Modeling Abstract Numeric Relations Using Concrete Notations

David Landy (dhlandy@gmail.com)
Department of Psychology, University of Richmond
Richmond, VA, 23173

David Brookes (dtbrookes@gmail.com)
Florida International University
11200 SW 8th St, CP 204
Miami, FL 33199

Ryan Smout (ryan.smout@richmond.edu)
Department of Psychology, University of Richmond
Richmond, VA, 23173

Abstract

Abstract relational reasoning is a core component of human thinking. While abstract relations are understood using a wide variety of methods, the formal algebraic equation is among the most powerful and general mechanisms for representing relational statements. It has often been assumed that the means by which expressions represent relations are purely semantic, and are encoded in an abstract syntax that governs the use of notation without regard to the details of its physical structure (Anderson, 2005; Hegarty, Mayer, & Monk, 1995). In contrast, we propose an image of equation construction that highlights the role of concrete physical relations in mediating the interpretation of equations. In this account, construction processes involve a structural alignment across representation systems. Alignment biases reasoners toward the selection of representations that maintain the concrete structure of source representations. We demonstrate that this approach accounts naturally for a variety of previously reported phenomena in equation construction, and correctly predicts several new phenomena.

Keywords: Relational reasoning, analogy, psychology, education, problem solving

Introduction

Formal mathematics is among the most powerful means we have for dealing with abstract relational assertions. While mathematics uses a wide variety of tools to represent relations, for the last 500 years, the mathematical expression or equation as written in the formalism of algebraic notation has been the most ubiquitous and recognizable way to express abstract relationships. Understanding the use of formal symbol systems is thus an important part of understanding relational reasoning more generally.

Given the importance of algebraic formalisms in mathematics generally as well as in middle-school and high-school mathematics curricula, it is perhaps surprising that many people have great trouble constructing and interpreting even basic expressions written in it (e.g., Koedinger, Alibali, & Nathan, 2008). In particular, reasoners have great difficulty constructing and making sense of *relational equations*—equations that assert a quantitative relationship between two entities. When Clement, Lochhead, and Monk (1981) asked undergraduate engineering students to write an equation representing the natural language expression “There are six times as many students as professors,” they found that 37% percent of their participants made errors. The great majority of those errors reversed the appropriate relation, (e.g., $6S=P$ instead of $6P=S$). Subsequent research established clearly the difficulty of these and other relational equations (e.g., Clement, 1982; Mestre & Lochhead, 1983; Hegarty, Mayer, & Monk, 1995; Martin & Bassok, 2005).

In this paper, we provide an account of the source of reversal errors motivated by perceptually grounded theories of notation interpretation and use. Errors on relational equations result from a structural mismatch between the surface structure of typical relational comparison statements in English and the structure of default (multiplication) equations. Before turning to our account, we review existing accounts for the difficulty of relational equations.

Two systems accounts of equation construction

Most research in relational equations posit two distinct strategies: a normative reasoning process, which leads from situation descriptions to correct relational equations, and some sort of shallow surface strategy which leads to errors (e.g., Clement, 1982; Fisher, Borchert, & Bassok, in press; Hegarty, Mayer & Monk, 1995; Martin & Bassok, 2005). On this kind of account, successful reasoning relies on building a mental model of the semantic relations specified in a situation description. The relations encoded in the mental model can then be semantically converted into the relations of notational algebra, and this semantic alignment is used to construct a problem representation.

Two systems accounts also posit a second route, in which reasoners use a heuristic or shortcut instead of modeling. The most commonly discussed second route is *left-right transcription*. Reasoners using this strategy would directly transcribe words from a story problem into mathematical symbols in a left-right manner, without regard for the meaning of the mathematical symbols (Clement, 1982; Clement, et al., 1981; Hegarty, Mayer, & Monk, 1995).

Fisher, Borchert, and Bassok (in press) proposed a two systems account that relies on their demonstration of two facts. First, they persuasively demonstrate that inexpert algebraists often consider multiplication equations as a ‘default’ algebraic expression, and treat division equation more-or-less exclusively as denoting arithmetic operations. Second, they speculate that reliance on multiplication as a standard form leads students into reversals, because multiplication equations’ left-right structure affords (incorrect) left-right transcription. Fisher et al. evaluated

Table 1: Ways to phrase quantitative relational assertions in English.

