10-2018

English Learners and Mathematical Word Problem Solving: A Systematic Review

Virginia Clinton
University of North Dakota, virginia.clinton@und.edu

Deni Lee Basaraba

Candace Walkington

Follow this and additional works at: https://commons.und.edu/efr-fac
Part of the Science and Mathematics Education Commons

Recommended Citation
Clinton, Virginia; Basaraba, Deni Lee; and Walkington, Candace, "English Learners and Mathematical Word Problem Solving: A Systematic Review" (2018). Educational Foundations and Research Faculty Publications. 6.
https://commons.und.edu/efr-fac/6

This Book Chapter is brought to you for free and open access by the Department of Educational Foundations and Research at UND Scholarly Commons. It has been accepted for inclusion in Educational Foundations and Research Faculty Publications by an authorized administrator of UND Scholarly Commons. For more information, please contact zeinebyousif@library.und.edu.
ENGLISH LEARNERS AND MATHEMATICAL WORD PROBLEM SOLVING: A SYSTEMATIC REVIEW

Virginia Clinton1,∗, Deni Lee Basaraba2 and Candace Walkington3
1Department of Educational Foundations and Research, University of North Dakota, Grand Forks, ND, US
2Bethel School, Eugene, OR, US
3Department of Teaching and Learning, Southern Methodist University, Dallas, TX, US

ABSTRACT

Successful solution of mathematics word problems (MWP) requires students to be able to understand the language of the MWP, which may be particularly challenging for English Learners (ELs). In this chapter, we review 21 empirical studies about specific linguistic features of MWP, and the effects of modifying linguistic features on mathematics word problem-solving performance. Results of our review indicated that a variety of linguistic features has differential effects on the mathematics word problem-solving performance of ELs (compared to non-ELs), and that the effects vary by linguistic feature and grade level. Additionally, the effects of modifying the linguistic features of items were mixed, with some studies indicating positive effects, some indicating negative effects, and others indicating mixed effects across different groups of students. We include recommendations for future research, particularly the need to test the effects of modifying specific linguistic features while holding other features constant. We conclude with implications for practice, both for test developers, who have direct control over the language of MWP, and for teachers, who can use this information to scaffold their mathematics instruction.

∗ Corresponding Author Email: Virginia.clinton@und.edu.
**ENGLISH LEARNERS AND MATHEMATICS WORD PROBLEM SOLVING: A SYSTEMATIC REVIEW**

Mathematical word problems (MWPs) in which numbers and mathematical operations are presented in text, are frequently used in mathematics instruction and assessment (Son & Kim, 2015; Walkington, Clinton, & Shivraj, 2018). Solving MWPs is a complex cognitive process because both language and mathematical skill are necessary (Wang, Fuchs, & Fuchs, 2016). Moreover, the process of solving MWPs is complicated by the fact that the language of mathematics is multi-semiotic and is comprised not only of words, but also of numerals, mathematical symbols, and visual representations (e.g., tables, charts, diagrams, etc.), to create meaning (Fang, 2012). This multi-semiotic nature contributes to the complexity of mathematics for both ELs and native English-speaking students (English-only; EOs) (Solano-Flores, Barnett-Clarke, & Kachchaf, 2013). Given these issues, it is not surprising that there are performance gaps between English Learners (ELs) and EOs on mathematics assessments (U.S. Department of Education, 2017).

Numerous studies have examined linguistic features of MWPs that are particularly challenging for ELs, as well as the effects of linguistic modification of MWPs, on ELs’ problem-solving performance (e.g., Abedi & Lord, 2001; Banks, Jeddeeni, & Walker, 2016; Beal & Galan, 2015; Johnson & Monroe, 2004; Lee & Randall, 2011; Martiniello, 2009). Findings from these studies are helpful for understanding how the language of word problems relates to the mathematical problem-solving performance of ELs. However, a cohesive review of the findings is lacking in the literature. Such a review would be helpful for designing word problems and providing scaffolding for ELs during mathematics instruction. The purpose of this systematic literature review is to provide a cohesive understanding of how linguistic features of MWPs relate to problem-solving performance by ELs. We focus only on the *linguistic* (or language-based) features of MWPs rather than symbols, numerals, or visuals, as we hypothesize language in particular may create challenges for ELs when solving MWPs.

**LITERATURE REVIEW**

MWPs may help students learn mathematical ideas by grounding them in real-world contexts, thereby making abstract concepts more concrete (Goldstone & Son, 2005). For example, the mathematical concept of division may be grounded in a word problem about dividing money among siblings. Grounding the mathematical operations within a verbal MWP may promote understanding, as students can connect the mathematical concept(s) (e.g., division) with familiar situations (e.g., sharing with their siblings; Koedinger, Alibali, & Nathan, 2008). In addition, MWPs are frequently used in mathematics assessments
(Walkington, Sherman, & Petrosino, 2012). According to Walkington, Clinton, and Shivraj (2018), 90% of the problems in two standardized mathematics assessments used for testing in the United States were MWPs. Comprehending the language (and linguistic structure) of a word problem in English is likely more difficult for ELs than EOs (Lesaux & Kieffer, 2010). Therefore, it would be helpful to know what particular linguistic features (e.g., the types of words used, syntactical structures) may facilitate or interfere with ELs comprehending and subsequently solving MWPs.

To solve a MWP successfully, students must coordinate multiple sources of information (Kintsch, 1998). Based on Kintsch’s construction-integration theory of text comprehension, reading and understanding the text of MWPs involves three levels of representation: the surface, the textbase, and the situation model. The surface structure involves the exact wording and syntax. The reader uses the words and syntax in the surface structure to create propositions (ideas or concepts) in the text. The reader joins these propositions together to form a textbase of meanings. Consider, for example, the MWP presented in the left panel of Figure 1 below, and the propositions derived from this MWP on the right:

<table>
<thead>
<tr>
<th>Proposition 1: Car 1 from Minneapolis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposition 2: The rate of Proposition 1 is 60 mph</td>
</tr>
<tr>
<td>Proposition 3: The direction of Proposition 1 is west</td>
</tr>
<tr>
<td>Proposition 4: Car 2 from Minneapolis</td>
</tr>
<tr>
<td>Proposition 5: Proposition 4 is two hours later than Proposition 1</td>
</tr>
<tr>
<td>Proposition 6: The direction of Proposition 4 is west</td>
</tr>
<tr>
<td>Proposition 7: The locations of Proposition 4 and Proposition 1</td>
</tr>
<tr>
<td>Proposition 8: Highway is same</td>
</tr>
<tr>
<td>Proposition 9: The rate of Proposition 4 is 70 mph.</td>
</tr>
<tr>
<td>Proposition 10: Car 2 overtakes Car 1</td>
</tr>
<tr>
<td>Proposition 11: How long for Proposition 10</td>
</tr>
</tbody>
</table>

Figure 1. Sample MWP and corresponding textbase (adapted from Nathan et al., 1992).

The textbase formed by these propositions is integrated with general world knowledge and personal experiences to create the situation model, an in-depth mental representation of the text composing the actions and relations in the text (Kintsch, 1998). The corresponding situation model for this MWP would likely consist of a representation of two cars driving on a highway with one eventually passing the other. An accurate rewording of a text would differ from the original in terms of surface representation but be similar in terms of textbase and situation model representation. For example, if the
first sentence of the MWP were restated as “Someone is driving west in a vehicle from Minneapolis at a rate of 60 miles per hour;” the surface representation in terms of the exact words and syntax would be different but the textbase and situation model representation would remain the same. In MWPs, the computation and symbolic operations, knows as the problem model, must be developed as well. For the example, the equation to solve the mathematical operations would be in the problem model. Thus, successful word problem solving requires students to coordinate their representations in their situation and problem models (Kintsch & Greeno, 1985; Nathan, Kintsch, & Young, 1992).

Developing a surface representation of MWPs can be particularly challenging due to the complexity of the mathematics register (Zevenbergen, Hydge, & Power, 2001), which is how language conveys meaning in a manner specific to mathematics (Halliday, 1978). One aspect of the mathematics register is technical vocabulary words (e.g., divisor, numerator; Schleppegrell, 2010). ELs may struggle with learning the technical vocabulary of mathematics if it is taught in English because it takes language to learn language. If ELs lack proficiency in English and they receive explanations in English, it may be difficult for them to understand the definitions for, or explanations of, technical vocabulary (Leung, 2005). The mathematics register also involves words that are used differently in mathematical contexts (e.g., face, product, left; Pimm, 1994). The different meanings of everyday words in mathematics may be especially difficult for ELs (k, 1995), because second language learners often initially understand and use words with multiple meanings (i.e., polysemous words), such as face and place, in terms of their most common use (Crossley, Salsbury, & McNamara, 2010). For example, ELs would be more likely to understand the word place to describe general vicinity (more common use) than function (less common use; Crossley et al., 2010).

