This is an uncorrected version of the article. The published version can be found in the *Journal of Educational Psychology* (2013) DOI: 10.1037/a0013843

The Contributions of Working Memory and Executive Functioning to Problem Representation and Solution Generation in Algebraic Word Problems

Abstract

Solving algebraic word problems involves multiple cognitive phases. We used a multi-task approach to examine the extent to which working memory and executive functioning are associated with generating problem models and producing solutions. We tested 255 11-year-olds on working memory (Counting-Recall, Letter-Memory, Keep-Track), ability to inhibit inappropriate responses (inhibition: Numeric-Stroop, Stop-Signal), mental flexibility (switching: Number-Letter, Plus-Minus), English literacy, and algebraic problem solving skills (problem representation, solution generation and other sub-components). Working memory explained about a quarter of the variance in both representation and solution formation. Literacy explained an additional 20% of the variance in representation formation. Ability to discern quantitative relationships explained an additional 10%. Our findings go beyond a demonstration of association between working memory and problem solving accuracy. They show that success in word problems are particularly reliant on ability to decode and assign mathematical operators to quantitative relationships: skills that draw heavily on working memory resources.

Keywords: Working memory, algebra, executive function, reading comprehension, word problems

The Contributions of Working Memory and Executive Functioning to Problem Representation and Solution Generation in Algebraic Word Problems

In Singapore, problem solving is the main focus of the primary or elementary mathematical curriculum (see Figure 1). Algebraic thinking is an essential part of the curriculum and is introduced early in the form of word problems. It serves as a vehicle for teaching pattern recognition, quantitative comparison, and operation reversal. These foundation skills provide a bridge to symbolic algebra, which is taught in the secondary or high school years.

Insert Figure 1 about here

Although Singaporean students performed well in international comparisons of mathematics achievement (Martin, Mullis, Gonzalez, & Chrostowski, 2004; Mullis, Martin, Gonzalez, & Chrostowski, 2004), not unlike other countries, there are significant individual differences. Previous studies have documented strong relationships between individual differences in working memory and performance in arithmetic word problems (Kail & Hall, 1999; Passolunghi & Pazzaglia, 2004; Passolunghi & Siegel, 2001; Swanson, 2006; Swanson, Cooney, & Brock, 1993). Lee, Ng, Ng, and Lim (2004) extended this work to algebra word problems and found a similar relationship.

Algebra word problems are often complex. Generating a solution often involves multiple cognitive components occurring in different phases of the problem solving process. In this study, we focused on an issue regarding the role of working memory that has not been fully addressed: to what extent is working memory involved in different phases of problem solving?

Working Memory and Components of Problem Solving

Being able to read and understand problems are important first steps in problem solving (e.g., Blessing & Ross, 1996; Bobrow, 1968; Cummins, Kintsch, Reusser, & Weimer, 1988; Mayer & Hegarty, 1996; Nathan, 1992; Kintsch & Greeno, 1985; Reusser, 1990). Because working memory is known to predict individual differences in reading comprehension (Daneman & Merikle, 1996), it is possible the correlation between working memory and word problem solving is due largely to the former's role in reading comprehension. Indeed, a number of studies have shown that correlations between working memory and problem solving accuracy are substantially lower when they have been corrected for differences in reading comprehension (Fuchs et al., 2006; Kail & Hall, 1999; Swanson et al., 1993).

Lee et al. (2004) examined this issue by modelling the relationship between working memory, problem solving accuracy, and English literacy. As expected, working memory predicted literacy; literacy also predicted problem solving accuracy. Of interest was that even after controlling for literacy, working memory still contributed independently to problem solving accuracy. This suggests that the contribution of working memory cannot be attributed solely to its relationship with reading comprehension. This finding is consistent with observations showing that children with difficulties in reading do not always perform poorly in mathematics (e.g., Geary, Hoard, & Hamson, 1999; Geary, Hamson, & Hoard, 2000; Hanich, Jordan, Kaplan, & Dick, 2001). Furthermore, differences in working memory profiles amongst children with poor academic performances in English versus mathematics have been documented (Lee & Peh, 2008).

If working memory resources are important for processes other than reading comprehension, what are the other processes that draw on these resources? This is a particularly important question for pedagogy. Just knowing that working memory predicts mathematical problem solving is of limited utility. Given the multi-componential nature of word problems, pedagogical interventions are more likely to succeed if they are targeted at particularly challenging components of problem solving.

What are the components of problem solving? Much of the relevant pedagogical literature was influenced by Polya's (1957) seminal treatise. Polya identified four phases in problem solving. First, problem solvers have to understand the question and identify the kind of information that is needed. Second, they need to ascertain the connections between information that is known versus unknown. This information assists in producing a problem solving heuristic. The third phase involves the execution of the heuristic. Polya identified the final phase as reviewing the heuristic and its outcome for further improvement.

Similar phases are often advocated in cognitive investigations of mathematical problem solving (Briars & Larkin, 1984; Kintsch & Greeno, 1985; Mayer & Hegarty, 1996; Nathan, 1992; Reusser, 1990). Mayer and Hegarty (1996), for example, specified four processing phases: translation, integration, planning, and execution. They argued that one approach to successful problem solving was to understand the question and transform relevant information into an integrated mathematical representation. Such representations, in the form of equations or another kind of schematic, guide subsequent phases in problem solving.

In our study, we examined the contributions of working memory to different phases of algebraic problem solving by first contextualising the phases to the local curriculum. In Singapore, children are introduced to start-unknown problems involving algebraic thinking in the fifth grade. The example below is typical.

John and Peter have $298.

John has $92 more than Peter.

How much money does Peter have?

The problem is start-unknown and algebraic in that the problem solving process starts with unknown quantities (Bednarz & Janvier, 1996): We know neither the amount owned by John nor Peter. Children in primary schools are taught to solve such problems with a variety of heuristics that do not involve symbolic algebra; the most popular of which is "the model method". This heuristic offers a systematic procedure for depicting quantitative relationships amongst protagonists. Relevant information is depicted in a schematic format (see Ng & Lee, 2005; Ng, Lee, Ang, & Khng, 2006 for detailed descriptions of the heuristic). In Figure 2, for example, two bars are drawn to depict the amount of money owned by John and Peter. A longer bar was drawn for John to depict his greater wealth. Note that all available quantitative specifications are depicted in the model.

Insert Figure 2 about here

Once a schematic representation is constructed, children need to derive and execute a set of mathematical procedures that lead them to the solution. In primary schools, children are taught to use arithmetic procedures such as unwinding -- starting from known quantities, reversing relevant mathematical operations until the unknown is found -- to get from the schematic to the solution (Ng & Lee, 2005).

Although the model method does not require knowledge of symbolic algebra, it requires algebraic thinking or an ability to understand, depict, and manipulate complex quantitative relationships amongst multiple protagonists. We investigated the contributions of working memory to these processes by dividing the problem solving process in two and devising two tasks to measure relevant competencies. The "Representation Formation" task provided a measure of children's competency in the initial phases of problem solving. Specifically, it measured children's understanding of word problems and their ability to construct schematic representations. The "Solution Formation" task targeted children's ability to derive and execute procedures that lead them from schematic to solution. An "overall accuracy" test measured children's ability to answer a set of algebraic questions from start-to-finish.

In addition to the Representation and Solution Formation tasks, we devised several componential tasks that targeted different aspects of algebraic problem solving. These tasks were derived from areas of known difficulties, some were derived from Mayer and Hegarty's (1996) framework; others were based on those used in Sakamoto (1998). The "Question Understanding" task assessed children's ability to identify information that is known versus unknown. It tested children's knowledge of what they were asked to find in a problem. The "Relational" tasks measured children's ability to understand the quantitative relationships specified in questions (e.g., more than, less than, and as many times as). In one Relational task, children were presented with inconsistent relational statements for which the required mathematical operator was the opposite to that usually required, e.g., addition when a "less than" relationship is specified. Previous studies show this to be a particular area of difficulty (e.g., Lewis & Mayer, 1987; Pape, 2003; Verschaffel, De Corte, & Pauwels, 1992). A second Relational task tested children's ability to identify quantitative relationships. We also included a computation task, which contained arithmetic questions stated in a purely numeric format. It tested children's abilities to solve questions involving the four operators, and the use of fractions.

Executive Functioning

In addition to working memory, we explored the roles played by other executive functions. Switching and inhibition are commonly identified executive functions in the literature. Switching can be viewed as being synonymous with mental flexibility and refers to the ability to move between alternative sets of mental operations. Inhibition refers to the ability to resist interference from competing responses or processes (Nigg, 2000; Friedman & Miyake, 2004). A third type of commonly encountered function, updating, is closely related to working memory capacity (Miyake et al., 2000; St Clair-Thompson & Gathercole, 2006). It refers to the ability or capacity to refresh and maintain information in working memory. In this study, we used both traditional complex span and updating measures to index working memory.

