Mediterranean Journal for
Research in Mathematics Education
Vol. 14, 43-57, 2015

Relationships between Student Performance on Arithmetic Word Problems Eye-Fixation Duration Variables and Number Notation: Number Words vs. Arabic Numerals

CSABA CSÍKOS: University of Szeged, Institute of Education, Szeged, Hungary

JÁNOS STEKLÁCS: Kecskemét College, Teacher Training Faculty, Kecskemét, Hungary

ABSTRACT: *This study investigated 4th grade students' eye-movements during arithmetic word problem solving. The sample consisted of 24 students (13 boys and 11 girls, mean age 10 years and 5 months). The students solved a 2x2 system of four tasks: (1) addition–numerals, (2) subtraction–number words, (3) subtraction–numerals, and (4) addition–number words. Besides performance and response time, fixation duration variables were computed: fixation duration on the text (FDT) and fixation duration on the number areas (FDN), and fixation duration on the keyword (FDK). Significant correlations were found between FDN and FDK, both for Task 1 and Task 3, but not for Task 2 and Task 4, suggesting that the number format played a significant role in the problem solving process. Our research may yield new results about the practical educational use of word problems with different number notations.*

Key words: *Eye-movement, Arithmetic, Number notation.*

INTRODUCTION

To what extent mathematics education fulfills the role of fostering useful mathematical knowledge is often measured by means of word problems. Csíkos, Kelemen, and Verschaffel (2011) highlight the importance of word problems in mathematics classroom practice by drawing attention to their significance in skill application, i.e., word problems may (or should) be a means of applying mathematical knowledge and skills in real-world like situations. Some mathematical word problems engage higher-order thinking skills by requiring students to build genuine mathematical models of an everyday situation (Verschaffel, Greer, & De Corte, 2000). Other types of word problems can be solved by means of only using a superficial solution strategy (see Verschaffel & De Corte, 1997), in other words, by searching for figures in the text, and connecting them with an arithmetic operation. In between these two sides of the coin, an interesting type of word problems requires students to search for figures in the text, but mechanically executing an arithmetic operation closely associated with a keyword in the text will lead to wrong results. In this research we used word problems of the compare type (a comparison between two quantities is required) using a so-called “inconsistent” keyword due to which the superficial solution strategy would fail.

Inconsistent Word Problems of the Compare Type

During the elementary school years, compare word problems are of special importance. Compare word problems have their steady place as a distinct category beside other types of simple arithmetic word problems. Early classifications identified four clusters of simple arithmetic word problems: change, combine, compare and equalize types (see Radatz, 1983; Stigler, Fuson, Ham, & Sook Kim, 1986); in recent publications combine, compare and change types are distinguished and defined (see Jitendra, Griffin, Deatline-Buchman, & Sczesniak, 2007; Riley & Greeno, 1998). In compare word problems, two values of a variable are given, and there is a relational statement connecting those two values. There is an extensive body of research about the difficulties compare word problems cause for elementary school children (see for example Mwangi & Sweller, 1998).

Inconsistent word problems are a special type of arithmetic word problem, where students are required to execute one arithmetic operation (e.g., addition), and there is a keyword presented in the text, which is a relational term inconsistent with the required operation (e.g., “less” when addition is the required operation (van der Schoot, Bakker Arkema, Horsley, & van Lieshout, 2009)). Here the term “required” refers to the strategy of connecting the two values of the text with one basic operation. A simple example of an inconsistent compare word problem is the following:

The oldest man in the village is 112 years old. He is 8 years older than his wife. How old is his wife?

In the above word problem the basic operation that connects the two numbers found in the text is subtraction, whereas the relational term “older” would have been consistent with addition (while the relational term consistent with subtraction would be “younger”).

According to Hegarty, Mayer, and Green’s (1992) results among undergraduate students, inconsistent word problems take more time to solve than problems with a consistent keyword, and research provided evidence for the longer fixations needed for inconsistent word problems (Verschaffel, De Corte, & Pauwels, 1992).