Moreover, there are syntactical complexities in academic language specific to mathematics that should also be noted. For example, mathematical language uses nominalizations to create abstract “things” that can participate directly in mathematical reasoning (e.g., to sum becomes a virtual “thing” in MWPs, such as the sum of given numbers; Fang, 2012). In addition, the language in MWPs may ask students to complete relatively abstract actions, such as estimate the outcome of a given situation or isolate a specific variable to solve a MWP (Haag, Heppt, Stanat, Kuhl, & Pant, 2013).

In addition, it is likely that the difficulty of MWPs for ELs may increase with the length of the problem given that ELs tend to have more difficulty understanding English texts than EOs do (Lesaux & Kieffer, 2010). However, length, in terms of the number of words or sentences, sentence length, and phrase length, has also been noted as a source of difficulty in problem solving for students in general (Haag et al., 2013; Walkington, Clinton, Ritter & Nathan, 2015).
Research Methodologies

Psychometric methods have been used to investigate linguistic features of MWPs particularly challenging for ELs, such as differential item functioning (DIF; e.g., Loughran, 2014) and differential bundle functioning (DBF; Banks et al., 2016). DIF procedures involve comparing two groups of students, such as ELs and EOs, and identifying items on which there are differences in performance when the two groups have been matched on skill and should be performing similarly (Clauser & Mazor, 1998). Following this, DIF can be used to identify items on which ELs perform differently than EOs for reasons other than mathematical skill, such as linguistic features (Martiniello, 2009). DBF is similar to DIF, except bundles (i.e., groups) of items are examined (Banks, 2013). DIF and DBF are useful for determining which MWPs are more difficult for ELs than EOs, but it can be difficult to discern why student had difficulty with a particular MWP (i.e., which particular linguistic feature of the MWP contributed to the different performance based on whether English is L1 or L2).

Qualitative approaches in which ELs are interviewed about what they find challenging or helpful in MWPs (Celedón-Pattichis, 1999) or asked to think-aloud their cognitive processes as they work through problems (Martiniello, 2008) illuminate potential explanations, such as linguistic features, for challenges ELs may have with MWPs. However, one limitation of qualitative methodologies is that their labor-intensive nature limits the number of students that can participate in a given study, which makes findings difficult to generalize (Johnson & Onwuegbuzie, 2004).

One shared limitation of DIF, DBF, and qualitative methods is that causal claims cannot be made regarding which linguistic features of MWPs are particularly challenging for ELs. In contrast, causal claims can be made when experimental methods are used, such as when ELs are randomly assigned to receive items with different levels of linguistic complexity but the same mathematical operations (e.g., Tan, 1998). For example, Abedi and Lord (2001) modified items from a standardized assessment by changing multiple linguistic features (e.g., familiarity of non-mathematics vocabulary, verb tense, conditional and relative clauses, among others) and found that ELs performed better on items with simplified linguistic features compared to items that did not include the simplified features. In addition to permitting causal claims, the findings from experimental methods may be generalized to a broader population of students. However, experimental methods are not a good approach for exploring possible, previously-unidentified linguistic features that may relate to problem-solving performance for ELs. Therefore, a holistic approach to examining linguistic features that may be particularly challenging for ELs with multiple research design methods is ideal. In this systematic review, the findings from studies using a variety of design methods are examined and compared, to provide a comprehensive understanding of the topic. Two research questions guided our review.
1. What linguistic features of mathematics word problems are particularly challenging for ELs?
2. How does modifying the language of mathematics word problems influence the mathematics problem-solving performance of ELs?

This chapter differs from the other two chapters in this edited volume focused on the linguistic features of mathematics and science items that may interfere with ELs’ performance (Noble, Kachchaf, & Rosesbery, 2018) and the systematic review of empirical studies of second language acquisition in mathematics (Baker, Basaraba, Polanco, & Sparks, 2018). For example, the results of the systematic review by Noble and her colleagues are organized with respect to whether the studies (a) included in the review tested specific hypotheses about the linguistic features of items on ELs’ performance and (b) met specific methodological criteria. Additionally, linguistic features included in the Noble et al. chapter were grouped into three broad categories – word-level, sentence-level, and item-level features – whereas in this chapter we examine linguistic features in terms of the surface, textbase, and situation models (Kintsch, 1998). In addition, the majority of the studies included in the Baker et al. review focused on instruction or intervention-level efforts to support ELs’ mathematical understanding, not on the specific linguistic features of mathematics assessment items that may contribute to differences in student performance.

**METHOD**

A systematic review procedure was conducted for studies that examined the linguistic features of MWPs with ELs as participants (following Moher, Liberati, Tetzlaff, Altman and the PRISMA Group, 2009). The inclusion criteria for this systematic review were the following: (1) the measures used included MWPs, (2) the study included student performance empirical data (either qualitative or quantitative), (3) all of the participants were ELs or separate analyses were reported for ELs, and (4) specific information about the linguistic features - and analysis of those features regarding problem-solving performance– was included. See Figure 2 for a flowchart describing the literature search.

First, in September of 2017, searches for relevant literature were conducted using terms in the following databases: Scopus, ERIC (Educational Resources Information Center), ProQuest, and PsychINFO. The search terms used were “English,” “language,” “mathematic*,” and “problem*” (with * as a joker). These searches led to 2,547 records (Abstrackr was used for screening; Wallace, Small, Brodley, Lau, & Trikalinos, 2012). In addition, we conducted a hand-search of the following journals: The Bilingual Research Journal, Educational Assessment, International Journal of Testing, Applied Measurement in Education, International Journal of Bilingual Education and Bilingualism, Applied
In preparation for analysis, descriptive information for each of the 21 reports was identified: author, year of publication, participant grade level, number of participants, measures, linguistic features examined or manipulated, type of mathematics problems, and type of analyses conducted. In addition, these reports were coded with respect to whether the research questions were clearly stated, data analysis information was reported, and whether the research design was appropriate to address the research question(s).

Figure 2. Flow chart of study selection process.
RESULTS AND DISCUSSION

In the sections that follow, we synthesize the findings from the reports that met the aforementioned criteria for inclusion in this systematic review. To the extent possible, we group the findings from these reports by topic as it relates to each research question in an effort to emphasize the similarities and dissimilarities in the methodologies and results obtained. In addition, we summarize the results of each study included in our review in tabular format for ease of interpretation.

Features of MWPs That Are Challenging for ELs

Across the studies, a variety of linguistic features related to the surface structure (i.e., the exact words and phrasing; Kintsch, 1998) of MWPs were frequently observed as contributing to the mathematics word problem-solving experiences of ELs. These features included (a) vocabulary, (b) multiple meaning words (i.e., homophones and polysemous words), (c) the length of the problem, and (d) an aggregate of multiple linguistic features. We present a summary of these studies and their results in Table 1.

Vocabulary

Understanding the vocabulary is necessary to make interconnections among propositional ideas to construct a textbase and have the background knowledge to develop a situation model (Cain & Oakhill, 2014). In addition, knowing mathematics vocabulary is necessary for selecting the correct mathematical operational for creating the problem model (Ambrose & Molina, 2014; Verzosa & Mulligan, 2013). Lager (2006), for example, explored whether specific linguistic features of MWPs (including vocabulary) had negative effects on the mathematics word problem-solving performance of ELs in Grades 6 and 8 in two ways: (a) total count of correct responses and (b) think-alouds in which students were asked to describe what was confusing about the MWP, what specific phrases in the MWP meant, and how they arrived at their answer. Although the total correct scores did not lend insight into the specific vocabulary that may have contributed to incorrect responses, qualitative follow-up analyses of patterns in student responses shed light on misconceptions students held about specific mathematics vocabulary terms. For example, in response to an item asking students to identify a pattern and draw the next two iterations of that pattern, one-third of the students who responded incorrectly drew incorrect responses that did not hold to the pattern, indicating that they did not yet understand the concept of pattern. In the Martiniello (2008) study in Grade 4, students were asked: (a) whether they could understand the MWP in English, (b) to rephrase the
MWP in English or Spanish, (c) to identify any aspects of the English text they could not understand, and (d) to determine what the MWP was asking them to do (even if they could not fully understand the text). For example, one MWP problem described how Tamika must spin an even number on a spinner identical to the one pictured and to determine whether the probability of Tamika spinning an even number is certain, likely, unlikely, or impossible. Student think-aloud data revealed that some students’ lack of knowledge of the vocabulary words used in the distractors, such as certain, likely, and unlikely, impeded their ability to correctly solve the problem.

