Examining Academic Language in Mathematics Test Items for English Language Learners

Guillermo Solano-Flores

University of Colorado at Boulder

Rachel Prosser

University of Colorado at Boulder

Carne Barnett-Clarke

WestEd

Maxie Alexandra Gluckman

University of California, Los Angeles

Paper presented at the annual meeting of the American Educational Research Association Denver, April 30-May 4. Symposium, Rethinking the Mathematics and Language Connection for Latinos/as and ELLs: Findings from the *Math Pathways & Pitfalls* Project.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305K050050 to WestEd. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.

Examining Academic Language in Mathematics Test Items for English Language Learners

Guillermo Solano-Flores

University of Colorado at Boulder

Rachel Prosser

University of Colorado at Boulder

Carne Barnett-Clarke

WestEd

Maxie Alexandra Gluckman

University of California, Los Angeles

Abstract

This paper reports on an investigation which examined academic language in mathematics tests for English language learners (ELLs). The investigation is part of a broader project, *Math Pathways and Pitfalls (MPP)* .We investigated whether MPP tests intended to assess mathematical content knowledge and tests intended to assess mathematics academic language differ in their mathematical academic language load (ALL). To achieve this goal, we developed a conceptual framework on mathematical academic language and a rubric for coding academic language in mathematics test items. Our conceptual framework identifies five academic language dimensions: symbolic; lexical; analytical; visual; and register. Two independent coders coded the items according to a double blind review procedure. These coders coded the items in sequences determined randomly with the intent to control for the effects of fatigue and practice. From this coding, we were able to determine whether the coding categories were understood consistently by independent coders and to identify any statistically significant differences in ALL between the items that assess mathematics content knowledge and those that assess mathematical academic language. We found that MPP effectively generated items that differed on their emphasis of academic language. Content knowledge (CL) items and mathematical language (ML) items were distinguishable by the frequency of types of their ALLs. The ALL of ML items was significantly greater than the ALL of CK items.

Academic language has been defined as the language used in "professional books, characterized by specific linguistic features associated with academic disciplines" (Scarcella, 2003, p. 9). It includes both the register of a discipline—which is defined in terms of the meanings underlying the grammar and words that are specific to a given discipline within a specific context (Halliday, 1978)—and the ways in which language is used to socialize within the context of a discipline, for example, to build an argument, formulate a problem, express disagreement, or discuss a topic (see Solano-Flores, in press; Solano-Flores & Trumbull, 2008).

The role that academic language plays in effectively supporting ELLs to achieve academically cannot be overestimated. It is well known that developing basic communication and conversational skills in a second language is very different from developing the vocabulary, expressions, and discursive forms that are specific to a discipline in that language (Cummins, 1981). Also, it is well known that the time needed to develop the academic language needed to understand and learn the content of that discipline and succeed academically may take between five and seven years of schooling (Hakuta, 2000).

Given the complex role of communication skills as an agent for learning, attending to the characteristics of academic language is especially relevant to ensuring effective mathematics instruction. For example, classroom mathematical conversations involve both communicating to learn mathematics and learning to communicate mathematically (Brenner, 1998; Khisty, 1995; Lampert & Cobb, 2003; Webb, 1991). Unfortunately, while current mathematics standards documents (e.g., National Council of Teachers of Mathematics, 2000) recognize communication as critical to organizing and consolidating mathematical thinking, they are silent about the relation of communication and academic language in ELL instruction. Yet, formally addressing academic language is necessary if ELLs are to benefit from mathematics instruction.

While it is recognized that the language of testing has characteristics that are not shared with typical classroom discussion or daily conversation (Abedi & Lord, 2001), available literature that addresses the conceptual underpinnings of academic language (e.g., Bailey & Butler, 2003; Scarcella, 2003) pays considerably more attention to academic language in the context of teaching than in the context of testing. However, we know that

> ...performing well on a standardized test requires ELLs to know more than the content area assessed by the test or the language in which it is administered. It also requires from them the use of the register of that discipline and the register of tests. This combined register is defined by the activity in which individuals are engaged at a given time (e.g., taking a test). Among other things, this register differs from other registers on features such as semantics (e.g., *root* has different meanings in colloquial language and in mathematics); word frequency (e.g., *ion* is mostly restricted to the content of science); idiomatic expressions (e.g., the phrase *None of the above* is used almost exclusively in multiple-choice tests); notation (e.g., *A divided by B* is represented as *A/B*); conventions (e.g., uppercase letters are used to denote variables); syntactical structures (e.g., the structure of multiple choice items in which an incomplete sentence [the stem] is followed by several phrases [the options]); and ways of building arguments (e.g., *Let A be an integer number*). (Solano-Flores, 2006, p. 2363).