Phrasing Type	Examples	Variable Type	Equation Model	Aligned?
Direct Comparison	There are four screws for every nail.	Count	4N=S	No
	There are five rhinos for every six elephants.	Weight	5R=6E	Yes
	Sally has seven more coats than hats.	Count	7+H=C	No
Hypothetical Comparison	If there were four nails for every nail there really is, there would be as many screws as nails.	Count	4N=S	Yes
	If Sally had seven more hats, she would have as many hats as coats.	Count	7+H=C	Yes
	If there were three magazines for every magazine there actually is, there would be as many magazines as journals.	Weight	3M=J	No
Operation	Multiplying the number of nails by four yields the number of screws.	Count	4N=S	Yes
	Six times the number of rhinos is the number of elephants times five.	Weight	6E=5R	No

their proposal by asking participants to construct division models of relational statements, as in

$$\frac{P}{6} = S$$

In this case left-right transcription is impossible, ostensibly forcing students to engage in more sophisticated modeling. Indeed, Fisher et al. report a drastic decrease in the number of reversal errors when divisions are required. In one study, 29% of multiplication equations were reversals, compared to only 8% of division equations.

Our account builds on Fisher et al.'s conclusion that students prefer to use multiplication operations in algebra equations, and indeed our account also predicts that a greater willingness to use division equations would correspond to increased success in relational equation modeling. However, our proposal for *why* reasoners are more successful with division equations in their task is quite different from Fisher et al.'s.

Structural alignment of concrete relations

The pattern of errors in relational equation construction is quite robust, and has been replicated many times. However, it is not clear that the two systems account is the best way to explain these data. First, reversals are often produced even when transcription seems impossible. For instance, Mestre and Lochhead (1983), analyzing across several studies involving different populations of college students in Israel, the US, and Japan, reported that reversal rates either stayed the same or increased when the verbal statement "In one school, there are six times as many students as there are teachers," was replaced with an aerial photograph of 5 cows and 1 pig in a field. Furthermore, students are often more successful at solving problems phrased in natural language than in algebraic syntax (Koedinger & Nathan, 2004; Koedinger, Alibali, & Nathan, 2008). These results suggest that the difficulty students face in solving such problems lies in their ability to work with algebraic notation, not in their ability or willingness to interpret relationships.

We think that the explanation for the difficulty of relational statements lies in a greater appreciation for the ways we interact with formal expressions. Landy & Goldstone (2007; see also Kirshner & Awtry, 2004) argue that our use of notation is intimately bound up with the concrete physical structure of the algebraic notation—people use proximity, for instance, as a strong cue to the binding structure of formal expressions.

We propose the *concrete alignment view* as a novel explanation of difficulties students have in constructing and interpreting relational algebraic expressions. On this account, students construct equations through constructing representations and using relational alignment (Gentner, 1983) to reidentify terms, objects, and relations across representations, using as guides both semantic features and the concrete relations that inhere in physical notations. Errors can crop up in either the construction or the mapping processes, but the concrete alignment view attributes the primary source of reversals to the mapping stage.

The concrete alignment view accounts for past results by noting the general mismatch between the physical structure of comparison statements in English and (multiplication) statements in algebra, illustrated in Table 1. English comparatives such as "There are six students for every professor" bind through textual proximity and phrasal structure the relating quantity "six" to "students." Algebraic multiplication statements in which variables represent counts make the opposite binding.

Table 1 illustrates that rephrasing the comparison as an operation automatically binds the relation to objects in the same manner as a multiplicative equation. In a similar way, when Fisher et al. (in press) instructed students to write division equations, they asked them to construct equations that naturally matched the physical structure of the direct comparison. On our interpretation, typical relational problems are hard because the structure of the phrasing mismatches the structure of problem, and aligning structures whose concrete features mismatch is generally hard.