Other researchers have used larger samples of students and other analytic approaches to explore the effects of vocabulary on ELs’ mathematics word problem-solving performance (Loughran, 2014; Shaftel, Belten-Kocher, Glassnapp & Poggio, 2006; Wolf & Leon, 2009). Shaftel and colleagues (2006) for example, had approximately 8,000 students in Grades 4, 7, and 10 solve MWPs to determine whether linguistic features of MWPs affect MWP difficulty and if specific linguistic features, such as vocabulary, had the greatest impact on student performance. Regression analyses revealed that mathematics vocabulary had a statistically significant negative effect on item difficulty across all three grades. Interestingly, math vocabulary was the only statistically significant negative predictor that spanned the three grades; it was one of five significant negative predictors for Grade 4 students, one of two significant negative predictors for Grade 7 students, and the only significant predictor for Grade 10 students, confirming the importance of vocabulary to solve mathematical problems. Table 1 includes all the predictors that might affect student mathematical performance.

Wolf and Leon (2009) used DIF analyses to explore whether items grouped by common linguistic feature (e.g., vocabulary) functioned differently depending on students’ English language proficiency. In addition, they explored whether specific linguistic features were correlated with DIF for items that were characterized as “relatively easy” (i.e., answered correctly by non-ELs at least 75% of the time) or “not easy”. Results indicated that the total number of academic vocabulary words (general academic vocabulary words, context-specific academic vocabulary words, and technical vocabulary words) was significantly negatively correlated with performance on items that had been categorized as relatively easy across three comparative groups of students (EOs, ELs with high levels of English proficiency based on the state label of “advanced” or above, and ELs with low levels of English proficiency based on the state label of “intermediate” or below). Wolf and Leon also used DBF to group items that had similar proportions of language to non-language features and examined the effects of the proportion of language on the item-level performance of ELs and non-ELs. Results indicated that the negative effects of academic vocabulary were greater for ELs with low English proficiency than for ELs with high English proficiency.
Table 1. Studies examining linguistic features of mathematics word problems

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Purpose</th>
<th>Sample</th>
<th>EL Status Determination</th>
<th>Linguistic Features</th>
<th>Research Design</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banks, K., Jeddeeni, A., &amp; Walker, C. M. (2016)</td>
<td>Explore whether ELs and non-ELs with equal math proficiency differed in their chances of responding correctly to bundles of math word problems with specific item features</td>
<td>Participants: 23,285 (1063 EL, 22,222 non-EL); Grades 7-8</td>
<td>Not reported</td>
<td>Passive voice Conditional clauses Relative clauses Multilingualistic (multiple features)</td>
<td>Items were bundled into subtests that had similar linguistic features in item stems Matching subtests used to place ELs and non-ELs into equal proficiency groups Parametric (Differential Test Functioning) and non-parametric (SIBTEST) analyses used</td>
<td>No persistent differential bundle functioning (DBF) against ELs observed; effects of passive voice and conditional clauses were amplified against ELs (e.g., effects on individual items were negligible but significant when combined across items) Presence of DBF did not bias mean total correct scores</td>
</tr>
<tr>
<td>Beal, C. R., &amp; Galan, F. C. (2015)</td>
<td>Investigate whether (a) text complexity has an effect on the MWPS of ELs and non-ELs, and (b) ELs and non-ELs differ in terms of their goals, motivation, and effort during MWPS</td>
<td>Participants: 442 (209 ELs, 223 non-ELs); Grade 9</td>
<td>California English Language Development Test (CELDT)</td>
<td>Text readability (grammatical complexity, vocabulary frequency)</td>
<td>MWPs presented in context of Intelligent Tutoring System (ITS) ANOVAs used to explore differences between language groups and text types (easy-text, hard-text)</td>
<td>Compared to non-ELs, ELs had longer MWPS response times, made more incorrect answer attempts, used more hints per problem, and were less likely to answer correctly All students took longer on hard-text problems, made more incorrect answer attempts, and were less likely to answer correctly;</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Purpose</td>
<td>Sample</td>
<td>EL Status Determination</td>
<td>Linguistic Features</td>
<td>Research Design</td>
<td>Results</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>----------------------------------------------</td>
<td>---------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Celidon-Pattichis, S. (1999)</td>
<td>Explore how ELs use language to negotiate mathematical meaning while solving English and Spanish MWPs and learn more about what problem-solving strategies they use</td>
<td>Participants: 9; Grades 6-8&lt;br&gt;Items: 10 (5 English; 5 Spanish)</td>
<td>Language Assessment Battery</td>
<td>Language of word problems – Spanish (L1) or English (L2)</td>
<td>Think-a louds conducted with students, first in English and 3 weeks later in Spanish</td>
<td>Use of homophones created confusion for some (but not all) students&lt;br&gt;Most common problem-solving pattern observed was students reading the MWP at least twice and then translating key words into Spanish. Some students indicated that ignoring irrelevant words/information was helpful</td>
</tr>
<tr>
<td>Lager, C. A. (2006)</td>
<td>Explore language challenges when making meaning from MWPs focused on linear patterns</td>
<td>Participants: 221 (133 ELs; 88 non-ELs); Grades 6 &amp; 8&lt;br&gt;Items: 9 (could be solved in English or Spanish)</td>
<td>Not reported</td>
<td>Language of word problems – Spanish or English&lt;br&gt;Unfamiliar vocabulary</td>
<td>Think-a louds conducted with students in English, Spanish, or English and Spanish&lt;br&gt;Students asked to highlight unknown words/phrases</td>
<td>Language factors that contributed to ELs’ lower proficiency scores included: (a) lexical problems, (b) new technical vocabulary, (c) lack of awareness about confusing words/phrases, (d) polysemy, and (e) unknown language (and concepts)</td>
</tr>
<tr>
<td>Lee, M. K., &amp; Randall, J. (2011)</td>
<td>Investigate sources of differential item functioning (DIF) in MWPs between ELs and non-ELs</td>
<td>Participants: Over 64,000 (2,844 ELs, 1,314 former ELs, 60,000+ non-ELs)</td>
<td>Not reported</td>
<td>Lexical features (frequency &amp; impact of general academic vocabulary)&lt;br&gt;Grammatical features (passive voice, nominalizations, modals, conditional clauses, relative clauses, number of words &amp; sentences per problem, number of words per sentence)</td>
<td>Experts rated items for lexical &amp; grammatical complexity&lt;br&gt;Differential Item Functioning (DIF) analyses using matched sampling MANOVAs (using ratings of lexical/grammatical complexity)</td>
<td>11 items identified as showing potential DIF with respect to thresholds or factor loadings&lt;br&gt;Fewer items showed DIF between former ELs and non-ELs&lt;br&gt;Majority of items exhibiting DIF were from one strand of the mathematics assessment (Data Analysis, Statistics, and Probability)&lt;br&gt;Average ratings of lexical and grammatical complexity for items were low&lt;br&gt;Effects of lexical/grammatical complexity on MWPS varied for ELs and non-ELs</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Purpose</td>
<td>Sample</td>
<td>EL Status Determination</td>
<td>Linguistic Features</td>
<td>Research Design</td>
<td>Results</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>--------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Leon, R. E.</td>
<td>Explore the effects of the presence of MWPs on the performance of Hispanic/Latino students with learning disabilities while solving MWPs in English and Spanish</td>
<td>Participants: 41 Hispanic ELs with LD; 9-14 years old Items: 48 (24 English; 24 Spanish)</td>
<td>Language Assessment Scales</td>
<td>Presence of extraneous information Language of word problems – Spanish or English</td>
<td>Students listed to audio-recordings of MWPs during two separate sessions</td>
<td>Significantly higher mean total correct scores for MWPs with no extraneous information Effect of extraneous information on MWPS was similar for English and Spanish (for addition and subtraction items)</td>
</tr>
<tr>
<td>Loughran, J. M.</td>
<td>Investigate (a) the presence of DIF in MWPs favoring non-ELs over ELs (after controlling for SES), (b) whether factors of linguistic complexity predict DIF, and (c) whether schematic representations attenuate DIF</td>
<td>Participants: 59,775 (3,286 ELs; 56,489 non-ELs); Grades 4 &amp; 8; English, Spanish Items: 109 Grade 4; 119 Grade 8 (English, multiple-choice)</td>
<td>Not reported</td>
<td>Descriptive features (number of words/ sentences) Grammatical features (sentence type, prepositional phrases, dependent adjective clauses, modals, passive verbs) Lexical features (general academic vocabulary, polysemous words, cognates) Schematic representations</td>
<td>Logistic regression DIF used to determine predictive contribution of demographic variables DIF statistics regressed on linguistic complexity features and schematic representation variable</td>
<td>Magnitude of uniform DIF was low for Grades 4 &amp; 8, indicating that items favored non-ELs more than ELs Grade 4: Number of adjective clauses and polysemous words were significant predictors of DIF; items with more adjective clauses favored non-ELs and items with more polysemous words favored ELs Grade 8: Number of prepositional phrases, adjective clauses, polysemous words not related to math, and cognates were significantly correlated with uniform DIF; items with more of these features tended to have lower uniform DIF in favor of ELs Number of adjectives a was a significant predictor of DIF in favor of ELs</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Purpose</td>
<td>Sample</td>
<td>EL Status Determination</td>
<td>Linguistic Features</td>
<td>Research Design</td>
<td>Results</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Martiniello, M. (2008)</td>
<td>Examine linguistic features of MWPs that show DIF disfavoring ELs</td>
<td><strong>DIF Study:</strong> 68,839 (3,179 ELs); Grade 4; 39 English multiple-choice items</td>
<td>Not reported</td>
<td>Complexity in structural relationships among words, phrases in sentences in MWPs</td>
<td><strong>DIF Study:</strong> Standardization and Mantel-Haenszel analyses</td>
<td>DIF Study: 9 items identified as having slight-moderate DIF and 1 item identified as moderate-large DIF; 50% of items exhibiting DIF were from the Data Analysis, Probability, and Statistics strand of the state math assessment</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Think-Aloud Study:</em> 24 ELs; Grade 4</td>
<td></td>
<td>Vocabulary frequency and familiarity</td>
<td><strong>Think-Aloud Study:</strong> Decoding errors coded while students read problem aloud; students asked if they could understand the problem in English, to rephrase the problem in English or Spanish, to identify text they could not understand, and to figure out what the MWP was asking them to do</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Main effect of linguistic complexity on DIF was positive (controlling for presence/type of symbolic complexity)</strong> – items with greater linguistic complexity showed DIF favoring non-ELs</td>
<td></td>
</tr>
<tr>
<td>Martiniello, M. (2009)</td>
<td>Explore the impact of symbolic and visual representations in MWPs on the relation between linguistic complexity and DIF for ELs and non-ELs</td>
<td><strong>Participants:</strong> 68,839 (3,179 ELs); Grade 4; Items: 39 English multiple-choice</td>
<td>Not reported</td>
<td>Linguistic complexity composed of grammatical complexity (number of clauses, noun phrases, verbs, verb phrases) and lexical complexity (frequency of word use) Clauses coded for syntactic function and order</td>
<td>Experts rated items for grammatical and lexical complexity</td>
<td>Main effect of linguistic complexity on DIF was positive (controlling for presence/type of symbolic representations) – items with greater linguistic complexity showed DIF favoring non-ELs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>See Martiniello (2008) for description of think-aloud coding</strong></td>
<td></td>
</tr>
</tbody>
</table>
### Table 1. (Continued)

