The Mathematics Educator
2024 Vol. 32, No. 1

Semantic Structure of Word Problems: A Content Analysis

Julia E. Calabrese and Jamaal Young

Solving word problems is a standard part of the mathematics classroom. However, many students struggle to solve word problems due to the variety of problem types. One word problem can require a different solution approach than another with few semantic changes. The Common Core State Standards Initiative released a taxonomy for classifying different types of word problems. The present study is a content analysis of three widely distributed mathematics textbooks (i.e., one from three of the most popular textbook publishers). The researchers analyzed the textbooks to classify the types of word problems they contained according to the taxonomy. Results indicate that word problems that were identified as difficult for students to master in previous research now appear less frequently in mathematics textbooks. The findings also show that variation in semantic structure is not proportionate within a single textbook; however, category frequency is similar across textbooks.

Word problems enable us to see mathematics as a meaningful subject rather than an interactive practice. Students learning how to solve word problems are faced with the task of translating between concrete and abstract information (Gagne, 1983). When students are faced with the application side of mathematics, they experience the content in its natural form, highlighting the material's relevance in their own lives (National Council of Teachers of Mathematics, 2000). However, students of all ages, abilities, and grade levels struggle to correctly solve word problems (Pongsakdi et al.,

Julia E. Calabrese *is a Postdoctoral Research Associate at the University of Texas at Arlington. Her research interests include mathematical problem posing, preservice teacher education, and word problem structure.*

Jamaal Young *is an associate professor of mathematics and STEM education at Texas A&M University. His primary research focuses on using systematic reviews and Meta-analysis to examine the effects of opportunity structures related to instruction (e.g., teacher quality, access to technology, or out-of-school time activities) on STEM learning outcomes.*

2020). This trend could be explained by two factors. First, solving word problems is more challenging than the algorithmic aspect of problem solving alone (Almolhodaei, 2002). In other words, the ability to solve word problems requires more than procedural knowledge. Students must not only have the mathematical skills necessary to solve the problem but also the ability to decode the context within the word problem itself to see the mathematics that lies underneath (Fuchs et al., 2015; Sepeng & Madzorera, 2014). Second, word problems come in many forms with variations in semantics (Daroczy et al., 2015). Prior research has indicated that student word problem-solving challenges vary across the problem types, with students tending to struggle more with certain categories of problems compared to others (Arsenault & Powell, 2022; Koedinger & Nathan, 2004; NRC, 2009). Thus, it is important for teachers to include a variety of word problems as a regular part of their mathematics curriculum.

To address the notable challenges in student word problem proficiency it is important to expose students to a variety of word problems early and often. The structure and delivery of mathematics instruction is in a constant state of flux as the field continues to evolve to meet the needs of the next generation of learners. Although many schools have opted for digital resources, textbooks remain the cornerstone of K-12 instruction. This is evidenced by the large expenditures made on K-12 textbooks in the U.S, which continues to grow (Curcic, 2023). Hence, the content and organization of many classroom resources utilized in K-12 settings are often derived from popular textbooks in circulation.

Mathematics learners often encounter word problems, which are among the most difficult problems to solve (Verschaffel et al., 2020). Yet, word problems play a central role in mathematics education, serving as a bridge between abstract mathematical concepts and real-world applications. The centrality of word problems in mathematics education goes beyond just teaching mathematical concepts. Word problems promote critical thinking, real-world application, and the development of skills that are crucial for success in various aspects of life (Verschaffel et al., 2014). Thus, mathematics education continues to recognize and support the incorporation of word problems into the curriculum to contribute to a more effective and relevant mathematics education. This is more notably seen in the inclusion of word problems as the cornerstone of most mathematics textbooks currently in circulation.

Textbooks are artifacts that represent the curriculum implemented within the classroom (Valverde et al., 2002). Furthermore, they often dictate student exposure to content (Cai & Jiang, 2017; Yang & Sianturi, 2022) and have been correlated with student performance (Törnroos, 2005). More specifically, Yang and Sianturi (2022) found that different textbooks exhibited different types of problems and connected this variation to variation in mathematics performance. Therefore, one determinant of the variety of word problems presented in the mathematics classroom is the variety of word problems present in mathematics textbooks. The variety of content within mathematics textbooks essentially represents the different content that students are exposed to in the mathematics classroom. Textbooks are designed to reflect the priorities and philosophies of the textbook authors and contributors. Thus, understanding how different textbook authors prioritize specific types of word problems is essential to unpacking how textbook content may contribute to the consistent and notable word problem struggles facing students in the U.S. and around the world in the mathematics classroom (Xin, 2007). Thus, the purpose of this study is to examine three widely distributed textbooks from top textbook publishers to determine the variation of the problem types in elementary mathematics textbooks.

The results of this content analysis have important implications for teaching, learning, and success in mathematics because prior research has identified the most problematic problem types (e.g., NRC, 2009), but the distribution of these problem types across textbooks remains under-examined. In the present study, we analyzed the variations in semantics among different types of word problems presented in highly circulated textbooks to better understand how word problems vary in terms of semantics and structure across the mathematics curriculum. This is important because exposure to a variety of word problems early and often improves student proficiency in solving them (Woodard et al., 2012). We also explored how different textbooks contribute to the variety of word problems presented in the mathematics classroom as a means to understand how textbook content may contribute to the consistent struggles students face in solving word problems. The variation in word problem-solving challenges is influenced by the types and structures of word problems presented in textbooks. Thus, different types and structures of word problems contribute to varying levels of difficulty for students. The first step to developing policies and programs to address this trend is to understand how problem types and structures are distributed across textbooks, hence our focus on the variation of problem structures across different textbooks. Finally, we wanted to revisit the role of textbooks in shaping mathematics education. By examining how textbooks from various publishers differ in the types of word problems they include, we hope to spark a broader conversation regarding the prioritization of specific types of mathematics content in general, and how word problems specifically in textbooks can contribute to students' struggles in mathematics.

Word Problem Solving

Problem solving remains one of the most challenging tasks for learners, yet data indicate that the benefits of a strong problem-solving background are mentally as well as practically important. The ability to solve problems, not education or test scores, predicts employment and wages in adulthood (Gross et al., 2009; Hanushek & Woessmann, 2008). Unfortunately, problem solving remains a complicated subject within mathematics education research. As stated by Lester (1994), "Problem solving has been the most discussed, but also might be one of the least understood topics in mathematics curriculum in the U.S." (pp. 661–662). In the present study, we focus on a specific aspect of problem solving — word problems.

