-Barnes, 2001, p. 539). Effective learning is more likely to occur when students are allowed to use their linguistic resources in full.

Tension between languages. The use of multiple languages in the classroom creates a tension between languages in the sense that different languages have different functions and, therefore, different sets of advantages and resources for constructing mathematical knowledge. For example, in post-apartheid South Africa's multilingual classrooms, the use of students' native languages can be encouraged with the purpose of facilitating the construction of arguments in mathematical conversations. However, mathematical register is available only in the language of instruction.

This tension parallels the process known in sociolinguistics as *language contact*, which refers to "the outcomes for speakers and their languages when new languages are introduced into a speech community (Mesthrie & Leap, 1999, p. 248)." Language contact occurs as a consequence of an increased interaction "between people from neighbouring territories who have traditionally spoken different languages. But, more frequently, it is initiated by the spread of languages of power and prestige via conquest and colonisation" (Mesthrie & Leap, 1999, p. 248).

Languages interact dynamically in a process that involves power, status, and utility. While code-switching can be a valuable resource in these classrooms, official language policies and teachers' conceptions of language and personal goals can conflict in the classroom practice, creating personal, practical, and contextual dilemmas for teachers. South African teachers who are interested in promoting mathematical conversation among their students through the use of their native languages may also be interested in providing opportunities for them to develop both mathematical language and informal English—a language which they value as a social asset (see

Setati, 2002). Since most of the mathematical language is available only in English, it needs to be modeled by the teacher through English. However, modeling English academic language may take valuable time that could be spent in mathematical conversation.

In one of the lessons in multilingual classrooms, Setati (2002) observed that the use of English produced also the dominance of procedural discourse: Students responded in procedural discourse when the teacher asked a conceptual question or remained silent until the teacher asked a procedural question. Setati attributes this finding to the fact that the two forms of discourse have different sets of linguistic and mathematical knowledge demands. While the former can be constructed through simple memorization of the sequence of actions that need to be performed to solve a problem, the latter requires from the learner to understand the reasons underlying that sequence of actions. These observations confirm the notion that mathematical intellectual practices are social practices, not simple cognitive routines (O'Connor, 1998). Just the choice of a language can impose, under a given set of circumstances, what is learned and how it is learned.

Adler (1995; 1998) notes that dilemmas involving language choice with regards to supporting mathematics learning are not necessarily problematic and should be seen as "sources of praxis" whose analysis can help teachers to improve their skills. Gutiérrez, Baquedano-Lopez, & Tejeda (1999) focus on the potential for learning of these *hybrid spaces*—as they call the environments in which students are allowed to use multiple forms of language. To them, hybridity and diversity are resources, not challenges; hybrid spaces are environments in which "social, political, material, cognitive, and linguistic [tensions] are sites of rupture, innovation, and change that lead to learning" (Gutiérrez, Baquedano-Lopez, & Tejeda, 1999, p. 287).

Though effective, some strategies intended to address the challenges derived from teaching multilingual classes have their own sets of challenges. Adler uses the term, *mathematical* *language teaching* to refer to teaching in which "language itself, and particularly talk, becomes the object of attention in the mathematics class and a resource in the teaching and learning processes [...] [and includes] being more explicit about instructions for tasks and more careful in [the] use of mathematical terms and [the] expression of ideas" (Adler, 1999, p. 48). While the use of this approach may be beneficial to all students regardless of their linguistic background, it also may produce an excessive attention to students' mathematical verbalizations and neglect mathematical conversation. Effectively teaching in linguistically diverse classroom requires sophisticated skills among educators, who need to walk a fine line between teaching content and providing the linguistic support needed by students to learn that content.

Language influences in the interpretation of mathematical problems. The body of research on the linguistic structure of number names and its influence on performance in counting (e.g., Miller & Stigler, 1987; Miura, 1987; Miura, Okamoto, Vlahovic-Stetic, Kim, & Han (1999) is a good example of research that addresses mathematics learning and teaching with a perspective of language as a system. The transparency with which the naming of numbers reflects the logic behind the base-10 number system varies considerably across cultures (see Saxe, 1988). For example, in English, *eleven, twelve, thirteen, twenty, thirty,* and *forty* do not express clearly the number of units of a decade. In contrast, the names for the same numbers in Chinese and other Asian languages are literally *ten-one, ten-two, ten-three, two tens, three tens,* and *four tens*.

Miura and Okamoto (2003) examine those and other differences between languages and propose that languages provide their users with different sets of supports for the development of mathematics understanding. These differences can affect the ease with which children learn to count and shape the kind of initial exposure they have to mathematics in the context of formal instruction (see Kilpatrick, Swafford, & Findell, 2001). Because languages are governed by different sets of rules, exact equivalence of the same problem across languages is virtually impossible (see Greenfield, 1997). An example of these rules is the use of *ko*, a numeral classifier which is used in Japanese when counting small, round objects. Miura and Okamoto discuss the word problem: *Joe has 6 marbles. He has 2 more than Tom. How many does Tom have?* When translated into Japanese, the problem reads like: *Joe has 6 (ko) marbles, 2 (ko) more than Tom. How many (ko) does Tom have?*

Once the referent is established, the noun (marble) can be omitted, but the corresponding numeral classifier (ko) has to be used in the remainder of the problem. Thus, in Japanese, the problem is understood as, *Joe has 6 (small, round thing) marbles, 2 (small, round thing) more than Tom. How many (small, round thing) does Tome have?* Miura and Okamoto argue that, in Japanese, "numbers in isolation, as in "2 more than Tom" are not an abstract quantity. *Ko* acts as a concept signifier that makes problems more concrete by producing a more vivid representation of the word problem. Thus, the same problem may not pose the same set of challenges in different languages.

Formal Views

Language as a Structure

Views of language as a structure are observed in research that examines how mathematical problem understanding is influenced by their linguistic features. *Structure* refers to the organization of text (see Crystal, 1997). The vocabulary (technical and non-technical) and the syntactic complexity of mathematical problems are examined by means of judgmental procedures which focus primarily on grammar constituents (e.g., propositional phrases, verbs, determiners) as units of analysis.

Research that addresses the structural aspects of language faces the challenge that the set of features that make up formal mathematical discourse are so distinctive (e.g., the frequent occurrence of passive voice and nominalization [see Morgan, 1998]), that in some cases mathematical discourse may be confounded with mathematical content. Test writers and test reviewers continuously face the dilemma of using a discourse that is consistent with the discourse of the discipline and using a discourse that does not pose unnecessary reading demands to test takers but may not have the level of abstraction that is perceived as inherent to mathematical reasoning.

Here I discuss two closely related aspects of language as a structure, grammar and semantics. The former refers to the structural complexity of language as a predictor of item difficulty. The latter refers to the relation between wording and semantic relations in mathematics word problems.

Grammar. Research on the linguistic complexity of test items addresses the concern that verbal and reading ability have a significant influence on student performance in mathematics tests (Thurber, Shinn, & Smolkowski, 2002). It is argued, for example, that word problems have characteristics that make them different from other text materials due to their unique style, the abundance of lexical terms, and the scant continuity of ideas across sentences (Davis, 1991; Ferguson & Fairburn, 1985).

A great deal of the process of test development has to do with refining the wording of items (see Solano-Flores & Shavelson, 1997). Not surprisingly, even a small change in the wording of an item may affect the semantic structure of test items and, therefore, the way in which students interpret them (De Corte, Verschaffel, & Pauwels, 1990; Shorrocks-Taylor, & Hargreaves, 1999).

In their hope to count with handy tools for examining the linguistic complexity of text in tests, every now and then researchers, practitioners, and even test developers turn to readability formulas created based on counts of such features as sentence length and number of syllables—which are assumed or have been observed to be good predictors of reading level (Gunning, 2003) for a given population and for a specific type of text. Unfortunately, readability formulas have serious limitations derived from the fact that they tend to ignore important factors involved in text comprehension, such as word meaning and the complexity of sentence construction (see Crystal, 1997). To be properly used (but without loosing sight of their limitations), they need to be developed from sufficiently large samples of text of the same kind as the text to be examined and with large samples of individuals who are representative of the target population of readers (see Harrison, 1999). Since, by definition, test items consist of small segments of text, any measure of the readability of test items is objectionable (Paul, Nibbeling, & Hoover, 1986).