Van der Shoot et al. (2009) revealed among 5th and 6th grade children that besides inconsistency, another factor called “markedness” makes the solution process more difficult. Markedness refers to the phenomenon that words expressing the ‘negative’ quality of antonymous pairs of words are semantically more complex, therefore their presence may lengthen the solution process. For instance, the appearance of “less than” or “smaller than” may make the solution process more difficult.

Verschaffel, De Corte, and Pauwels (1992) showed that this consistency-inconsistency effect could be detected in data gathered among 3rd grade students, but not among university students. These seemingly contradictory results could be reconciled when university students faced real challenges when solving compare problems, i.e. when they had to find the solution to the tasks rather than just state what operations had to be computed.

Stimulus Modality in Arithmetic Comparisons

"Number reading is … architecturally similar to word reading" (Cohen, Dehaene, & Verstichel, 1994, p. 279). According to the triple code theory developed by Dehaene and summarized by Dehaene, Molko, Cohen, and Wilson (2004), dedicated brain circuits are engaged in recognizing the number of objects in a set. This model suggests that when solving simple compare word problems, different left and right segments of the brain (intraparietal sulcus) are activated. Different malfunctions in arithmetic computations can be associated with different neural correlates (Dehaene, Piazza, Pinel & Cohen, 2003), and this suggests that there are different neural coding systems for number words and Arabic numerals, and proper functioning is associated with appropriate number representation in case of both types of coding (Dehaene et al., 2004).

Number comparison is a prerequisite to solving compare word problems. Research on number comparison using event-related potential (ERP) suggests that the time needed for determining whether a quantity is above or below five does not interact with number notation (see Temple & Posner, 1998).

Another prerequisite component of solving compare word problems is the execution of simple arithmetic operations such as addition or subtraction. In a research with young Hungarian adults, ERP brain activity results suggested that simple one-digit addition required more time when presented in word number format (Szűcs & Csépe, 2004). Similarly, Rayner (1998) summarizes results from an eye-movement study: when reading numbers, fixation times vary with the number of syllables and with the frequency and magnitude of the numbers.

In previous eye-movement research on inconsistent word problems, stimulus materials contained numerical values in the Arabic numeral format (De Corte et al., 1990; Hegarty et al., 1992; van der Schoot et al., 2009). Since the effort needed for solving compare word problems can be measured by means of eye-fixation durations, De Corte et al. (1990) suggested several different time-related measures attributable to different difficulties such word problems may cause. One of these possible measures is the duration of fixations on textual components and on the numbers of the word problem. Within the textual components, the keywords of the inconsistent word problems are of special importance.

The Current Practice in the Textbooks

In Hungary, the currently used textbooks and task booklets overwhelmingly use numerals to represent numbers in word problems. Arithmetic word problems containing number words instead of numerals may find their place in further research and in educational practice, too. One of the authors of one of the most widely used 4th grade Hungarian mathematics textbooks (I. Libor, personal communication, March 11, 2013) proposed three factors that may explain this phenomenon: (1) According to the rules of Hungarian spelling, numbers can be written either in Arabic numeral format or in words, but in *common practice* (italicized by us) numbers that can be written in a short phrase are written in words (e.g., ten million). (2) Teaching practice suggests that using number words makes a word problem more difficult and therefore cardinal numbers are usually written in Arabic numeral format whereas ordinal numbers and simple fractions (where the enumerator is 1) are written as number words in word problems. A brief analysis based on a sample series of word problems published on the website of the largest textbook publisher (Szöveges feladatok, 2015) shows that out of 25 sample tasks, only one contained numbers words for cardinal numbers and another one mixed the Arabic cardinal numbers and the number words. Ordinal numbers were written as number words. (3) The textbook review process has a bias towards the "common practice", i.e. towards the use of Arabic numerals in word problem texts.

As part of this section on the current textbook practice, for the sake of international comparison, two key elements of the transparency of Hungarian number words will be given. A recent summary from Cankaya, LeFevre, and Sowinski (2012) contains examples for naming two- and three-digit numbers in German, Dutch, English, French, Czech, Basque, and Chinese. Compared to them, Hungarian composition of number words between 10 and 1000 is rather similar to the Chinese way of naming two- and three-digit numbers, and the authors call this system "regular". A second characteristic is that numbers below 2000 are written as one single number word in Hungarian.