Findings from both word problem studies (Passolunghi, Cornoldi, & De Liberto, 1999; Passolunghi & Siegel, 2001) and arithmetic studies (Bull, Johnston, & Roy, 1999; Bull & Scerif, 2001; Clair-Thompson & Gathercole, 2006) suggest inhibitory abilities may contribute to arithmetic problem solving. However, a recent study by Swanson (2006) found inhibition predicted arithmetic calculation but not problem solving performance in the early elementary years. Furthermore, the contribution of inhibition may be task dependent. Questions containing irrelevant information (as used in Passolunghi & Pazzaglia, 2004) will likely place additional demands on stimuli selection and inhibition. In Singapore at least, this type of problem is not commonly used. Instead, teachers tend to focus on teaching the fundamentals using questions in which all the given information is relevant. In this context, would inhibition still play a role?

Both inhibition and switching may be important for the representation of problems (Mayer & Hegarty, 1996). In this phase, information extracted from a word problem is integrated with pre-existing knowledge to form a conceptual problem model (Kintsch, 1998). For questions containing inconsistent relational statements, Pape (2003) argued that children need to be able to invert and to understand "x more than y is z" is semantically the same as "x fewer than z is y" (p. 400). Such abilities to invert and consider the truth value of different propositions are likely to be related to mental flexibility, or the ability to switch between different mental representations. In this study, we used the Relational tasks to gauge children's abilities to consider such relationships.

Research Questions and Hypotheses

We examined the relationship between individual differences in higher cognitive capabilities and problem solving by administering to children a battery of working memory, executive functioning, and algebraic measures. We expected working memory to predict performances in the overall accuracy task and in both the Representation and Solution Formation tasks. A question of interest was whether working memory would account for individual differences in overall accuracy above and beyond that explained by the Representation and Solution Formation tasks. Because the two tasks divide the typical word problem in two, it is possible the contribution of working memory will become negligible after its contributions to these tasks are taken into account. Alternatively, working memory might contribute to the ability to integrate or perform the two tasks in an iterative fashion (Kintsch, 1998; Koedinger & Nathan, 2004).

When compared with a measure of general literacy, the Question Understanding and Relational tasks allowed us to test whether comprehending algebraic questions is simply a matter of possessing general literacy skills or whether skills specific to the comprehension of algebraic questions are needed (Kintsch, 1991). We also examined whether any of the componential tasks were particularly reliant on working memory resources.

Regarding executive functioning, we examined whether inhibitory abilities were related to problem solving accuracy. We made no specific hypotheses regarding whether inhibitory abilities or mental flexibility were more important for the componential tasks. Much of the literature on the relationship between inhibition and switching specifies a functional dependency between the two (e.g., Miyake et al., 2000). Here we tested whether inhibition and switching predicted performances on the relational tasks.

Method

Participants

255 Primary 5 children (132 boys, M_{age} of 11.2 years, SD = .36) were recruited from public schools located in the western zone of Singapore. Children participated with parental consent. Schools that participated in this study were located in middle to lower middle class areas. We conducted a power analysis using *Power and Precision* (Borenstein et al., 2001) with parameters based on findings from Lee et al. (2004). We expected 40% (minimum) of the total variance in algebraic performance to be explained by ten predictors in the regression model. With Type 1 error set at .05, the present sample size has 85% power in detecting a set of three predictors that explains at least 3% of the variance in algebraic performance.

In Singapore, Primary 5 classes were streamed according to pupils' performances in English, "mother-tongue", and mathematics. The sample included children from the middle stream, which contained the most variability in mathematical performance. Apart from mother-tongue classes, all other lessons were conducted in English. By Primary 5, pupils could be considered functionally bilingual. For about half of the children, the "mother-tongue" was effectively a second language because English was spoken at home. The remainder spoke Mandarin Chinese, Malay, or Tamil at home. We administered measures of vocabulary and reading comprehension as indices of literacy.

Materials and Procedure

Children were tested as part of a larger study on executive functioning and were administered a battery of working memory, executive functioning, English reading comprehension, vocabulary, performance IQ, arithmetic, and algebraic tasks. The tasks were divided into 2 sets and were administered over two sessions. Depending on school schedule and children's availability, the sessions were separated by 2 to 7 days. The paper-and-pencil executive functioning tasks, the vocabulary and block design subtests from the Wechsler Intelligence Scale for Children (WISC, Wechsler, 1991), and the reading comprehension tasks (from the Wechsler Objective Reading and Language Dimensions, WORLD, Rust, 2000) each took approximately 15 minutes to administer. Executive functioning tasks that were computerised took about 15 to 20 minutes per task.

Algebraic tasks. The overall accuracy test was modified from Lee et al. (2004) and contained ten algebraic, start-unknown questions drawn from the Primary 4 to Primary 6 curriculum. Children were asked to use the model method to solve these questions. Because the curriculum was highly standardised across all schools in Singapore, this method was taught in all schools and was the most commonly used heuristic for questions included in the task (Ng & Lee, 2005). Requiring children to use the same heuristic reduced differences that may be caused by the degree to which different heuristics are reliant on working memory or executive resources (Lee et al., 2007).

To provide comprehensive coverage of the types of questions encountered in the curriculum, a range of questions utilising different quantitative concepts and involving different numbers of protagonists were used. Four questions utilised the terms "more than" or "less than"; half of these involved two protagonists, the remainder involved three. In four other questions, we utilised multiplicative relationships; again involving either two or three protagonists. The remaining two questions were more complex and required two sets of inter-related schematics to depict quantitative relationships amongst protagonists (a full list of all algebraic tasks and their classification can be found in the supplemental section of the journal website). Responses were coded as either right or wrong (range: 0 to 10, K-R 20 = .87).

To obtain separate estimates of children's abilities to (a) translate and create explicit schematic representations for word problems and (b) plan, execute, and compute a solution from such schematics, we devised the Representation and Solution Formation tasks. The Representation Formation task is, in essence, the initial steps in solving a word problem. Children were asked to use the model method to depict information in five algebraic (start-unknown) questions. The questions mirrored the structures of those used in the overall accuracy task. They provided a measure of children's abilities to integrate information regarding known and unknown, and to depict them in a schematic representation. To reduce inter-correlation and to attenuate potential difficulties associated with question understanding, information regarding known states and problem goals were provided in the same format as that used in the Question Understanding task (see below). Schematics were coded as right or wrong (range: 0 to 5, K-R 20 = .76).

In the Solution Formation task, children were asked to construct step-by-step procedures to solve a number of algebraic problems. We used five word problems that were structurally identical to those used in the overall accuracy and Representation Formation tasks. To attenuate difficulties associated with question understanding and problem representation, we provided information regarding known states and problem goals, and a correctly drawn schematic. One mark was awarded for each correct solution (range: 0 to 5, K-R 20 = .84).

Four componential tasks targeting procedures deemed important for algebraic problem solving were also constructed. In the Question Understanding task, children were presented with two arithmetic (end-unknown) practice problems. This was followed by three arithmetic (end-unknown) and five algebraic (start-unknown) word problems. The mathematical structures of these questions mirror those used in the overall accuracy test and utilised "more than", "less than", and multiplicative relationships. Half of the questions involved two protagonists; the remainder involved three. Each question had two parts. Children were asked to identify (a) the known and unknown quantities, and (b) the problem goal. No computation was required. One mark was awarded for each correct response (range: 0 to 16, K-R 20 = .80).

The Relational task had two sections. The first section (Relational A) used simple arithmetic word problems to assess children's understanding and utilisation of relational information, i.e., "more than", "less than", "as many times as". Children were presented with twelve simple one-step arithmetic word problems and were asked to compute the solutions. Questions were phrased in a consistent or inconsistent manner. For inconsistent questions, stated relational terms are opposite to the required mathematical operator, e.g., "less than" when addition is required. Because we were interested primarily in the mathematical expressions used to generate the solutions, children were asked to show their workings. A response was deemed as correct as long as the correct mathematical operator was used (range: 0 to 12, Cronbach's α = .78).

The second section (Relational B) focused on children's understanding of relational information in the context of algebraic problems. The ten questions contained two or three protagonists and specified "more than", "less than", and multiplicative relationships. The mathematical structure of these questions mirrored those used in the overall accuracy test. To discourage calculation, children were not provided with sufficient information to compute the answers. Children were presented with two response alternatives and were asked to identify the quantitative relationship specified in each question (range: 0 to 10, Spearman-Brown = .64).

The computation task contained 20 arithmetic questions depicted in mathematical expressions. Questions required the use of all four arithmetic operators and contained both whole numbers and fractions. One mark was awarded for each correct solution (range = 0 to 20, K-R 20 = .66).

Literacy. The Wechsler Objective Reading and Language Dimensions (Rust, 2000) is a locally normed instrument designed to measure reading and listening skills. Because we were primarily concerned with children's ability to decode written information, only the reading comprehension subtest was used. This test consisted of a series of short, printed passages and orally presented questions. Children were asked to respond orally to each question. Responses were scored on a three point scale. The dependent measure was a standardised score that reflected both the quality and quantity of correct responses (range = 0 to 38). We also administered the vocabulary subtest from the Wechsler Intelligence Scale for Children – Third Edition (Wechsler, 1991).