In an attempt to identify some principles that can guide the process of test development and test adaptation (especially for students with limited proficiency in the language in which tests are administered), researchers have investigated how student performance on science and mathematics test items is affected by linguistic complexity, which is defined in terms of the frequency of technical vocabulary, verb phrases, conditional clauses, relative clauses, and the like (Abedi & Lord, 2001). There is evidence that the linguistic simplification of items can reduce the score gap attributable to language proficiency between English language learners (ELLs) and native English speakers. However, the effects of this form of testing accommodation are moderate (Abedi, Lord, Hofstetter, & Baker, 2000; Abedi, Hofstetter, & Lord, 2004).

Shaftel, Belton-Kocher, Glasnapp, and Poggio (2006) observed that some indicators of the complexity of mathematics items were better predictors of item difficulty for English language learners (ELLs) in earlier school grades than for students in higher school grades. However, they also found that mathematical vocabulary was the only common predictor of item difficulty across grades. Their findings confirm the notion that technical terms pose serious linguistic challenges to students in their attempt to learn or demonstrate conceptual understanding. These findings also speak to the elusive nature of language, whose structural complexity does not account entirely for problem difficulty.

The linguistic properties of test items interact with the contextual information they provide and the students' own experiences. In mathematics word problems, contextual information used with the intent to make them meaningful may be interpreted by test takers in many unexpected ways. Take as an example the Lunch Money item (National Assessment of Educational Progress, 1996), whose linguistic features we (e.g., Solano-Flores & Trumbull, 2003) have examined extensively:

> Sam can purchase his lunch at school. Each day he wants to have juice that costs 50¢, a sandwich that costs 90¢, and fruit that costs 35¢. His mother has only $1.00 bills. What is the least number of $1.00 bills that his mother should give him so he will have enough money to buy lunch for 5 days?

Among the many potential linguistic challenges identified in this item is that "only $1.00 bills" could be interpreted by students in three ways: as restricting the number of dollar denominations (as in *His mother has only dollar bills*), as restricting the number of dollar bills (as in *His mother has only dollars*), and as restricting the amount of money (as in *His mother has only one dollar*). We also observed that, in their responses to this item, some students living in poverty used survival strategies (e.g., giving up on the sandwich, suggesting Sam to ask his mother to give him more money) rather than mathematical strategies such as adding the costs of the sandwich, the juice, and the fruit, then multiplying the result by 5, and then rounding the result to the next higher integer. Linguistic challenges like this cannot be detected unless a careful process of review is used that combines strategies like having the students read the items aloud, interviewing them about their interpretations of the items, and examining their written responses.

Other formal approaches for examining the linguistic complexity of items focus on the syntactic structure of sentences. For example, by using a combination of graph theory (see Harary, 1969) and structural linguistics-based sentence parsing procedures (see van Gelderen, 2000; Veit, 1999), it is possible to detect unnecessary complexity in the structure of sentences (Solano-Flores, Trumbull, & Kwon, 2003). This complexity is reflected by properties such as the number of levels and branches and the number and types of nodes in the graph that represents the structure of a sentence. Figure 2 shows two sentences from items from the National Assessment of Educational Progress (1996) public release with different syntactic complexities. One of the sentences is from the Lunch Money item, discussed above. Clearly, the second sentence is more complex than the first and is likely to be considerably more challenging to test takers.

Figure 2

Formal and logical approaches for analyzing problems should be used judiciously. They are valuable approaches for developing and reviewing mathematics items only when they are used in combination with verbal protocols, cognitive interviews, and other empirical data collection procedures.

Semantics. Research that addresses the semantic aspect of mathematical problems aims at identifying how their formal properties are related to their cognitive demands. Efforts of this kind are in line with the tradition of using formal, logical approaches with the intent to enumerate solution strategies—and their degree of correctness—for a given type of mathematical problem (Brown & Burton, 1978) and approaches intended to define problem universes in ways that test items can be logically generated according to a set of generation rules (Hively, Patterson, & Page, 1968) or through mapping sentences (Bormouth, 1970) or other representational devices.

In examining the complexity of items, judges may experience difficulty distinguishing content from linguistic complexity in mathematics test items. For example, in an investigation of the linguistic complexity of mathematics items, we (Kidron & Solano-Flores, 2006) asked teachers to rank a set of word problems according to their level of difficulty. We observed that the teachers' judgments on the complexity of the mathematics skills needed to solve the problems were influenced by the complexity with which they were worded. Because of this perception, a problem like

How much money can Lara save in four days if she saves $3.75 everyday?

could be perceived as assessing more complex mathematics skills than if it read:

Lara saves $3.75 everyday. How much money can she save in four days?

The series of investigations by De Corte and his associates (see De Corte & Verschaffel, 1991; De Corte, Verschaffel, & De Win, 1985; De Corte, Verschaffel, & Pauwels, 1990) illustrates research that investigates the semantic aspect of mathematical problems. These investigations show how the linguistic properties of items can be modified systematically with the intent to assess performance on different types of problems within a given domain. In one of them, De Corte, Verschaffel, and De Win (1985) studied the effect of rewording on the performance of students on three types of word mathematics problems referred to as, *change*, *combine*, and *compare* problems—basic categories from the classification of addition and subtraction problems originally proposed by Riley, Greeno, & Heller (1983). In change problems, an event changes the value of a quantity (e.g., *Joe won 3 marbles. Now he has 5 marbles. How many marbles did Joe have in the beginning?*); in combine problems, two amounts are combined (e.g., *Tom and Ann have 9 nuts altogether. Tom has 3 nuts. How many nuts does Ann have?*); and in compare problems, two amounts are compared (e.g., *Pete has 8 apples. Ann has 3 apples. How many apples does Pete have more than Ann?*). (Note 1)

De Corte and his colleagues gave problems of the three types to Grade 1 and Grade 2 students in two series. In the first series, the problems were not reworded. In the second series, the problems were reworded by making explicit the semantic relations of their components. For example, after rewording, the problem of Joe and the marbles read: *Joe had some marbles. He won 3 more marbles. Now he has 5 marbles. How many marbles did Joe have in the beginning?* The students' scores and the proportions of correct responses on the reworded problems were significantly higher than the students' scores and the proportions of correct responses on the problems that had not been reworded. These results indicate that problems involving the same kind of arithmetic operation can have different degrees of difficulty due to differences in their underlying semantic structures. Word problems can be reworded to make their semantic relations more explicit without affecting their underlying semantic and mathematical structures, thus making it easier for young students to understand those problems and provide solutions (see De Corte & Verschaffel, 1991).

From a broader perspective, these results also show that even a small modification of the wording of a problem may produce a substantial change in its semantic structure, its difficulty, and the way in which students interpret it. It is because of this complex interaction of linguistic features, that De Corte and his colleagues' findings might not be generalizable to more complex problems. As can be seen from comparing the original version and the reworded version of Joe and the marbles, an increased amount of text and an increased set of reading demands may be the price of making the semantic relations among the components of word problems more explicit. This increase in reading demands may be even much greater for problems that are substantially more complex.

Language as a Factor

Views of language as a factor are observed in research that uses categories of linguistic groups and treatment conditions, mainly in the context of assessment and, more specifically, in large-scale testing. *Factor* refers to the fact that language is seen as something that needs to be controlled or accounted for in order to obtain accurate measures of mathematics achievement. Score differences between groups are critical to devising strategies intended to reduce the effect of language as an extraneous variable (e.g., Abedi & Lord, 2004; Abedi, Hofstetter, Baker, & Lord, 2001).