In this paper, we address the need for effective approaches to examining the linguistic features of test items (Solano-Flores, 2008). We report on an investigation that examined the linguistic features of items used to evaluate the impact of *Math Pathways and Pitfalls*, a mathematics curriculum developed by WestEd with the purpose of providing English language learners (ELLs) with support for strengthening their understanding of difficult mathematical concepts and improving their capacity to use mathematics academic language (Clarke, 2007a, b, c). These tests covered mathematics content knowledge (CK) and mathematics academic language (ML).

For the purposes of this investigation, academic language was defined as the set of terms, syntactical structures, discursive styles, notation conventions, forms of graphic representation, and ways of asking questions and obtaining student responses in mathematics tests.

Critical to our investigation are two concepts. The first is, *academic language constituent* (ALC)—a discernible academic language feature that requires the test taker to know representation or usage conventions in order to access the content of the item. The second concept is *academic language load* (ALL)—the number of ALCs identified in an item.

We asked, *How are the CK and ML test items different, in terms of their academic language constituents?* To answer this question, we developed a system for coding the ALCs of the items. Then, we compared the CK and ML items in two ways. First, to determine whether the structure of the academic language demands of the items varied across content and grade, we examined the proportional frequencies of the observed ALCs belonging to each of the five categories. Second, we computed academic language load (ALL)—defined as the number of academic language constituents observed in an item—and performed a series of analyses of variance (ANOVA) to examine if any ALL differences observed in items of different content and grade were statistically significant.

Methods

Test Materials

The tests covered two content areas in two grades: Grade 4-mathematics content knowledge (CK) (35 items); Grade 4-mathematics academic language (ML) (32 items); Grade 5-CK (35 items); and Grade 5 ML (30 items).

We coded a total of 132 items from four tests used in *Math Pathways and Pitfalls* from the 2006 school year. These tests were given to students before and after implementation of the MPP curriculum. For each of the two grades (4 and 5), one test intended to assess CK and the other test intended to assess ML. Table 1 shows the number of items for each test.

Table 1. Number of test items by test and grade.

Test	Grade 4	Grade 5	Total
Mathematics content knowledge (CK)	35	35	70
Mathematics academic language (ML)	32	30	62
Total	67	65	132

Categories of Academic Language Constituents

To compare the linguistic features of CK and ML test items, we examined each MPP test items and identified the different ALCs present. Next, as Table 2 shows, we grouped those ALCs into five categories: (1) symbolic, (2) lexical, (3) analytical, (4) visual, and (5) register. While these categories are mutually exclusive, several ALCs from one or more categories may be present in the same item.

Symbolic ALCs are notational conventions used in mathematics to express variables, magnitude, precision, proportion, units, operations, and relationships. This category includes symbols commonly used in mathematics such as the operational signs (+, -), as well as representing fractions (⅔).

Lexical ALCs are everyday language terms also used in mathematics. As with science terms (see Osborne and Wellington, 2001), the meaning of terms in the context of mathematics may be similar to or different from their everyday meaning. Terms of a given register found in both everyday life and within the context of a discipline may have the same meaning or different meanings.

Analytical ALCs are specific to the field of mathematics and refer to mathematical concepts. This category includes aspects of the specific way of speaking about mathematics. Students must be able to master various expressions and syntactical structures in order to successfully communicate as mathematicians.

Visual ALCs are non-textual representations of shapes, position, and mathematical ideas involving relationships such as functions and proportions.

Finally, the register category of ALCs refers to the wide variety of testing register and mathematical discourse, including the terms, syntax, and discursive structures that are specific to mathematics in tests. This category includes cloze questions, question phrases, and noun phrases.

Coding Procedures

Two coders were trained to identify the academic language constituents shown in Table 2. Then, they were asked to independently and dichotomously code the presence or absence of each ALC in all 132 MPP items. The sequence in which the items were given to the coders was randomly assigned with the intent to control for the effects of fatigue or practice. To ensure that the coding decisions were not influenced by the coder's knowledge of the origin of the items, any information on grade or content was concealed.

Altogether, the coders made a total of 670 positive coding decisions (i.e., they coded a given ALC as present in one of the items). Of these 670 positive coding decisions, 144 were discrepant coding decisions in which one of the two coders coded the presence of an ALC and the other did not. Thus, the proportion of discrepant coding decisions relative to the total of coding decisions gives an inter-coder agreement of .785, which is sufficiently high to support the use of data based on independent coding.