Solving word problems addresses a key aspect of students' problem solving — application. The goal of mathematics education is for students to be able to extend classroom knowledge into outside contexts. Thus, mathematics instruction has evolved to encourage connections to real-world contexts (i.e., word problems; Ashlock, 2006). Additionally, solving word problems can help foster motivation and enhance creative thinking (Verschaffel et al., 2000), two important skills that could further support problem-solving ability. As such, it is important for educators to understand how best to teach students to solve word problems. Researchers have provided evidence of practices that improve students' word problem-solving skills (e.g., Jitendra et al., 2015). However, there is a need for further research on improving strategies for teaching students to solve word problems (Verschaffel et al., 2020). To our knowledge, there are a limited number of studies that have examined the distribution of word problem types in popular mathematics textbooks as a pedagogical consideration for improving teaching word problem solving (Singh et al., 2020). Therefore, we contend that examining mathematics textbooks to assess variety in word problem semantics may provide insight into the diversity of word problems present in students' regular mathematics learning.

Word Problem Taxonomy

Solving word problems is a point of tension for many students. One possible reason for this is that the solution approach to solving a word problem depends on the semantic structure of the problem itself (Hershkovitz & Nesher, 2003; Rosenthal & Resnick, 1974). By simply changing what information is given or left unknown, a completely new subcategory of problem is formed (Riley et al., 1984). In fact, there are so many variations of problem types that students may struggle to keep them straight, making the process of understanding the relationships between numbers in a particular problem particularly difficult (Vergnaud, 1982). To solve word problems, students need to be able to translate the problem's semantic form into its equation representation, including the proper operation and location of the variable or unknown.

There has been ample research regarding the semantics of word problems. Carpenter et al. (2015) established a taxonomy for classifying one-step addition and subtraction problems, organized into a matrix. The National Research Council (NRC, 2009) and the Common Core State Standards Initiative (CCSSI, 2023) have slightly modified this taxonomy. Within the addition and subtraction taxonomy, there are three overarching categories, or situations, of problems: 1) Change situations, 2) Combine situations, and 3) Compare situations (Nesher et al., 1982). Each of these contains subcategories based on the location of the unknown information. Permutations of the categories and subcategories create 12 distinct problem types.

Change Situations

Addition and subtraction can be used to model change. Change, or Add To and Take From, situations are used to show the modifying of a set over time. Specifically, Change Plus, or Add To, situations are used to depict the increasing of a set, and Change Minus, or Take From, situations depict the decreasing of a set (Carpenter et al., 2015). The three pieces of information that make up these problems are the start, change, and result. Because each of these pieces can become the unknown information and because there are two choices of operation, there are a total of six subcategories of Change situations (Riley et al., 1984). Using the CCSSI (2023) naming scheme, the six subcategories and their equation representations can be seen in Table 1.

Combine Situations

At first glance, Combine and Change situations seem similar. After all, the equation representations are very similar (CCSSI, 2023). However, there are two key differences. Combine, or Put Together/Take Apart, situations, emphasize the addends rather than the total (NRC, 2009). Furthermore, while Change situations represent the increase or decrease of one set of items, Combine situations represent a collection of subsets (Vergnaud, 1982). This is the same as the concept of part-part-whole (Carpenter et al., 2015) which helps identify the equals sign as a relationship between two numbers rather than signifying an end result. Although Combine situations can be separated into Put Together situations and Take Apart situations (NRC, 2009), this is not explicitly shown in the CCSSI taxonomy; thus, the two types will be subsequently referenced as one type. As represented by the CCSSI, these situations are modeled in Table 2.

Compare Situations

In Compare situations, addition and subtraction are used to identify the exact differences between two amounts. In other words, when given two amounts, rather than simply identifying the larger one, the student can calculate how much larger it is (NRC, 2009). Similarly, if given the difference in the size of two sets and the size of one of the sets, the student can calculate the size of the other set (Riley et al., 1984). Whereas Change and Combine situations modify a single set or establish a relationship between a set and its subsets, respectively, Compare situations emphasize the comparison of two distinct sets (Carpenter et al., 2015). Compare situations can be broken down into More and Less/Fewer situations (NRC, 2009). Both types are addressed in the CCSSI taxonomy under one overarching category, though they are not separated at all in the Carpenter et al. (2015) version of the taxonomy. The CCSSI representations of the different Compare situations can be seen in Table 3.

Table 1

Change Plus and Change Minus Situations

Add To - Result Unknown	Add To - Change Unknown	Add To - Start Unknown
Six people were riding the bus. Three more joined them. How many people are on the bus?	Six people were riding the bus. Some more joined them. Now there are ten people riding the bus. How many people joined the first six?	Some people were riding the bus. Three more joined them. Then there were ten people riding the bus. How many people were on the bus before?
6 + 3 = ?	6 + ? = 10	? + 3 = 10
Take From - Result Unknown	Take From - Change Unknown	Take From - Start Unknown
Joni got four balloons for her birthday. Two of them popped. How many balloons does Jon have now?	Joni got four balloons for her birthday. Some of them popped. Then there were two balloons. How many balloons popped?	Joni got some balloons for her birthday. Two of them popped. Then there were two balloons. How many balloons did Joni get for her birthday?
4 - 2 = ?	4 - ? = 2	? - 2 = 2

Table 2

Combine Situations

Put Together/Take Apart - Total Unknown	Put Together/Take Apart - Addend Unknown	Put Together/Take Apart - Both Addends Unknown
Five jelly donuts and three sprinkle donuts are in a box. How many donuts are in the box?	Eight donuts are in a box. Five are jelly and the rest are sprinkle. How many donuts are sprinkle?	Mario has four pencils. How many can he put in his backpack and how many in his desk?
5 + 3 = ?	5 + ? = 8	4 = ? + ??