Aims of the Current Research

Combining the ideas from these two previous lines of research, we investigated the role of keywords and the role of the notation of numbers in the text in solving inconsistent word problems. The aim of the investigation is to reveal the relative (if any) effects of the following factors on performance and on different eye-fixation measures: the arithmetic operation to be computed and number notation (Arabic vs. number words).

We hypothesize that (1) the duration of fixations on different parts of the word problems, i.e. text area (excluding keyword area) and number area, and (2) the notation of the number presented in the text (number words vs. Arabic numbers) will have an effect on students' performance. Furthermore, it is hypothesized that (3) number notation and the operation to be computed in a task will have an effect on reaction time and on fixation duration.

METHODS

Sample

The sample consisted of 24 students (13 boys and 11 girls, mean age 10 years and 5 months). They all attended the same school in a large county seat town and had diverse social-economic status backgrounds. The heterogeneity of the sample in terms of their family background may strengthen the generalizability of our results. The sample – although being so-called convenient sample – has an appropriate size for applying various statistical methods with the aim of documenting the phenomenon to be investigated. Also, this sample size allowed for providing uniform experimental settings throughout the experiment. None of the participants suffered from eye disease or from dyscalculia.

Tasks

The students solved four experimental tasks and two buffer tasks. The four experimental tasks comprised a 2x2 system with the arithmetic operation needed for an effective solution as one variable, and the number format as another variable: Task 1: addition – numerals; Task 2: subtraction – number words; Task 3: subtraction – numerals; Task 4: addition – number words. All tasks contained an inconsistent keyword (e.g., shorter, when addition was required). The four tasks were presented on the screen of the eye tracking system in three lines. The two statements about the quantities formed two lines, and the third line contained the question. The three lines were center-aligned. The English translation of the tasks (based on the translation – re-translation method in developing the translated text) is shown in Table 1.

Table 1

Task System of the Four Word Problems Used in the Experiment

Task	Text	Operation	Modality
1	John has 115 books. He has 8 books less than Grete has. How many books does Grete have?	addition	numerals
2	The oldest man in the village is one hundred and twelve years old. He is eight years older than his wife. How old is his wife?	subtraction	number words
3	The highest pyramid in Egypt is 137 m high. It is 20 meters higher than the Great Lighthouse. How many meters high is the Great Lighthouse?	subtraction	numerals
4	The running track is one hundred and twenty-five meters long. It is seventeen meters shorter than our street. How long is our street?	addition	number words

The first buffer task was a very easy warm-up task of the consistent type that contained numbers in the Arabic numeral notation and required students to simply add two numbers. The next four tasks formed the core part of the research and the last task (the second buffer task) was a puzzle-like, 'tricky' word problem, which is not discussed in the current analysis but seems to be worth being analyzed from a linguistic, social and socio-psychological viewpoint.

Procedure and Measures

Data collection was done in the school. Before the experiment started, children received information about the general aim of obtaining information about their mathematical task solving abilities. The students were then individually tested in a quiet room. They were given a few minutes to look around and examine computer-like screen of the eye tracking system and then the investigation began with the usual calibration process. Eye movements were registered with a Tobii T120 eye tracker.

Having received the answer to each task, the research assistant switched to the next task by pressing the space button on her computer, and then noted down the student's answer without giving any feedback as to whether the answer was right or wrong. Having completed all tasks, the student had to leave the room and go to another room without the chance of informing his or her classmates about either the tasks or the procedure.

Five dependent variables were computed for each task: (1) performance, (2) response time (RT), total fixation duration time (TFD), and fixation time on (3) the text components (excluding the keyword) (FDT), on (4) number areas (FDN), and on (5) the keyword (FDK). Since RT and TFD are very closely correlated (Pearson-correlations between .83 and .99), they could be used interchangeably.