Working memory and executive functioning tasks. Although definitions of updating, switching, and inhibition are relatively consistent across studies, there is variation in the way they are operationalised. In this study, we adopted the classification used by Miyake et al. (2000). Tasks such as Keep Track and Letter Memory were defined as simple or purer measures of updating and were deemed less likely to impose demands on other executive functions. With the exception of a complex span task, we used simple tasks to index executive functioning. On account of the participants' age and the logistic demands imposed by the algebraic tasks, we used a subset of the nine tasks used by Miyake et al. (2000). We selected tasks that were more age appropriate and had higher correlations with the constructs being measured. Apart from minor changes to render the tasks simpler for children and the use of a numeric version of the Stroop task, task procedures were kept similar to those used in Miyake et al. (2000).

All simple tasks were computerised and administered using E-Prime (Schneider, Eschman, & Zuccolotto, 2002). We used estimates from two updating tasks -- Keep Track and Letter Memory -- together with data from the Counting Recall task to estimate working memory capacity. We used two simple tasks to index inhibition (Numeric Stroop and Stop Signal) and switching (Number-Letter, Plus-Minus). The tasks are described below.

Keep Track was used to measure updating capacity; children were administered word lists and were asked to remember the last exemplars from specified semantic categories. In each trial, the names of two to four semantic categories were presented. Fifteen words, including 3 exemplars from each of five possible categories (countries, names, animals, sports, and occupations), were presented serially and in random order for 1500 ms each. The target semantic categories remained on screen at all times. Children were asked to remember the last word presented in each category and to write down these words at the end of the trial. Prior to the experimental trials, children were given five minutes to study the exemplars in each category to ensure they knew the category to which each exemplar belonged. In addition, children were given three practice trials that contained 7 or 15 exemplars from one or two categories. In the first block, children performed three experimental trials with two target categories. The second block contained three trials with three target categories; the third block had three trials with four target categories. The total number of words that were recalled accurately served as the dependent measure (range: 0 to 27, Cronbach's α = .60)

The Letter Memory task was also included as an updating measure; children were administered lists of letters -- presented serially for 2000 ms per letter -- and were asked to recall the last 4 letters on each list. The number of letters presented in each trial (5, 7, 9, or 11) varied randomly across trials. This information was not disclosed to children to increase the likelihood that they continued updating till the end of each trial. In total, 12 experimental trials were administered. Prior to the experimental trials, two sets of practice trials were administered, each consisting of 5 trials. In the first set, children were asked to recall the last 2 letters presented in the lists and were told the number of letters that would be shown in the first three trials: 2, 4 or 5. In the remaining trials, the number of letters to be shown was not revealed. In the second set, children were asked to recall the last 3 letters in the list. Similar to the first set, children were told only the number of letters to be shown in the first 3 trials: 3, 5, or 6. The dependent measure was the proportion of letters recalled correctly within each trial, summed across trials (Cronbach's α = .66).

For Counting Recall, we used a standardised version from the Working Memory Test Battery for Children (Pickering & Gathercole, 2001). Children were presented with a series of dot arrays that contained 4, 5, 6, or 7 dots. They were asked to count the dots and when all the arrays in a given trial had been presented, to recall the number of dots on each card in the order they were presented. Three practice trials were presented to determine each child's starting span. The dependent measure was the number of trials correctly recalled (range 11 - 37). Gathercole, Pickering, Ambridge, and Wearing (2004) reported test-retest reliability of .61.

The Numeric Stroop and Stop Signal tasks were used to index inhibition. In Miyake et al. (2000), the letter-colour Stoop was used as an index of inhibitory ability. The numeric version was used in this study because the numeric but not the letter-colour version was previously found to predict arithmetic performance (Bull & Scerif, 2001). Although the antisaccade task was found to have the highest factor loading in Miyake et al., it was not used here because measures of cognitive or behavioural inhibition were deemed to be better predictors of algebraic performance than were measures of oculomotor inhibition.

In the Numeric Stroop task, each child was administered three conditions. The first condition used non-numeric stimuli (e.g., @ @) and provided a baseline measure. Arabic numerals were used in both the congruent and incongruent trials. For the former, the number of numerals corresponded with their numeric value (e.g., 4 4 4 4, 2 2). For incongruent trials, the quantity of numerals and their numeric values did not correspond (e.g., 2 2 2). Only 1, 2, 3, and 4 were used to ensure children could subitize and did not have to count. We administered 12 practice trials, followed by 4 blocks of 24 experimental trials. In each experimental block, an equal number of trials from each of the 3 conditions were presented in random order. The dependent measure was a reaction time (RT) score (estimated reliability = .47), obtained by subtracting the mean RT for the baseline condition (Cronbach's α= .86) from that of the incongruent condition (Cronbach's α = .84).

The Stop signal task consisted of 3 blocks. The first two blocks, 48 trials each, were designed to build up a prepotent categorisation response. Unlike Miyake et al. (2000), who used one block of 48 trials, an additional block was used to give children additional experience and to strengthen the categorisation response. In each trial, children were presented with one word and were asked to determine whether it should be categorised as animal or non-animal. All words were balanced for length and word frequency. Instructions emphasised both speed and accuracy. In the third block (192 trials), children were asked not to respond when they heard a computer-emitted tone but otherwise to keep performing the same categorisation task as before. These tones appeared on 48 randomly selected trials. The time at which the stop signal occurred was adjusted for each child by taking their mean RT from the second block and subtracting 225 ms. For all trials, children were given up to 1500 ms to respond by button press. Prior to the experimental trials in the first block, children were given 25 non-signal practice trials and 25 signal practice trials. The dependent measure was the proportion of trials in which a stop signal was sounded, but for which the child failed to inhibit a categorization response (K-R 20 = .92).

The Number-letter was used by Miyake et al. (2000) to measure switching abilities. A number-letter pair (e.g., G3, K8) was presented in one of four corners on the computer screen. When the number-letter pair appeared at the top of the screen, children were asked to indicate whether the number was odd or even. When it appeared at the bottom of the screen, they were instructed to indicate whether it contained a consonant or vowel. There were three blocks of trials; the first two blocks contained 32 trials each and did not require task switching. They involved odd/even or vowel/consonant judgement only. In the third block (128 trials) -- divided into four sets -- the number-letter pairs were presented in a clockwise direction. On alternate trials, children had to switch between the two types of categorization responses. Ten practice trials were included in each of the pure blocks; 20 practice trials were added in mixed block. Children responded by pressing one of two response keys using their dominant hand. Local switch cost was computed from the third block by subtracting the average RT of the switch from the non-switch trials (estimated reliability = .63).

In the Plus-minus task, a series of two digit numbers (printed in red or blue) were presented on the computer screen. When red digits were shown, children were instructed to add 2 to that number and to type their answers on a numeric keypad. When blue digits were presented, children were asked to subtract 2. In Miyake et al.'s (2000) study, participants added or subtracted 3; we used 2 on account of the children's age. The task consisted of 4 practice trials followed by 3 blocks of 30 experimental trials. In the first block, all the numbers were red. In the second block, all the numbers were blue. In the third block, a mix of red and blue numbers was shown in a randomised order. Children were instructed to complete the task quickly and accurately. RT based local switch costs were computed as per the Number-Letter task. Because there was only one mixed block, we could not generate a reliability estimate for the difference score. Averaged Cronbach's α for trials contributing to local switch cost is .76.

Results

Data Screening

For RT based tasks, data were screened for extreme scores on two levels. Initial screening involved removal of trials that exceeded RT thresholds. These were determined by inspection of overall RT distributions for each task. RTs outside the following bands were discarded: < 400 ms or > 2799 ms for the Stroop task, < 500 ms or > 6999 ms for Number-Letter, and > 9000 ms for Plus-Minus. After a mean RT has been computed for each participant for each task, data from the whole sample went through a second screening. Individual RT with values exceeding the sample mean by three *SD* was replaced with a value at three *SD*. Non-RT based measures only went through level two screening: Scores that were more than three *SD* beyond their respective means were replaced by values at three *SD*.

With the exception of Question Understanding and computation, the algebraic measures exhibited only mild departure from normality (skewness < -.86, see Table 1 for details). For Question Understanding, we applied an inverse transformation to reduce skewness. For computation, we used a logarithmic transformation. For ease of interpretation, we presented all means and correlation ratio using the non-transformed variables (see Table 1. To correct for multiple comparisons, all correlation ratios were evaluated at the Bonferroni adjusted level of .0005).

Insert Table 1 about here

Below, we report findings from the various algebraic tasks. The main focus is to ascertain the degree to which performance on the overall accuracy task is related to competencies on the componential tasks. This is followed by analyses of the relationships between working memory, executive functioning, and algebraic performance. All regression and canonical models were All models evaluated at the Bonferroni adjusted level of .002.