Table 3
Compare Situations

Compare - Difference Unknown	Compare - Bigger Unknown	Compare - Smaller Unknown
Priya has two rings. Jo has four rings. How many more rings does Jo have than Priya? OR Priya has two rings. Jo has four rings. How many fewer rings does Priya have than Jo?	Jo has two more rings than Priya. Priya has two rings. How many rings does Jo have? OR Priya has two fewer rings than Jo. Priya has two rings. How many rings does Jo have?	Jo has three more rings than Priya. Jo has five rings. How many rings does Priya have? OR Jo has five rings. Priya has three fewer rings than Jo. How many rings does Priya have?
1 + ? = 5 5 -2 = ?	1 + 2 = ? 2 + 1 = ?	3 - 2 = ? ? + 2 = 3

Varying Levels of Difficulty

The structure of a word problem can greatly impact the difficulty in solving it. This has been found even when comparing two problems that require the same operation (Riley et al., 1984). Change situations are often the easiest of the problem types and Compare situations are often the most difficult (Nesher et al., 1982). This may be due to the fact that in addition to needing to think operationally, Compare situations require students to think relationally. In other words, not only do students have to look at the differences in sizes, but they also need to separately analyze the relationships between the given quantities (NRC, 2009). Furthermore, regardless of problem type, the level of difficulty of a word problem is directly related to the location of the unknown information (Riley et al., 1984), as certain situations pose a more natural solution for children than others (Vergnaud, 1982). For instance, when the unknown information is the starting information, this tends to pose more difficulty for students. These factors combined provide insight into the importance of the semantics of word problems.

Each of the problem types described previously is standard for elementary mathematics. Even though the CCSSI displays the full taxonomy on their website, little reference is made regarding it in mathematics education literature. In fact, much of the existing literature and research on varying problem types was published before the introduction of the CCSSI (e.g., Carpenter et al., 2015; De Corte & Verschaffel, 1987), yet the inclusion of the taxonomy in national education guidelines suggests that it is still relevant today. Additionally, due to students' continued difficulty with deciphering contextual aspects of word problems (Fuchs et al., 2015; Sepeng & Madzorera, 2014), it seems pertinent that students should be adequately exposed to the various word problem types (Woodward et al., 2012) to potentially increase comfort with all types and levels of difficulty. Historically, it has been shown that textbooks are considered an influential tool in determining what material students are exposed to (Johansson, 2006). Thus, exploring current elementary mathematics textbooks should provide some insight into curriculum presently taught in classrooms. The purpose of this study is to examine three widely distributed textbooks from top textbook publishers to determine the variation of the problem types in elementary mathematics textbooks. Specifically, the researchers aim to answer the following research questions:

RQ1. What is the prevalence of Common Core addition and subtraction word problem semantic structures in select mathematics textbooks?

RQ2. How does the prevalence of Common Core addition and subtraction word problem semantic structures differ across select mathematics textbooks?

Methods

Data in this study were examined in four steps. Each of the steps is described in detail in the subsequent sections.

Step 1: Textbook Selection

The researchers first identified the three leading publishers of elementary mathematics textbooks in the United States. It was determined these were Houghton Mifflin Harcourt, McGraw Hill, and Savvas (Pearson). These companies account for 85% of K–12 textbooks published in the United States (Hazard Owen, 2020). Then the researchers contacted each publisher individually to inquire which of their elementary mathematics textbooks is the most widely distributed in Texas. This was due to the fact that this study took place in the context of a larger study investigating word problem types in the work of second-grade students and preservice teachers within the state of Texas. Therefore, the researchers also selected the second-grade edition of the recommended textbooks. The selected textbooks are as follows: from McGraw Hill, My Math (Carter et al., 2018), from Savvas, enVisionmath2.0 (Charles et al., 2017), and from Houghton Mifflin Harcourt, Texas GO Math! (Dixon et al., 2015). These textbooks are representative of the most used textbooks in elementary mathematics classrooms in the United States.

Step 2: Sample Selection

To analyze the textbooks in the second step, the researchers selected a representative sample chapter from each book. The chapters were selected based on the concentration of addition and subtraction word problems presented, as well as the chapters' overall dedication to word problem instruction. In other words, the researchers selected the chapter in each book that specifically focused on addition and subtraction and included word problems. In both the McGraw Hill and Savvas books, this was the first chapter. For the McGraw Hill book, the chapter was titled "Apply Addition and Subtraction Concepts" and included a review of addition and subtraction as well as an entire section on word problems. For the Savvas book, the chapter was titled "Fluently Add and Subtract Within 20" and contained a section on word problems. In the Houghton Mifflin Harcourt textbook, the selected chapter was Module 5, titled "Basic Facts." This chapter was selected because it focused on one-step addition and subtraction problems and was most comparable to the other two selected chapters.

Step 3: Word Problem Screening

The researchers examined the word problems of each sample chapter. First, the researchers determined whether each word problem fit within the taxonomy for one-step addition and subtraction word problems (CCSSI, 2023). To meet this requirement, the problem had to contain only two numbers and require addition or subtraction when translated directly from word problem to equation. Problems that were multiple-choice format or asked if a given answer was correct were acceptable if they met the previous requirements of fitting into the taxonomy. Though not a specific requirement for classification, all selected problems contained only one- or two-digit numbers.

After compiling a list of the selected problems, the researchers reviewed and coded each problem according to the CCSSI taxonomy. An overview of the coding scheme is presented in Table 4. First, both researchers analyzed the Houghton Mifflin Harcourt book individually. Then the researchers compared their codes to find interrater agreement ($k = 94.3\%$). The researchers then discussed and reconciled discrepancies in the codes. To code the remaining two textbooks, one researcher analyzed the Savvas book while the other analyzed the McGraw Hill book.

Table 4

CCSSI Taxonomy Coding Scheme

Add To - Result Unknown	Put Together/Take Apart - Total Unknown
Add To - Change Unknown	Put Together/Take Apart - Addend Unknown
Add To - Start Unknown	Put Together/Take Apart - Both Addends Unknown
Take From - Result Unknown	Compare - Difference Unknown
Take From - Change Unknown	Compare - Bigger Unknown
Take From - Start Unknown	Compare - Smaller Unknown
	Does not fit within the taxonomy

Step 4: Textbook Analysis

Borrowing from the medical literature, we determined the prevalence of specific word problem structures as a measure of the variation present within and between textbooks. We first defined prevalence based on the equation presented below to examine the representation of Common Core addition and subtraction word problem structures.

$$Prevalence = \frac{\#\ of\ word\ problems\ with\ a\ specific\ semantic\ structure}{Total\ \#\ of\ word\ problems\ in\ sample}$$

Thus, the closer the ratio between the number of specific semantic structures to the total number of word problems in the sample is to 1, the greater the prevalence. We calculated overall and within textbook prevalence for word problems in the selected textbooks to characterize the prevalence of particular word problem semantic structures. We did this by summarizing the frequency data for the sample and generating distribution tables for overall and within problem structure prevalence. “Overall” represents the broad problem structure category (i.e., Add To, Take From, Put Together/Take Apart, or Compare) and within describes the subcategorical problem types (e.g., Add To - Result Unknown).

