Are diagrams always helpful tools? Developmental and individual differences in the effect of presentation format on student problem solving

Julie L. Booth\(^1\)* and Kenneth R. Koedinger\(^2\)

\(^1\)Temple University, Philadelphia, USA
\(^2\)Carnegie Mellon University, Pittsburgh, USA

**Background.** High school and college students demonstrate a verbal, or textual, advantage whereby beginning algebra problems in story format are easier to solve than matched equations (Koedinger & Nathan, 2004). Adding diagrams to the stories may further facilitate solution (Hembree, 1992; Koedinger & Terao, 2002). However, diagrams may not be universally beneficial (Ainsworth, 2006; Larkin & Simon, 1987).

**Aims.** To identify developmental and individual differences in the use of diagrams, story, and equation representations in problem solving. When do diagrams begin to aid problem-solving performance? Does the verbal advantage replicate for younger students?

**Sample.** Three hundred and seventy-three students (121 sixth, 117 seventh, 135 eighth grade) from an ethnically diverse middle school in the American Midwest participated in Experiment 1. In Experiment 2, 84 sixth graders who had participated in Experiment 1 were followed up in seventh and eighth grades.

**Method.** In both experiments, students solved algebra problems in three matched presentation formats (equation, story, story + diagram).

**Results.** The textual advantage was replicated for all groups. While diagrams enhance performance of older and higher ability students, younger and lower-ability students do not benefit, and may even be hindered by a diagram’s presence.

**Conclusions.** The textual advantage is in place by sixth grade. Diagrams are not inherently helpful aids to student understanding and should be used cautiously in the middle school years, as students are developing competency for diagram comprehension during this time.

An abundance of evidence exists on the cognitive benefits of different types of external representations, including that the addition of relevant diagrams to text enhances learning (Mayer, 1989, 2005). However, several theories of diagrammatic reasoning caution that the benefit of diagrams is dependent on their relevance for the task at hand.

*Correspondence should be addressed to Julie L. Booth, 1301 W. Cecil B. Moore Avenue, Philadelphia PA 19122, USA (e-mail: julie.booth@temple.edu).

DOI:10.1111/j.2044-8279.2011.02041.x
the context of the representation, and the skill of the user (for a review, see Acevedo Nistal, Clarebout, Elen, Van Dooren, & Verschaffel, 2009). For example, representations that are not appropriate or compatible with the task that the user is asked are unlikely to be helpful (Meyer, 2000), as they may not support the necessary types of cognitive processing required for the task (Greeno & Hall, 1997). Diagram utility may also depend on the spatial grouping of information that will be used together (Larkin & Simon, 1987), the degree of contiguity between the text and associated diagram (Mayer, 2005), and a lack of complete redundancy between the text and diagram (Chandler & Sweller, 1991; Kalyuga, Chandler, & Sweller, 1998). The effects of these characteristics may be more or less pronounced depending on the prior knowledge and spatial ability of the user (Mayer & Gallini, 1990; Mayer, Steinhoff, Bower, & Mars, 1995). However, user characteristics may influence effective use of even compatible, well-designed diagrams. For example, domain experts can interpret diagrams within their domain more accurately than novices (Ainsworth, 2006), as they can more easily identify critical components of the diagram (Larkin & Simon, 1987) and connections among diagram components and the represented situation (Narayanan & Hegarty, 1998). Domain-general diagrammatic reasoning and spatial skills may be also crucial for drawing inferences from the spatial layout of the diagram (Larkin & Simon, 1987), and familiarity with the general form and components of the representation may also influence correct interpretation (Ainsworth, 2006). Diagrams that are aligned with these learner characteristics and are designed to support the cognitive processes of those learners are more likely to yield deep comprehension of the represented information (Butcher, 2006; Davenport, Yaron, Klahr, & Koedinger, 2008).

One context in which diagrams may be especially useful is in supporting students’ understanding of key concepts within a domain. A commonly proposed strategy for improving learning involves creating a smooth transition from what students already know, which is often more concrete, to the desired knowledge, which is often more abstract. For example, Piaget’s theory of cognitive development suggests that it is necessary to provide concrete experience from which young students can abstract higher order concepts (Kamii, 1974); a number of empirical studies support this hypothesis that transitioning from grounded representations to more abstract ones is an effective instructional technique (Bransford, Brown, & Cocking, 1999; Goldstone & Son, 2005; Koedinger & Anderson, 1998; Moreno & Mayer, 1999; Nathan, Kintsch, & Young, 1992; Nathan & Koedinger, 2000; Romberg & de Lange, 2011; Schwartz & Black, 1996). For example, instructors might ensure that early experiences are concrete, in order to promote connections between the facets of a representation and their real-world counterparts, and then fade the concreteness so learners are able to use more abstract representations and transfer their knowledge more easily to different situations (Goldstone & Son, 2005).

Much of the relevant empirical work in mathematics focuses on the transition from concrete physical representations, or manipulatives, to abstract, symbolic ones in elementary school (e.g., Kennedy & Tipps, 1994; Sowell, 1989). However, bridging instruction may also be important later in development, when students are transitioning from arithmetic to algebraic thinking. Despite the common belief that word problems are inherently more difficult than equations, solving simple algebra problems (e.g., those that refer to the variable only once) in a textual format can actually be easier for high school (typically aged 15–18) (Koedinger & Nathan, 2004) and college students (Koedinger, Alibali, & Nathan, 2008) than solving them in an equivalent equation format. However, for more complex problems (e.g., those that refer to the variable twice), story problems
are harder to solve than matched problems in equation format (Koedinger et al., 2008). These results suggest that instruction may be more effective if it builds understanding and skill with abstract symbolic representations on top of existing competence with more concrete textual representations.