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Purpose</th>
<th>Sample</th>
<th>EL Status Determination</th>
<th>Linguistic Features</th>
<th>Research Design</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampson, S. O. (2005)</td>
<td>Explore whether (a) performance on an algebra assessment is confounded by linguistic complexity, (b) item difficulties vary between ELs and non-ELs, and (c) differences in item-level performance can be attributed to linguistic complexity of the items</td>
<td>Participants: 444 high-school students (51 ELs) Items: 31 English multiple-choice items</td>
<td>Language Assessment Scales (LAS)</td>
<td>Complexity of items determined by following features: item length; unusual, unfamiliar, or low-frequency words; ambiguous words; irregularly spelled words; compound and/or complex sentences; comparative structures; prepositional phrases; conditional phrases; complex noun phrases; passive voice</td>
<td>Rasch model to generate student-ability estimates DIF analyses to examine functioning of items</td>
<td>Items with large DIF contained the following features: multiclausal complex structures, long phrases, limited syntactic transparency, unfamiliar vocabulary, polysemous words, American cultural references, lack of 1:1 correspondence between syntactic boundaries of clauses and text layout</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Purpose</td>
<td>Sample</td>
<td>EL Status Determination</td>
<td>Linguistic Features</td>
<td>Research Design</td>
<td>Results</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>--------</td>
<td>-------------------------</td>
<td>--------------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>Shaftel et al. (2006)</td>
<td>Determine whether (a) linguistic features of MWPs affect their difficulty, (b) these linguistic features affect ELs and SWD disproportionately compared to general student sample, and (c) specific linguistic features have the greatest impact on student performance</td>
<td>8,000 (328 – 905 ELs per grade level); Grades 4, 7, 10</td>
<td>Not reported</td>
<td>Frequency counts of the following: words; words with &gt; 6 letters; sentences; prepositional phrases; relative pronouns; slang, idiomatic, or multiple-meaning words/phrases; homophones; uses of passive voice; clauses; complex verb forms; infinitive verb phrases; pronouns; unusual or difficult mathematics vocabulary; conditional constructions; comparative constructions; references to American holidays; references to American cultural events</td>
<td>Expert review of items for linguistic complexity ANOVA to evaluate whether there was a grade x student group interaction for overall item means Regression analyses to examine whether item linguistic characteristics predicted item difficulties (across grades and groups of students)</td>
<td>Linguistic features showed statistically negative significant effects of item difficulty for all students: Grade 4: prepositions, ambiguous words, complex verbs (3+ words), pronouns, math vocabulary Grade 7: comparative terms, math vocabulary Grade 10: math vocabulary Fourth graders were more influenced by test item language than seventh and tenth graders</td>
</tr>
<tr>
<td>Walkington et al., (2018)</td>
<td>Explore whether key readability factors are differentially associated with performance on MWPs for students from different demographic subgroups and whether those results change when no longer controlling for mathematics achievement</td>
<td>Participants: 744,180 (300,060 speaks language other than English); Grades 4 &amp; 8</td>
<td>Not reported</td>
<td>Word concreteness Pronoun density Presence of 2nd person pronouns Word count</td>
<td>Descriptive variables used to control for characteristics of individual problems in the mixed-effects logistic regression models included: problem type, problem difficulty, problem complexity, grade level, and content domain</td>
<td>Larger negative associations observed between word count and accuracy for students who spoke a language other than English at home</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Purpose</td>
<td>Sample</td>
<td>EL Status Determination</td>
<td>Linguistic Features</td>
<td>Research Design</td>
<td>Results</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>--------</td>
<td>-------------------------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td>Wolf, M. K., &amp; Leon, S. (2009)</td>
<td>Explore (a) language characteristics of state assessments and the extent to which language varies across states and grades; (b) the extent to which these items function differently for ELs and subgroups of ELs, (c) whether the DIF items are associated with the language demands of the items, and (d) what types of linguistic complexity are most associated with DIF for ELs</td>
<td>Participants: 18,019 (10,303 High ELs, 7,168 Low ELs); Grades 4, 5, 7, 8</td>
<td>Not reported</td>
<td>Frequencies of linguistic features: length, academic vocabulary, grammatical features, cohesion, and sentence type Holistic ratings of linguistic features: form of presentation, visual features, reliance on language</td>
<td>Trained raters used linguistic protocol to code linguistic features DIF analyses Differential Bundle Functioning (DBF) to detect whether groups of certain items function differently depending on students’ language proficiency Correlational analyses to examine relation between item difficulty, linguistic rating scores, and DIF statistics</td>
<td>0 – 8 items in the math assessments exhibited DIF for ELs (compared to non-ELs) but between 4 – 16 items exhibited DIF for Low ELs (compared to non-ELs) Correlations were low to moderate between the ratings of language features and signed uniform DIF by item difficulty; trends in correlations were evident for the “relatively easy” items, indicating significant uniform DIF against the focal group (ELs and EL subgroups) Magnitude of uniform DIF against ELs increased as the linguistic complexity of the item bundles increased (e.g., ELs would be predicted to score 8 percentage points lower per item than non-ELs on the most linguistically complex items)</td>
</tr>
</tbody>
</table>
Overall, these studies indicate that academic and technical vocabulary words tend to be negatively associated with the mathematics word problem-solving performance of ELs (Lager, 2006; Lee & Randall, 2011; Sampson, 2005). Moreover, the presence of difficult vocabulary words is also negatively associated with mathematics word problem-solving performance across both EOs and ELs (Shaftel et al., 2006; Walkington et al., 2015, 2018). In contrast, words that ELs are more likely to understand, such as Spanish-English cognates for L1 Spanish-speakers (e.g., impossible/imposible, combination/combinación), may support their problem-solving performance (Martiniello, 2008). However, results of these studies also demonstrate that the presence of cognates in MWPs does not guarantee that ELs will accurately understand the content, particularly in cases where MWPs use cognates that ELs are less likely to encounter in everyday language (e.g., igual instead of idéntico; Martiniello, 2008) or in the case of false cognates (e.g., pan, which refers to a metal container for cooking food in English but refers to a loaf of bread in Spanish). In general, the surface structure feature of difficult vocabulary words in MWPs could result in students (ELs or EOs) missing key idea(s) needed to form the textbase, situation model, and problem model required to successfully solve the problem (Kintsch, 1998; Nathan et al., 1992).