We then examined frequencies for specific problem structures, calculated 95% confidence intervals for prevalence based on the overall proportions of problem types, and presented the distribution in tables. We calculated confidence intervals using the formulas below, where p'represents the estimated proportion of success or sample proportion of success, *x* is the number of successes (i.e., particular types of word problems), and *n* is the number of trials (i.e., the number of word problems examined). In our analysis, p' is equivalent to the prevalence.

$$p' = \frac{x}{n} \qquad p' \pm z\sqrt{\frac{p'(1-p')}{n}}$$

It is important to note that in instances where there were fewer than five successes or failures, the confidence intervals were calculated using Stata’s built-in exact methods procedure. Then, using the stock graphing option in Microsoft Excel, we plotted the overall and within problem structure prevalence separately. A lack of overlap between pairs of 95% confidence interval plots provides a visual indication of statistically significant prevalence differences. Moreover, confidence intervals are an important indicator of the precision of the prevalence point estimates. Unfortunately, due to the data's aggregate nature, a chi-square analysis was inappropriate due to the violation of statistical assumptions. For example, because the data were aggregated from several textbooks, word problems are not mutually exclusive within and beyond problem categories.

However, chi-square analyses were used to examine statistically significant differences in word problem structures across the selected textbooks. We determined whether to conduct chi-square tests based on the overlap between confidence bands. If the overlap was less than 25%, we considered the categories sufficiently distinct to warrant further analysis. This threshold helps ensure that we only perform chi-square tests with a clear indication of potential differences, thus reducing the risk of Type I errors from multiple testing. Using this criterion, we selectively applied chi-square tests only when necessary, focusing our statistical efforts on categories where substantial differences were evident. This approach reduces the statistical corrections needed and ensures that our findings are robust and relevant.

To address prevalence differences across the three textbooks, we conducted a three-step analysis process. First, we plotted 95% confidence intervals for the prevalence of each category of word problems. Then, based on the overlap between confidence bands present, we determined whether a chi-square test was appropriate. If substantial overlap was not present (i.e., less than 25%), we conducted a chi-square test of independence to assess the association between word problem prevalence and each textbook. By only conducting chi-squared analyses for selected semantic structures, we reduced the number of statistical corrections necessary by only conducting tests that were warranted. If the chi-square test indicated a statistically significant relationship, then the odds ratio (OR) effect sizes were calculated using the textbook with the least prevalence as the referent for the other two textbooks post hoc.

The OR is often misinterpreted and underutilized in the social sciences. The OR is one of several statistics used to assess the risk of a particular outcome if a certain factor is present (Schmidt & Kohlmann, 2008). This analysis uses the OR to estimate how much more likely a particular semantic problem structure will be prevalent within a textbook. In the context of the present study, we interpret ORs as follows: OR = 1, no relationship; OR < 1, less likely to include word problem type; OR > 1, more likely to include word problem type.

Researchers suggest that interpreting the magnitude of the OR is somewhat ambiguous. For example, depending on the user's chosen cut points, the magnitude of a variable's effect may be exaggerated (Cohen & Chen, 2009). Therefore, in the present study we used benchmarks based on those proposed for interpreting Cohen's d. According to Chen et al. (2010), OR values of 1.68, 3.47, and 6.71 are equivalent to Cohen's small, medium, and large effects, respectively. In addition, we calculated 95% confidence intervals to determine whether the OR values were statistically significantly different from 1. If the OR statistically significantly diverges from 1, then the 95% confidence interval will not include 1.

Our combined use of 95% CIs and ORs provided nuanced insights into the prevalence of word problem types across textbooks. Visual inspection of CIs allowed us to preliminarily identify categories with potential differences that could be further tested using chi-square tests. The ORs offered a clear, interpretable measure of effect size, enabling us to quantify and compare the likelihood of word problem types across textbooks. These approaches offer several advantages over simpler descriptive statistics. First, 95% confidence intervals support precision and clarity because CIs provide a clear measure of estimate precision. Second, by using 95% CIs and ORs, we reduced the need for multiple comparisons, minimizing the risk of Type I errors. Finally, ORs support effect size quantification because ORs offer a robust measure of the strength of associations. These statistical methods support robust, interpretable, and meaningful comparisons across textbooks.

Results

The researchers classified each of the problems from the sample chapters according to problem type. Table 5 below shows the frequencies of each of the problem types by textbook. Comparing the frequencies of both the categories and subcategories provides insight into the distribution of the word problem semantic structures within elementary mathematics textbooks.

The frequency of appearance of the various semantic structures varied by type. The most frequent classification of problems was the Combine situations (50), followed closely by the Take From (47) situations. Note that oftentimes Add To and Take From problem types are combined to make Change situations. In this context, Change situations (81) would be the most frequently occurring classification across the textbooks. Finally, the Compare situations appeared least frequently (27). Figure 1 shows the distribution of situations across the textbooks.

Figure 2 presents the confidence interval plots for each of the four categories' overall semantic problem structure prevalence. We plotted 95% confidence intervals for each category of word problem's prevalence to assess the variability and overlap between textbooks visually. The 95% confidence interval is a standard choice in social sciences because it provides a balance between confidence in the results and the risk of Type I errors (false positives). This interval indicates that if the study were repeated numerous times, we expect the true prevalence to fall within this range 95% of the time.

The length of the individual bands reflects the precision of the point estimates. The figure suggests that the precision of the point estimates across the categories are relatively similar because the bands' lengths are somewhat similar. The plots also provide a means to compare the relative magnitude of each point estimate. The data indicate that Compare problems are statistically significantly less prevalent in the dataset based on the lack of overlap between the Compare problems' confidence band and the confidence bands for Take From problems and Put Together problems. However, based on the overlapping confidence bands for Add To and Compare problems, we conclude that there are no statistically significant differences between their relative prevalence in the dataset. We would be remiss not to note that the only way to assess statistically significant differences definitively would be through formal hypothesis testing; however, as stated earlier, this is inappropriate given the nature of the data. Thus, we used confidence intervals and visual inspection (American Psychological Association, 2020; Cumming, 2013).