Failure to translate story problems into usable internal representations (De Corte, Verschaffel, & De Win, 1985; Kintsch & Greeno, 1985; Zawaiza & Gerber, 1993) or to produce appropriate mathematical representations of the problem (Heffernan & Koedinger, 1997; Koedinger & Nathan, 2004) can each preclude successful problem solving. For such difficulties in representation, providing a diagram could facilitate solution by effectively scaffolding the representation process. External representations (e.g., diagrams, graphs, tables, etc.) are recommended tools for math instruction (National Council of Teachers of Mathematics [NCTM], 2000), and empirical work from the field of mathematics education provides evidence of performance-enhancing benefits of diagrams – Hembree’s (1992) meta-analysis concluded across 16 studies that students provide more correct answers to word problems when they contain accompanying diagrams. However, recent studies have called into question the effectiveness of diagrams for word problems. For example, De Bock, Verschaffel, Janssens, Van Dooren, and Claes (2003) report that diagrams may actually be harmful for high school students learning geometry. Further, a detrimental effect of diagrams was found for both strong and weak fifth grade math students when solving arithmetic word problems (Berends & Van Lieshout, 2009); in some situations, solving problems with even a well-designed diagram may increase the cognitive load placed on students (Berends & Van Lieshout, 2009; Lee, Ng, & Ng, 2009).

Diagrams are often used in math instruction in Singapore (Beckmann, 2004) and Japan (Murata, 2008), two countries in which mathematics achievement is consistently outstanding by world standards (National Center for Education Statistics, 2003). The style of diagrams used in these countries, sometimes called tape diagrams (Murata, 2008), strip diagrams (Beckmann, 2004), or bar models (Hoven & Garellick, 2007), uses strings of objects and/or strips of different lengths to represent the magnitude of and relationships between the quantities in the problem. Like algebraic equations (but unlike symbolic arithmetic problems), these diagrams are not meant to help users carry out operations, but to help them decide what operations to use and to understand why those operations are conceptually sound (Beckmann, 2004). Tape diagrams are also found in textbooks and some educational software (e.g., Carnegie Learning, 2009) in the United States, but their use is less frequent and inconsistent (Murata, 2008). There is, however, evidence that American students can use these diagrams to solve algebraic word problems that would ordinarily be quite challenging for them (Koedinger & Terao, 2002). For example, when given a tape diagram with a word problem, students in their study answered it correctly 71% of the time. On comparable problems without diagrams, middle school students were only 4% correct (Bednarz & Janvier, 1996) and college algebra students were only 54% correct (Koedinger & Alibali, 1999). Thus, a potential motivation for including a diagrammatic representation with a problem might be to help younger students or those with lower ability to solve the problem, as they would be the ones less likely to be able to represent the problems accurately on their own.

The present study
Experiment 1 had three purposes. The first was to replicate the verbal or textual advantage with younger students (Koedinger and Nathan, 2004; Koedinger et al.,
2008) by presenting middle school students (typically aged 12–14) with difficult algebra problems (e.g., start-unknown and systems of equations problems) in equation and story formats. We also extend these findings by investigating developmental differences in the textual advantage, as older students’ increased familiarity with equations may improve their problem solution in that format.

The second purpose was to determine whether a diagrammatic advantage also exists for students learning algebra. Previous research does not yield a conclusive prediction: diagrams may prove to be beneficial (Hiebert & Carpenter, 1992; Koedinger & Terao, 2002), but could be futile or even harmful (Berends & Van Lieshout, 2009; De Bock et al., 2003; Gravemeijer, 1994; Uttal, Liu, & DeLoache, 2006). Alternatively, diagrams may be helpful for some students and not others (e.g., Ainsworth, 2006; Larkin & Simon, 1987). We examine whether diagrams are useful in the context of algebraic story problems and investigate developmental differences in their effectiveness.

The third purpose was to examine individual differences in the textual and diagrammatic advantages based on students’ mathematics ability level. The optimal type of presentation may vary for students with different background knowledge (cf. Kalyuga et al., 1998). As low-ability students have particular difficulty representing word problems (Montague, Bos, & Doucette, 1991), diagrams that eliminate or obviate the necessity of creating a problem representation could be particularly beneficial for them. Alternatively, high-ability students (domain experts) may have greater success than low-ability students (domain novices) at interpreting the diagrams, in which case high-ability students may show greater benefit from diagrams (Kozma, 2003; Lowe, 2003).

In Experiment 1, we presented sixth, seventh, and eighth grade students with algebra problems in each of three presentations types: equation, story, and story + diagram. The main predictions were that the textual advantage would be replicated (story problems would be easier than equations) and that a diagrammatic advantage would be evident (adding diagrams to story problems would facilitate solution); individual differences in grade and math ability were expected for each type of advantage.

EXPERIMENT I

Method

Participants
Participants were 373 students [121 sixth grade (age 12), 117 seventh grade (age 13), 135 eighth grade (age 14)] drawn from an ethnically diverse middle school in the American Midwest in which 25% of attending students were African American, 7% Asian, 62% Caucasian, and 5% Latino; 40% of attending students were from low-income families. All classes used Connected Mathematics, a reform-based curriculum, which begins developing algebraic skills in sixth grade, with increasing attention given to algebra in seventh and eighth grade (Lappan, Fey, Fitzgerald, Friel, & Phillips, 2002). Students are exposed to both symbolic and spatial representations throughout the curriculum, though it does not introduce the type of diagrams tested in this study.