Multiple-Meaning Words

Homophones (i.e., words that have the same pronunciation but different meanings and polysemous words (i.e., words that have multiple meanings) comprised another category of word-level linguistic features examined as a potential source of difficulty for students when solving MWPs (Celedón-Pattichis, 1999; Lager, 2006; Loughran, 2014; Martiniello, 2008, 2009; Shaftel et al., 2006). Given the complexity associated with understanding words that have multiple meanings and the need to be able to use the context within which the words are presented to help derive their meaning, it might be hypothesized that multiple-meaning words are persistent contributors to the difficulty of MWPs. However, our review of studies indicates this is not necessarily the case. Of the six studies that included multiple-meaning words as a potential predictor of MWP difficulty, multiple-meaning words were negatively associated with performance on MWP in three studies (Lager, 2006; Martiniello, 2008, 2009), mixed effects were observed in two studies (Celedón-Pattichis, 1999; Loughran, 2014), and no effects were observed in one study (Shaftel et al., 2006). In the following paragraphs, the findings are organized based on the direction of effects and potential reasons for the inconsistencies are presented at the end.

Findings from three studies indicated negative effects associated with multiple-meaning words in MWPs. Lager (2006), in her think-aloud study with 221 students in Grades 6 and 8 (133 ELs), concluded that polysemous words were one linguistic factor that contributed to the significantly lower scores obtained by ELs when solving nine MWPs compared to EOs. In particular, examination of the words and phrases highlighted by students during the think-alouds emphasized the negative role these words and phrases
played during students’ problem-solving process. For example, asking students to record the *figure number* \( (N) \) in a table as one step in identifying a pattern was a challenging phrase for ELs because they did not understand (or only partially understood) the meaning of the words *figure* and *number*. Misunderstandings of the word *figure* in the context of the phrase *figure number*, which is a nominal phrase to describe the process of naming a figure (e.g., Figure 1), and not understanding whether *number* referred to nominality, ordinality, or cardinality is thought to have led three different misconceptions of what the problem was asking and resulted in incorrect solutions to the problem (Lager 2006). Martiniello (2008, 2009), also observed that polysemous words common to mathematics that students are likely to encounter, such as *one* and *off*, contributed to the challenges ELs experienced when trying to understand the MWPs, such as the word *one* in the sentence “To win a game, Tamika must spin an even number on a spinner identical to the *one* below.” It was not uncommon for students to interpret this *one* as the numeral one (more common usage of this polysemous word) than to interpret *one* as a pronoun that referred to the number pictured in the spinner.

On the other hand, Celedón-Pattichis (1999), and Loughran (2014) observed inconclusive effects of multiple-meaning words on students’ mathematics word problem-solving. Celedón-Pattichis (1999), for example, conducted think-alouds with nine Spanish-speaking ELs in Grades 6-8 in which students were asked to solve five English MWPs and to then participate in a think-aloud interview; students were asked to solve the same five MWPs that the researcher had translated into Spanish and to participate in a second think-aloud interview three weeks later. Although Celedón-Pattichis observed that the mixing of language functions associated with the word *can* within the text of a MWP created challenges for some students (i.e., *can* as a noun or as an auxiliary verb), she observed more frequently errors associated with students’ incorrect translations of homophones (when solving the English MWPs) as a factor that contributed to errors ELs made during the problem-solving process. When asked, most students indicated that translating the English MWPs into Spanish before working on the problem was one of their key strategies for solving the English MWPs.

However, incorrect translations by students created a faulty understanding of what the MWPs were asking. Examples of errors created during this process included translating *than* in a comparison problem (i.e., *How many more children came to the pool on Saturday than on Sunday?*) to *entonces* (then; i.e., *How many more children came to the pool on Saturday then on Sunday?*), or reading *many* as *money* (i.e., *The number of known asteroids is about 1,600. Astronomers believe that about 20 times that many exist. How many asteroids do astronomers think exist?*). In these examples, incorrect translations of key words in the MWPs made by students while reading and solving the problem resulted in false homophones and made it challenging – if not impossible – for them to develop an accurate situation model for the problem.
For example, the use of the false homophone *then* and omission of the word *than*, which is critical for understanding the comparison problem schema in the question “*How many more children came to the pool on Saturday than Sunday?*”, made accurate understanding of what the problem was asking for virtually impossible. Not all students made these errors, however, despite the fact that all participating students commented that translating key words in the MWP into their native language (Spanish) was a critical part of their problem-solving process. The observed lack of consistency in the translation errors, combined with the fact that students performed similarly on the problems in their L1 of Spanish or L2 of English, led Celedón-Pattichis to conclude that language did not have a significant effect on ELs’ mathematics word problem-solving performance. However, the findings should be interpreted with caution given that there were only nine students in this study.

Loughran (2014) also observed mixed effects for multiple-meaning words, although in this study the differential effects appeared to be more systematic as they varied by grade level. In this study, polysemy was included as one linguistic complexity feature in the prediction of DIF on MWPs on a state assessment solved by almost 60,000 students (3,286 ELs) in Grades 4 and 8. Descriptive analyses of item-level features revealed that, on average, Grade 4 items contained more multiple-meaning words than Grade 8 items, even though Grade 8 items were, on average, more linguistically complex. Results of multiple regression analyses revealed that multiple-meaning words were a significant predictor of DIF favoring ELs compared to EOs in Grade 4 indicating that multiple-meanings words were actually better solved by ELs than EOs ($\beta = -0.24, p < .01$), but were not a predictor of DIF in Grade 8. However, Loughran excluded polysemous words related to mathematics from the category of multiple-meaning words. This is noteworthy because of previous findings that polysemous mathematics vocabulary (e.g., *place*, *face*) is a source of difficulty for ELs (Celedón-Pattichis, 1999; Lager, 2006; Martiniello, 2008, 2009). For example, a student who uses the meaning of *face* in terms of a body part rather than mathematically would interpret an unintended proposition that would lead to the formation of an inaccurate textbase, situational model, and problem model. Moreover, because polysemy was hand-coded and inter-rater reliability was not reported (Loughran, 2014), the results should be interpreted with caution.

Shaftel et al., (2006) examined the performance of approximately 8,000 students in Grades 4, 7, and 10 (approximately 200 items per grade), using multiple regression to determine the relative contribution of specific linguistic item-level features to item difficulties for general students, ELs, and students with disabilities (SWD). Results indicated that while homophones were not a significant predictor of item difficulty in any of the three grades examined, there was a statistically significant difference in the effect of ambiguous words on item difficulty for Grade 4 only. However, according to the coding scheme used for the linguistic features, this category of words included not only multiple-
meaning words but also slang, idiomatic, or ambiguous words; consequently it was unclear what were the specific features of these words that affected MWP solving by ELs.

Collectively these results indicate that the effects of multiple-meaning words on MWP performance is – at best – unclear. One reason for the inconsistent findings could be variations in L2 language proficiency. The appropriate meaning of a polysemous word is discerned from the context (Williams, 1992). Some ELs may have not understood the other words well enough in the MWPs to understand the context to discern the meaning of the polysemous words and connect propositions into an accurate textbase representation (Kintsch, 1998). Unfortunately, the studies reviewed did not report language proficiency in English, which limits interpretations of their findings.