Views of language as a factor also are observed in research aimed at detecting and minimizing item bias, systematic performance differences that are attributable to group membership, not the construct being measured (see Schmeiser & Welch, 2006). Approaches derived from item response theory (see van der Linden & Hambleton, 1997; Yen & Fitzpatrick, 2006) are common in research on testing across cultural groups (e.g., van de Vijver & Tanzer, 1997) and testing of groups in different languages (Ercikan, 2002; Cook & Schmitt-Cascallar, 2005). The use of groups is "a statistical device, used because potential bias is uncovered by aggregating evidence across test takers within such groups" (Camilli, 2006, p. 228).

Here I discuss three aspects of research that addresses language as a factor: classifying individuals into population group categories and assigning them to testing conditions, and incorporating language variation into research designs.

Population group categories. Language can be addressed by referring it to the characteristics of the populations who speak (or do not speak) a given language, rather than the characteristics of that language. (Note 2) Individuals are classified into broad linguistic group categories (e.g., English speakers, French speakers), according to a small number of levels of proficiency in a language (e.g. "limited English proficient," "English proficient") or according to a small number of categories intended to describe histories of language development (e.g., "monolingual," "native English speaker").

Examples of research that uses population groups with the intent to address language come from the field of international test comparisons such as TIMSS (Trends in Mathematics and Science Study) and PISA (Programme of International Student Assessment), in which different linguistic groups are tested with the same sets of items. The use of item response theory (a psychometric theory of scaling; see van der Linden & Hambleton, 1997) allows detection of biased items—items which are said to function differently across linguistic groups because the performance of the two linguistic groups is not equivalent after controlling for the difference in the overall ability measured (see Allalouf, 2003; Camilli & Shepard, 1994; Hambleton, 2005; Sireci & Allalouf, 2003). There is evidence that differential item functioning can result from very subtle ways in which tests are translated and which affect the equivalence of items across languages. Even the way in which a single word is translated may influence this differential functioning (Ercikan, 1998). Detecting and correcting the origin of this differential functioning may require the use of cognitive interviews with students from the target populations (Ercikan, 2002).

Important differences between international comparisons and the testing of linguistic minorities can be noticed in the ways in which the characteristics of persons are used to group them into linguistic group categories. In the case of international comparisons, linguistic groups are distinguished naturally as a result of their nationalities and their languages of instruction. (Note 3) By contrast, official definitions of "limited English proficient," like that shown in Figure 3 have many possible interpretations. While this definition acknowledges the multi-faceted nature of language (i.e., it includes the ability to speak, read, write, and comprehend English) and the consequences of not being proficient in English, it does not provide objective criteria for making sound classification decisions on who should be included in the "English language learner" category and who should not. As a consequence of this vagueness, this definition of ELL is likely to be operationalized based on proxy, demographic variables. (Note 4)

Figure 3

An additional problem of this way of defining a linguistic group has to do with its comparability across states. Since different states use different tests to measure English proficiency (National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs, 2006), the category, ELL has not the same meaning, which poses some problems of comparability and equivalence of measures of language proficiency.

Serious limitations of language proficiency measures make one wonder the extent to which they should be used to make instructional or testing decisions about students. The first limitation has to do with the fact that "language proficiency" is a complex construct that is highly dependent on context. A person can be proficient in a second language for some contexts, not others (e.g., Bachman, 1990; Canale, 1983; De Avila, 1988; Grosjean, 1985; Hymes, 1972; MacSwan, 2000). What counts as communicative competence in a given context may not count as communicative competence in another context (see Romaine, 1995; Trumbull & Farr, 2005). In addition, different tests of language proficiency emphasize different language skills (e.g., García, McKoon, & August, 2006). As a consequence, a measure of language proficiency based on a given test may be generalizable only to the set of situations and tasks that are similar to the situations and tasks used by that test.

The second limitation is that, due to their different migration histories, formal education background, and many other reasons, each bilingual individual has a unique pattern of language dominance across the four language modes (listening, speaking, reading, writing) (see Baker, 2006; Bialystok, 2001; Durán, 1989; Solano-Flores & Trumbull, in press). As a consequence, a measure of language proficiency may give inaccurate information about an individual's actual competences in a language.

Testing conditions. An important aspect of the research that addresses language as a factor consists of assigning individuals to different testing conditions such as the language used in a test (e.g. English, Spanish), the language mode in which the test is given to students (e.g., orally, in printed form), or the language mode in which students provide their responses to that test (verbal responses, written responses). A testing condition also may be produced by modifying the linguistic properties of a test with the intent to reduce its linguistic demands (e.g., simplifying the wording of items, including glossaries with word-to-word translations) or from modifying properties of the test that are not related to language but which are thought to be relevant to cognitively processing language (e.g., administering a test with no completion time limit).

A body of research on the use of testing accommodations for ELLs has explored a wide variety of testing conditions intended to reduce the linguistic demands of mathematics and science test items (see the reviews by Abedi, Hofstetter, & Lord, 2004; and Sireci, Li, & Scarpati, 2003). ELL students who receive and who do not receive a form of accommodation are compared to non-ELL students as to their test scores. If the accommodation is effective, then its impact should be reflected as a reduction in the score differences between ELL and non-ELL students; also, the test scores should be higher for ELLs who received the accommodation than the scores for ELLs who did not receive the accommodation. Also, non-ELL students who receive the accommodation are compared with non-ELL students who do not receive the accommodation. If the accommodation truly operates on the linguistic demands of tests, then the scores obtained by non-ELL students with and without the accommodation should not differ substantially.

Not surprisingly, language-related accommodations appear to be more effective than accommodations unrelated to language in reducing the score gap between ELL and non-ELL students (Abedi, 2002; Abedi & Hehri, 2004). More specifically, the linguistic simplification of items appears to be the most effective form of accommodation, although the score differences between ELL and non ELL students are moderate (Abedi, Lord, Hofstetter, & Baker, 2000). In addition, this form of accommodation does not benefit non-ELL students, which indicates that the linguistic simplification of items is not a threat to the comparability of scores of ELL and non-ELL students (Rivera & Stansfield, 2004).

The effectiveness of testing accommodations may be limited by the lack of fidelity with which they are implemented. For example, while the literature reports with some level of detail the procedures used to create and provide accommodations, the individuals in charge of providing accommodations may not have the qualifications needed (e.g., translation skills, sensitivity to subtle but important dialect variation, knowledge of the mathematical register in the ELL student's native language) to provide them properly. To complicate matters, states vary tremendously as to the types of ELL testing accommodations they use and the kinds of provisions they have for their implementation (see Rivera, Collum, Willner, & Sia, 2006). Thus, a treatment condition such as "test in the student's native language" may mean many different things.

Creating categories of language proficiency and assigning individuals to testing conditions may be appropriate. What is not appropriate is to overestimate the accuracy of measures of language proficiency, or to underestimate the conditions that hamper proper implementation of testing conditions.

Language variation and research design. Assessing measurement error due to language factors rather than pretending that language variation can be controlled for by using a few (and sometimes dubious) categories of language proficiency, may be a more effective approach to addressing language in testing. Guided by these reasonings, we (Solano-Flores & Li, 2006) have examined language and dialect as sources of measurement error. We have used a design in which ELL students are tested with the same set of mathematics items in two languages. Rather than testing students in bilingual formats, the intent of this design is to examine score variation across languages and to determine how many items are needed to obtain dependable measures of academic achievement when the students are tested in English and when they are tested in their native language.

By using generalizability theory—a psychometric theory of measurement error developed as an extension of analysis of variance (Brennan, 1992; Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Shavelson & Webb, 1991)—we (Solano-Flores & Li, 2006) have been able to identify the amount of score variation due to the interaction of student, item, rater, and language in mathematics, open-ended tests. Our results show that the performance of ELLs in mathematics tests is instable both across items and across languages.

More specifically, our results indicate that, in addition to individual differences in the mathematical skills measured, each ELL student has a unique set of strengths and weaknesses in both his/her first language and his/her second language; also, in addition to its intrinsic cognitive and academic demands, each item poses a specific set of linguistic demands in each language. Our results also indicate that, due to language variation, localities may vary tremendously as to both the language that is more appropriate to use to test ELLs and the number of items needed to obtain dependable scores. This variation occurs even among ELLs classified within the same level of English proficiency.