The concrete alignment view is compatible with a wide variety of published results. To evaluate it further, we tested several combinations of phrasing and operations not previously reported. Here we report results of three variations on this theme: participants read a problem in one of the above phrasings, and constructed an equation.

Experiment 1

Method

Participants 16 undergraduates attending the University of Richmond received partial course credit for participation.

Design We constructed sixteen relational equation problems. Target items were separated by a multidigit arithmetic problem that served as a distracter. Each target described in a short paragraph (2-4 sentences) two sets of similar objects (e.g. screws and nails) on opposite sides of a balanced scale. The critical sentence described the numerical relationship using two relatively prime constants. Participants filled appropriate numbers and variables into an *equation frame* consisting of an operation and *equals sign*.

The test problems varied along three dimensions: phrasing (direct comparison or operation), equation frame format (multiplication or division), and variable type (count or weight). The "rhino" sentences from Table 1 provide examples of the form of the direct and operation comparison statements used in Experiment 1.

On *weight* problems, students wrote an equation using variables to represent the *weight* of each object, rather than the number of objects. Mathematically, this has the effect of inverting the concrete relations of the correct equation without affecting those in the text. For instance, if there are four nails for every screw, and the total weight nails and screws is equal, then each screw weighs as much as four nails. *Count* problems asked participants to construct more typical equations in which variables stand for set sizes.

Predictions The basic prediction of the concrete alignment account is quite simple: accuracy will be highest when the text places closely together terms that should be placed closely together in the correct equation—that is, when the equation *aligns* with the text (see table 1). Note that on division problems, alignment is reversed from multiplication problems. Since each experimental dimension reverses the alignment of text and equation, this model predicts a three-way interaction between phrasing, variable type, and equation frame.

The two strategies account also makes clear predictions. On problems with a multiplication equation frame, participants may engage in left-right transcription, so accuracy should depend on whether that transcription is correct (transcription yields correct count equations in the operation condition, and weight equations in the direct comparison condition). On division frame problems, participants will be unable to engage in left-right transcription, and so will be forced to model. The difficulty of the problem will depend on how difficult it is to extract the relevant information from the text. Crucially, on this account comparison phrasing and problem type act roughly independently, so that these factors should not interact in determining performance.

This prediction provides a strong way to discriminate the two strategies account, which predicts main effects of phrasing and variable type when division problems are constructed, from the concrete alignment view, which predicts a three-way interaction.

Results

Results were analyzed using nested mixed-effects logistic regression models, including main effects of phrasing, frame, and variable type, as well as interactions. Items on which participants wrote equations or notes outside the provided equation frame were excluded from analysis.

Reversal rates are shown in Figure 1. The model including a three-way interaction between phrasing, equation frame, and variable type fit the data better than the model including only main effects by a likelihood ratio test, ($\chi^2(4) = 76$, p<.001), and than the model including only two-way interactions ($\chi^2(1) = 58$, p<.001). This was also the best fitting model overall.

The concrete alignment view uniquely predicts an interaction between variable type and phrasing for division problems. We separately explored this two-way interaction by computing mixed-effect logistic regressions using just division problems. The model that included a two-way interaction term fit the data substantially better than a model including only main effects ($\chi^2(1) = 6.9$, p<.01). Examination of this model revealed that weight problems were associated with more reversals than number problems (e^β=10.4, z=2.8, p<.01), and operation language led to marginally more errors than direct comparison (e^β=5.2, z=1.9, p~.06). The interaction was also significant, such that problems that were both weight problems and expressed using operation phrasing were solved with fewer errors than problems that contained just one of these.

Discussion

The pattern of results closely matched the predictions of the concrete alignment view. Reversal rates on problems in which relation binding in English mismatched that of mathematics were very high (averaging 76%); when the concrete features of the two situations aligned, the reversal rate was just 13% on average. The pattern is weaker in division than in multiplication. The simplest account for this difference is suggested by Fisher et al.: multiplication is the default pattern in algebra, so many students may initially have conceived a multiplication expression, and converted it mentally in order to fit it into the required frame.