In spite of the reasonably high inter-grader coefficient of .785, the coders were asked to discuss and resolve their discrepancies and produce a compiled version of their coding. This allowed us to document the origin of the discrepancies and assess the factors that hamper the coding of academic language used in tests. In all cases, the origins of the discrepancies were minor differences in the ways in which the coders interpreted some ALCs.

After the coding discrepancies were resolved, there were 607 positive coding decisions distributed across 132 items. This version of the coding was used in our analyses.

Structure of the Academic Language Demands Across Tests

An analysis of the relative frequencies of the different types of ALCs observed in the tests allows examination of the structure of the linguistic demands posed by the items from the four tests. Table 2 and Figure 1 show the percentages of ALC categories observed by test and grade. Consistent differences in the relative frequencies of ALC categories can be observed between CK and ML items. Whereas lexical, analytical, and register ALCs were more frequent among ML items than CK items, symbolic and visual ALCs were more frequent among CK items than ML items. These differences are, in general, consistent across grades and appear to indicate that CK and ML items were constructed based on different sets of academic language properties.

Table 2. Percentage of ALCs observed by test, grade, and ALC category.

Grade	Test items	Number of observations	Category				
			Symbolic	Lexical	Analytical	Visual	Register
4	CK (n=35)	141	46.81	12.77	19.15	14.89	6.38
	ML (n=32)	171	26.32	29.24	22.81	5.26	16.37
5	CK (n=35)	141	60.99	12.77	5.67	14.18	6.38
	ML (n=30)	154	31.17	29.87	18.18	3.90	16.88

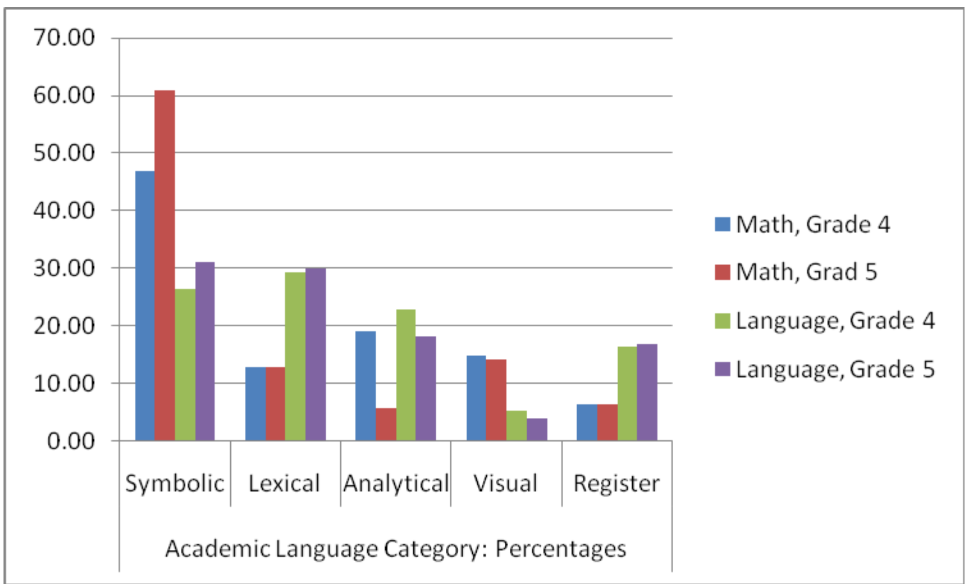

Figure 1. Percentage of ALCs observed by test, grade and ALC category.

Academic Language Load

We constructed a measure of the frequency of academic language features observed in the items, *academic language load* (ALL), which we define as the number of ALCs observed in an item. We computed the ALL for each item and performed an analysis of variance (ANOVA) to examine if any ALL differences observed between CK and ML items were statistically significant.

Table 3. Mean academic language load measures on the mathematics content knowledge (CK) and mathematics academic language (ML) tests by grade.

Grade	Test items	n	Mean	s.d.
4	CK	35	4.03	1.071
	ML	32	5.34	1.450
5	CK	35	4.03	1.403
	ML	30	5.13	1.665

Table 3 shows the mean ALL measures obtained for items from the four tests. The table shows that mean ALLs are higher for ML items than CK items in both the Grade 4 and Grade 5 tests. A factorial, test x grade ANOVA revealed statistically significant differences due to test and no statistically significant differences due to the grade or the interaction of test and grade (p=.000; eta squared=.164). ML items had a significantly greater academic load than CK items.