RESULTS

The results will be presented according to the three research questions we proposed. Before doing so, some statistical analyses are required to ensure the reliability and validity of our analyses.

Uniformity of the length of texts

In order to eliminate a factor of text length that might possibly cause difficulties in interpreting the results, we compared the number of words in each task. The numbers of words in the tasks were: 16, 15, 21, and 17, respectively. Uniform distribution can be assumed; Chi-square = 1.20, $p = .75$.

Normal distribution assumption

Before showing the results of various quantitative statistical methods, the assumption of normal distribution required for several analyses has been investigated. Although t-tests and ANOVA are robust to the violation of the normal distribution assumption (Schmider, Ziegler, Danay, Beyer, & Bühner, 2010), due to the sample size we took care special attention to provide reliable analyses. According to the one-sample Kolmogorov-Smirnov-tests, the assumption of the normal distribution of the quantitative variables being involved in the next analyses can be hold in all but two cases. It should be noted that for four tasks and for five eye-fixation measures it meant in eighteen cases the assumption of the normal distribution could be held. The two exceptional cases were: fixation duration on the text (FDT) and fixation duration on the numbers (FDN) in Task 3 ($p = .03$, and $p = .01$, respectively). For the performance measures the dichotomous nature of that variable prevented us from having normal distribution variables, therefore McNemar-tests will be used.

Descriptive statistics of the basic quantitative measures

The results show that one of the two tasks with number words (Task 4) proved to be more difficult than the other three tasks. The percentages of correct solutions for the four tasks were 46%, 46%, 50%, and 21%, respectively. Mauchly's *W* statistics showed that the variances of the differences in the six possible pair-wise comparisons can be considered equal ($W = .73$, $p = .23$). Consequently, the univariate repeated measures statistics can be used in the analyses without corrections. Table 2 shows the basic descriptive statistics of the task solution process and Table 3 summarizes the results of paired-samples t-tests for each pair of tasks.

Table 2

Descriptive Statistics of the Task Solution Process – Mean Values (SD in Parentheses) of the Main Variables

Task	Perf.	TFD	TFC	FDK	FDT	FDN
1	.46 (.51)	25.01 (13.78)	68.67 (31.83)	1.33 (1.11)	10.56 (4.24)	12.25 (8.79)
2	.46 (.51)	29.72 (18.90)	84.29 (41.22)	1.82 (1.24)	14.91 (6.61)	14.15 (14.60)
3	.50 (.51)	26.83 (19.68)	79.83 (43.62)	1.71 (1.78)	18.89 (13.02)	7.44 (6.83)
4	.21 (.41)	30.70 (14.21)	91.83 (38.77)	2.29 (2.08)	14.40 (7.44)	15.07 (8.60)

Note: (1) Perf. = mean solution rate (i.e., percentage of correct solutions); TFD = total fixation duration; TFC = total fixation count; FDK = fixation duration on the keyword; FDT = fixation duration on the text (other than the keyword); FDN = fixation duration on the numbers. (2) Fixation duration values are given in seconds.

Table 3

Results of the Paired-Samples t-tests (p in Parentheses) for Pair-Wise Task Measure Comparisons

Task pair	TFD	TFC	FDK	FDT	FDN
1 – 2	1.48 (.15)	*2.08 (0.05)*	1.26 (.22)	*3.70 (<.01)*	.81 (.43)
1 – 3	.53 (.60)	1.31 (.20)	.77 (.45)	*3.17 (<.01)*	*3.23 (<.01)*
1 – 4	*2.53 (.02)*	*3.26 (<.01)*	*2.12 (.05)*	*2.90 (<.01)*	*2.26 (.03)*
2 – 3	1.27 (.22)	.60 (.55)	.31 (.76)	1.70 (.10)	*3.79 (<.01)*
2 – 4	.37 (.71)	1.25 (.22)	1.17 (.26)	.57 (.58)	.37 (.71)
3 – 4	1.05 (.31)	1.27 (.22)	1.15 (.26)	1.80 (.09)	*4.43 (<.01)*

Note: (1) Mean = mean solution rate (i.e., percentage of correct solutions); TFD = total fixation duration; TFC = total fixation count; FDK = fixation duration on the keyword; FDT = fixation duration on the text (other than the keyword); FDN = fixation duration on the numbers. (2)Fixation duration values are given in seconds. (3) Significant values are italicized.