Measures and procedure
The three items that were the focus of this study were embedded in a written comprehensive algebra assessment that included a variety of other items used for
studies on student understanding of equality and variability, which have been published elsewhere (Alibali, Knuth, Hattikudur, McNeil, & Stephens, 2007; Knuth, Alibali, McNeil, Weinberg, & Stephens, 2005; Knuth, Stephens, McNeil, & Alibali, 2006). The assessment was administered to students early in their fall semester by the classroom teacher during normal class time; individual students were randomly assigned to receive one of three alternate forms. Each form contained the same three problems, but varied the presentation format: equation, story, or story + diagram. Each form contained one problem for each presentation format, and each problem appeared once in each of the three presentation formats across the three forms. One item was a single-reference problem and one a double-reference problem, similar to those used by Koedinger et al. (2008); the remaining problem used two variables. Figure 1 shows the three problems in each format.

For approximately one-third of the students (n = 128: 43 sixth, 41 seventh, 44 eighth grade), parents consented to collection of national percentile rankings for the math section of the students’ most recent TerraNova tests. For sixth and seventh graders, scores were taken from the current year; the school district does not administer the TerraNova test in eighth grade, thus, eighth graders’ scores were from the previous year. Percentile rankings on the Terra Nova test correlated positively with student accuracy on the study problems, R(128) = .46, p < .001.

Results

Textual advantage
A 3 (grade: 6, 7, 8) × 2 (presentation: equation, story) ANOVA on correct responses yielded a main effect of presentation format F(1,371) = 31.95, p < .01, ηp^2 = .08; more correct responses were given when problems were in story form (24%) than in equation form (9%). There was no main effect of grade or grade by presentation interaction.

Diagrammatic advantage
A 3 (grade: 6, 7, 8) × 2 (presentation: story, story + diagram) ANOVA on correct answers yielded a main effect of grade F(2,370) = 3.17, p < .05, ηp^2 = .02. Least Significant Differences (LSD) post hoc analyses revealed that eighth graders answered more problems correctly (30%) than sixth graders (21%). No main effect of presentation format or interaction was found.

Differences by problem type
Because Koedinger et al. (2008) revealed that the advantages of different presentation formats could vary based on problem difficulty, we examined differences in accuracy by problem. Table 1 shows accuracy data on each problem type for the full sample

---

1 Analyses were conducted both across all three presentation formats and separately for the textual and diagrammatic advantages. The pattern of results from both analyses were the same, thus, as the purpose of the paper is to test two particular comparisons: story versus equation (to examine the textual advantage) and story versus story + diagram (to examine the diagrammatic advantage); separate ANOVAs are presented for each comparison.

2 Data were analysed using both parametric and non-parametric tests (when available), which yielded similar results. Due to the large sample size in the study and the fact that appropriate non-parametric tests did not always exist (e.g., there is no non-parametric version of two-factor, repeated measures tests), we only report results from the parametric tests.
<table>
<thead>
<tr>
<th>Equation</th>
<th>Story</th>
<th>Story + Diagram</th>
</tr>
</thead>
</table>
| Solve the equation below to find the value of N:  
\[(N - 45) \div 3 = 20.50\] | Mom won some money in a lottery. She kept $45 for herself and gave each of her 3 sons an equal portion of the rest. If each son got $20.50, how much did Mom win?  
(You can use the picture below to help you solve the problem.) | Mom won some money in a lottery. She kept $45 for herself and gave each of her 3 sons an equal portion of the rest. If each son got $20.50, how much did Mom win?  
(You can use the picture below to help you solve the problem.) |
| Find values of S and C that make these equations true:  
\[3S + 2C = 58\]  
\[2S + 3C = 52\] | John bought 3 t-shirts and 2 baseball caps for $58. Sue bought 2 t-shirts and 3 baseball caps for $52. What is the cost of one shirt? What is the cost of one baseball cap?  
(You can use the picture below to help you solve the problem.) | John bought 3 t-shirts and 2 baseball caps for $58. Sue bought 2 t-shirts and 3 baseball caps for $52. What is the cost of one shirt? What is the cost of one baseball cap?  
(You can use the picture below to help you solve the problem.) |
| Solve the equation below to find the value of N:  
\[N - \frac{1}{5} \times N = 30\] | Molly bought a coat on sale. It was \(\frac{1}{5}\) off the original price. She paid $30. What was the original price of the coat?  
(You can use the picture below to help you solve the problem.) | Molly bought a coat on sale. It was \(\frac{1}{5}\) off the original price. She paid $30. What was the original price of the coat?  
(You can use the picture below to help you solve the problem.) |

*Figure 1.* Problems displayed in each presentation format.
Table 1. Experiment 1: Accuracy by condition for full sample and each grade level for each problem