From a more general perspective, our results show that it is possible to use psychometric models that are consistent with the notion of language variation, and with the notion that each bilingual individual has a unique pattern of language dominance. Notice how, in a design like the one described above, the effect of language proficiency on performance in tests is not addressed by comparing ELL students with their mainstream counterparts.

Concluding Comments

In this paper, I have presented a conceptual framework for examining research on language and mathematics education. According to the framework, there are four views of language that influence how language is thought of and treated: as a process, as a system, as a structure, and as a factor. Views of language as a process and as a system can be called, functional because they focus on the role of language as a phenomenon that influences mathematical communication and the development of mathematical knowledge. Views of language as a structure and as a factor can be called, formal because they focus on patterns of linguistic features of mathematical problems or types of linguistic groups. No view of language is better than the others; each emphasizes a particular aspect of language.

While a focus on different aspects of language is a natural consequence of the fact that different areas of research investigate different issues, that does not necessarily imply that views of language cannot be used in combination. Yet my analysis shows that functional views are almost exclusive of research on teaching and learning (including classroom informal assessment) and formal views are almost exclusive of research and practice in large-scale testing. This divide may confirm the notion that multidisciplinary work is greatly needed in educational research (see Lee, 1999, 2002; Lee & Fradd, 1998; Pellegrino, Chudowsky, & Glaser, 2001).

Future research work should pay more attention to dialect diversity in mathematics education. Practices in the field of mathematics teaching are influenced by inaccurate assumptions about dialect. While the devastating consequences of devaluing non-standard dialects in the classroom have been discussed extensively (Brisk, 2006; Delpit, 1995; Wolfram, Adger, & Christian, 1999), more research is needed to determine effective ways to train teachers to create respectful, inclusive learning environments that make mathematical knowledge accessible to all students

Moschkovich (2007) observes that educators tend to base their views about language and learners on vocabulary, on the multiple meanings of words, or on discourse. These views are associated with the views they have about bilingual individuals, respectively as deficient, as facing more difficulties than monolingual students in learning mathematical register and dealing with multiple meanings, and as individuals whose competencies and resources may be comparable to the competencies and resources of mainstream students. According to Moschkovich, only when they have a view of language as discourse, teachers are capable of viewing instruction as a means for uncovering the bilingual students' competences and building mathematical knowledge upon those competences.

Lauren Young (personal communication) has suggested the notion of *mathematical language knowledge* as a form of knowledge that professional development should address along with content mathematical knowledge and pedagogical mathematical knowledge. The contribution of this notion may be worth exploring because it appears to address, in its right dimension, the relevance of language in mathematics teaching and learning. Developing a sophisticated view of language is unlikely to occur simply from being provided with certain language principles. For example, properly addressing the tension between languages in multilingual classrooms takes sophisticated teaching skills, accurate knowledge of language issues, and a favorable attitude towards linguistic diversity.

Although conventional professional development activities can change individuals' stated beliefs about language and its relationship to the teaching of disciplinary knowledge, those changes are not necessarily reflected in teachers' practices (Lee, Hart, Cuevas, & Enders, 2004). A deep transformation of teacher's practices takes place only when there is opportunity for reflection and insight and extensive, continuous support from colleagues and facilitators (Lee, 2004; 2005). Thus, in addition to the multiple facets of language and its complex relation to mathematics, the role of culture in the mental representation of mathematics (Stigler & Baranes, 1988-1989) and the social dimension of mathematical thinking (Moses, 2001) need to be considered in order to accurately specify the domain of mathematical language knowledge. This kind of knowledge involves attitudes, beliefs, and thinking and practice that go far beyond basic language principles.

As with research on teaching, future research on mathematics assessment work should pay more attention to dialect diversity. The relevance of dialect as a fairness and validity issue in testing may have been underestimated in the past due to the fact that dialect differences are more subtle and less obvious than language and language proficiency level (see Freedle, 2003). In addition, because of the lack of appropriate knowledge on language, non-standard dialects may not be properly considered in testing practices. The use of standard English is frequently invoked by test developers as a proof that dialect variation is properly addressed. The underlying assumption is that tests are fair if they are written in standard English because standard English is the dialect that everybody understands. However, current thinking in the field of sociolinguistics recognizes that dialect is, in reality, the dialect of the segment of the population of a society which has social and economic power (Halliday, 1978). The linguistic features of a test written in Standard English reflect the totality of the features of the dialects used by the privileged segment of the society but only a portion of the features of dialects used by other groups (Solano-Flores, 2006).

There is evidence that subtle but important issues of dialect can be detected and properly addressed if teachers are allowed to participate in the process of test review, by discussing the items at length and adapting the linguistic features of mathematics tests to the characteristics of the language (either English or their ELL's first language) used in their communities (Solano-Flores, Speroni, & Sexton, 2005). This finding underscores the importance of a simply but frequently ignored fact—that language, as a social phenomenon, can be properly addressed only through social participation.

All these facts speak to the need for a deeper understanding of language issues among researchers, practitioners, and policy makers. If we take seriously the goal of making mathematics accessible to all students, and if mathematics teaching and assessment are to be a coordinated effort, then we need to have clarity about the possibilities and limitations of each view of language across the different aspects of mathematics education. Hopefully, this paper has contributed to address these needs.

Notes

Note 1. The text of the three word problems used as examples is taken verbatim from De Corte, Verschaffel, and De Win (1985).

Note 2. The linguistic group categories used in research and practice in mathematics assessment are determined based on tests of language proficiency or language development whose construction may be based on views of language as a system or as a process. However, this discussion is about how language is addressed in research on mathematics assessment, not how language is measured to classify students according to language proficiency.

Note 3. In some cases, countries participating in international comparisons use two language versions to test different linguistic groups within them (e.g., K. O'Connor & Malak, 2000).

Note 4. Criteria for determining when a student is no longer an ELLs also may be based on erroneous assumptions about language development. Such is the case of the time of schooling as a criterion for deciding when an ELL should be assumed to be able to take tests in English. In spite of the debate over the distinction proposed by Cummins in the early 1980's between the concept of basic interpersonal skills—the skills involved in conversational fluency—and cognitive academic language proficiency—the linguistic proficiency needed to succeed academically—(see Cummins, 2003; MacSwan & Rolstad, 2003; Rivera, 1984), there is consensus among specialists that it is not reasonable to expect that adequate measures of academic achievement can be obtained for ELLs after a short period of immersion in a second language (Guerrero, 2004; Hakuta, 2001; Hakuta, Buttler, Witt, 2000). As discussed in the paper, academic language involves much more than vocabulary—it also involves skills such as negotiating meaning, constructing an argument, or expressing disagreement (Echevarria & Short, 2002; Scarcella, 2003).

References

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometric issues. *Educational Assessment, 8*(3), 231–257.

Abedi, J. (2004). The No Child Left Behind act and English language learners: Assessment and accountability issues. *Educational Researcher*, *33*(1), 4-14.

Abedi, J., & Hejri, F. (2004). Accommodations for students with limited English proficiency in the National Assessment of Educational Progress. *Applied Measurement in Education, 17{4),* 371-392.

Abedi, J. Hofstetter, C., Baker, E., & Lord, C. (2001). *NAEP math performance and test accommodations: Interactions with student language background.* CSE Technical Report No. 536. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74 (1),* 1-28.

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

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice, 19*(3), 16-26.

Adler, J. (1998). A language of teaching dilemmas: Unlocking the complex multilingual secondary mathematics classroom. *For the Learning of Mathematics, 18*(1), 24-33.

Adler, J. (1995). Dilemmas and a paradox—secondary mathematics teachers' knowledge of their teaching in multilingual classrooms. *Teaching and Teacher Education, 11*(3), 263-274.