Table 5
Frequency of Appearance of Problem Types

Situation	Problem Type	Savaas	McGraw Hill	HMHco	Total
Change	Add To - Result Unknown	7	7	5	19
	Add To - Change Unknown	4	4	6	14
	Add To - Start Unknown	0	0	1	1
	Take From - Result Unknown	14	13	3	30
	Take From - Change Unknown	2	6	4	12
	Take From - Start Unknown	2	1	2	5
Combine	Put Together/Take Apart - Total Unknown	10	12	11	33
	Put Together/Take Apart - Addend Unknown	4	4	5	13
	Put Together/Take Apart - Both Addends Unknown	3	1	0	4
Compare	Compare - Difference Unknown	11	0	7	18
	Compare - Bigger Unknown	1	0	0	1
	Compare - Smaller Unknown	2	0	6	8
	Does not fit within the taxonomy	0	1	0	1

Note: HMHco = Houghton Mifflin Harcourt

Figure 1
Percentages of Problem Situations by Textbook

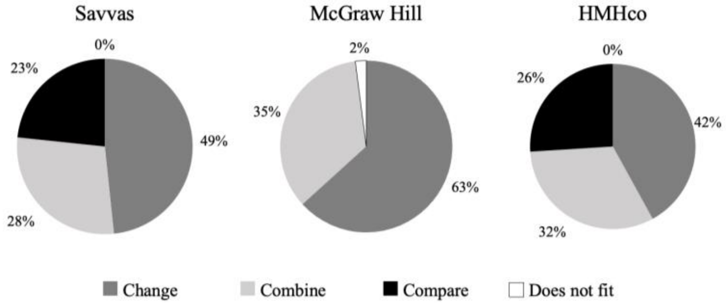
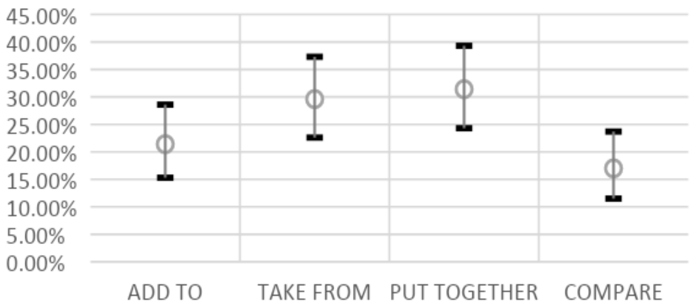


Figure 2
95% Confidence Interval Plots for Overall Semantic Problem Structure



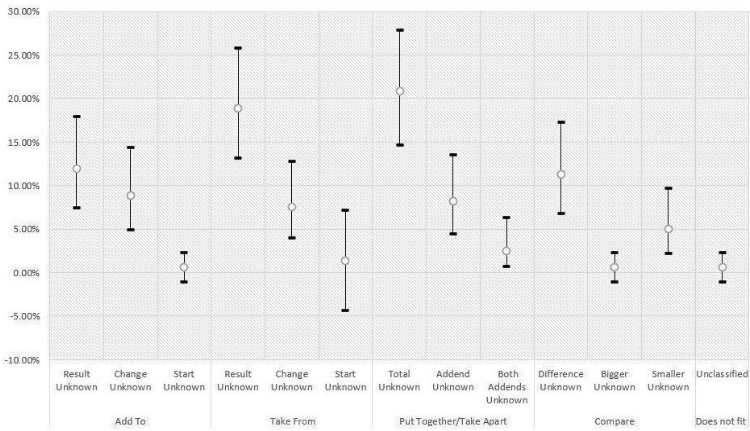
Prevalence by Category

A similar procedure was used to assess the prevalence of semantic problem types within categories (see Figure 3). From left to right, we plotted the 95% confidence intervals for the prevalence of the four semantic number sentence categories (i.e., Add To, Take From, Put Together/Take Apart, and Compare). Within the Add To category, Result Unknown was most prevalent, and Start Unknown was the least prevalent. Additionally, based on the width of the Start Unknown confidence band, we conclude that the Start Unknown point estimate is more precise compared to the other subcategories in the Add To group.

It is important to note that the width or precision of confidence bands is inversely related to sample size, thus as the sample size becomes larger the measures become more precise. But the width is directly related to variability, thus large variability would mean wider bands. In the case of the start unknown, there was a small sample size, but the prevalence was consistent (i.e., low variability), thus more precise. Furthermore, compared to the other confidence bands, the lack of overlap between the Start Unknown confidence band and the other bands indicates statistically significant differences.

Figure 3

95% Confidence Interval Plots for Word Problem Type Prevalence Within Subcategories



The data presented within the Take From category indicate that the point estimates are similar in their precision based on their confidence bands' relative lengths. Result Unknown problems were the most prevalent and appear to be statistically more prevalent than Change Unknown and Start Unknown problems based on the relative lack of independent confidence bands. Similar to the data presented earlier for Add To problems, the Take From problem data also indicate that the Start Unknown subcategory is the least prevalent. However, the differences do not appear to be statistically significantly smaller than the Change Unknown subcategory, based on the overlap between the pair of confidence bands.

The next category was the Put Together/Take Apart problem category. Applying 95% CIs to Put Together/Take

Apart Problem Structure, as well as other categories, helps us identify meaningful differences between textbooks. This method offers a visual and statistical way to gauge the precision of our prevalence estimates and assess if the observed differences are likely due to sampling variability or represent true differences in textbook content. Within this category, there was slightly more divergence in the relative precision of the confidence bands. Based on the widths of the confidence bands, the Total Unknown subcategory was the least precise, followed by the Addend Unknown and Both Addends Unknown problem subcategories. The Total Unknown problem type was the most prevalent and appears to be statistically significantly more prevalent than the Addend Unknown and Both Addends Unknown subcategories. The Both Addends Unknown subcategory was the least prevalent but did not appear to be statistically significantly less prevalent than the Addend Unknown subcategory.