<table>
<thead>
<tr>
<th>Problem</th>
<th>Overall</th>
<th>Equation</th>
<th>Story</th>
<th>Story + diagram</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lottery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full sample</td>
<td>49%</td>
<td>22%</td>
<td>64%</td>
<td>61%</td>
</tr>
<tr>
<td>Grade 6</td>
<td>45%</td>
<td>20%</td>
<td>63%</td>
<td>51%</td>
</tr>
<tr>
<td>Grade 7</td>
<td>44%</td>
<td>18%</td>
<td>58%</td>
<td>54%</td>
</tr>
<tr>
<td>Grade 8</td>
<td>57%</td>
<td>26%</td>
<td>70%</td>
<td>76%</td>
</tr>
<tr>
<td>T-shirt</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full sample</td>
<td>4%</td>
<td>0%</td>
<td>5%</td>
<td>7%</td>
</tr>
<tr>
<td>Grade 6</td>
<td>2%</td>
<td>0%</td>
<td>3%</td>
<td>2%</td>
</tr>
<tr>
<td>Grade 7</td>
<td>5%</td>
<td>0%</td>
<td>8%</td>
<td>8%</td>
</tr>
<tr>
<td>Grade 8</td>
<td>5%</td>
<td>0%</td>
<td>4%</td>
<td>11%</td>
</tr>
<tr>
<td>Sale</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full sample</td>
<td>9%</td>
<td>5%</td>
<td>4%</td>
<td>17%</td>
</tr>
<tr>
<td>Grade 6</td>
<td>2%</td>
<td>0%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>Grade 7</td>
<td>16%</td>
<td>8%</td>
<td>8%</td>
<td>32%</td>
</tr>
<tr>
<td>Grade 8</td>
<td>8%</td>
<td>7%</td>
<td>2%</td>
<td>16%</td>
</tr>
</tbody>
</table>

and separately by grade for each presentation format. First, to determine problem-type differences in difficulty level, we conducted a 3 (grade: 6, 7, 8) × 3 (problem: lottery, t-shirt, sale) ANOVA with repeated measures on problem. This analysis revealed a main effect of problem, such that the lottery problem (single reference: a single variable is included only once in the equation) was easier than the sale (double reference: a single variable is included twice in the equation) and t-shirt problems (two variables: two separate variables are referred to in the problem), \( F(1,370) = 211.55, p < .001, \eta^2_p = .36 \). The main effect of grade \( [F(2,370) = 3.99, p < .05, \eta^2_p = .02] \) and problem by grade interaction \( (F(2,370) = 5.49, p < .01, \eta^2_p = .03) \) were also significant, but follow-up repeated measures ANOVAs conducted separately by grade confirmed that the pattern of difficulty was the same for each grade level (sixth grade, \( F(1,120) = 84.72, p < .01, \eta^2_p = .41 \); seventh grade, \( F(1,116) = 30.54, p < .001, \eta^2_p = .21 \); eighth grade, \( F(1,134) = 114.59, p < .001, \eta^2_p = .46 \)).

To determine whether the textual advantage existed for each problem, we conducted a 3 (grade: 6, 7, 8) × 2 (presentation: equation, story) ANOVA for each problem. For both the lottery problem and the t-shirt problem, there was a significant main effect of presentation, with accuracy higher in the story format than the equation format (lottery, \( F(1,244) = 55.34, p < .001, \eta^2_p = .19 \); t-shirt, \( F(1,242) = 6.35, p < .01, \eta^2_p = .03 \)). However, the main effect of presentation for the sale problem was not significant \( [F(1,242) = 0.06, \text{ns}, \eta^2_p = .00] \), and no other main effects or interactions were found. These results replicate Koedinger et al.’s (2008) finding that stories are easier than equations for single-reference problems (i.e., lottery), but that stories and equations were equally challenging for double-reference problems (i.e., sale).

A parallel set of 3 (grade: 6, 7, 8) × 2 (presentation: story, story + diagram) ANOVAs were conducted for each problem to test for the diagrammatic advantage. Main effects of grade were found for both the lottery \( (F(2,242) = 3.36, p < .05, \eta^2_p = .03) \) and sale problems \( (F(2,244) = 7.70, p < .001, \eta^2_p = .06) \). Results for the sale problem (double reference) also yielded a main effect of presentation \( (F(1,244) = 12.19, p < .001, \eta^2_p = .05) \), with accuracy higher on the story + diagram format than the story format. The presentation by grade interaction was also significant \( (F(1,244) = 3.65, \eta^2_p = .03) \).
Table 2. Exp 1: Distribution and analysis of high- and low-ability students by test form and grade

<table>
<thead>
<tr>
<th>Sample</th>
<th>Form A</th>
<th>Form B</th>
<th>Form C</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>High ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 6</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>$X^2(4, N = 64) = 5.17, ns$</td>
</tr>
<tr>
<td>Grade 7</td>
<td>4</td>
<td>10</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Grade 8</td>
<td>5</td>
<td>11</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Low ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 6</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>$X^2(4, N = 64) = 7.39, ns$</td>
</tr>
<tr>
<td>Grade 7</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Grade 8</td>
<td>5</td>
<td>3</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

$p < .05$, $\eta^2 = .03$). Follow-up $t$ tests conducted separately for each grade level on the sale problem indicated that story + diagram and story formats were equally challenging for sixth graders ($t(80) = 0.00$, ns, $d = .00$), but that accuracy was higher for the story + diagram format than the story format for seventh ($t(76) = 2.78$, $p < .01$, $d = .65$) and eighth graders ($t(88) = 2.33$, $p < .05$, $d = .50$). No analyses involving presentation format were significant for the lottery or t-shirt problems.

**Individual differences in the textual and diagrammatic advantages**

For the subset of students for whom math achievement scores were obtained, we computed the median national percentile ranking and grouped students by whether they were of high (greater than or equal to the median) or low (below the median) mathematical ability. Approximately, half of the students in each grade fell into the high-ability group. Table 2 shows the distribution of students in the high- and low-ability groups in each grade to test form; chi-squared tests yielded no significant differences in distribution.