These problems were clearly quite challenging for our participants. The within-subjects design also meant that participants were faced with one difficult problem after another, and may not have felt motivated to model each situation individually. Additionally, the operation statements were written in a manner that could easily be transcribed, which, while it indeed matched the patterning

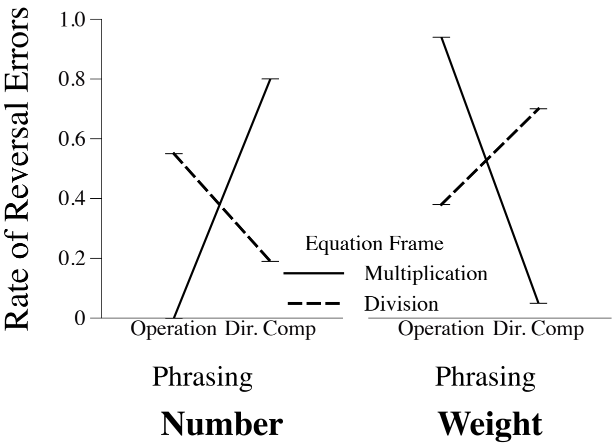

Figure 1: Reversal errors in Experiment

of algebra, was also fairly unnatural as an English sentence. These properties may have inclined students away from meaningful semantic modeling, or encouraged an arithmetic approach. Experiment 2 controls for these problems by giving participants extremely simple problems associated with a high degree of success, phrased in the relatively natural hypothetical comparison structure.

Experiment 2

Method

Participants 100 undergraduates attending the University of Richmond received partial course credit in exchange for participation. 5 participants did not complete the study, or did not fill their responses into the blanks as instructed. These participants were eliminated from analysis.

Design and Procedure Participants completed a written test containing 32 story problems, half of which were targets and half unrelated distracters. Each target item described in a single sentence the relative quantities of two sets of similar objects. Test problems varied along two factors: *comparison type* and *phrasing*.

The type of comparison could be *more* or *fewer*. For instance, if in the target situation there are eight more nails than screws, then half of all participants saw the comparison statement "there are eight more nails than screws", while half saw "there are eight fewer nails than screws." Each participant translated 8 comparisons of each type.

Problems were phrased as either direct comparisons or hypothetical comparisons. Direct comparisons relate unequal quantities, as in "Alex has five more pocket watches than wristwatches". Hypothetical comparisons describe what would make the two sets equal, e.g., "If Alex had five more pocket watches she would have as many pocket watches as wristwatches."

Participants used experimenter-specified variables to fill in an equation frame that modeled the described situation. The frame included three blanks on the left side of an equals sign, and one blank on the right. All comparison statements could be modeled by either an addition or a subtraction equation; the participant chose which operation to use.

Predictions On the concrete alignment view, hypothetical comparisons should be formalized more readily than direct comparisons, because their concrete and semantic properties align well with the same correct response.

The two strategies account does not make a strong prediction about accuracy in this case. On direct comparison statements, some reasoners will be 'lured' into making left-right transcriptions, while others will engage in modeling. In the hypothetical comparisons, left-right transcription is blocked by the phrasing, but students who engage in modeling will arguably have a harder time doing so, since the relations involved are less transparent in English. Thus, which phrasing is associated with higher error rates will depend on whether more students are lured into transcription in the direct comparison, or more students make modeling errors in the hypothetical.

However, the two models do make different predictions about the specific pattern of responses, because the constant-variable patterning of a "more than" statement is slightly dispreferred in algebra notation. That is, if the sentence "Sally has seven more coats than hats" is transcribed, the result is "7 + c = h" even though the version "c + 7 = h" is slightly preferred (MacGregor & Stacey, 1993). While students engaging in left-right transcription may correct this and other minor oddities in producing expressions, they may not. On the other hand, there is no pressure in the two systems account for students to invert expressions in this way in the hypothetical comparison. Thus, on the two systems account reasoners are more likely to invert in the direct than the hypothetical comparison.

Just the opposite conclusion follows from the concrete alignment account. Forming correct equations from direct comparison statements requires ignoring either physical or semantic cues, making matching the physical pattern less likely than in the hypothetical case, in which physical relationships are a sound guide to correct responding.