Algebraic tasks

To examine the extent to which performances on the Representation and Solution Formation tasks predicted performances in the overall accuracy task, we conducted a standard regression analysis (A summary of all regression analyses can be found in Table 2). The two measures accounted for 76% of variance in overall accuracy, $F(2, 254) = 390.97$, $p < .001$. Both tasks provided unique contributions ($sr^2_{\text{Representation Formation}} = .08$, $sr^2_{\text{Solution Formation}} = 10$). These findings show that the two tasks are successful in capturing key processes in algebraic problem solving.

Insert Table 2 about here

Children did very well on the Question Understanding task with performances approaching ceiling. Performances on the arithmetic versus algebraic questions did not differ, with both at 93%. Because the linguistic and mathematical structures of the translation task were the same as the overall accuracy task, this finding suggests many children were able to identify the base elements in the questions: what is known versus unknown and what they were asked to find.

Children's performances were more variable on Relational A and B. As expected, performances were more accurate on the direct than on the indirect questions, $t(254) = 12.82$, $p < .001$. Although the computational task contained arithmetic items with larger operands than were typically used in the curriculum, the children had little difficulty with most of the items. Performances were at or near ceiling for several operations (all of which have a range of 0 to 3): addition ($M = 2.76$, $SD = .48$), subtraction ($M = 2.68$, $SD = .57$), and multiplication ($M = 2.80$, $SD = .42$). Children found division (range: 0 – 4, $M = 2.56$, $SD = .92$) and operations involving fractions more difficult (range: 0 - 7, $M = 5.99$, $SD = 1.32$).

All four componential scores correlated reliably with the overall accuracy score (see Table 1). Together, they predicted 63% its variance, $F(4, 250) = 107.36$, $p < .001$. Although much of the explanatory power was shared, both Relational A and B provided unique contributions to the model, $sr^2 = .08$ and .09 respectively. Neither Question Understanding nor computation provided unique contributions ($sr^2 = .006$ and .008, respectively).

The componential scores were also correlated reliably with performances on Representation and Solution Formation. Because there were multiple componential tasks and the Representation and Solution Formation tasks were correlated, we conducted a canonical correlation analysis to examine their interrelations. Canonical analysis is similar to multiple regression with the exception that there are multiple predicted as well as criterion variables. Similar to multiple regression, values on each variable are weighted. But rather than applying weights that minimise observed and predicted variables, weights that maximise correlations amongst the predictor and criterion sets are used (Thompson, 1984).

The canonical analysis produced two canonical correlations. The first correlation showed that the two types of scores were strongly correlated, r_c = .827 (68% overlapping variance), χ^2 (8) = 293.97, p < .001. The second correlation was not reliable and was not interpreted. Results from a redundancy analysis showed that the componential scores accounted for 61% of variance in the variate containing Representation and Solution Formation. To examine the stability of these estimates, we conducted a series of sensitivity analyses by deleting consecutively one variable from each variate. The findings show reasonable stability with small variations in loadings and coefficients across analyses (see Table 3 for canonical loadings, coefficients, and results of the stability analysis).

Insert Table 3 about here

Working memory and algebraic performance

Overall accuracy. Performance on the overall accuracy task was correlated highly with both Counting Recall and the other working memory/updating measures (see Table 1). A standard regression showed that the three measures accounted for 26% of variation in the overall accuracy measure, $F(3, 251) = 29.55$, p < .001. The next set of analyses examined the relationships between the various measures of algebraic competencies and working memory.

Representation and Solution Formation. The Representation and Solution Formation tasks contain, in essence, typical algebraic problems that have been truncated and presented separately in two parts. For this reason, we expected competencies on these two tasks to mirror closely performance on the overall accuracy task. Indeed, we found these tasks accounted for 76% of variance in the overall accuracy task. Although there is no reason to suspect the overall accuracy task requires any more mathematical knowledge than do the Representation and Solution Formation tasks, performing processes required for both Representation and Solution Formation may require additional executive resources. To test this hypothesis, we performed a hierarchical regression. Overall accuracy was regressed to performance scores on the Representation and Solution Formation tasks (entered as a single block in the first step) and the working memory measures (entered in the second step). Working memory failed to account for additional variance in the overall accuracy score, $\Delta R^2 = .003$.

Next, we examined whether working memory contributed to performances in the Representation Formation and Solution Formation tasks. Results from a separate standard regression showed that working memory accounted for a large amount of variance in both tasks, $R^2 = .23$ and $.27$ respectively. Because these values may be affected by correlation amongst Representation and Solution Formation, we conducted a canonical correlation between these measures and the measures from the three working memory tasks. Two canonical correlations were found. The first canonical correlation had a value of .54: indicating a 29% overlap in the variance amongst the two canonical variates (one computed from the working memory measures and the other from the Representation and Solution Formation measures), $\chi^2(6) = 87.06$, $p < .001$. The second canonical correlation was not reliable. A redundancy analysis showed that 25% of variance in the variate containing Representation and Solution Formation was explained by the working memory measures. Cross-loadings show that Representation and Solution Formation have similar correlations with working memory, .48 and .52 respectively. These findings show that though working memory does not account for additional variance in overall accuracy beyond that of Representation and Solution Formation, it predicts reliably performance on these constituent tasks.

Componential tasks. Because the componential tasks accounted for large proportions of variance in the overall accuracy as well as in the Representational and Solution Formation tasks, we conducted a canonical correlation on the working memory and componential measures. Three canonical correlations were produced. The first correlation showed a close relationship between the two variates, $r_c = .58$ (34% overlapping variance), $\chi^2(12) = 104.36$, $p < .001$. The remaining canonical correlations were not reliable and were close to zero. Low scores on the working memory measures were associated with low scores on the componential tasks. A redundancy analysis showed that 19% of variance on the componential tasks was explained by the working memory measures. Cross loadings showed that performance on Relational A ($r = -.51$) was more closely correlated with working memory than were Relational B ($r = -.42$), Question Understanding ($r = -.39$), or Computation ($r = .42$, the value is positive due to logarithmic transformation). Results from both the stability analysis and analyses involving the transformed variables showed similar patterns.

Working Memory, Literacy, and Representational Formation

Reading and understanding the question are vital components of the problem solving process. The next analyses examined the extent to which Representation Formation is related to working memory, literacy, and two of the componential measures that indexed children's understanding of the questions. Based on findings from Lee et al. (2004), we treated working memory as an upstream capability that affects performances in domains more heavily influenced by experience and schooling. For this reason, we hierarchically regressed Representation Formation to the working memory scores first, followed by the two literacy measures in the second block, Question Understanding in the third block, and Relational B last. As reported earlier, working memory scores explained 23% of variation in Representation Formation, $F(3, 251) = 25.03$, $p < .001$. Addition of the literacy measures explained an additional 22% of the variance, $F_{change}(2, 249) = 50.23$, $p < .001$. Because children are unlikely to proceed to considerations regarding quantitative relationship unless they can understand basic specifications such as what they are asked to find, we entered scores from Question Understanding followed by those from Relational B. The former accounted for an additional 3% in variation, $F_{change}(1, 248) = 15.17$, $p < .001$. Relational B accounted for an additional 9%, $F_{change}(1, 247) = 51.80$, $p < .001$. These findings suggest there is more to understanding word problems than general literacy.

To examine the specificity of working memory to performance on algebraic measures, we conducted a parallel analysis in which the literacy scores were entered prior first. The literacy measures explained 39% of variance, F (2, 252) = 80.31, $p < .001$. The working memory measures explained an additional 6%, $F_{change}(3, 249) = 9.42$, $p < .001$. Of interest was that Keep Track failed to provide unique contribution when literacy was added to the model. This suggests dependency between the two measures. One possibility is that participants with better vocabulary may have found it easier to classify words into semantic categories: a component of the Keep Track task.

Working Memory, Computation, and Solution Formation

Computing a solution from a problem representation requires planning and arithmetic skills. Prior to computation, participants need to decide on the arithmetic operators required and the sequence in which they are to be used. For more sophisticated problem solvers, this involves the construction of equations. For typical Primary school children, this involves constructing a series of arithmetic expressions, which are then solved by unwinding: working backwards from known quantities, and undoing or reversing arithmetic operations (Ng & Lee, 2005). We used performance on Relational A to gauge participants' ability to select the correct arithmetic operators in both straight-forward and difficult word problems. Computation provided a measure of arithmetic competence.

We used a hierarchical regression to examine the extent to which Solution Formation is reliant on working memory resources and computation skills. The Solution Formation score was regressed to the working memory measures in the first block, followed by Relational A and computation in separate blocks. Working memory alone explained 27% of variance in Solution Formation, $F(3, 251) = 30.85$, $p < .001$. Relational A explained an additional 28%, $F_{change}(1, 250) = 158.17$, $p < .001$. The computation score failed to provide significant improvement, $\Delta R^2 = .6\%$, $F_{change}(1, 249) = 3.44$, $p = .07$.