**Textual advantage**

A 3 (grade: 6, 7, 8) $\times$ 2 (presentation: equation, story) $\times$ 2 (math ability: high, low) ANOVA on correct answers yielded a main effect of presentation format, $F(1,122) = 7.09$, $p < .01$, $\eta^2 = .06$; as in the parallel full sample analysis, students gave correct answers more frequently for story problems (29%) than equations (15%). There was also a main effect of math ability, $F(1,122) = 23.20$, $p < .01$, $\eta^2 = .16$; students with high math ability solved more problems correctly (33%) than peers with low math ability (11% correct). There were no other significant main effects or interactions. Thus, there were no individual differences in the textual advantage; story problems were easier than equations for all students, and high math ability increased the likelihood of solving either type of problem correctly.

**Diagrammatic advantage**

A parallel 3 (grade: 6, 7, 8) $\times$ 2 (presentation: story, story + diagram) $\times$ 2 (math ability: high, low) ANOVA assessed individual differences in the diagrammatic advantage. Similar to the full sample analysis, a main effect of grade was found ($F(2,122) = 4.96$, $p < .01$, $\eta^2 = .08$), as was a trend towards an interaction between the grade and presentation format variables, $F(2,122) = 2.31$, $p = .10$, $\eta^2 = .04$. Proportion of correct answers
increased with grade, and sixth graders tended to answer more story problems correctly than story + diagram problems (25% vs. 17%, respectively), while the opposite was true for seventh and eighth graders (43% and 46% correct on story + diagram problems vs. 39% and 23% on story problems). This suggests that some improvement in diagram use is happening as students mature and/or experience more math instruction. Not surprisingly, there was also a main effect of math ability $F(1,122) = 5.47, p < .05, \eta^2_p = .04$. High-ability students answered more problems correctly (39%) than low-ability peers (26%).

The interaction among the three variables was also significant, $F(2,122) = 3.42, p < .05, \eta^2_p = .05$ (Figure 2). Separate 3 (grade) × 2 (presentation) repeated measures ANOVAs were then conducted for each ability group. The analysis for high-ability students yielded only a main effect of grade, $F(2,61) = 3.43, p < .05, \eta^2_p = .10$. In contrast, the analysis for low-ability students yielded trends towards main effects of presentation (20% [story] vs. 33% [story + diagram]; $F(1,61) = 3.14, p < .10, \eta^2_p = .05$) and grade ($F(2,61) = 2.45, p < .10, \eta^2_p = .07$), and a significant presentation by grade interaction, $F(2,61) = 8.17, p < .001, \eta^2_p = .21$. Paired-sample $t$ tests comparing the story and story + diagram formats were then conducted separately for the low-ability students in each grade. Low-ability sixth graders scored higher on story problems (27% correct) than story + diagram problems (4% correct; $t(25) = 2.29, p < .05, d = .53$). The pattern was opposite for low-ability eighth graders (10% (story) vs. 55% (story + diagram); $t(19) = 2.93, p < .01, d = .98$); no difference was found for low-ability seventh graders (22% (story) vs. 33% (story + diagram); $t(17) = 1.37$, ns, $d = .33$).

**Figure 2.** Experiment 1 performance on story and story + diagram problems by grade for low- and high-ability students.

**Discussion**

Results from Experiment 1 replicated the textual advantage in a younger student population: algebra story problems are easier for middle school students to solve than matched equations. This same pattern holds for both single-reference and two-variable problems. Previous findings suggest that for college students, the effect is reversed: double-reference problems are easier in equation format than story format (Koedinger et al., 2008). This effect was not found directly for middle school students in the present study, likely because the students are not yet familiar with complex equations; however,
it is important to note that for double-reference problems, stories and equations are equally challenging.

This experiment also explored the role of diagrams in algebraic problem solving. Despite a number of reasons why diagrams should be beneficial, this study found no general advantage for including diagrams with story problems, except for with the more difficult double-reference problem. However, students’ ability level and age appear to mediate potential benefits of diagrams. Sixth grade students tended to solve more problems correctly in the story format but seventh and eighth graders solved more problems correctly when a diagram accompanied the story. Low-ability sixth graders appeared to be particularly confused by the diagram, but low-ability eighth graders benefited from its presence.

Experiment 1 provides initial evidence of a diagrammatic advantage for some students, but suggests that students must develop some diagram-relevant capability before any such benefits appear. In Experiment 2, we use a longitudinal design to examine whether and how the sixth graders from Experiment 1, who did not show a diagrammatic advantage, come to benefit when they reach seventh or eighth grade. We investigate the interaction between grade level and presentation format (story vs. story + diagram), and also focus, in particular, on how the performance of low-ability sixth graders changes over time.

An additional goal of Experiment 2 was to evaluate the types of errors students make to begin to explore the nature of any observed benefit. One possible mechanism is that by eliminating the need for students to represent the problem, diagrams may make the problems seem easier, and students may be more likely to attempt to solve them. Alternatively, the diagram may provide necessary supports that afford students a better conceptual framing of the problem. For example, the diagram may make the relations between problem components explicit, reducing students’ need to discern those relations from the words (Larkin & Simon, 1987). The distribution of elements in the diagram also provides visual cues about the appropriate size of the answer (Nunes, Schliemann, & Carraher, 1993; Rittle-Johnson & Koedinger, 2005), and makes it readily apparent that certain tempting solution paths are not valid (Larkin & Simon, 1987) (See Figure 3), which should reduce the conceptual errors students make in solution attempts.  