**Problem Length**

Generally speaking, the length of a MWP was negatively associated with ELs’ performance on MWPs (Lee & Randall, 2011; Leon, 1992; Martiniello, 2008, 2009; Walkington et al., 2018; Wolf & Leon, 2009). This finding was consistent across the multiple metrics used to quantify length including (a) number of words (Lee & Randall, 2011; Walkington et al., 2018 Wolf & Leon), (b) number of sentences (Lee & Randall, 2011; Wolf & Leon, 2009), and (c) phrase length (Martiniello, 2008, 2009). For example, Lee and Randall found that as the number of words in a MWP increased, performance on the MWP decreased for ELs compared to EOs. On the other hand, Shaftel et al. (2006) did not find that the number of words or sentences was related to the problem-performance of ELs or EOs in Grades 4, 7, or 10. One possible explanation for these inconsistent findings could be due to whether the problem is longer due to helpful or irrelevant information. This is discussed further in the section on problem length modifications.

**Aggregate Measures of Linguistic Complexity**

Other studies combined multiple linguistic features as a single composite metric to examine the general role of linguistic complexity (Lee & Randall, 2011; Martiniello, 2009) or readability (Beal & Galan, 2015) in students’ mathematics word problem-solving performance. For example, Martiniello (2009) created a composite score of linguistic complexity ratings from three scores: (a) micro-analytic ratings of linguistic complexity based on a coding scheme developed to identify grammatical elements of complexity (e.g., number of clauses, noun phrases, verbs, and verb phrases), syntactic functions, and order of clauses; (b) ratings of items for linguistic and lexical complexity by experts. Analyses of 39-items in a mathematics assessment revealed that, after controlling for the presence and type of symbolic representations in the items, linguistic complexity was a significant predictor of DIF. Items that had higher levels of linguistic complexity favored EOs over ELs, while items with lower levels of linguistic complexity favored ELs over EOs.
Table 2. Studies examining the effectiveness of linguistic modifications to mathematics word problems

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Purpose</th>
<th>Sample</th>
<th>EL Status Determination</th>
<th>Linguistic Features</th>
<th>Research Design</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abedi, J., &amp; Lord, C. (2001)</td>
<td>Compare the performance of ELs and non-ELs on MWPs and investigate whether modifying the linguistic structure of items affected student performance</td>
<td>Student Perception Study: 36 (native speakers or English, Spanish, Cambodian, Vietnamese); Grade 8</td>
<td>Not reported</td>
<td>Unfamiliar/infrequent words revised Passive verbs changed to active verbs Conditional clauses replaced with separate sentences or order of conditional and main clause was changed Relative clauses removed or recast Complex questions changed to simple questions Abstract/impersonal presentations made more concrete</td>
<td>Student Perception Study: Students read original &amp; modified items, asked to indicate which they would prefer to solve, to identify words that might be confusing, and to explain why the problem they chose seemed easier Accuracy Study: 2 booklets of 10 original, 10 linguistically modified, and 5 low-language items were created; 2-factor ANOVA used with booklet and EL classification as IVs</td>
<td>Students preferred the revised, linguistically modified items over the original items because they were easier to read and the vocabulary was more familiar Small, statistically significant higher mean score observed on the linguistically modified items compared to the original items Overall, ELs demonstrated greater rates of improvement on the modified problems compared to EOs; percentage of improvement varied by SES status and math course enrollment</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Purpose</td>
<td>Sample</td>
<td>EL Status Determination</td>
<td>Linguistic Features</td>
<td>Research Design</td>
<td>Results</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Barbu &amp; Beal, 2010</td>
<td>Compare the performance of ELs on MWPs constructed to systematically vary with respect to linguistic complexity and mathematics difficulty</td>
<td>Participants: 41 ELs; Grades 6-7 Items: 8 (2 easy math/easy English, 2 easy math/hard English; 2 hard math/easy English; 2 hard math/hard English)</td>
<td>Not reported</td>
<td>Mathematics difficulty: Varied operations (easy = addition/subtraction; hard = multi-digit multiplication/division) Text Difficulty: Changed vocabulary and grammatical structure of MWP while holding word count constant</td>
<td>Interviews conducted with each student Interviewer (a) read MWP aloud, (b) asked student to identify operation needed to solve (providing hints as needed), (c) asked student to rate difficulty of words, (d) asked student to solve the problem, and (e) asked student to identify most difficult problem Two-way repeated measures ANOVAs for problem-solving and student ratings of difficulty</td>
<td>Problem Solving: Main effects of text difficulty observed on (a) students’ identification of the correct operation required to solve, (b) number of hints provided by interviewer, and (c) number of computational errors made Significant interaction between mathematics difficulty and text difficulty on number of computational errors made Student Ratings: Main effects of math difficulty and text difficulty on student ratings Significant interaction between math difficulty and text difficulty (items with challenging math operations and complex language rated as most difficult)</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Purpose</td>
<td>Sample</td>
<td>EL Status Determination</td>
<td>Linguistic Features</td>
<td>Research Design</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Bernardo, A. B. (1999)</td>
<td>Investigate the effects of two external variables (rewording of MWP and language) and two internal variables (grade level and academic achievement) on student performance on MWPs</td>
<td>Participants: 283 (L1 – Filipino, L2 – English); Grades 2-4</td>
<td>Not reported</td>
<td>Reworking problems to make known and unknown quantities more explicit</td>
<td>ANOVA with language and wording of problems as within-group variables and grade level and academic achievement as between-group variables</td>
<td>Students performed significantly higher on the reworded problems and on problems that were in their L1 (despite receiving mathematics instruction in L2) Benefit of rewording was greater when problems were written in L1 compared to L2</td>
</tr>
<tr>
<td>Hofstetter, C. H. (2003)</td>
<td>Explore what school- and classroom-level factors are most related to ELs’ performance on MWPs and whether these factors influence student performance when test accommodations are applied</td>
<td>Participants: 849 (676 ELs; 173 non-ELs); Grade 8</td>
<td>Student language background questionnaire</td>
<td>Modified English accommodation: Unfamiliar/infrequent words revised to more familiar words, passive verbs changed to active verbs, shortened long nominal, complex questions revised to be more simple Spanish accommodation: Direct translation of English items</td>
<td>Multi-level modeling regression analyses</td>
<td>On average, ELs who received mathematics instruction in their native language (Spanish) scored lower on all 3 assessment forms than those who received instruction in English Of ELs who received math instruction in Spanish, students who completed the Spanish translated items had higher scores than those who completed the English items ELs who received mathematics instruction in English scored lower on the Spanish-translated items than on the English items</td>
</tr>
</tbody>
</table>
Table 2. (Continued)