The last analysis took into account the relationship between Representation and Solution Formation. Although we had separate measures for children's competency in these areas, one could argue that a child who was unable to understand information depicted in a word problem was unlikely to proceed to solution formation. We addressed this issue by regressing Solution Formation to Representation Formation. Representation Formation explained 57% of variance in Solution Formation. From this model, we computed a residual score that captured variance in Solution Formation not explained by Representation Formation. The residual score was then regressed to the working memory and componential measures. Regarding the componential measures, Relational A was related to Question Understanding and Relational B. Although Relational A focuses on children's abilities to select an appropriate mathematical operator to represent a stated quantitative relationship, they first have to understand and identify the nature of this relationship. This latter is measured by Relational B. We addressed these inter-relations with a hierarchical regression model: the Solution Formation residual score was regressed to working memory, Question Understanding, Relational B, Relational A, and Computation with each componential score entered in a separate block.

Keep Track, Question Understanding, and the two Relation measures were reliably correlated with the residual score (rs > .23, ps < .0005). The working memory measures explained 7% of variance in the residual score, $F(3, 251) = 5.87$, $p = .001$. Question Understanding, explained another 6% of variance, $F_{change}(1, 250) = 16.23$, $p < .01$. Of the remaining predictors, only Relational A provided marginally reliable explanatory power, $\Delta R^2 = .03$, F_{change} (1, 248) = 9.04, $p = .003$). Much of the explanatory power of the working memory measures was captured by their correlation with the componential measures: they no longer provided unique contributions when the componential measures were entered into the model.

Inhibition and switching

The bivariate correlations showed that neither the inhibitory nor switch measures were reliably correlated with either the working memory or any of the algebraic measures.

Discussion

Not surprisingly, working memory was predictive of performance on the overall accuracy test and accounted for a quarter of all variation. This finding is consistent with previous studies that focused on mathematical word problem solving (e.g., Lee, Ng, Ng, & Lim, 2004; Passolunghi & Pazzaglia, 2004; Swanson, 2006). One of our main concerns was the extent to which working memory was involved in different components of problem solving. Data from the Representation and Solution Formation tasks showed that performances on both tasks were similarly related to the working memory scores, which accounted for approximately a quarter of the variation in both tasks.

Canonical analyses of the componential tasks suggest that Relational A, which indexes children's ability to utilise relational information by selecting the correct mathematical operation, is more strongly associated with working memory capacity. Nonetheless, cross loadings between working memory and the other componential measures are also substantial ($r > .39$).

These findings show that working memory contributed to both (a) the text decoding and schematic construction components (Representation Formation) and (b) the solution and computation components (Solution Formation) of algebraic problem solving. A potential critique of this conclusion is that the representation and solution components may be iterative or inter-related. When participants are given a word problem, they may construct a schematic representation, test-run some problem solving procedures, and re-assess the schematic when they encounter a problem. If this is the case, estimating the relationship between working memory and the two components separately may not capture the full relationship.

We addressed this issue in a series of follow-up analyses. The first of which showed that working memory did not explain variance in overall accuracy above and beyond that explained by

Representation and Solution Formation. Such variance should capture additional processes that are not required when Representation and Solution Formation are engaged on their own. For this reason, the finding suggests that even if children did engage representational and solution generation processes in an iterative fashion, it did not impose additional working memory demands. This was surprising as having to perform the whole task should impose additional demands.

A plausible explanation is that even in the overall accuracy task, children were asked to draw schematic representations. Perhaps having to do so provided an interim goal not unlike explicitly truncating the task in two. Another explanation is that in proposing an iterative process, the assumption is that the children formulate and deliberate on interim products as they go along. However, it is possible the process is more proceduralised than we thought and the children's approach is more linear.

A related concern is whether the Representation and Solution Formation tasks (and, indeed, the other algebraic tasks) really do reflect competencies in different components of algebraic problem solving. Although the various tasks are strongly correlated, there is sufficient divergence amongst the various measures to suggest they index different aspects of problem solving. Representation and Solution Formation, for example, both provided unique contributions (8% to 10%) in explaining variance in overall accuracy. The two relational measures -- A and B -- also provided unique contributions (8 to 9%) in explaining overall accuracy.

Summary. These findings provide a finer grade analysis of the role of working memory in word problem solving. Consistent with previous findings on its role in reading comprehension (Daneman & Merikle, 1996), the findings show that it plays an important role in text decoding and in constructing a schematic representation of the problem. In this study, we asked all participants to use the model method and effectively demanded the construction of a schematic. In a wider context, not all problem solvers would bother with a schematic representation. Some will go straight into processes associated with Solution Formation (Kintsch & Greeno, 1985). Our data suggest that processes associated with both representational and solution generation are strongly correlated with working memory capacity. Although much of this demand seems to be placed on processes shared amongst Representation and Solution Formation, working memory also explained 7% of variance that was unique to Solution Formation.

Looking deeper: Relationships with the Componential Tasks

Representation Formation. An aim of this study was to investigate the relationship between working memory, literacy, and performance on word problems. Kintsch (1991) argued that learning to solve word problems "involves learning to use ordinary language in a special way". This suggests that ability to comprehend words and sentences, as used in normal discourse, may not be sufficient for deciphering word problems. To investigate this hypothesis, we used performance on the Representation Formation task as a more proximal index of processes involved in understanding and constructing meaningful schematic representations. Not surprisingly, general literacy explained a large proportion of variance (22% above that predicted by working memory alone). Of interest was that indices of children's understanding of basic elements in the questions, and of the quantitative relationships amongst protagonists also explained additional variance (12%).

Although phrases that are used for specifying quantitative relationships -- e.g., "more than", "less than" -- are commonly found in common discourse, our findings suggest that their comprehension draws on additional capabilities. These findings are consistent with Kintsch's (1991) explanation, which posits that words used in algebraic questions are often used in more restricted and specific manners than they are in daily discourse.

Solution Formation. Although Solution Formation involved computation, our analysis shows that little of its variance is explained by the computation measure. Our initial analyses showed that working memory and children's ability to select the appropriate mathematical operator (as indexed by Relational A) was more predictive of performance in Solution Formation than was computation. However, we were concerned that these findings may not be specific to processes unique to the generation of solutions. To what extent is working memory predicting processes shared between Solution and Representation Formation? Furthermore, to what extent is children's ability to select the appropriate mathematical operator affected by their understanding of the underlying quantitative relations?

To address these issues, we calculated a residual score that better indexed processes unique to the generation of solution. We also used hierarchical regression to separate children's ability to select the appropriate mathematical operator, from their understanding of the underlying relation. In addition to knowing who has more than whom, abilities to assign the correct mathematical operator require syntactical knowledge of arithmetical equations. Not surprisingly, in these analyses, the contributions of both working memory and ability to select the correct operator was reduced. Working memory explained 7% of variance. The contribution of operator selection was marginal; it explained 3% of variance in the residual.

In the cognitive literature, much effort has gone into examining processes involved in the acquisition of computational skills. However, for our children at least, computation explained little variance in problem solving accuracy. This finding is consistent with previous findings (Cummins et al., 1988) and is perhaps not surprising given that most 11 year olds in Singapore would have had ample practice in arithmetic computation.

In relation to the role of working memory, of interest was the finding that amongst the componential tasks, ability to select the correct operator was most strongly associated with the working memory measures. This suggests that working memory may be most involved when participants consider the mathematical operations needed for computing a solution. This, in turn, contributes to performance in Solution Formation.

Working memory and Algebraic Word Problem Solving

Why is working memory related to algebraic problem solving? A number of computer simulations of word problem solving (Bobrow, 1968; Cummins et al., 1988; Reusser, 1990; LeBlanc

& Weber-Russell, 1996) have shown processes associated with both representation and solution formation are open to mechanisation. But mechanisation should not be equated to automaticity. A recent fMRI study shows that even amongst adults highly competent in algebraic word problems (over 90% accuracy on tested items), transformation of simple one-step start-unknown problems activated areas associated with working memory processes (Lee et al., 2007). A capacity based explanation of the relationship between working memory and word problem solving would perhaps specify the latter as a complex controlled tasks. Such tasks, by definition, require attentional or working memory resources. Although attractive in its simplicity, such explanations lack specificity. Our present findings show that there are specific components of problem solving – representational formation and ability to select the correct operator -- that are particularly resource intensive.

Although we do not have a conclusive explanation of why these components are particularly resource intensive, we were attracted to an explanation inspired by Thevenot, Devidal, Barrouillet, and Fayol (2007). They argued that understanding word problems require real time construction of mental models. They argued against earlier models (Kintsch & Greeno, 1985), which suggested that cue responsive schema, stored in long term memory, are involved in solving such problems. Instead, they favoured models suggesting that temporary situation models – stored in working memory – are more critical in problem solving (Nathan, Kintsch, & Young, 1992; Reusser, 1990).