Pair-wise comparisons of the performance rates required us to use the Wilcoxon-tests. These comparisons showed that the solution rate on Task 4 is significantly lower than on Task 2 and Task 3 ($Z = 2.12$, $p = .03$; $Z = 2.11$, $p = .04$, respectively).

Hypothesis 1: Connection between performance and fixation duration measures

In hypothesis 1, we proposed that the fixation duration measures (TFD, FDT, FDK, FDN) have an effect on students' performance on the task. Correlations between the five different task measures were computed. There were no significant correlations between performance and FDN or FDK, except for Task 3, where we found a significant negative correlation between performance and FDT on the one hand ($r = -.44$, $p = .03$), and between performance and RT on the other ($r = -.40$, $p =.05$).

Both tasks with number words (Task 2 and 4) required longer fixation duration than tasks with numerals (Task 1 and 3). The difference was significant between Task 1 and Task 4. As for total fixation counts, Task 1 required significantly fewer fixations than Tasks 2 and 4. As for the fixation duration measures, the differences in favor of Task 1 (in the sense of shorter fixation compared to the other tasks) can be in part attributed to the slightly shorter text of the word problem. However, differences in fixation duration on the number components of the word problem indicate that word numbers may require longer time to read than numerals while reading the text and solving the word problem.

Hypotheses 2 and 3: The role of number notation

To test the effects of number notation (Hypotheses 2 and 3), 2x2 repeated measures ANOVAs were computed. The first factor is number notation and the second is the operation (addition vs. subtraction). Table 4 shows the eta-squared effect sizes of the two factors for each dependent variable.

Table 4

Results of Two-Way Repeated Measures ANOVA and Effect Sizes of the Number Notation (Numeral vs. Number Word) and of the Operation (Addition vs. Subtraction) on the Dependent Variables

Dependent variable	Factor in ANOVA repeated measures	F(1,23)	p	Eta-squared (%)
Response time	Number notation	10.507	.004	31.4
	Arithmetic operation	1.943	.177	(7.8)
	Interaction	0.774	.388	(3.3)
FDT	Number notation	0.020	.888	(0.1)
	Arithmetic operation	9.015	.007	29.1
	Interaction	9.190	.006	29.5
FDN	Number notation	18.391	<.001	44.4
	Arithmetic operation	2.726	.112	(10.6)
	Interaction	3.425	.077	(10.3)
FDK	Number notation	3.126	.091	(12.4)
	Arithmetic operation	0.127	.725	(0.6)
	Interaction	2.463	.131	(10.1)

Note: FDT = fixation duration on the text (other than the keyword); FDN = fixation duration on the numbers; FDK = fixation duration on the keyword. Non-significant eta-squared values are shown in parentheses.

Table 4 shows the outstandingly relevant role of number notation in response time and this was due to the effect of notation on fixation duration on numbers. However, no such strong effect of number notation was found for performance, fixation duration on text components and on fixation duration on keyword. Interestingly, the arithmetic operation to be computed had a significant effect on fixation duration on text components, with higher FDT values for subtraction problems.

To further investigate the role of number notation, correlations between fixation duration measures within a task are shown in Table 5.

Table 5

Correlations Between FDT, FDN and FDK Variables Within a Given Task

		FDN (for the given task)	FDK (for the given task)
Task 1	FDT	.742	.800
	FDN	1	.690
Task 2	FDT	.505	.575
	FDN	1	(.068)
Task 3	FDT	.901	.935
	FDN	1	.830
Task 4	FDT	.469	.739
	FDN	1	(.218)

Note: Coefficients in parentheses are not significant at the p = .05 level. FDT = fixation duration on the text (other than the keyword); FDN = fixation duration on the numbers; FDK = fixation duration on the keyword.