The final category presented in Figure 3 is the Compare category. Differences in the precision of the point estimates are present within the Compare category, based on the relative differences between the widths of the confidence bands. Regarding point estimate relative magnitudes, the data in Figure 3 suggest that the Difference Unknown subcategory was the most prevalent. However, the Difference Unknown subcategory appears to only be statistically significantly more prevalent than the Bigger Unknown subcategory of word problems. Finally, each point estimate's precision varies, with the Difference Unknown subcategory being the least precise, followed by the Smaller Unknown and then the Bigger Unknown subcategories. These data characterize the prevalence of different categories of semantic word problem types.

There were many similarities and differences in the prevalence of the semantic structures across the textbooks. Table 6 shows the frequency of the structure categories for each textbook. For the Add To and Put Together/Take Apart categories, the frequencies were nearly identical. However, for the other categories, there were noticeable differences in frequency. Specifically, the Houghton Mifflin Harcourt textbook has half as many Take From problems as the other books, and the McGraw Hill textbook has no Compare problems within the selected chapter.

Table 6
Word Problem Category Frequency by Textbook

Taxonomy Classification	Savvas	McGraw Hill	HMHco	Total
Add To	11	11	12	34
Take From	18	20	9	47
Put Together/Take Apart	17	17	16	50
Compare	14	0	13	27

Figure 4 presents the first in a series of four 95% confidence interval plots comparing the prevalence of specific semantic word problem structures across the three selected textbooks. The data in Figure 4 suggest that the relative precision of the point estimates was similar, based on the similar widths of the three textbooks' corresponding confidence bands. The Houghton Mifflin Harcourt textbook appears to have the highest prevalence of Add To problems, followed by the McGraw Hill textbook and then the Savvas textbook. The substantial overlap observed between the three confidence bands indicates that statistically significant differences are highly unlikely; thus, a Pearson chi-square analysis was not performed.

The 95% confidence intervals for the Take From category prevalence across textbooks (see Figure 5) also suggest that the point estimate precision is similar, based on the confidence band widths. The McGraw Hill textbook has the highest prevalence, followed by the Savvas and then the Houghton Mifflin Harcourt textbooks. A chi-square test of independence was completed to assess statistically significant textbook prevalence differences. The relationship between textbook and prevalence of Take From problems was statistically significant: X2 (2, N = 159) = 6.196, p = .045. OR effect sizes were calculated post hoc to assess the magnitude of the relative differences using the Houghton Mifflin Harcourt textbook as the referent, or control, because it has the least prevalence of Take From problems. The results indicated that the odds of the Savvas textbook including Take From problems compared to the Houghton Mifflin Harcourt textbook were almost double: OR = 1.95 (95%CI 0.79-4.84). Because the confidence interval includes 1, the aforementioned effect size is not statistically significant.

Figure 4
95% Confidence Interval Plots for Add To Problem Structure Prevalence Across Textbooks

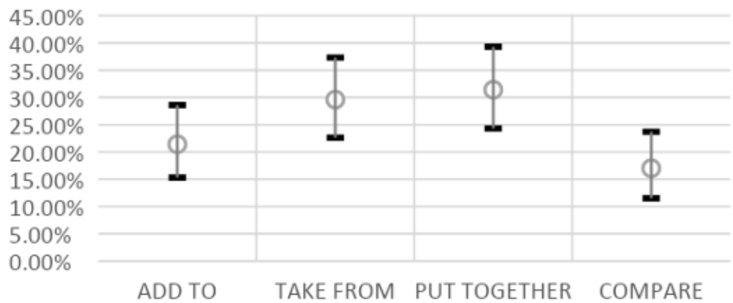
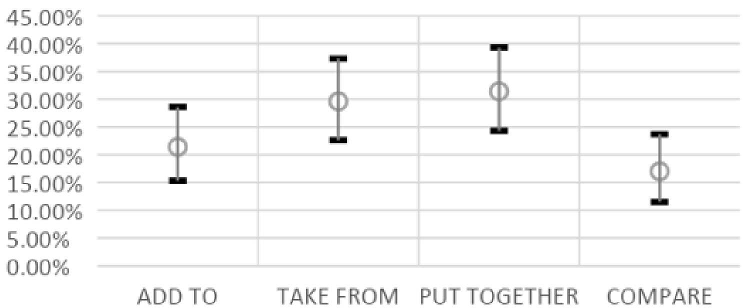


Figure 5
95% Confidence Interval Plots for Take From Problem Structure Prevalence Across Textbooks



However, the OR effect sizes for the differences between the McGraw Hill textbook and the Houghton Mifflin Harcourt textbook indicate that the odds of the former including Take From problems compared to the latter is more than triple: OR = 3.14 (95%CI 1.25-7.88). Because the confidence interval does not include 1, the effect size is statistically significant. Thus, the odds are larger than expected due to chance.

Figure 6 presents the 95% confidence interval plots comparing the prevalence of Put Together/Take Apart problems across the selected textbooks. The data in Figure 6 suggest that the point estimates' relative precision was similar, based on the similar widths of the corresponding confidence bands. The McGraw Hill textbook appears to have the highest prevalence of Put Together/Take Apart problems, followed by the Houghton Mifflin Harcourt textbook, then the Savvas textbook. The substantial overlap observed between the three confidence bands indicates that statistically significant differences are highly unlikely; thus, a Pearson chi-square analysis was not performed.

Figure 6
95% Confidence Intervals for Put Together/Take Apart Problem Structure Prevalence

Because the McGraw Hill textbook does not include any Compare problems in its designated word problems section, we could not construct 95% confidence interval plots for the Compare problems. Likewise, because the prevalence was zero, and given the relatively large prevalence across the other textbooks, statistically significant differences were highly likely. The relationship between textbooks and prevalence of Compare problems was statistically significant: X2 (2, N = 159) = 14.63, $p < .001$. Moreover, it was impossible to calculate OR effect sizes using the McGraw Hill textbook as the referent. Therefore, OR effect sizes were calculated for Compare problem prevalence within the Savvas and Houghton Mifflin Harcourt textbooks. The results indicate that the prevalence of Compare problems across the Savvas and Houghton Mifflin Harcourt textbooks were nearly identical: OR = 1.15, (95% CI = .48-2.76).