Allalouf, A. (2003). Revising translated differential item functioning items as a tool for

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Baker, C. (2006). *Foundations of bilingual education and bilingualism, Fourth edition.* Clevedon, UK: Multilingual Matters

Barwell, R. (2005). Ambiguity in the mathematics classroom. *Language and Education, 19*(2), 118-126.

Bialystok, E. (2001). *Bilingualism in development.* Cambridge, UK:Cambridge University Press.

Bialystok, E. (2002). Cognitive processes of L2 users. In V. J. Cook (Ed.). *Portraits of the L2 user.* Buffalo, NY: Multilingual Matters, Ltd. pp. 145-165.

Black, P. (1993). Formative and summative assessment by teachers. *Studies in Science Education, 21*, 49–97.

Bormuth, J. R. (1970). *On the theory of achievement test items.* Chicago: University of Chicago Press.

Brennan, R. L. (1992). *Elements of generalizability theory.* Iowa City, IA: The American College Testing Program.

Brenner, M.E. (1998a). Development of mathematical communication in problem solving groups by language minority students. *Bilingual Research Journal, 22*, 103-128.

Brenner, M.E. (1998b). Meaning and money. *Educational Studies in Mathematics, 36*, 123-155.

Brenner, M. E. (1994). A communication framework for mathematics: Exemplary instruction for culturally and linguistically diverse students. In D. McLeod (Ed.), *Language and learning: Educating linguistically diverse students* (pp. 233-267). Albany, NY: State University of New York.

Brisk, M. E. (2006). *Bilingual education: From compensatory to quality schooling, Second edition*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Brown, J. S. & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science, (2)*2, 71-192.

Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge, UK: Cambridge University Press.

Bialystok, E. (2001). *Bilingualism in development: language, literacy and cognition*. Cambridge: Cambridge University Press.

Burkhardt, H., & Schoenfeld, A. H. (2003). Improving educational research: toward a more useful, more influential, and better funded enterprise. *Educational Researcher, 32*(9), 3-14.

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational Measurement, Fourth Edition*. (pp. 221-256). Westport, CT: American Council on Education and Praeger Publishers.

Camilli, G., & Shepard, L.A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: SAGE Publications.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards, & R. Schmidt (Eds.), *Language and communication* (pp. 2-27). London: Longman.

Cook, L. L, & Schmitt-Cascallar, A. P. (2005). Establishing score comparability fo tests given in different languages. In: R. K.Hambleton, P. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 139-169). Mahwah, NJ: Erlbaum.

Corson, D. (2001). *Language diversity and education*. Mahwah, NJ: Lawrence erlbaum Associates, Publishers.

Coulmas, F. (2005). *Sociolinguistics: The study of speakers' choices*. Cambridge, UK: Cambridge University Press.

Crawford, J. (2000). *At war with diversity. US language policy in an age of anxiety*. Clevedon, UK: Multilingual Matters.

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.

Crystal, D. (1997). *The Cambridge encyclopedia of language, Second edition*. Cambridge, UK: Cambridge University Press.

Cummins, J. (2003). BICS and CALP: Origins and rationale for the distinction. In C. B. Paulston & G. R. Tucker (Eds.,) *Sociolinguistics: The essential readings* (pp. 322-328). Malden, MA: Blackwell Publishing, Ltd.

Davis, A. (1991). The language of testing. In. K. Durkin & B. Shire (Eds.), *Language in mathematical education: Research and practice* (pp. 40-47). Buckingham, UK: Open University Press.

De Avila, E. A. (1988). Bilingualism, cognitive function, and language minority group membership. In Rodney R. Cocking & Jose P. Mestre, *Linguistic and Cultural Influences on Learning Mathematics. Hillsdale*, NJ: Erlbaum, 101-121.

De Corte, E., & Verschaffel, L. (1991). Some factors influencing the solution of addition and subtraction word problems. In. K. Durkin & B. Shire (Eds.), *Language in mathematical education: Research and practice* (pp. 17-30), Buckingham, UK: Open University Press.

De Corte, E., Verschaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology, 77*(4), 460-470.

De Corte, E., Verschaffel, L., & Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders' eye movements. *Journal of Educational Psychology, 82*, 359-365.

Delpit, L. (1995). *Other people's children: Cultural conflict in the classroom.* New York, NY: New Press. (pp. 48-69).

Durán, R. P. Testing of linguistic minorities. (1989). In R. L. Linn, (Ed.), *Educational measurement, third edition.* (pp. 573-587). New York: American Council on Education-Macmillan Publishing Company.

Echevarria & Short, 2002. *Using Multiple Perspectives in Observations of Diverse Classrooms: The Sheltered Instruction Observation Protocol (SIOP).* Center for Research on Education, Diversity, and Excellence. http://crede.berkeley.edu/tools/policy/siop/1.3doc2.shtml. Retrieved from the world wide web, February 25, 2007.

Ercikan, K. (1998). Translation effects in international assessment. *International Journal of Educational Research, 29*, 543-553.

Ercikan, K. (2002). Disentangling sources of differential item functioning in multi-language assessments. *International Journal of Testing*, 2, 199-215.

Farr, M., & Ball, A.F. (1999). Standard English. In B. Spolsky (Ed.), *Concise encyclopedia of educational linguistics*. Oxford, UK: Elsevier.

Fasold, R., & Connor-Linton, J. (2006). Introduction. In J. Fasold & J. Connor-Linton (Eds.), *An introduction to language and linguistics*. Cambridge, UK: Cambridge University Press.

Ferguson, A.M., & Fairburn, J. (1985). Language experience for problem solving in mathematics. *Reading Teacher, 38*, 504-507.

Fishman, J. A. 1965. Who speaks what to whom and when? *Linguistique*, 2, 67-88.

Freedle, R. O. (2003). Correcting the SAT 's ethnic and social-class bias: A method for reestimating SAT scores. *Harvard Educational Review, 73*(1), 1-43.

Fuson, K. C. (1991). Children early counting: Saying the number-word sequence, counting objects, and understanding cardinality. In. K. Durkin & B. Shire (Eds.), *Language in mathematical education: Research and practice* (pp. 27-39). Buckingham, UK: Open University Press.

Garcia, E. E. (1993). Language, culture, and education. *Review of Research in Education, 19*, 51-98.

García, G. E., McKoon, G., & August, D. (2006). Language and literacy assessment of language-minority students. In August, D. & Shanahan, T. (Eds.), *Developing literacy in second-language learners: report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers. (pp.597-626).

Greenfield, P. M. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist, 52*(10), 1115-1124.

Greeno, J. (1989). For the study of mathematics epistemology. In R. Charles & E. Silver (Eds.), *The teaching and assessing of mathematical problem solving* (pp. 23-31). Reston, VA: National Council of Teachers of Mathematics.

Grosjean, F. (1985). The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual and Multicultural Development*, 6, 467-477.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language 36,* 3-15.

Guerrero, M. D. (2004). Acquiring academic English in one year: An unlikely proposition for English language learners. Urban Education, 39(2), 172-199

Gunning, T. (2003). The role of readability in today's classrooms. *Topics in Language Disorders, 23*, 175-189.

Gutiérrez, K. D., Baquedano-Lopez, P. & Tejeda, C. (1999). Rethinking diversity: Hybridity and hybrid language practices in the third space. *Mind, Culture, and Activity 6*(4), 286-303.

Hakuta, K. (2001). *How long does it take English learners to attain proficiency?*. University of California Linguistic Minority Research Institute. Policy Reports. Santa Barbara: Linguistic Minority Research Institute. Available at: http://repositories.cdlib.org/lmri/pr/hakuta

Hakuta, K., Butler, Y.G., & Witt, D. (2000). *How long does it take for English learners to attain proficiency?* Policy Report 2000-1. Santa Barbara, CA: University of California Linguistic Minority Research Institute.

Halliday, M. A. K. (1978). Language as social semiotic: The social interpretation of language and meaning. London: Edward Arnold.