The differences between the correlation coefficients are significant in any case when the smaller coefficient is either .218 or .068 and the bigger coefficient is greater than or equal to .690. Table 5 suggests that the connections between fixation duration length variables are of different strengths, revealing the possible role of number modality in fixation duration on different parts of the word problem.

In the cases of Task 2 and Task 4 (where the numbers were given in the number word format) correlations between fixation duration on numbers and fixation duration on keyword proved to be non-significant. Since correlations between other fixation duration variables were significant in all tasks, this peculiarity needs further interpretation. The significant correlations between fixation durations on text, numbers and keyword (FDT, FDN and FDK) indicate that the more time a student spends on a given part of the word problem text, the more time he or she spends on other types of text components. The exception is the correlations between fixation duration time on the number word and fixation duration on the keyword. This can be interpreted as follows: when the number word notation is used, the more time a student spends on the number word, the longer he or she will fixate on text components of the word problem except for on the keyword component. This may furthermore indicate that albeit longer fixation time is needed for the number word component, the keyword component will not require more fixation time. What is curious is the increased fixation duration on the text components when the numbers are given in number word format.

DISCUSSION AND CONCLUSION

This research provided evidence about the role that number notation may play in arithmetic word problem solving. The main novelty of this research was focusing on number notation. Surprisingly few studies have examined the role of number notation in an educational context. One possible reason of the scarcity of such research is that in real classroom settings there is a lot of variation in student characteristics that may interfere with task-related variables. For example, as De Corte et al. (1990) pointed out, research on arithmetic word problem solving should use tasks in which students' reading and computational skills play a negligible role. In our current study, one of these two challenges has been eliminated (i.e., the tasks were well within the computational skills of 10 year old students), and the other factor was an important component of the dependent variables (i.e., fixation duration measures on textual components).

Our results confirmed the hypothesis about the effects of number notation on different fixation duration measures. In accordance with our expectations based on the literature (Rayner, 1998), significantly longer response time and fixation duration on numbers is needed when using number word notation. However, number notation proved to have no important role in performance, fixation duration on text components, and in fixation duration on keyword. Since longer fixation duration on text components and longer reaction time are associated with a lower level of performance (albeit the correlation was significant only for Task 4), but number notation has no significant effect on performance, the longer FDT and RT in the tasks involving number words may point to a "compensation" effect. It means that the longer time needed for the completion of the problem is not accompanied by worse performance when the number element of the word problem is presented in a word format.

In the light of the importance of so-called common practice when choosing number notation in school word problem texts, research should focus on the possible advantages or disadvantages of different number notations. Namely, research should examine the advantages or disadvantages of using number words instead of Arabic numbers in the text of word problems. Of course, it is better to be able to solve arithmetic compare word problems with either Arabic numerals or number words than to be able to solve word problems with only one of these two notations. Since in this experiment the number notation did not play a crucial role in students' performance and in their previous classroom practice students encountered mainly Arabic numerals in word problems, two possible explanations can be considered.

First, the current classroom practice sufficiently develops students' word problem solving skills with either notation. Second, being unusual, the use of number words may alter the task solving process. According to this second explanation, the gains students possibly attain from the unusual number notation may raise awareness in a thinking process that would have otherwise been (over-) automatized. The possibility of modifying the thinking process is supported by the non-significant correlations between fixation duration on numbers and fixation duration on keyword in Task 2 and Task 4 (number word notation). Since there is still a significant correlation between fixation durations on the number word and on the text components, these results might be interpreted as follows. While solving arithmetic compare word problems students slow down their reading on the number words and on the text components in general but not on the keyword of the task. Albeit the keyword plays an important role in deciding which arithmetic operation the student should choose, it seems that these keywords can fulfill their role in tasks with number words in a relatively short time.

As a final conclusion about the practical implication of our results, we encourage textbook writers and teachers to construct and use arithmetic compare word problems with diverse notations of numbers. Further research with bigger samples may provide evidence about how different number notations change the task solving process. Additionally, cross-cultural (cross-linguistic) studies can reveal to what extent performance and number notation are independent of each other.