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Purpose</th>
<th>Sample</th>
<th>EL Status Determination</th>
<th>Linguistic Features</th>
<th>Research Design</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson, E., &amp; Monroe, B. (2004)</td>
<td>Examine the effects of a simplified language accommodation on MWPs for ELs and SWD and whether simplifying language effects the psychometric characteristics of the assessment</td>
<td>Participants: 1,232 (34 EL); Grade 7 Items: 20 English items (16 multiple-choice, 4 open-ended)</td>
<td>Not reported</td>
<td>Guidelines for simplified language by Kopriva (2000) were followed (e.g., items shortened, revised for active voice, used high-frequency words) and reviewed by assessment experts</td>
<td>Teachers asked to randomly distribute assessment forms to students and identify on the test if the student was an EL, SWD, or receiving general education services</td>
<td>No significant effects observed for simplified language accommodation For ELs, only 6/20 of the simplified language items were easier</td>
</tr>
<tr>
<td>Kiplinger et al. (2000)</td>
<td>Examine the effectiveness of language accommodations on ELs and students with special needs</td>
<td>Participants: 1,198 (311ELs); Grade 4 Items: 24 English items (16 multiple-choice, 8 constructed response)</td>
<td>Student language background questionnaire; Language Assessment Scale (LAS)</td>
<td>‘Simplified’ version: MWPs simplified by making changes to linguistic structures and non-mathematical vocabulary ‘Glossary’ version included definitions of non-mathematics vocabulary for words thought to be unnecessarily difficult</td>
<td>ANOVAs used to evaluate the effect of linguistic modification and English proficiency on MWPs</td>
<td>ELs in the second quintile on the LAS performed significantly better on the ‘Glossary’ version than the Original version ELs in the fourth quintile performed significantly better on the Simplified version than the Glossary or Original versions</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Purpose</td>
<td>Sample</td>
<td>EL Status Determination</td>
<td>Linguistic Features</td>
<td>Research Design</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Sato et al. (2010) | Determine whether (a) the effects of linguistic modifications to MWPs vary across subgroups of students (ELs, non-English Language Arts proficient ELs (NEPs), and English Language Arts proficient ELs (EPs)) and (b) there is evidence of DIF across the three subgroups for students with comparable mathematics proficiency | Participants: 4,617 (1,214 ELs); Grades 7-8  
Items: 50 English multiple-choice | Student language background survey; California English Language Development Test (CELDT)  
Modified verb tense (e.g., past/future tense to present tense; passive to active)  
Graphics/text added for clarity | Revise/remove unfamiliar/infrequent English words  
Simplified complex grammatical structure  
ANOVAs using total correct scores and IRT estimates of mathematics ability  
DIF analyses to detect subgroup differences on items not attributable to math content knowledge | Linguistically modified items were (1) reviewed by experts, (2) used in cognitive labs, and (3) pilot tested prior to large-scale study | Significant differences in 1-PL ability estimates for ELs, NEPs, and EPs; post-hoc analyses revealed the effect of linguistic modification was significantly different for EL and EP students (favoring ELs)  
Mean differences in performance on two item sets across four scoring approaches (raw scores, 1-PL, 2-PL, or 3-PL) were greatest for ELs  
No items demonstrated moderate to high DIF when comparing NEPs and EPs, but one item in the original set and two items in the linguistically modified set exhibited moderate to high DIF when comparing ELs and EPs |
| Tan, J. (1998)      | Investigate the effects of native language and rephrasing on MWPs and attitudes toward mathematical tasks | 104 (52 ELs; 52 non-ELs); Grade 6 | Not reported  
High frequency vocabulary (including mathematics vocabulary), Simplified sentence structure | Students randomly assigned to original/rephrased MWPs  
ANOVAs using total correct scores | ANOVAs using total correct scores  
Main effect of rephrasing MWPs was significant for non-ELs, but not for ELs  
Interaction between rephrasing and language status (EL/non-EL) was not significant |
Beal and Galan (2015) created an overall readability index that included grammatical complexity and vocabulary frequency. Items were then categorized into two groups: items that required reading skills higher than a Grade 7 reading level and lower than a Grade 7 reading level. Participants were 442 Grade 9 students (209 ELs) who were asked to solve MWPs with the support of an intelligent tutoring system. Results indicated that ELs had longer response times when solving MWPs, used more hints, made more incorrect response attempts, and were less likely to answer correctly than EOs. Moreover, EOs were more successful in solving MWPS when the readability level was above Grade 7 compared to ELs.

Other features of linguistic complexity such as the use passive voice and conditional clauses in MWPs has also been negatively associated with problem-solving performance (Banks et al., 2016). A plausible explanation is that second language readers often assume that the first noun of a sentence is the subject of that sentence (VanPatten, 2004, 2007). In conditional clauses, for example, sentences typically begin with an “If” statement that states an expectation that an event will or will not occur, conditional on the behavior described in the first part of the clause (e.g., If you study for a test, then you will get a good grade). Thus, the syntax of the word can lead to misunderstanding the intended meaning of the sentence (Abedi, 2011) making it more difficult for students to construct an accurate textbase (Kintsch, 1998).

**Modifications to the Language of MWPs**

Results from the nine studies that addressed our second research question, although not conclusive, suggest that modifying the language of an MWP by making alternations to linguistic features such as (but not limited to) verb tense, voice, and complex clause structures, (e.g., conditional and relative clauses) can improve the mathematics performance of ELs. In the sections that follow, we synthesize first the studies that suggested modifying language can have positive effects on ELs’ mathematics word problem-solving performance (highlighting the various methods for modifying language), followed by summaries of studies that suggest language modifications may not be beneficial for all ELs. We then describe findings from studies in which researchers investigated the effectiveness of modifying the problem length in MWPs for supporting ELs’ mathematics performance. We also present a summary of these studies in Table 2.

**Positive Effects of Modifying Language**

In each of these studies, the surface structure of words and syntax were modified. In three studies, researchers opted to replace low-frequency words with high-frequency words with which ELs are more likely to be familiar (Barbu & Beal, 2010; Hoffstetter, 2003; Sato, Rabinowitz, Gallagher, & Huang 2010). Sato et al. (2010) revised or replaced
unfamiliar or infrequent English words (including unfamiliar sociocultural references and/or idioms) with more common English words, simplified complex grammatical structures, and modified verb tense. Findings revealed beneficial effects for ELs and for non-English Arts proficient EOS (NEPs), but not for English Language Arts proficient EOs (EPs), suggesting that the linguistic modifications were helpful for students who lack English language proficiency regardless of whether English was their first language. In other studies that simplified syntactic complexity such as changing verbs from the passive voice to the active voice (e.g., Abedi & Lord, 2001; Hofstetter, 2003; Sato et al., 2010) or removing conditional clauses, findings indicated that ELs demonstrated a slightly greater percentage of improvement from the original to linguistically modified items than EOs (See Table 2 for details).

Barbu and Beal (2010) examined the effects of modifying the vocabulary and grammatical structure of MWPs while holding the number of words constant. Students were asked not only to solve the MWPs, but were also asked to rate the perceived mathematics and text difficulty of each MWP to see if they could distinguish the sources of difficulty in MWPs. Results indicated statistically significant main effects associated with text difficulty on the correct identification of the operation required to successfully solve MWPs ($\eta^2 = 0.018$). In addition, students were more likely to identify the correct operation in problems that had easier text. A significant interaction was observed between mathematics and text difficulty in the number of computational errors made ($\eta^2 = 0.108$), indicating that students made more computational errors while trying to solve linguistically complex MWPs that required more complex mathematical operations (e.g., multiplication and division), compared to less complex problems. These findings were corroborated by ELs’ perceptions that the linguistic complexity of a MWP affected their ability to solve the problem.

Mixed Effects of Modifying Language

Two studies indicated that the modifications benefited only some (but not all) ELs (Abedi & Lord, 2001; Kiplinger, Haug, & Abedi, 2000), while two other studies observed no effects on ELs’ mathematics word problem-solving performance as a result of modifying the language of MWPs (Johnson & Monroe, 2004; Tan, 1998). Abedi and Lord (2001), for example, observed that while almost all Grade 8 students (except those in Honors Algebra) exhibited an improvement in the percentage of items correct when responding to the linguistically modified items (compared to the original, non-modified items), the percentage of improvement observed varied widely depending on the math course in which students were enrolled. In particular, students enrolled in lower-level math classes (e.g., “low-level” math and “average-level” math) demonstrated higher rates of improvement (6.7% and 6.6%, respectively) than students enrolled in higher-level math classes (e.g., “high-level” math and Algebra; 0.4% and 0.7%, respectively). Students enrolled in ESL math classes (including bilingual and sheltered English classrooms) only
demonstrated 0.9% improvement when responding to the linguistically modified items. This finding is interesting because it might be expected that students who were receiving additional language supports (e.g., native-language supports and/or the use of specific language scaffolds to increase the accessibility of mathematics content) in these math classes might benefit more from linguistically-modified items. Differential effectiveness of linguistic modifications was also observed by Kiplinger et al. (2000), who observed that only ELs with higher-levels of English proficiency benefited from linguistic modifications. The authors suggest this was likely due to floor effects (i.e., items were so difficult that only ELs with higher levels of English proficiency could understand them well enough to benefit from the linguistic modifications). A floor effect was also the likely cause of the null results observed by Tan (1998). In this study (Tan, 1998), the interaction effect between rephrasing MWPs and language status (EL/EO) was not significant. However, given that ELs only answered 30% of the items correctly, it is difficult to determine whether linguistic modifications helped students on some items and not others, or whether it was irrelevant.

Johnson and Monroe (2004) postulated that there may be two reasons for the observed lack of positive effects for linguistic modifications. One is that the sample of ELs included in linguistic modification studies is often small, which could lead to a lack of power to detect an effect. A second reason is that the modifications provided may be insufficient or may inadvertently make the MWPs more difficult. For example, researchers may have opted to retain key vocabulary terms that ELs may not be familiar with, such as the concise phrase figure number (N), which was noted as a factor that was noted as interfering with ELs’ comprehension of MWPs in other studies (Lager, 2006; Martiniello, 2008).