We explore age- and ability-related differences in no response and conceptual errors in Experiment 2.

EXPERIMENT 2

Method

Participants
Participating in this study were 84 students who had participated in Experiment 1 as sixth graders, including 23 of the 26 low-ability sixth graders and 14 of the 17 high-ability sixth graders.

Procedure
The procedure was identical to that in Experiment 1. In the Fall of their seventh and eighth grade school years, students were given the same form of the test they had taken in sixth grade. Thus, they completed the same problems with the same numbers and in the same format all three years.
<table>
<thead>
<tr>
<th>Conceptual error made on story problem</th>
<th>Diagram supports correct conceptual strategy</th>
</tr>
</thead>
</table>
| 9. Mom won some money in a lottery. She kept $45 for herself and gave each of her 3 sons an equal part of the rest. If each son got $20.50, how much did Mom win? | 9. Mom won some money in a lottery. She kept $45 for herself and gave each of her 3 sons an equal part of the rest. If each son got $20.50, how much did Mom win?  
You can use the picture below to help you solve this problem  
45 Mom kept  
\[ \frac{45}{3} \text{ each son got} \]  
\[ 15 \]  
\[ 15 \text{ How much Mom won} \]  

| 10. Mally bought a coat on sale. It was 1/5 off the original price. She paid $30. What was the original price of the coat? | 10. Molly bought a coat on sale. It was 1/5 off the original price. She paid $30. What was the original price of the coat?  
You can use the picture below to help you solve this problem  
30 paid  
\[ \frac{1}{5} \text{ off} \]  
\[ \frac{30}{5} \]  
\[ 6 \]  
\[ 6 \text{ Original Price} \]  

---

Figure 3. Examples of conceptual errors made by middle school students on story problems (left) and conceptually correct strategies guided by the presence of a diagram (right).

Results

Textual advantage

A 3 (grade: 6, 7, 8) \( \times \) 2 (presentation: equation, story) repeated measures ANOVA on correct answers yielded a main effect of presentation, \( F(1,83) = 4.61, p < .05, \eta^2 = .05 \). More correct responses were given for problems in story form (22% correct) than in equation form (11%). No main effect of grade or interaction was found.

Diagrammatic advantage

A 3 (grade: 6, 7, 8) \( \times \) 2 (presentation: story, story + diagram) repeated measures ANOVA on correct answers yielded a main effect of grade \( F(2,82) = 5.75, p < .05, \eta^2 = .07 \); probability of correct responses increased with grade. There was no main effect of presentation, but the grade \( \times \) presentation interaction was significant, \( F(2,82) = 5.75, p < .05, \eta^2 = .07 \) (Figure 4). Follow-up repeated measures ANOVAs conducted separately by presentation revealed no grade differences within the story presentation, but performance on story + diagram problems improved with grade \( F(1,83) = 9.85, p < .01, \eta^2 = .11 \).

Development of a diagrammatic advantage in low-ability students

To examine how low-ability sixth graders’ performance on the two types of problems changed over time, a 3 (grade: 6, 7, 8) \( \times \) 2 (presentation: story, story + diagram)
repeated measures ANOVA for the 23 low-ability students yielded a main effect of grade and an interaction between the grade level and presentation format variables, both $F(1,22) = 11.73, p < .01, \eta_p^2 = .35$ (Figure 5). Probability of correct responses increased with grade, and separate follow-up repeated measures ANOVAs on grade for each presentation format revealed that low students improved with age on diagram problems ($F(1,22) = 11.73, p < .01, \eta_p^2 = .35$) but not story problems ($F(1,22) = 2.10, ns, \eta_p^2 = .09$).

**Dissecting the diagrammatic advantage: Differences in error patterns by grade and math ability**

Following Koedinger and Nathan (2004), we conducted a qualitative analysis of the types of errors students made on story and story + diagram problems. Cases in which students did not attempt a problem at all were coded as no response errors. Errors that indicated a failure to correctly translate the problem situation into an appropriate solution path (e.g., adding two numbers that should be multiplied) were coded as conceptual errors;
Table 3. Exp 2: Frequency of error types by grade on story and story + diagram problems for all students, low-ability students, and high-ability students

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sixth grade</th>
<th>Eighth grade</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Story</td>
<td>Diagram</td>
<td>Story</td>
<td>Diagram</td>
</tr>
<tr>
<td>All students</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No response</td>
<td>41%</td>
<td>46%</td>
<td>21%</td>
<td>19%</td>
</tr>
<tr>
<td>errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conceptual</td>
<td>58%</td>
<td>60%</td>
<td>63%</td>
<td>47%</td>
</tr>
<tr>
<td>errors&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No response</td>
<td>39%</td>
<td>65%</td>
<td>26%</td>
<td>22%</td>
</tr>
<tr>
<td>errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conceptual</td>
<td>64%</td>
<td>75%</td>
<td>65%</td>
<td>44%</td>
</tr>
<tr>
<td>errors&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No response</td>
<td>50%</td>
<td>29%</td>
<td>21%</td>
<td>29%</td>
</tr>
<tr>
<td>errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conceptual</td>
<td>43%</td>
<td>60%</td>
<td>64%</td>
<td>40%</td>
</tr>
<tr>
<td>errors&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *Percentage computed out of attempted problems (e.g., for students in sixth grade, 58% of attempted story problems, or 34% of all story problems, contained a conceptual error).

this is to be distinguished from errors in computation that occurred when a student tried to solve a problem using a correct solution path (e.g., correctly decided to multiply the two numbers, but made an error in carrying out the multiplication).