Such temporary models may be of particular import in explaining our findings. When we inspected children's problem solving scripts, we typically found a single neatly drawn schematic per problem. Only in 266 out of a total of 2550 questions did we see signs of corrections having been made. This suggests that children may be using schematics not as testing grounds for their understanding of problems, but as final representations that they present for examination. If this is the case, much of the evaluation and testing of alternative schematics must have occurred mentally. Similar to the temporary situation models advanced by Thevenot et al. (2007), the generation of such mental schematics will likely be highly working memory intensive.

Inhibition and Switching

Previous work conducted by Passolunghi and her colleagues (e.g., Passolunghi et al., 1999) suggested that inhibition should predict problem solving accuracy. In contrast, our data show that neither the inhibition nor switch measures were correlated with performance on the algebraic tasks. Why did inhibition and switch efficiency not predict accuracy in the present study? Apart from age and other curricular differences that may have influenced findings, a major difference is that we did not include irrelevant information in our algebraic questions. The use of such information in Passolunghi's study may have increased demands on participants' ability to select and disregard unneeded information. Another issue that may have affected the findings is strategy selection. In a related study conducted with older children, who were learning new heuristics for solving algebraic problems, the same inhibitory measures predicted likelihood of intrusion from previously learned heuristics (Khng & Lee, under review). In the present study, we asked children to use only the model method. This constraint in strategy choice may have further reduced demands on inhibition.

Limitations, Caveats, and Suggestions for Future Studies

Why did we not use non-overlapping componential measures? In designing the componential tasks, we reduced dependencies by providing children with the necessary prerequisite information. Nonetheless, there were common processes, such as reading, for which this approach proved ineffective. For both Relation A and B, we considered using graphical depictions of quantitative relationships for which text decoding would be less critical. However, such modifications make little sense because a critical aspect of what we intended to measure was children's ability to decipher quantitative relationship depicted in word problems. For these reasons, we relied on statistical corrections to account for much of the interrelations amongst the componential measures. These methods allow us to preserve the essence of processes involved in word problem solving. Also, they do not present the same risks inherent in the use of de-contextualised questions with which children are unfamiliar.

We ameliorated ceiling effects on the Question Understanding and subscales of the computation measures by transformation. In fact, similar problems were identified in pilot testing. As a result, item difficulty was increased. For both tasks, we were limited by potential confounds associated with further increases in difficulty. The computation task, for example, used the same arithmetic operations as questions used in the overall accuracy task. The version we used already contained operands larger than those used in the curriculum. Although we could have further increased difficulty by using larger operands and more operations per question, such changes would likely be contaminated by abilities that are incidental to arithmetic abilities, e.g., exercise of care in computation. For this reason, we interpreted ceilings associated with both tasks as indicative of children not having significant problems with arithmetic computation or understanding the questions.

A more general issue that affects the measurement of relationships between switching, inhibition, and the algebraic tasks problem solving is related to the reliability of executive functioning measures. As noted by Miyake (2000; 2004) and others, these measures tend to have relatively low internal reliability. This is an important issue as low correlation between switching, inhibition, and algebraic performance could be due to low task reliability rather than the absence of relationship. Miyake and his colleagues (2000, 2004) argued that reliability for the executive tasks may be low because participants use different strategies, even within a session, to perform these tasks. Also of note is that many switch and inhibition measures rely on difference scores for generating inhibitory or switch costs. Such scores tend to have low internal reliability (Cohen & Cohen, 1983). As a case in point, though the Stroop task is a well established measure, its estimated reliability was low. This is despite high reliability for the incongruent and neutral measures from which the difference score was estimated. Others have also noted a similar differentiation and have argued that reliability estimates for the contributing conditions may serve as better indicators of reliability than does reliability of difference scores (Eide, Kemp, Silberstein, Nathan, & Stough, 2002).

We urge further development on the measurement of executive functions. There are few measures of executive functioning that do not rely on difference scores. Of those that do, some are complex tasks that require functions other than solely inhibition or mental flexibility (e.g., the Wisconsin Card Sorting Test, Heaton, 1993). Perhaps because of its nature, we know of no switch tasks that do not rely on difference scores: this is a major challenge for future research.

To what extent are findings specific to the Singapore context? We suggested that working memory did not predict overall accuracy above and beyond what was predicted by Representation and Solution Formation because children may be approaching the questions in a linear, step-by-step fashion. Pedagogical practices vary from classroom to classroom. Some teachers adopt a more algorithmic approach and emphasise practice on true-and-tested heuristics. Such an approach may encourage children to produce proceduralised responses when asked to answer typical problems, such as those used in this study. In contrast, pedagogical environments in which children are encouraged to adopt a more investigative approach may see problem solving proceeding in a more iterative fashion. If this is indeed the case, such environments will likely place even greater demands on working memory. This is an empirical question that awaits further investigation.

Pedagogical Implications

The findings suggest three avenues for intervention. First, as pointed out by Kintsch (1991), understanding algebra word problems require more than being able to read. Additionally, children need to be made aware of the specific ways in which terms like “more than” and “less than” are used in the context of word problems. Activities designed to show the importance of identifying the beginning and end state of objects being compared in these relationships may be useful.

Second, our findings show that success on the overall accuracy task is closely related to working memory capacity. One way to improve performance is to increase working memory capacity. Already, there are a number of demonstrations using computer training programmes that show promise (e.g., Klingberg et al., 2005). A remaining question is whether there are windows in children's development at which their working memory systems are particularly sensitive to intervention.

An alternative way to improve performance to teach children how to break complex questions into more manageable parts. A fear regarding such an approach is that it may be just as resource intensive to put the parts back together. Our findings suggest that this may not be a concern and that putting the parts together may not be particularly resource intensive. However, they also show that it is not sufficient to just break the problems into smaller components. Although we do not have absolute measures of working memory resources required by the whole problem versus its components, success on many of the componential tasks is also strongly associated with working memory capacity. What is an alternative approach?

The findings show that selecting an appropriate operation is cognitively demanding. Rather than tackling its working memory requirements directly, an alternative approach is to provide content specific pedagogical intervention. One reason children find this task difficult is the many-to-one relationship between word problems and their underlying mathematical structure. Using an example from Verschaffel and De Corte (1997), an arithmetic expression "$x + y$" can be expressed in many forms:

1. Adam has x marbles. Ben gives him y more marbles. How many marbles has Adam now?
2. Adam has x marbles. Ben has y marbles. How many marbles do they have altogether?
3. Adam has x marbles. Ben has y marbles more than Adam. How many marbles has Ben?
4. Adam has x marbles. Adam has y marbles less than Ben. How many marbles has Ben?

Although exposure to such questions is part of primary children's curriculum, they are often experienced independently and not as a set. For this reason, children may not be aware of their inter-relation. A better approach is to construct activities that allow children to experience such questions as a set. Children can be asked to select the appropriate mathematical operation to represent the information in each problem. Rather than focusing on generating a solution, children can be encouraged to discuss why the same mathematical expression can be represented by very different word problems. More sophisticated pupils can be asked to construct word problems that have the same mathematical structure, but different wordings. These approaches utilise the idea of reversibility to raise children's awareness that the same arithmetic expression may arise from structurally different word problems, and conversely, that the same key phrases, "more than" or "less than" do not always map onto addition and subtraction.

Conclusions

Previous studies have shown that working memory explains a significant amount of variation in algebraic word problem solving performance. This study provides a finer analysis of its role by decomposing algebra word problems into their components. Our findings suggest that understanding word problems and representing them in a schematic format requires more than general skills in reading comprehension. Children who succeeded in the overall accuracy task were better at decoding quantitative relationships amongst protagonists. Although using schematics may provide external memory cues that reduce working memory demands, our findings show that producing such schematics is also working memory intensive. We suggest focusing pedagogical intervention on helping children negotiate relational terms used in word problems. In particular, a better understanding of the vicissitude in the relationship between these terms and their corresponding mathematical operation will likely improve performance.

References

Bednarz, N. & Janvier, B. (1996). Emergence and development of algebra as a problem-solving tool. Continuities and discontinuities with arithmetic. In N. Bednarz, C. Kieran, & L. Lee (Eds.), *Approaches to Algebra* (pp. 115-136). Dordretch, The Netherlands: Kluwer.

Blessing, S. B. & Ross, B. H. (1996). Content Effects in Problem Categorization and Problem Solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 792-810.

Bobrow, D. G. (1968). Natural language input for a computer problem solving system. In M. Minsky (Ed.), *Semantic information processing* (pp. 146-226). Cambridge, MA: MIT Press.

Borenstein, M., Rothstein, H., Cohen, J., Schoenfeld, D., Berlin, J., & SPSS (2001). Power and Precision (Version 2.1) [Computer software]. Englewood, NJ: Biostat.

Briars, D. J. & Larkin, J. H. (1984). An Integrated Model of Skill in Solving Elementary Word Problems. *Cognition and Instruction, 1,* 245-296.

Bull, R., Johnston, R. S., & Roy, J. A. (1999). Exploring the roles of the visual-spatial sketch pad and central executive in children's arithmetical skills: Views from cognition and developmental neuropsychology. *Developmental Neuropsychology, 15,* 421-442.