Effects Associated with Modifying Problem Length

Modifying the problem length in MWPs has been another approach for examining the effects of linguistic modifications on ELs. For example, Leon (1992) investigated the contribution of extraneous information to the MWP performance of 41 ELs with learning disabilities on MWPs that varied on three experimental conditions: (a) language of the problem (i.e., English or Spanish); (b) operation (i.e., addition or subtraction); and (c) extraneous information (i.e., presence or absence). Results of within-subject t-tests revealed statistically significant differences between items with extraneous information and without extraneous information favoring the latter on addition and subtraction problems (p < .05). Other studies that increased the problem length to add helpful contextual information intended to improve clarity, ELs in Grades 2-4 performed better on the reworded (more explicit, but longer) MWPs than the original MWPs, F (1, 277) = 95.27, p < .001 (Bernardo, 1999). For example, a MWP problem originally written as, “Rico and Pat have 9 candies altogether. Rico has 3 candies. How many candies does Pat have?” was revised to read, “Rico and Pat have 9 candies altogether. 3 of these candies belong to Rico. The rest belong to Pat. How many candies does Pat have?” However,
results from length modifications by adding clarity (Bernardo, 1999; Leon, 1992; Sato, 2010), a glossary (Kiplinger et al., 2000), or shortening clauses (Hofstetter, 2003), did not come to conclusive findings about whether these modifications benefit ELs because the studies involved multiple modifications to the MWPs in addition to length such as simplified wording.

Thus, simply decreasing the problem length of MWPs for ELs is not a clear-cut method for improving performance. Additional language may be useful for ELs if it adds clarity (as in Bernardo, 1999; Sato et al., 2010) that can help with connecting propositions together into a coherent textbase (Kintsch, 1998; Nathan et al., 1992). For example, dividing long, complicated sentences into multiple, shorter simple sentences may be particularly beneficial for comprehension (Abedi 2011). In contrast, extraneous information may make it more difficult to use background knowledge appropriately to create the situation model and apply problem-solving strategies (Cook, 2006).

**LIMITATIONS AND FUTURE DIRECTIONS FOR RESEARCH**

One area needing future examination is a more precise understanding of which specific linguistic features are challenging to ELs. Most of the studies on linguistic modifications to MWPs involved changing multiple linguistic features in each MWP (Abedi et al., 2001, Barbu & Beal, 2010; Hofstetter, 2003; Johnson & Monroe, 2004; Sato et al., 2010; Tan, 1998). To better understand how to design MWPs, it would be helpful to test out modifications of specific linguistic features while holding other features constant. Testing modifications of specific linguistic features in isolation would help refine modification approaches. In addition, testing modifications based on specific linguistic features would illuminate whether the linguistic feature in isolation creates challenges or if it is how linguistic features may build on each other that create challenges for students. Challenging linguistic features often co-occur; for example, more difficult words are often found in longer sentences (Abedi, 2011). Testing word difficulty and sentence length modifications separately would inform which linguistic feature is more challenging for ELs and how to best modify MWPs (i.e., should the focus of modification be reducing sentence length or reducing word difficulty?).

Many of the research studies in this review are also limited in terms of the background information provided about their EL participants. ELs are a diverse group with notable variation not only in their English language proficiency (Hwang, Lawrence, Mo & Snow, 2015; Master, Loeb, Whitney, & Wyckoff, 2016), but also in their native language proficiency, which is influenced by multiple contextual factors (e.g., parental education, socioeconomic status, country of origin, etc.; Dürgunoglu & Goldenberg, 2011). Most of the studies did not include a measure of English language or reading proficiency (for exceptions see Hofstetter, 2003; Kiplinger et al., 2000), despite previous work indicating...
moderate to strong relationships between language and/or reading proficiency and mathematics word problem-solving performance (Adelson, Dickinson, & Cunningham, 2015; Henry, Nistor, & Baltes, 2014; Purpura & Ganley, 2014; Purpura & Reid, 2016; Vukovic & Lesaux, 2013). Including English language and reading proficiency measures in future studies would provide a more nuanced view of linguistic features that contribute to the challenges ELs face when solving MWPs.

Additionally, none of the reviewed studies examined how computational skills in mathematics measured through symbolic items without words (e.g., \(2 + 2 = \square\)) interacted with linguistic features and problem-solving performance. ELs with strong computational skills may more easily develop a problem model of the MWP than their peers with weaker computational skills; therefore, word problem solving may be less dependent on linguistic features for ELs with strong computational skills.

Finally, only four studies have empirically explored the effectiveness of providing ELs with a mathematics test that has been translated into their native language (Alt, Arizmendi, Beal, & Hurtado, 2013; Hoffstetter, 2003; Kempert, Saalbach, & Hardy, 2011; Robinson, 2010). Results from these studies have indicated that (a) ELs demonstrate higher rates of accuracy when solving MWPs translated into their native language (Alt et al., 2013; Robinson, 2010), (b) level of native-language dominance may influence student performance on some (but not all) mathematics assessments (Alt et al., 2013), and (c) that the effectiveness of translating items may be moderated by the language of mathematics instruction for older students (Hoffstetter, 2003). The effectiveness of providing ELs with translated items, however, is an under-explored option, perhaps because accommodations that provide students with linguistic supports (such as test translations or test adaptations) are most frequently prohibited by states (Rivera, Stansfield, Scialdone, & Sharkey, 2000). Given the increasing numbers of ELs in schools, however, and promising findings from studies of cross-linguistic transfer (in which students’ L1 supports the development of L2; Baker, Park, & Baker, 2012; Baker, Basaraba, & Polanco, 2016; Melby-Lervåg & Lervåg, 2011; Nakamoto, Lindsey, & Manis, 2012), further exploration of native-language translated items as one method to obtain more detailed information about ELs’ mathematics word problem-solving skills is warranted.

**Implications for Practice**

Teachers, test developers, and curriculum developers may use the findings from this systematic review to guide writing mathematics items for ELs. Across the different studies, confusing and/or difficult vocabulary words appeared to interfere with word problem-solving performance (e.g., Lager, 2006; Lee et al., 2011; Martiniello, 2008; Shaftel et al., 2006). When writing items, the use of familiar vocabulary (i.e., high frequency words) may be helpful for ELs because they are more likely to know the meanings of the words in the
problem. Being mindful of words with multiple meanings is also important; however, simply avoiding these words may not be best (Loughran, 2014). High frequency words are often polysemous words (Crossley et al., 2010) and the use of polysemous words in a manner distinctive to mathematics is often necessary in word problems (Pimm, 1994). Given these issues, clarifying which meaning of a word is intended in the word problem may help scaffold comprehension for EL students, especially if the intended meaning of a polysemous word is a less common definition.

Test developers should also be conscientious about the length in MWPs. Language that is irrelevant likely adds unnecessary information for ELs to process, which could diminish performance (e.g., Leon, 1992). In contrast, language that adds clarity or simplifies linguistic complexity (e.g., modifying an information dense sentence into two simpler sentences) would likely assist ELs in developing the textbase, situation model, and/or problem model (Nathan et al., 1992), improving performance (Ambrose & Molina, 2014).

We recognize that teachers are not always able to modify the language of MWPs. However, knowing which linguistic features may interfere with problem-solving performance can help teachers identify which MWPs ELs may need extra scaffolding. One approach that may be effective for assisting ELs with word problem solving performance is the use of schematic instruction (SI) that involves explicit strategies to comprehend linguistically complex problems (Driver & Powell, 2017). In SI, students learn how to identify and categorize known and unknown information in a MWP, which helps students write the correct equation to solve the problem (Driver & Powell, 2017). SI also involves schematic diagrams to help students visualize the information in the MWP (Jitendra, Harwell, Dupuis, & Karl, 2017) and may help ELs better understand MWPs (Loughran, 2014; Martiniello, 2009).

**CONCLUSION**

In this systematic review, we found that certain linguistic features, such as word difficulty and the amount of complex language, appeared to make mathematics word problem-solving more difficult for ELs. Linguistic modifications may be beneficial, but better understanding of the role of specific linguistic features is needed. The findings from this review may be helpful for test and curriculum developers to guide item writing and for teachers to identify features of MWPs that may be particularly challenging for ELs. This chapter also highlights the complexities of solving MWPs. Future research that provides more specific guidance on how to improve the understanding of MWP for English learners is warranted.

**ACKNOWLEDGMENTS**
We thank Chia Lin Chang and Kristina Syverson for their assistance screening abstracts. We also thank Chia Lin for her assistance with formatting the references.

REFERENCES

*Indicates study included in the systematic review.