Discussion and Conclusion

In this study, we analyzed the variation of word problem structures within three popular elementary mathematics textbooks. Specifically, we examined the prevalence and variation of Common Core addition and subtraction word problem semantic structures within and across these textbooks. The results of this study indicate that variation in problem structure does not appear to occur proportionately within a single textbook. However, the frequency of the problem categories appears to be similar across textbooks. In other words, students may not be seeing the 12 different problem types in equal amounts in any of the given textbooks. On the other hand, they may see comparable types of problems regardless of which textbook is chosen. Hence, educators and curriculum designers should be mindful of the semantic structures presented in textbooks. A more deliberate selection or design of instructional materials that offer a balanced exposure to different problem types can contribute to improved student proficiency in word problem-solving at differing levels of difficulty.

An interesting connection was found between problem structure frequency and problem difficulty, as noted in prior research. In all four of the overarching problem structures, the most common substructure had the unknown information as the end result (e.g., Take From - Result Unknown). In fact, when looking at the overall frequency of the 12 possible substructures, the following three substructures were the most common substructures in two out of the three books: Take From - Result Unknown, Put Together/Take Apart - Total Unknown, and Compare - Difference Unknown. It has previously been determined that students tend to have more difficulty solving problems if the result is part of the given information rather than the unknown (Riley et al., 1984), yet most of the problems in popular textbooks continue to reflect

Result Unknown substructures. Thus, our results suggest that most word problems presented in the textbooks analyzed had semantic sentences that matched the required calculation (i.e., Three chocolate chip cookies and four peanut butter cookies are in a box. How many cookies are in the box? or 3 + 4 = ?). This is a notable finding because textbooks guide instructional decisions, thus if less difficult semantic structures are presented more frequently, then subconsciously some teachers may not attend to student challenges with more difficult problem types by creating or assigning problems beyond those provided at the district level (i.e., textbooks).

A similar observation can be made regarding the frequency of problem categories. Of all the problem categories, Compare problems occurred least frequently. In prior research, these were previously labeled as the most difficult problem type (Nesher et al.,1982). Compare problems are especially problematic because the semantic structure requires the student to decode the proper computational number sentence based on a comparison of two quantities which is difficult for some teachers to identify, pose, and answer correctly (Calabrese & Capraro, 2022). Hence, increasing student exposure to these problem types seems warranted. Further research is needed to explore the impact of specific word problem structures on students' performance. This includes investigating how variations in semantic structures contribute to the challenges students face and identifying evidence-based instructional strategies that address these challenges.

The textbooks selected in this study are considered to be widely distributed across the state of Texas, though they are intended to be representative of the textbooks used in second-grade classrooms across the country. The results of this study indicate that there is currently a disproportionate representation of the 12 addition and subtraction word problem types. This proportion is reminiscent of prior findings on solution difficulty. That is, the word problems that had been previously labeled as more difficult to solve now have been found to be less prominent in mathematics textbooks. Educational policymakers should consider the implications of the study's findings when developing standards and guidelines for instructional materials. Advocacy for a more standardized approach to the inclusion of word problems in textbooks can lead to a more consistent and supportive learning experience for students. Textbook authors could then use this information for future consideration on whether students could benefit from a more balanced exposure to word problem types. To this end, collaboration among textbook publishers to align their approaches to word problem inclusion can contribute to a more coherent and effective mathematics curriculum. Moreover, shared guidelines and best practices can help create a more standardized and supportive learning environment for students across different educational settings.

In conclusion, the study sheds light on the importance of considering the semantic structure of word problems in elementary mathematics textbooks. Understanding the semantic structure of word problems is crucial to addressing the challenges students face in solving them. The findings suggest that different textbooks prioritize certain types of word problems, potentially contributing to the varied struggles students experience. The emphasis on specific problem types in textbooks reflects the content exposure students receive, influencing their problem-solving skills. Moreover, our findings reinforce the importance of recognizing the role of textbooks in shaping students' mathematical experiences and highlight the need for a more standardized approach to word problem inclusion in instructional materials. It is our hope that the results of the present study emphasize the need for collaborative efforts among educators, curriculum designers, researchers, and policymakers to improve the quality of mathematics education, with a specific focus on enhancing students' word problem-solving skills.

Limitations

As with any study, there are certain limitations to take into consideration while reviewing results. In the present study, the researchers only analyzed a single chapter of three textbooks in Texas. While these are the top-selling textbooks for the state for the selected grade level, they represent a very limited sample of textbooks within the United States. Furthermore, despite careful selection of the analyzed chapters within the selected textbooks, there may be further variation within the remaining chapters that were not analyzed in the present study. Thus, the audience should consider this sample before making further generalizations.

References

Almolhodaei, H. (2002). Students' cognitive style and mathematical word problem solving. *Journal of the Korea Society of Mathematical Education Series D: Research in Mathematical Education*, *6*(2), 171–182.

American Psychological Association. (2020). *Publication manual of the American Psychoogical Association* (7th ed.). American Psychological Association.

Arsenault, T. L., & Powell, S. R. (2022). Word–problem performance differences by schema: A comparison of students with and without mathematics difficulty. *Learning Disabilities Research & Practice*, *37*(1), 37–50. https://doi.org/10.1111/ldrp.12273

Ashlock, R. B. (2006). *Error patterns in computation: Using error patterns to improve instruction* (9th ed.). Macmillan.

Cai, J., & Jiang, C. (2017). An analysis of problem-posing tasks in Chinese and US elementary mathematics textbooks. *International Journal of Science and Mathematics Education*, *15*(8), 1521–1540. https://doi.org/10.1007/s10763-016-9758-2

Calabrese, J.E., & Capraro, M.M. (2022). Word problem taxonomy and problem posing: A professional development exploration, *Primary Mathematics*, *26*(2),16–20.

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (2015). *Children's mathematics: Cognitively guided instruction* (2nd ed.). Heinemann.

Carter, J. A., Cuevas,G. J., Day, R., Malloy, C., Altieri, M. B., Balka, D. S., Gonsalves, P. D., Grace, E.C., Krulik, S., Molix-Bailey, R. J., Moseley, L. G., Mowry, B., Myren, C. L., Price, J., Reynosa, M. E., Santa Cruz, R. M., Silbey, R. S., & Vielhaber, K. (2018). *My math: Grade 2* (Vol. 1). McGraw-Hill Education.

Charles, R. I., Bay-Williams, J., Berry, R. Q., III, Caldwell, J. H., Champagne, Z., Copley, J. V., Crown, W. D., Fennel, F., Karp, K., Murphy, S. J., Schielack, J. F., Suh, J. M., & Wray, J. A. (2017). *enVisionmath2.0*. Pearson Education.