In summary, the CK and ML items have different sets of linguistic features. Whereas, proportionally, CK items have more ALCs belonging to the symbolic and visual categories than the ML items, the ML items have more ALCs belonging to the lexical, analytical, and register categories. This pattern is consistent across the tests used in Grade 4 and Grade 5. Also, ML items had a greater ALL than CK items; these differences were statistically significant.

Summary and Conclusions

This document reports on an investigation that examined the linguistic features of test items used to evaluate the impact of *Math Pathways and Pitfalls*, a mathematics curriculum developed by WestEd with the purpose of providing English language learners with support for strengthening their understanding of difficult mathematical concepts and improving their capacity to use mathematics academic language.

We investigated how the linguistic features of those items were related to the performance of Grade 4 and Grade 5 students who participated in MPP or who received other forms of instruction. These students were given a mathematics content knowledge (CK) test and a mathematics academic language (ML) test before and after instruction. The Grade 4-CK, Grade 4-ML, Grade 5-CK, and Grade 5-ML tests had, respectively, 35, 32, 35, and 30 items.

We examined the features of all the items and developed a system for coding academic language constituents (features). Next, for each item, we coded the academic language constituents (ALCs) observed and compared CK and ML items in two ways. First, we examined the proportional distribution of the ALCs coded across the five academic language categories for each test. Second, we measured the academic language load (ALL) of the items by counting the number of ALCs observed in them. Also, we performed an analysis of variance (ANOVA) to examine if any ALL differences observed in the items of different content and grade were statistically significant.

We found that CK and ML items are distinguishable by the sets of their predominant ALCs. We observed that, proportionally, the CK items had more ALCs belonging to the symbolic and visual categories than the ML items; in contrast, the ML items had more ALCs belonging to the lexical, analytical, and register and discourse categories. This pattern was consistent across the tests used in Grade 4 and Grade 5.

We also computed academic language load, ALL, which we defined as the number of ALCs identified in the items. We observed that ML items had a greater AL than CK items. A series of ANOVAs showed that these differences were statistically significant. MPP effectively generated items that differed on their emphasis of academic language. CK and ML items are distinguishable by the frequency of types of their ALCs. The ALL of ML items is significantly greater than the ALL of CK items. Also, the pattern of magnitude of correlations between ALL and the items' difficulties are different for CK and ML items.

Our results show that it is possible to develop approaches for effectively assessing the linguistic characteristics of test items according to measurable properties. We hope that this approach will be used in the future by test developers and researchers as part of efforts to examine objectively the linguistic properties of items in the testing of ELLs.

References

Abedi, J., & Lord, C. (2001). The Language Factor in Mathematics Tests. *Applied Measurement in Education*, *14*(3), 219-234.

Bailey, A. L. & Butler, F. A. (2003). *An evidentiary framework for operationalizing academic language for broad application to K-12 education: A design document.* (CSE Tech. Rep. No. 611). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

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

Clarke, C. (2007a). *Evaluating Math Pathways & Pitfalls Practical Lessons Learned Math and Science Partnership.* Regional Conference Dallas, Texas February 6-8.

Clarke, C. (2007b). *Math Pathways & Pitfalls: Lessons for K-7 Students: Improving Mathematics Achievement for English Language Learners Executive Summary of Performance and Contributions*. U.S. Department of Education Grant Performance Report (ED 524B).

Clarke, C. (2007c). Mathematical Language Quiz Development.

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (Ed*.),* *Schooling and language minority students: A theoretical framework.* Evaluation, Dissemination and Assessment Center, California State University, Los Angeles.

Hakuta, K. (2000). How long does it take English learners to attain proficiency. University of California Linguistic Minority Research Institute Policy Reports (University of California, Multi-Campus Research Unit), eScholarship Repository, University of California. http://repositories.cdlib.org/lmri/pr/hakuta

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

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.

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.

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

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

Scarcella, R. C. (2003). *Academic English: A conceptual framework* (Rep. No. 2003-1). Santa Barbara: University of Califronia Linguistic Minority Research Institute.

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

Solano-Flores, G. (In Press). *Function and form in the research on language and mathematics education.* In, J. N. Moschkovich, *Language and Mathematics Education: Multiple Perspectives and Directions for Research,* Information Age Publishing Inc.

Solano-Flores, G., & Trumbull, E. (2008). In what language should English language learners be tested? In R. J. Kopriva. (Ed.), *Improving testing for English language learners.* New York: Routledge.

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