Halliday, M. A. K. (1993). Some grammatical problems in scientific English. In M. A. K. Halliday & J. R. Martin (Eds.), Writing science: Literacy and discursive power (pp. 69-85). Pittsburgh, PH: University of Pittsburgh Press.

Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In: R. K.Hambleton, P. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). Mahwah, NJ: Erlbaum.

Harary, F. (1969). *Graph theory.* Reading, Massachusetts: Addison-Wesley Publishing Company, Inc.

Harrison, C. (1999). Readability. In B. Spolsky (Ed.), *Concise encyclopedia of educational linguistics* (pp. 428-431). Oxford, UK: Elsevier.

Hively, W., Patterson, H. L., & Page, S. H. (1968). A "universe-defined" system of arithmetic achievement tests. *Journal of Educational Measurement, 5* (4), 275-290.

Hymes, D. (1972). On communicative competence. In J. B. Price, & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Harmondsworth, England: Penguin.

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

Kidron, Y., & Solano-Flores, G. (2006). *Formal and judgmental approaches in the analysis of test item linguistic complexity: A comparative study*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California, April 8-12.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.) and Mathematics Learning Study Committee, National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Kopriva, R., & Saez, S. (1997). Guide to scoring LEP student responses to open-ended mathematics items. Washington, DC: Council of Chief State School Officers.

Ladson-Billings, G. (1995). Making mathematics meaningful in multicultural contexts. In W. G. Secada, E. Fennema, & L. B. Adjian (Eds.) *New directions for equity in mathematics education*. (pp. 126-145). Cambridg, UK: University Press.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal, 27,* 29-64.

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

Lave, G., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge, UK: Cambridge University Press.

Lee, O. (1999). Equity implications based on the conceptions of science achievement in major reform documents. *Review of Educational Research, 69*(1), 83–115.

Lee, O. (2002). Promoting scientific inquiry with elementary students from diverse cultures and languages. *Review of Research in Education, 26,* 23–69.

Lee, O. (2004). Teacher change in beliefs and practices in science and literacy instruction with English language learners. *Journal of Research in Science Teaching, 41*(1), 65-93.

Lee, O. (2005). Adaptation of an instructional intervention in linguistically, culturally, and socioeconomically diverse elementary schools. Paper presented at the symposium on "Fidelity of Implementation" at the annual meeting of the American Educational Research Association, Montreal, Canada.

Lee, O., & Fradd, S. H. (1998). Science for all, including students from non-English language backgrounds. *Educational Researcher, 27*(4), 12–21.

Lee, O., Hart, J. E., Cuevas, P., & Enders, C. (2004). Professional development in inquiry-based science for elementary teachers of diverse student groups. *Journal of Research in Science Teaching, 41*(10), 1021-1043.

Leung, C. (2005). Mathematical vocabulary: Fixers of knowledge or points of exploration? *Language and Education, 19*(2), 127-135.

MacSwan, J. (2000). The threshold hypothesis, semilingualism, and other contributions to a deficit view of linguistic minorities. *Hispanic Journal of Behavioral Sciences, 22* (1), 3-45.

MacSwan, J., & Rolstad, K. (2003). Linguistic diversity, schooling, and social class: Rethinking our conception of language proficiency in language minority education. In C. B. Paulston & G. R. Tucker (Eds.,) *Sociolinguistics: The essential readings* (pp. 329-340). Malden, MA: Blackwell Publishing, Ltd.

Mesthrie, R., & Leap, W. L. (1999). Language contact 1: Maintenance, shift and death. In Mesthrie, R., Swann, J., Deumert, A., & Leap, W. L. (Eds.), *Introducing sociolinguistics*. (pp. 248-278). Philadelphia, PA: John Benjamins Publishing Company.

Meyerson, M. J., Ford, M. S., Jones, W. P., & Ward, A. W. (1991). Science vocabulary knowledge of third and fifth grade students. *Science Education, 75*, (4), 419-428.

Miller, K.F., & Stigler, J. W. (1987). Counting in Chinese: Cultural variation in a basic cognitive skill. *Cognitive Development, 2*, 279-305.

Mitchell, J. M. (2001). Interactions between natural language and mathematical structures: The case of "wordwalking". *Mathematical Thinking and Learning, 3*(1), 29–52.

Miura, I. T. (1987). Mathematics achievement as a function of language. *Journal of Educational Psychology, 81*, 109-113.

Miura, I. T., & Okamoto, Y. (2003). Language supports for mathematics understanding and performance. In A. J. Baroody & A. Dowker (eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Miura, I. T., Okamoto, Y., Vlahovic-Stetic, V., Kim, C. C., & Han, J. H. (1999). Language supports for children's numerical fractions: Understanding of cross-national comparisons. *Journal of Experimental Child Psychology, 74*, 356-365.

Morgan, C. (1998). *Writing mathematically: The discourse investigation*. London, UK: Falmer Press.

Morgan, C. (2005). Words, definitions and concepts in discourses of mathematics, teaching and learning. *Language and Education, 19*(2), 103-117.

Moschkovich, J. N. (2002). A situated and sociocultural perspective on bilingual mathematics learners. *Mathematical Thinking and Learning*, Special issue on Diversity, Equity, and Mathematical Learning, N. Nassir and P. Cobb (Eds.), 4(2&3), 189-212.

Moschkovich, J. (2006). Using two languages when learning mathematics. *Educational Studies in Mathematics, 64*(2), 121-144.

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

Moses, R. P. (2001). *Radical equations: Math literacy and civil rights*. Boston: Beacon Press.

Nagy, W. (1988). *Teaching Vocabulary to Improve Reading Comprehension*. Urbana, IL: National Council of Teachers of English.

Nasir, N. S. (2002). Identity, goals, and learning: Mathematics in cultural practice. *Mathematical Thinking and Learning, 4*(2&3), 213–247.

National Assessment of Educational Progress (1996). *Mathematics items public release*. Washington, DC: Author.

National Assessment of Educational Progress (2007). *NAEP Inclusion Policy*. Retrieved from world wide web, February 12, 2007. http://nces.ed.gov/nationsreportcard/about/inclusion.asp#accom_table

National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs (2006). *NCELA frequently asked questions.* Revised October 2006. Retrieved from the world wide web, February 25, 2007. http://www.ncela.gwu.edu/expert/faq/25_tests.htm

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

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

Nettle, D., & Romaine, S. (2002). *Vanishing voices: The extinction of the world's languages*. New York, NY: Oxford University Press, Inc.

No Child Left Behind Act of 2001 (2001). Pub. L. No. 107-110, 115 stat. 1961 (2002). Retrieved from the world wide web, February 25, 2007. http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Noonan, J. (1990). Readability problems presented by mathematics text. *Early Child Development & Care, 54*, 57-81.

O'Connor, M. C. (1998). Language socialization in the mathematics classroom: Discourse practices and mathematical thinking. In M. Lampert & M. Blunk (Eds.), *Talking mathematics: Studies of teaching and learning in school* (pp. 17-55). NY: Cambridge University Press.

O'Connor, K. M. y Malak, B. (2000). Translation and cultural adaptation of the TIMSS instruments. In M. O. Martin, K. D.Gregory y S. E. Stemler, (Eds.), *TIMSS 1999 technical repor.t* (pp. 89-100). Chestnut Hill, MA: International Study Center, Boston College.

Paul, D.J., Nibbelink, W.H., & Hoover, H.D. (1986). The effects of adjusting readability on the difficulty of mathematical story problems. *Journal of Research in Mathematical Education, 17,* 163-171.

Pellegrino, J. W., Chudowsky, N. & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Pimm, C. (1987). *Speaking mathematically: Communication in mathematics classrooms.* London, UK: Routledge & Kegan Paul Ltd.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics, 18*, 581-618.

Preston, D.R. (Ed.). (1993). *American dialect research.* Philadelphia: John Benjamins.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153-196). New York: Academic. Press.

Rivera, C. (1984). *Language proficiency and academic achievement.* Clevedon, UK: Multilingual Matters.