Bull, R. & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19,* 273-293.

Clair-Thompson, H. L. & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology, 59,* 745-759.

Cohen, J. & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences. (2nd ed.) Hillsdale, NJ: Erlbaum.

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20,* 405-438.

Daneman, M. & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review, 3,* 422-433.

Eide, P., Kemp, A., Silberstein, R. B., Nathan, P. J., & Stough, C. (2002). Test-Retest reliability of the emotional Stroop task: Examining the paradox of measurement change. *The Journal of Psychology, 136,* 514-520.

Friedman, N. P. & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of Experimental Psychology: General, 133,* 101-135.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M. et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98,* 29-43.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40,* 177-190.

Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and arithmetical cognition: a longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology, 77,* 236-263.

Geary, D. C., Hoard, M. K., & Hamson, C. O. (1999). Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology, 74,* 213-239.

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology, 93,* 615-626.

Heaton, R. K. (1993). Wisconsin Card Sorting Test (Version 2) [Computer software]. Odessa, FL.: Psychological Assessment Resources.

Kail, R. & Hall, L. K. (1999). Sources of developmental change in children's word-problem performance. *Journal of Educational Psychology, 91,* 660-668.

Kintsch, W. & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92,* 109-129.

Kintsch, W. (1991). A theory of discourse comprehension: Implications for a tutor for word algebra problems. In M. Carretero & M. L. Pope (Eds.), *Learning and instruction: European research in an international context, Vol. 3* (pp. 235-253). Elmsford, NY: Pergamon Press.

Kintsch, W. (1998). Comprehension: A paradigm for cognition. NY: Cambridge University Press.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K. et al. (2005). Computerized Training of Working Memory in Children With ADHD--A Randomized, Controlled Trial. *Journal of the American Academy of Child & Adolescent Psychiatry, 44,* 177-186.

Koedinger, K. R. & Nathan, M. J. (2004). The Real Story Behind Story Problems: Effects of Representations on Quantitative Reasoning. *Journal of the Learning Sciences, 13,* 129-164.

LeBlanc, M. D. & Weber-Russell, S. (1996). Text integration and mathematical connections: A computer model of arithmetic word problem solving. *Cognitive Science, 20,* 357-407.

Lee, K., Lim, Z. Y., Yeong, S. H. M., Ng, S. F., Venkatraman, V., & Chee, M. W. L. (2007). Strategic differences in algebraic problem solving: Neuroanatomical correlates. *Brain Research, 1155,* 163-171.

Lee, K., Ng, S. F., Ng, E. L., & Lim, Z. Y. (2004). Working Memory and Literacy as Predictors of Performance on Algebraic Word Problems. *Journal of Experimental Child Psychology, 89,* 140-158.

Lee, K. & Peh, Y. X. (2008). Working memory profiles of children with difficulties in English and/or mathematics. *Korean Journal of Thinking and Problem Solving, 18,* 21-41.

Lewis, A. B. & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology, 79,* 363-371.

Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004). TIMSS 2003 International Science Report: Findings From IEA's Trends in International Mathematics and

Science Study at the Fourth and Eighth Grades. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mayer, R. E. & Hegarty, M. (1996). The process of understanding mathematical problems. In R. J. Sternberg & T. Ben-Zeev (Eds.), *The nature of mathematical thinking.* (pp. 29-53). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49-100.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). TIMSS 2003 International Mathematics Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction, 9,* 329-389.

Nathan, M. J. K. (1992). A Theory of Algebra-Word-Problem Comprehension and Its Implications for the Design of Learning Environments. *Cognition & Instruction, 9,* 329.

Ng, S. F. & Lee, K. (2005). How primary five pupils use the model method to solve word problems. *The Mathematics Educator, 9,* 60-84.

Ng, S. F., Lee, K., Ang, S. Y., & Khng, F. (2006). Model Method: Obstacle or bridge to learning symbolic algebra. In W. Bokhorst-Heng, M. Osborne, & K. Lee (Eds.), *Redesigning Pedagogies* New York, NY: Sense.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin, 126,* 220-246.

Pape, S. J. (2003). Compare word problems: Consistency hypothesis revisited. *Contemporary Educational Psychology, 28,* 396-421.

Passolunghi, M. C. & Siegel, L. S. (2001). Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. *Journal of Experimental Child Psychology, 80,* 44-57.

Passolunghi, M. C., Cornoldi, C., & De Liberto, S. (1999). Working memory and intrusions of irrelevant information in a group of specific poor problem solvers. *Memory and Cognition, 27,* 779-790.

Passolunghi, M. C. & Pazzaglia, F. (2004). Individual differences in memory updating in relation to arithmetic problem solving. *Learning and Individual Differences, 14,* 219-230.

Pickering, S. J. & Gathercole, S. E. (2001). Working Memory Test Battery for Children. Kent: The Psychological Corporation.

Polya, G. (1957). How to solve it: A new aspect of mathematical method. (2nd ed.) Princeton, NL: Princeton University Press.

Reusser, K. (1990). From text to situation to equation: Cognitive simulation of understanding and solving mathematical word problems. In H. Mandl, E. De Corte, S. N. Bennett, & H. F. Friedrich (Eds.), *Learning & Instruction: European Research in an International Context* (pp. 477-498). Oxford, UK: Pergamon.

Rust, J. (2000). Wechsler Objective Reading and Language Dimensions (Singapore). London, UK: The Psychological Corporation.

Sakamoto, M. (1998). Existing Knowledge which determines Performance in Arithmetic Word Problems. *Psychologia, 41,* 102-113.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime reference guide. Pittsburgh, PA: Psychology Software Tools.

St Clair-Thompson, H. L. & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology, 59,* 745-759.

Swanson, H. L. (2006). Cross-Sectional and Incremental Changes in Working Memory and Mathematical Problem Solving. *Journal of Educational Psychology, 98,* 265-281.

Swanson, H. L., Cooney, J. B., & Brock, S. (1993). The influence of working memory and classification ability on children's word problem solution. *Journal of Experimental Child Psychology, 55,* 374-395.

Thevenot, C., Devidal, M., Barrouillet, P., & Fayol, M. (2007). Why does placing the question before an arithmetic word problem improve performance? A situation. *Quarterly Journal of Experimental Psychology, 60,* 43-56.

Thompson, B. (1984). Canonical correlation analysis: uses and interpretation. Beverly Hills, CA: Sage.

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, England: Psychology Press.

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.

Wechsler, D. (1991). Wechsler Intelligence Scale for Children - Third Edition. San Antonio, TX: Psychological Corporation.

Table 1

Means, Standard Deviations, and Correlation Ratios Amongst all Algebraic and Executive Functioning Measures

	Max	M	SD	Skew	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. OA	10	4.22	3.07	0.03															
2. RF	5	2.55	1.63	0.05	.808*														
3. SF	5	3.26	1.81	-0.68	.821*	.755*													
4. QU	16	15.00	1.63	-2.14	.424*	.426*	.516*												
5. Rel A	12	9.17	2.40	-0.86	.712*	.669*	.722*	.381*											
6. Rel B	10	8.03	1.83	-0.85	.699*	.656*	.646*	.368*	.622*										
7. Com	20	16.77	2.32	-1.17	.481*	.464*	.484*	.460*	.500*	.387*									
8. CR	42	96.08	13.74	-0.19	.326*	.331*	.313*	.219*	.308*	.255*	.244*								

9. LM	12	6.36	2.15	-0.20	.409*	.376*	.403*	.286*	.385*	.307*	.269*	.261*							
10. KT	27	14.61	3.36	-0.23	.390*	.356*	.422*	.295*	.419*	.364*	.352*	.356*	.355*						
11. ST	2799[a]	72.24	74.48	0.07	-.100	-.035	-.104	-.052	-.038	-.105	-.023	-.034	-.061	-.107					
12. SS	1.00	0.58	0.24	-0.31	-.086	-.073	-.116	-.116	-.130	-.054	-.097	-.148	-.166	-.105	-.064				
13. NL	6999[a]	725.58	397.66	0.46	.145	.158	.062	.042	.178	.115	.032	.095	.119	.059	.031	.038			
14. PM	9000[a]	-223.71	551.53	-0.19	.062	.026	-.012	.009	.013	.016	-.063	.010	-.070	.065	.042	-.053	.047		
15. VC	60	23.76	7.81	0.26	.560*	.562*	.504*	.325*	.500*	.492*	.342*	.277*	.212	.512*	-.105	-.003	.138	.107	
16. RC	38	25.41	4.22	-0.78	.524*	.549*	.496*	.354*	.499*	.507*	.383*	.299*	.269*	.505*	.049	-.041	.106	.087	.586*

Note. OA = Overall Accuracy, RF = Representation Formation, SF = Solution Formation, QU = Question Understanding, Rel A = Relational A, Rel B = Relational B, Com = Computation, CR = Counting Recall, LM = Letter Memory, KT = Keep Track, ST = Inhibition efficiency (Stroop), SS = Inhibition efficiency (Stop Signal), NL = Local switch cost (Number-Letter), PM = Local switch cost (Plus-Minus), VC =Vocabulary, RC = Reading Comprehension.