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—Simulation and Computation, 39*(4), 860–864. https://doi.org/10.1080/03610911003650383

Cohen, P., & Chen, H. (2009). How the reflection of linear correlation in odds ratios depends on the cut-off points. *Communications in Statistics-Simulation and Computation*, *38*(3), 610–620. https://doi.org/10.1080/03610910802592820

Common Core State Standards Initiative. (CCSSI, 2023). *Mathematics glossary: Table 1*. http://www.thecorestandards.org/Math/Content/mathematics-glossary/Table-1/

Cumming, G. (2013). The new statistics: A how-to guide. *Australian Psychologist*, *48*(3), 161–170. https://doi.org/10.1111/ap.12018

Curcic, D. (2023, June 13). *K-12 textbooks sales statistics.* https://wordsrated.com/k-12-textbooks-sales-statistics/

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, *6*, Article 348. https://doi.org/10.3389/fpsyg.2015.00348

De Corte, E., & Verschaffel, L. (1987). The effect of semantic structure on first graders' strategies for solving addition and subtraction word problems. *Journal for Research in Mathematics Education*, *18*(5), 363–381. https://doi.org/10.5951/jresematheduc.18.5.0363

Dixon, J. K., Burger, E. B., Larson, M. L., & Sandoval-Martinez, M. E. (2015). *Texas GO math!* (Vol. 1). Houghton Mifflin Harcourt Publishing Company.

Fuchs, L. S., Fuchs, D., Compton, D. L., Hamlett, C. L., & Wang, A. Y. (2015). Is word-problem solving a form of text comprehension? *Scientific Studies of Reading*, *19*(3), 204–223. https://doi.org/10.1080/10888438.2015.1005745

Gagne, R. M. (1983). Some issues in the psychology of mathematics instruction. *Journal for Research in Mathematics Education*, *14*(1), 7–18. https://doi.org/10.5951/jresematheduc.14.1.0007

Gross, J., Hudson, C., & Price, D. (2009). *The long term costs of numeracy difficulties*. Every Child a Chance Trust; KPMG; National Numeracy.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, *46*(3), 607–68. http://doi.org/10.1257/jel.46.3.607

Hazard Owen, L. (2020). *Common Core resources: Textbook & publisher resources*. National University Library System. https://nu.libguides.com/CommonCore/ccss_publishers

Hershkovitz, S., & Nesher, P. (2003). The role of schemes in solving word problems. *The Mathematics Educator*, *7*(2), 1–24.

Jitendra, A. K., Petersen-Brown, S., Lein, A. E., Zaslofsky, A. F., Kunkel, A. K., Jung, P. G., & Egan, A. M. (2015). Teaching mathematical word problem solving: The quality of evidence for strategy instruction priming the problem structure. *Journal of Learning Disabilities*, *48*(1), 51–72. https://doi.org/10.1177/0022219413487408

Johansson, M. (2006). *Teaching mathematics with textbooks: A classroom and curricular perspective* [Doctoral dissertation, Luleå University of Technology].

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The Journal of the Learning Sciences*, *13*(2), 129–164. https://doi.org/10.1207/s15327809jls1302_1

Lester, F. K. (1994). Musings about mathematical problem-solving research: 1970-1994. *Journal for Research in Mathematics Education*, *25*(6), 660–675. https://doi.org/10.5951/jresematheduc.25.6.0660

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*.

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. National Academies Press.

Nesher, P., Greeno, J. G., & Riley, M. S. (1982). The development of semantic categories for addition and subtraction. *Educational Studies in Mathematics*, *13*(4), 373–394. https://doi.org/10.1007/BF00366618

Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM*, *52*(1), 33–44. https://doi.org/10.1007/s11858-019-01118-9

Riley, M. S., Greeno, J. G., & Heller, J. I. (1984). Development of children's problem-solving ability in arithmetic. In H. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). Academic Press.

Rosenthal, D. J., & Resnick, L. B. (1974). Children's solution processes in arithmetic word problems. *Journal of Educational Psychology*, *66*(6), 817–825. https://doi.org/10.1037/h0021523

Schmidt, C. O., & Kohlmann, T. (2008). When to use the odds ratio or the relative risk? *International Journal of Public Health*, *53*(3), 165–167. http://doi.org/10.1007/s00038-008-7068-3

Sepeng, P., & Madzorera, A. (2014). Sources of difficulty in comprehending and solving mathematical word problems. *International Journal of Educational Sciences*, *6*(2), 217–225. https://doi.org/10.1080/09751122.2014.11890134

Singh, P., Yusoff, N. M., & Hoon, T. S. (2020). Content analysis of primary school mathematics textbooks and its relationship with pupils achievement. *Asian Journal of University Education, 16*(2), 15–25. https://doi.org/10.24191/ajue.v16i2.10286

Törnrros, J. (2005). Mathematics textbooks, opportunity to learn and student achievement. *Studies in Educational Evaluation*, *31*(4), 315–327. http://doi.org/10.1016/j.stueduc.2005.11.005

Valverde, G. A., Bianchi, L. J., Wolfe, R. G., Schmidt, W. H., & Houang, R. T. (2002). *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Kluwer.

Vergnaud, G. (1982). A classification of cognitive tasks and operations of thought involved in addition and subtraction problems. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 39–59). Routledge.

Verschaffel, L., Depaepe, F., & Van Dooren, W. (2014). Word problems in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 641–645). Springer.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Taylor & Francis.

Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM*, *52*(1), 1–16. https://doi.org/10.1007/s11858-020-01130-4

Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig P., Jitendra, A., Koedinger, K.R., & Ogbuehi, P. (2012). *Improving mathematical problem solving in grades 4 through 8*. National Center for Education Evaluation and Regional Assistance.

Xin, Y. P. (2007). Word problem solving tasks in textbooks and their relation to student performance. *The Journal of Educational Research*, *100*(6), 347–360. https://doi.org/10.3200/JOER.100.6.347-360

Yang, D. C., & Sianturi, I. A. J. (2022). Analysis of algebraic problems intended for elementary graders in Finland, Indonesia, Malaysia, Singapore, and Taiwan. *Educational Studies*, *48*(1), 75–97. https://doi.org/10.1080/03055698.2020.1740977