REFERENCES

Cankaya, O., LeFevre, J., & Sowinski, C. (2012). The influences of different number languages on numeracy learning. *Encyclopedia of Language and Literacy Development* (pp. 1-8). London, ON: Western University.

Cohen, L., Dehaene, S., & Verstichel, P. (1994). Number words and number non-words. A case of deep dyslexia extending to arabic numerals. *Brain, 117*, 267-279. doi:10.1093/brain/117.2.267

Csíkos, C., Kelemen, R., & Verschaffel, L. (2011). Fifth-grade students' approaches to and beliefs of mathematics word problem solving: a large sample Hungarian study. *ZDM – The International Journal on Mathematics Education, 43*, 561-571. doi:10.1007/s11858-011-0308-7

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

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology, 14,* 218-224. doi:10.1016/j.conb.2004.03.008

Dehaene, S., Piazza, M., Pinel P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20,* 487-506. doi: 10.1080/02643290244000239

Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye-fixations. *Journal of Educational Psychology, 84,* 76-84. doi:10.1037/0022-0663.84.1.76

Jitendra, A. K., Griffin, C. C., Deatline-Buchman, A., & Sczesniak, E. (2007). Mathematical word problem solving in third-grade classrooms. *The Journal of Educational Research, 100,* 283-302. doi:10.3200/JOER.100.5.283-302

Mwangi, W. & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction, 16,* 173-199. doi:10.1207/s1532690xci1602_2

Radatz, H. (1983). Untersuchungen zum Lösen einglekeideter Aufgaben. *Zeitschrift für Mathematik-Didaktik, 3/83,* 205-217.

Rayner, K. (1998). Eye-movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422. doi:10.1037/0033-2909.85.3.618

Riley, M. S., & Greeno, J. G. (1998). Developmental analysis of understanding language about quantities and of solving problems. *Cognition and Instruction, 5,* 49-101. doi:10.1207/s1532690xci0501_2

Schmider, E., Ziegler, M., Danay, E., Beyer, E., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against to the violations of the normal distribution assumption. *Methodology, 6,* 147-151. doi: 10.1027/1614-2241/a000016

Stigler, J. W., Fuson, K. C., Ham, M., & Kim, M. S. (1986). An analysis of addition and subtraction word problems in American and Soviet elementary mathematics textbooks. *Cognition and Instruction, 3,* 153-171. doi:10.1207/s1532690xci0303_1

Szöveges feladatok [Word problems] (2015). Retrieved January, 30, 2015, from http://www.ntk.hu/c/document_library/get_file?folderId=279536&name=DLFE-21081.pdf

Szűcs, D., & Csépe, V. (2004). Access to numerical information is dependent on the modality of stimulus presentation in mental addition: a combined ERP and behavioral study. *Cognitive Brain Research, 19,* 10-27. doi:10.1016/j.cogbrainres.2003.11.002

Van der Schoot, M., Bakker Arkema, A. H., Horsley, T. M., & Van Lieshout, E. C. D. M. (2009). The consistency effect depends on markedness in less successful but not successful problem solvers: An eye movement study in primary school children. *Contemporary Educational Psychology, 34,* 58-66. doi: 10.1016/j.cedpsych.2008.07.002

Temple, E., & Posner, M. I. (1998). Brain mechanisms of quantity are similar in 5-year-old children and adults. *Proceedings of the National Academy of Sciences, 95,* 7836-7841. doi:10.1073/pnas.95.13.7836

Verschaffel, L., De Corte, E., & Pauwels, A. (1992). Solving compare problems: An eye movement test of Lewis and Mayer's consistency hypothesis. *Journal of Educational Psychology, 84,* 85-94. doi:10.1037/0022-0663.84.1.85

Verschaffel, L., & De Corte, E. (1997). Word problems: A vehicle for promoting authentic mathematical understanding and problem solving in the primary school? In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 69-97). Hove, UK: Psychology Press.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse, The Netherlands: Swets & Zeitlinger.