Frequency of no response errors is computed out of all trials, correct or incorrect. In order to isolate the effects of each possible mechanism, frequency of conceptual errors is then computed out of all problems that were attempted (i.e., eliminating no response errors). We report error patterns for participants when they were in sixth grade (and least likely to benefit from diagrams) versus eighth grade (most likely to benefit), and then describe error patterns of students of varying math ability.

In eighth grade, students are equally likely to attempt diagram and story problems (19% vs. 21%, respectively), but are less likely to make conceptual errors on diagram problems than on story problems (47% vs. 63%; see Table 3, top right). In contrast, for sixth graders, diagrams do not reduce conceptual errors (60% vs. 58%) and, if anything, may increase no response errors (46% vs. 41%). When we separate the low- and high-ability sixth graders (Table 3, lower left), we find that the low-ability students are much more avoidant of diagram problems than story problems in sixth grade (65% vs. 39% no response errors). The high-ability students attempt diagram problems in sixth grade, but are not able to use them, as both low- and high-ability students do in eighth grade, to reduce conceptual errors.

**GENERAL DISCUSSION**

The longitudinal findings from Experiment 2 largely confirmed and strengthened the findings from Experiment 1. First, the textual advantage was replicated: story problems were easier than equations in both experiments, across all grade levels. The fact that there were no developmental or individual differences in the textual advantage is surprising, given that students experience increasingly more exposure to symbolic equations as they progress through the Connected Mathematics curriculum. However, the equations used in this study were, by design, difficult; perhaps students are developing symbolic capacity, but it is not transferring to more difficult equations. A textual advantage had
previously been demonstrated with high school (Koedinger & Nathan, 2004) and college students (Koedinger et al., 2008) on beginning algebra problems, but not for early elementary school students (Cummins, Kintsch, Reusser, & Weimer, 1988) on arithmetic problems. Thus, development of the textual advantage appears to occur between second grade and ninth grade, perhaps as reading comprehension improves. Results from the present study suggest that the textual advantage is in place by sixth grade, and that story problems may be useful as a transition for students learning to solve abstract, symbolic equations.

Evidence for a diagrammatic advantage was also found, but only for older and high-ability middle school students. Results from error analysis suggest that diagrams improve these students’ performance by increasing the likelihood that they will generate a conceptually correct understanding of the problem situation. Overall, diagrams are beneficial additions to story problems for more accomplished students, suggesting that their use as a transition to abstract, symbolic representations may be constructive for these students.

Unfortunately, results from both experiments also suggest that younger middle school students, and especially those with low math ability, do not benefit from the diagrams. Our error analysis suggests that the main barrier to successful diagram use in sixth grade was the inability to extract a correct conceptual understanding of the problem from the diagram. Whether this is due to misinterpretation of the diagram itself or a failure to accurately map the story problem to the diagram, the diagrams did not reduce conceptual errors.

Low-ability sixth graders also experience an additional barrier to success, as they were even less likely to attempt the diagram problem compared with the story problem. Unlike their more accomplished peers who were likely to find the diagram a helpful tool, for low students, diagrams made the problems less approachable than the story problems alone. This is perhaps not surprising, as low-ability students generally perceive problems as more difficult than high- or average-ability students and are thus more likely to shut down and not attempt the problems (e.g., Ericsson & Simon, 1980). In the present study, low-ability sixth graders may have given up on problems because they experienced or perceived an increase in cognitive load when examining the problem, perhaps due to an (incorrect) assumption that they would need to attend to both the diagram and the story simultaneously in order to solve the problem (Chandler & Sweller, 1991). Because they had limited diagram comprehension skills, their confusion over the diagram, it appears, led them to miss the fact that they could simply ignore it and work from the story alone.

Why did young, low-ability students, who potentially have the most to gain (Murata, 2004), fail to benefit from the diagrams, while older students found success? We join with Ainsworth (2006) and Larkin and Simon (1987) in asserting that a complex interaction of student abilities and task characteristics drives development of successful diagram use. In the case of the present study, it seems likely that sixth graders have not sufficiently developed the necessary component skills or capacities to reap the benefits of diagrams.

A number of changes - both developmental and instructional - occur between sixth and eighth grade that could influence diagram comprehension. First, students develop greater cognitive capacity during adolescence, including increases in working

3 Though a pure textual advantage between word problems and numerical problems has not been found for young students, beginning elementary students do fare better on word problems than on symbolic ones such as "a + _ = b" (Carpenter & Moser, 1984; De Corte & Verschaffel, 1981).
memory (Luna, Garver, Urban, Lazar, & Sweeney, 2004), which should reduce any deleterious effects of cognitive load, and executive control (Luciana, Conklin, Hooper, & Yarger, 2005). With increased capacity, students should be better able to coordinate the information in the diagram and problem statement and make decisions about where to focus; in fact, a recent study shows that success on diagram problems was indeed associated with increased working memory capacity (Lee et al., 2009). Formal reasoning skills, such as the ability to coordinate information gained from multiple sources or to reason logically about possible outcomes, also show considerable development across adolescence, enhancing students’ ability to solve problems in a variety of domains (Inhelder & Piaget, 1958; Marini & Case, 1994).