[a] Artificial ceiling imposed during data screening

* $p < 0.0005$

Table 2

Summary of Standard and Hierarchical Regression Analyses for Variables Predicting Performance on Algebraic Measures

Model	Order of entry of predictors into regression model	Predicted variable	R^2	ΔR^2	F
1	Representation Formation, Solution Formation	Overall accuracy	.76		390.97*
2	Question Understanding, Relational A, Relational B, Computation	Overall accuracy	.63		107.36*
3	Counting Recall, Letter Memory, Keep Track	Overall accuracy	.26		29.55*
4	1. Representation Formation, Solution Formation	Overall accuracy	.76		390.97*
	2. Counting Recall, Letter Memory, Keep Track			<.01	1.17
5	Counting Recall, Letter Memory, Keep Track	Representation Formation	.23		25.03*
6	Counting Recall, Letter Memory, Keep Track	Solution Formation	.27		30.85*
7a	1. Counting Recall, Letter Memory, Keep Track	Representation Formation	.23		25.03*
	2. Reading Comprehension, Vocabulary			.22	50.23*
	3. Question Understanding			.03	15.17*

	4. Relational B			.10	54.46*
7b	1. Reading Comprehension, Vocabulary	Representation Formation	.39		80.31*
	2. Counting Recall, Letter Memory, Keep Track			.06	9.42*
8	1. Counting Recall, Letter Memory, Keep Track	Solution Formation	.30		30.85*
	2. Relational A			.28	158.17*
	3. Computation			<.01	3.44
9	Representation Formation	Solution Formation	.57		334.75*
10	1. Counting Recall, Letter Memory, Keep Track	Solution Formation (residual)	.07		5.87*
	2. Question Understanding			.05	14.05*
	3. Relational B			.00	2.05
	4. Relational A			.03	8.49
	5. Computation			.00	0.19

Note. * $p < .001$

Table 3

Summary of Canonical Correlation Analyses for Algebraic and Executive Functioning Measures and Sensitivity Analysis for the Canonical Correlations

Variables	First canonical function: Loadings	First canonical function: Standardized coefficients	Results after deletion of one variable			
Analysis 1						
Set 1						
Question Understanding	-.66	-.28	omit.	-.71	-.69	-.66
Relational A	-.90	-.53	-.92	omit.	-.93	-.90
Relational B	-.83	-.36	-.86	-.90	omit.	-.83
Computation	.57	.08	.59	.62	.59	omit.
Percent of variance	.56		.65	.57	.56	.65
Redundancy	.39		.42	.34	.36	.45
Set 2						
Representation Formation	-.90	-.41	-.92	-.92	-.89	-.90
Solution Formation	-.96	-.65	-.95	-.95	-.97	-.97
Percent of variance	.87		.88	.88	.87	.87
Redundancy	.61		.57	.52	.56	.60
Canonical	.83		.81	.77	.80	.83

correlation									
				Analysis 2					
Set 1									
Counting Recall	-.63	-.32	omit.	-.72	-.71				
Letter Memory	-.78	-.52	-.82	omit.	-.87				
Keep Track	-.79	-.49	-.83	-.90	omit.				
Percent of variance	.55		.68	.67	.63				
Redundancy	.16		.18	.15	.15				
Set 2									
Representation Formation	-.89	-.38	-.87	-.89	-.93				
Solution Formation	-.97	-.69	-.98	-.97	-.95				
Percent of variance	.87		.86	.87	.88				
Redundancy	.25		.22	.19	.20				
Canonical correlation	.54		.51	.47	.48				
				Analysis 3					
Set 1									
Counting Recall	-.62	-.30	omit.	-.69	-.70	-.61	-.63	-.62	-.62
Letter Memory	-.76	-.49	-.79	omit.	-.87	-.73	-.75	-.76	-.78
Keep Track	-.82	-.54	-.86	-.92	omit.	-.85	-.83	-.82	-.80
Percent of	.54		.68	.66	.63	.54	.55	.55	.54

variance									
Redundancy	.18		.21	.18	.16	.17	.16	.18	.17
Set 2									
Question Understanding	-.68	-.29	-.66	-.63	-.74	omit.	-.73	-.68	-.70
Relational A	-.88	-.51	-.88	-.88	-.87	-.91	omit.	-.89	-.91
Relational B	-.73	-.18	-.74	-.75	-.70	-.76	-.79	omit.	.76
Computation[a]	.72	.30	.72	.76	.68	.75	.78	.73	omit.
Percent of variance	.57		.57	.57	.56	.66	.59	.60	.63
Redundancy	.19		.18	.15	.15	.20	.17	.19	.20
Canonical correlation	.58		.55	.52	.51	.56	.53	.57	.56

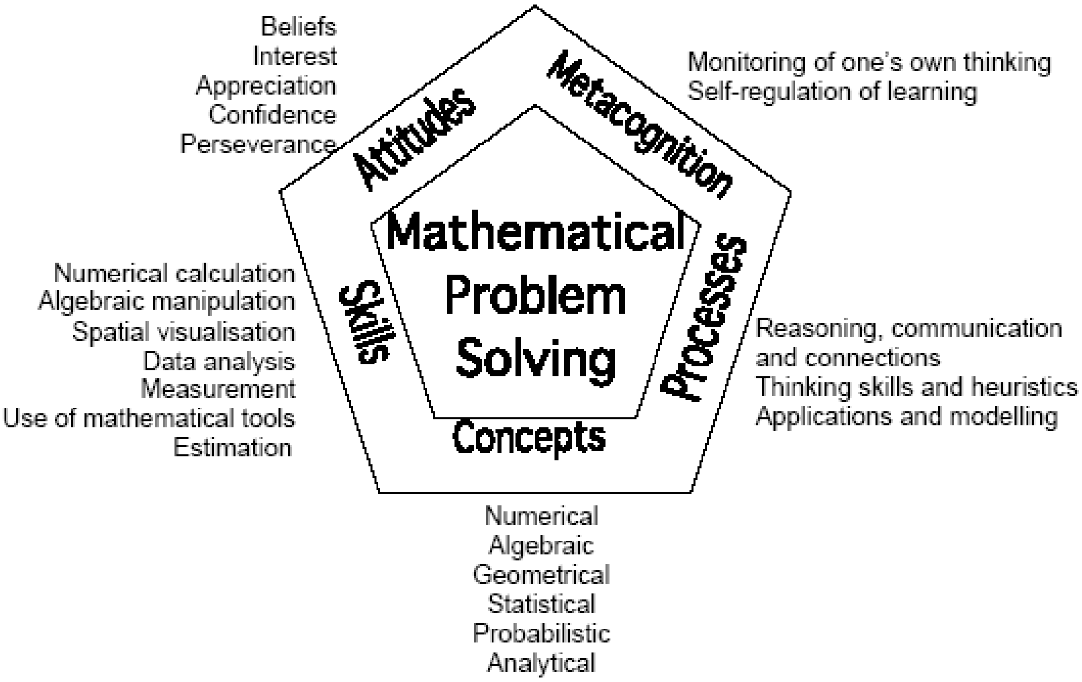
Note. omit. = omitted variable.

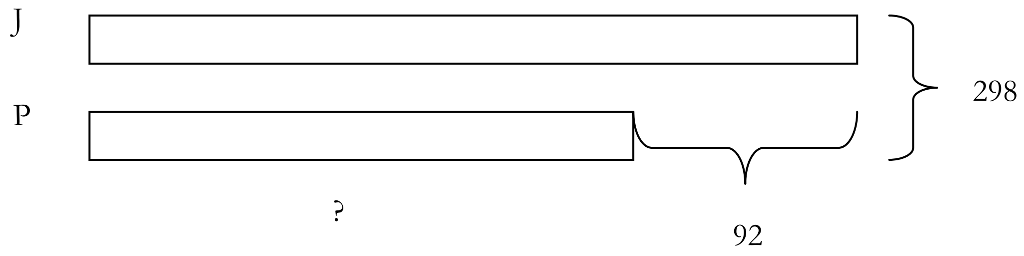
[a] The polarity of computation is the reverse of the others because a logarithmic transformation was applied.

Figure Caption

Figure 1. A graphical depiction of the Primary school mathematics curriculum in Singapore

Figure 2. A model representation of an algebraic question.


Beliefs
Interest
Appreciation
Confidence
Perseverance
Attitudes
Metacognition
Monitoring of one's own thinking
Self-regulation of learning
Mathematical
Problem
Solving
Numerical calculation
Algebraic manipulation
Spatial visualisation
Data analysis
Measurement
Use of mathematical tools
Estimation
Skills
Processes
Reasoning, communication
and connections
Thinking skills and heuristics
Applications and modelling
Concepts
Numerical
Algebraic
Geometrical
Statistical
Probabilistic
Analytical


J
P
298
?
92