Rivera, C., Collum, E., Willner, L. N. & Sia, Jr. Jose Ku. (2006). Study 1: An analysis of state assessment policies regarding the accommodation of English language learners. In Rivera, C. & Collum, E. (Eds.), *State assessment policy and practice for English language learners: A national perspective* (pp. 1-136). Mahwah, NJ: Lawrence Earlbaum Associates, Publishers.

Rivera, C., & Stansfield, C. W. (2004). The Effect of Linguistic Simplification of Science Test Items on Score Comparability. *Educational Assessment, 9*(3&4), 79-105.

Romaine, S. (1995). *Bilingualism*, 2nd ed. Malden, MA: Blackwell Publishing.

Rowland, T. (2000). *The pragmatics of mathematics education: Vagueness in mathematical discourse.* London, UK: Falmer Press.

Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment, 11*(3 & 4), 205–235

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring Teachers' Informal Formative Assessment Practices and Students' Understanding in the Context of Scientific Inquiry. *Journal of Research Science Teaching, 44*(1), 57-84.

Saxe, G. B. (1988). Linking language with mathematics achievement: Problems and prospects. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 47-62). Hillsdale, NJ: Lawrence Erlbaum Associates.

Scarcella, R. C. (2003). *Academic English: A conceptual framework.* Report 2003-1. Santa Barbara, CA: University of California Linguistic Minority Research Institute.

Schmeiser, C. B., & Welch, C. J. (2006). Test development. In R. L. Brennan (Ed.), *Educational Measurement, Fourth Edition.* (pp. 307-353). Westport, CT: American Council on Education and Praeger Publishers.

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective.* Mahwah, NJ: Erlbaum.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334-370). New York: MacMillan.

Schoenfeld, A. H. (2004). The math wars. *Educational Policy*, 18(1), 253-286.

Schoenfeld, A. H. (2006). Method. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning.* New York: MacMillan.

Setati, M. (2002). Researching mathematics education and language in multilingual South Africa. *The Mathematics Educator, 12*(2), 6-19.

Setati, M., & Adler, J. (2002). Between languages and discourses: Language practices in primary multilingual mathematics classrooms in South Africa. *Educational Studies in Mathematics 43*, 243–269.

Sfard, A. (2000). On reform movement and the limits of mathematical discourse, *Mathematical Thinking and Learning*, *2*(3), 157–189.

Sfard, A., & Cole M. (2003). *Literate mathematical discourse: What it is and why should we care?* Unpublished manuscript. Retrieved August 8, 2007, from http://lchc.ucsd.edu/People/mcole_bio.html

Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, G. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment, 11*(2), 105–126.

Shavelson, R.J., & Webb, N.M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

Shorrocks-Taylor, D. & Hargreaves, M. (1999). Making it clear: A review of language issues in testing with special reference to the mathematics tests at key stage 2. *Educational Research, 41*(2), 123-136.

Siegler, R. S. & Jenkins, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sireci, S. G., & Allalouf, A. (2003). Appraising item equivalence across multiple languages and cultures. *Language Testing, 20*, 148-166.

Sireci, S. G., Li, S., & Scarpati, S. (2003). *The effects of test accommodation on test performance: A review of the literature* (Research Report 485). Amherst, MA: Center for Educational Assessment.

Sloane, F., & Kelly, A. (2003). Issues in high stakes testing programs. *Theory into Practice, 42*(1), 12-18.

Solano-Flores, G. (2006). Language, dialect, and register: Sociolinguistics and the estimation of measurement error in the testing of English-language learners. *Teachers College Record. 108*(11), 2354–2379.

Solano-Flores, G., & Li, M. (2006). The use of generalizability (g) theory in the testing of linguistic minorities. *Educational Measurement: Issues and Practice 25*, 13-22.

Solano-Flores, G., & Shavelson, R. J. (1997). Development of performance assessments in science: Conceptual, practical and logistical issues. *Educational Measurement: Issues and Practice, 16*(3), 16-25.

Solano-Flores, G., Speroni, C., & Sexton, U. (2005). *The process of test translation: Advantages and challenges of a socio-linguistic approach*. Paper presented at the annual meeting of the American Educational Research Association. Montreal, Quebec, Canada, April 11-15, 2005.

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher, 32*, 3-13.

Solano-Flores, G., & Trumbull, E. (In press). In what language should English language learners be tested? To appear in: Kopriva, R. (Ed.), *Improving Large-Scale Achievement Tests for English Language Learners.* Lawrence Erlbaum.

Solano-Flores, G., Trumbull, E., & Kwon, M. (2003). *The metrics of linguistic complexity and the metrics of student performance in the testing of English language learners.* Symposium paper presented at the 2003 annual meeting of the American Evaluation Research Association. Chicago, IL, April 21-25.

Stigler, J. W., & Baranes, R. (1988-1989). Culture and mathematics learning. *Review of Research in Education, 15*, 253-306).

Thurber, R. S., Shinn, M. R., & Smolkowski, K., (2002). What is Measured in Mathematics Tests? Construct Validity of Curriculum-Based Mathematics Measures. *School Psychology Review, 31*, 498-513.

Trumbull, E., Farr, B. (2005). Introduction to language. In E. Trumbull & B. Farr (Eds.), *Language and learning: What teachers need to know* (pp. 1-32). Norwood, MA: Christopher-Gordon.

Valdés, G., & Figueroa, R. A. (1994). *Bilingualism and testing: A special case of bias*. Norwood, NJ: Ablex.

van der Linden, W. J., & Hambleton, R. K. (Eds.) (1997). *Handbook of modern item response theory*. New York, NY: Springer-Verlag.

van de Vijver, F., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 47*, 263-279.

van Gelderen, (2000). *A grammar of English: Sleeping in mine orchard, a serpent stung me.* Tempe, AZ: Arizona State University. Retrived April 10, 2003 from www.public.asu.edu/~gelderen/314text/index.htm.

Veit, R. (1999). *Discovering English grammar.* Needham Heights, Massachusetts: Allyn & Bacon.

Vygotsky, L. S. (1936). [1986]. *Language and thought.* Cambridge, MA, MIT Press.

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Wardhaugh, R. (2002). An introduction to sociolinguistics, fourth edition. Oxford, UK: Blackwell Publishing.

Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A. & Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The logic of everyday languages. *Journal of Research in Science Teaching, 38*(5), 529-552.

Webb. N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22*, 366-389.

Wellington, J., & Osborne, J. (2001). *Language and literacy in science education.* Buckingham, UK: Open University Press.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives, 10*(12). Retrieved, March 2, 2007 from the world wide web. http://epaa.asu.edu/epaa/v10n12/.

Wertsch, J. V. (1985). *Vygotsky and the social formation of mind.* Cambridge, Massachusetts: Harvard University Press.

William, D. (1999a). Formative assessment in mathematics. Part 1: Rich questioning. *Equals: Mathematics and Special Educational Needs, 5*(2), 15-18.

William, D. (1999b). Formative assessment in mathematics. Part 2: Feedback. *Equals: Mathematics and Special Educational Needs,* 5(3), 8-11.

Wolfram, W., Adger, C. T., & Christian, D. (1999). *Dialects in schools and communities.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Yen, W. M., & Pritzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational Measurement, Fourth Edition* (pp. 111-153). Westport, CT: American Council on Education and Praeger Publishers.

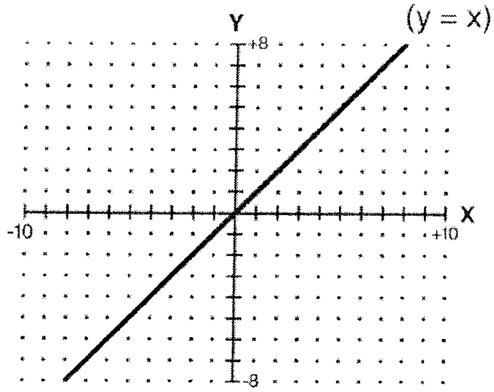
Figure captions

Figure 1. Steepness problem (Moschkovitch, 2006).