Instruction helps students build a foundation in pre-algebraic conceptual understanding during their middle school years. This increased domain expertise likely allows older students to make better inferences and connections between the different elements of the problem (Larkin & Simon, 1987; Narayanan & Hegarty, 1998); lack of knowledge about mathematical equality, fractions, or the relationship between number of items and cost could inhibit sixth graders’ performance. Further, as previously mentioned, the Connected Math curriculum aims to increase students’ representational fluency by helping students see connections between text and representations. Even though the style of diagrams presented in the current study are not included in the curriculum, increased exposure to instruction that illustrates the connections between different types of diagrams and text may account for much of the improvement seen in this study. It follows that with a sample that does not use Connected Math, even worse performance on diagram problems compared with story problems and less improvement over time in diagram problems may be found.

In contrast to younger students, older, more experienced students appear to have developed a number of skills necessary to take advantage of the beneficial diagrammatic features. It may be surprising to some readers that older middle school students (seventh and eighth graders in this sample) can benefit from diagrammatic representations without any substantial instruction on how to use diagrams in problem solving. On other hand, advocates for use of diagrams in mathematics education (NCTM, 2000) may be surprised that younger students (sixth graders) do not easily and naturally make use of diagrams. One limitation of this study is that we cannot determine much about why sixth graders are failing, in particular, whether there is a developmental barrier of some kind or whether there is particular prerequisite knowledge for diagram processing that they lack. Our results indicate, however, that further research is warranted to understand how this promising practice of Asian curricula might be better introduced in or before sixth grade to support pre-algebraic reasoning or whether it should be delayed to later grades.

Interestingly, there is evidence of similar trends for adolescents learning to comprehend graphs and maps, in which younger adolescents had more difficulty extracting implicit or conceptual information from the representations; a combination of increases in domain-specific content knowledge and age-related development of ‘graphicacy’ is purported to account for these changes (Postigo & Pozo, 2004). Is there a developmental threshold that should be reached before which any of these representations should be introduced with caution, or perhaps not used at all? There is a period of time in elementary school in which students must learn to read before they can be expected to gain new information from text (Chall, 1979). Similarly, perhaps there is a period of time before which students cannot be expected to glean new information from diagrammatic and graphical representations; they must first learn to comprehend and
use these representations. Further research is needed to tease apart the different aspects of development and knowledge acquisition that are occurring during the middle school years and determine which component skills or capacities are most crucial for successful use of diagrams and other visual representations.

**Improving students’ use of diagrams as problem solving aids**

If a representation (concrete, abstract, diagrammatic) is ever going to be helpful as a scaffold for learning and transfer, it needs to at least be easier. Can instruction improve students’ ability to use the diagrams when solving algebraic story problems? Research in both mathematics education and cognitive development suggests that students need effective instruction in order to use any type of external representations correctly (Fueyo & Bushell, 1998; Sowell, 1989; Uttal, Scudder, & DeLoache, 1997). Brief instruction on the nature of the particular visual representation may not be sufficient (Rittle-Johnson & Koedinger, 2001), but more effective representation-specific instruction could be time consuming, and may not transfer to other novel representations that students encounter. However, such instruction on different types of diagrams in the Connected Math curriculum may partially account for the growth found in the present study.

Having students construct diagrams to represent the story problems themselves may improve their use of given representations to solve problems. Middle school students are able to use representations they created themselves to successfully solve algebraic story problems (Koedinger & Terao, 2002), and constructing diagrams from scratch or completing partial diagrams has been shown to increase student learning in a variety of domains (e.g., Lewis & Mayer, 1987; see Van Meter & Garner, 2005 for a review). The construction process likely helps students learn to coordinate and integrate text with visual representations (Ainsworth, 2006; Easterday, Aleven, & Scheines, 2007), which may in itself improve students’ use of diagrams to solve algebra story problems and their subsequent ability to correctly represent the problem in an equation format. These types of activities also have the potential to increase students’ meta-representational fluency (diSessa & Sherin, 2000) such that students may come to build a broader understanding of representations in general, to understand the advantages and disadvantages of different representations, and to appreciate the usefulness and relevance of certain types of representations for problem solving (di Sessa, Hammer, Sherin, & Kolpakowski, 1991). Further research is needed to test whether such instruction can impact student learning to use diagrams for algebraic word problems.

Regardless of the type of instruction necessary to make the diagrams accessible to younger and low-ability students, such efforts may be well warranted. Crucial concepts can be learned via grounded, concrete experiences that help to scaffold students to more abstract understanding (Bransford et al., 1999; Goldstone & Son, 2005; Koedinger & Alibali, 1999; Moreno & Mayer, 1999; Schwartz & Black, 1996). Struggling students are arguably the most in need of this type of scaffolding, and may well be the reason that well-meaning teachers include diagrams as aids to understanding. However, in order to derive the intended benefit from the diagrams, these students must have gained enough familiarity and competence with diagrammatic conventions that adding diagrams enhances performance. Teachers (and even instructional designers) should implement simple formative assessments (e.g., a few quiz problems with and without diagrams) to determine when a particular kind of diagram may or may not be a scaffold for mathematical understanding for a particular student level. Instructors and designers should exercise caution when using diagrams in attempt to ease the transition to more
abstract, symbolic representations; those diagrams may confuse or hinder the progress of the very students who are in most need of assistance.

**Acknowledgements**

Funding for this research was provided by the Interagency Education Research Initiative (IERI), NSF Award Number REC-0115635, “Collaborative Research: Understanding and Cultivating the Transition from Arithmetic to Algebraic Reasoning”. Portions of this work were presented at the biennial meeting of the Society for Research in Child Development in Boston, MA, USA.

**References**


Received 14 April 2010; revised version received 3 June 2011