Figure 2. Syntactical structure of two sentences from NAEP Grade 4 mathematics items. S=sentence; N=noun; V=verb; Det=determiner; P=phrase; Mod=modifier; Adj=adjective; Aux=auxiliar; NP=noun phrase; VP=verb phrase; AP=adverb phrase; PP=prepositional phrase. From Solano-Flores, Trumbull, & Kwon (2003).

Figure 3. The No Child Left Behind definition of English language learner (No Child Left Behind act of 2001).

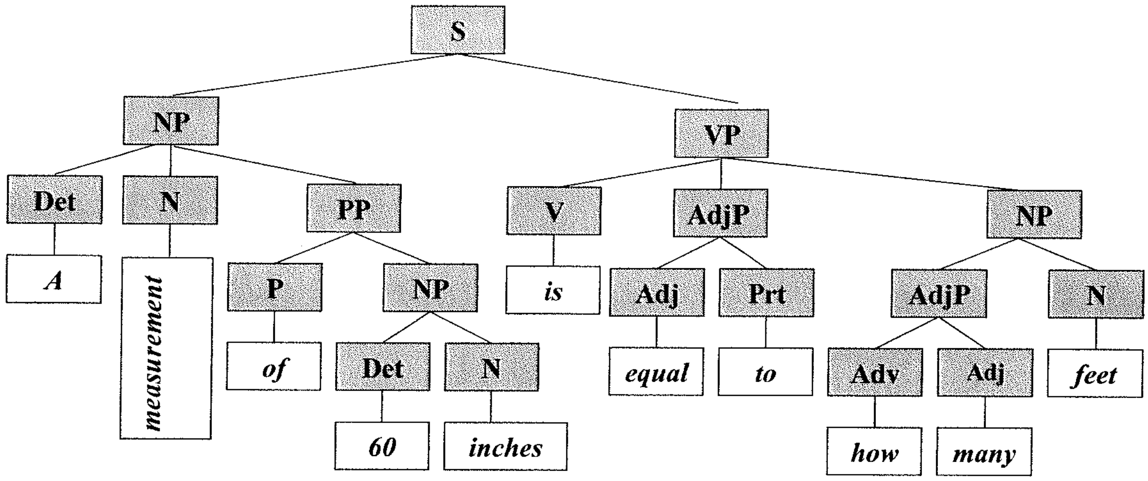
8a. If you change the equation y=x to y=-0.6x,
how would the line change?



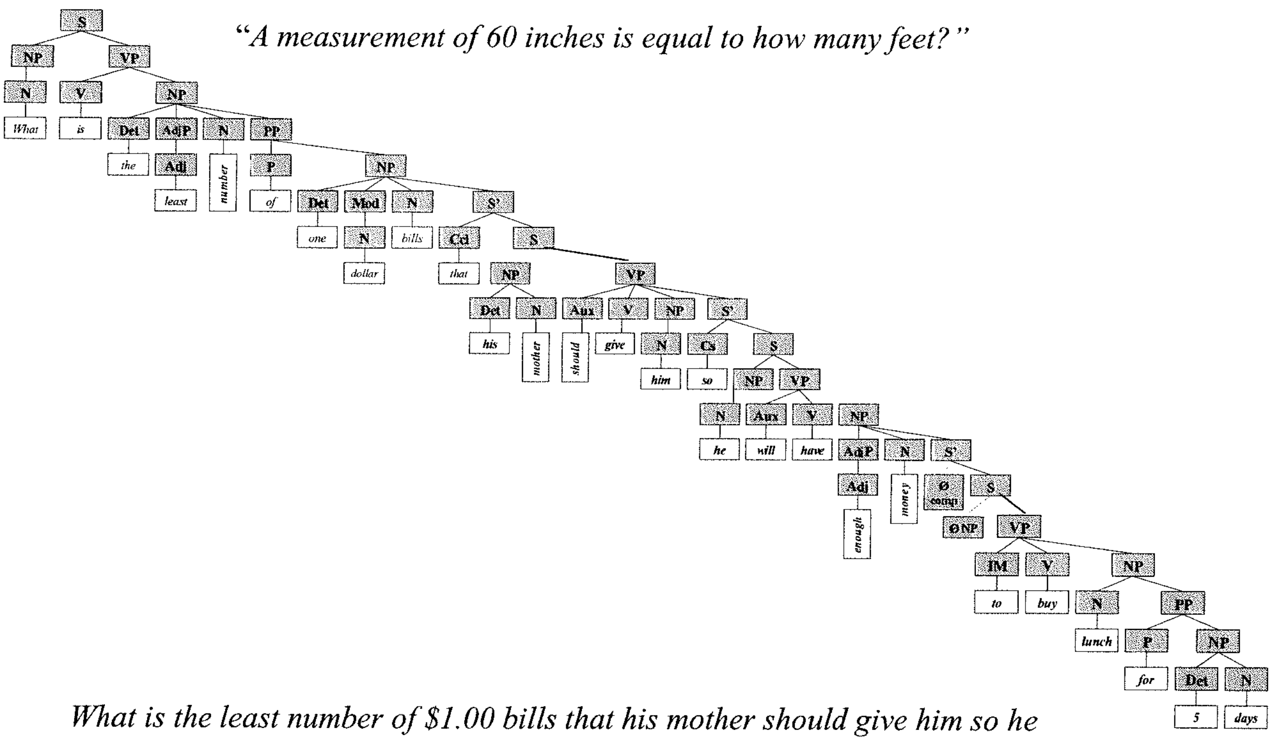
A. The steepness would change.
Why or why not?

____NO ____YES ___ STEEPER
___ LESS STEEP

Figure 1



"A measurement of 60 inches is equal to how many feet?"



What is the least number of $1.00 bills that his mother should give him so he will have enough money to buy lunch for 5 days?

Figure 2

"(25) LIMITED ENGLISH PROFICIENT.—The term 'limited English proficient', when used with respect to an individual, means an individual—

"(A) who is aged 3 through 21;

"(B) who is enrolled or preparing to enroll in an elementary school or secondary school;

"(C)(i) who was not born in the United States or whose native language is a language other than English;

"(ii)(I) who is a Native American or Alaska Native, or a native resident of the outlying areas; and

"(II) who comes from an environment where a language other than English has had a significant impact on the individual's level of English language proficiency; or

"(iii) who is migratory, whose native language is a language other than English, and who comes from an environment where a language other than English is dominant; and

"(D) whose difficulties in speaking, reading, writing, or understanding the English language may be sufficient to deny the individual—

"(i) the ability to meet the State's proficient level of achievement on State assessments described in section 1111(b)(3);

"(ii) the ability to successfully achieve in classrooms where the language of instruction is English; or

"(iii) the opportunity to participate fully in society.

Figure 3

Table 1. Views of language in research on language and mathematics.

	Functional views		Formal views	
	As a process	As a system	As a structure	As a factor
Role of language	Means for understanding	Resource for knowledge construction	Agent of problem complexity	Extraneous variable
Themes	Development and cognition	Social interaction and communication	Organization and difficulty	Condition and control
Unit of analysis	Individual learner, classroom	Classroom, community	Problem type	Group
Language modes	Speaking and writing	Classroom conversation	Reading (printed text)	Reading and writing
Key concepts	Meaning, register	Language, dialect, discourse, bilingualism, multilingualism, language choice, language contact	Grammar constituents, semantic structure	Language proficiency, testing conditions, bias
Areas of research	Language influences in the development of mathematical knowledge	Linguistic diversity in mathematics education	Influence of the linguistic features of mathematics problems on student performance and problem solving strategies	Effect of language differences on the accuracy of measures of mathematics achievement
Theories and disciplines	Sociocultural theory, constructivism, historical cultural theory, discourse theory, cognitive psychology, cognitive anthropology, cultural anthropology, sociolinguistics		Psychometrics, item response theory, cognitive psychology, structural linguistics	