Results

The main results are presented in Figure 2. Overall accuracy was very high (M=.9, SE=.01). Logistic regression models were evaluated with error rate as the dependent measure, and phrasing (direct vs. hypothetical comparison) and comparison type (more vs. fewer) as independent factors. The model including both factors fit better than a model including only phrasing ($\chi^2(1) = 6.1$, p=0.01), and one including only comparison type ($\chi^2(1) = 11.2$, p<.001), and was the best-fitting model overall. Problems containing the "more" language were solved more readily than problems using the "fewer" relation ($e^\beta=1.6$, $z=2.4$, $p=.02$); hypothetically phrased problems were solved more easily than direct problems ($e^\beta=1.82$, $z=3.3$, $p=.001$).
A logistic analysis regressing relative ordering of the constant and symbol against phrasing did improve the fit over a null model ($\chi^2(1) = 15.6$, p<0.001; see Figure 3).

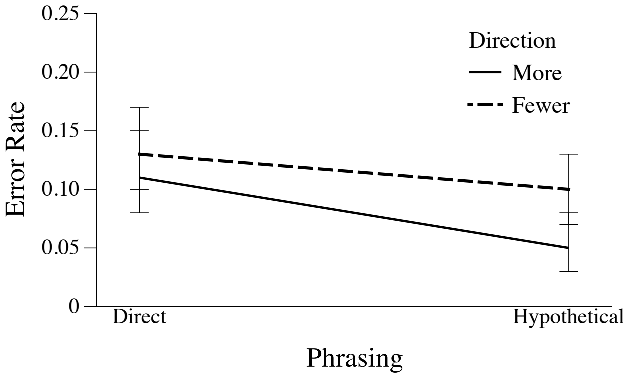

Figure 2: Error rates in Experiment 2. Errors bars on all graphs depict standard errors.

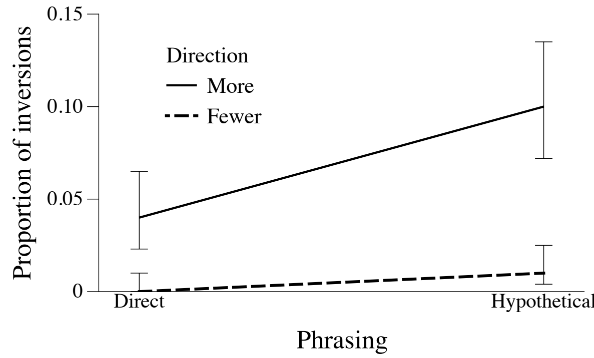

Figure 3: Proportion of responses that were inverted in Experiment 2.

Inversions were more common in the hypothetical than the direct comparisons. Not surprisingly, inversions were also more common on "more" problems than on "fewer" problems; a model that fit main effects of both comparison type and phrasing provided a better fit than one containing just phrasing ($\chi^2(1) = 71$, $p<0.0001$) and one containing just comparison ($\chi^2(1) = 19.7$, $p<0.0001$). Including the interaction did not improve the fit ($\chi^2(1) = 1.2$, $p\sim.27$).

It is possible that the use of physical spacing on hypothetical comparison problems was an artifact of the within-subjects nature of the design; participants may have noticed early on that physical spacing was a reliable guide, and employed it on later problems in consequence. An examination of just the first problems seen by each participant indicated that this was not the case: considering only the first problem seen by each participant, inversions were still more common on hypothetical than direct problems (the model including both factors fit better than the model just including comparison type, $\chi^2(1) = 11.8$, $p<.001$), and was the best-fitting model overall.

Discussion

As predicted by the concrete alignment view, hypothetical comparatives were easier for participants to correctly solve. More interestingly, inversions were more frequent in this condition than in the direct comparison condition. Rather than transcription accounting for errors, in this case "transcription" is selectively used when it is most likely to lead to correct responses. This difference cannot be accounted for by a two systems account, which views transcription as a shortcut to avoid complex thinking. However, it is quite natural in the concrete alignment account in which surface features contribute as a core component of the general modeling process.

The two systems view might accommodate some results of Experiments 1 and 2 by generalizing the simple heuristic shortcuts to include alignments of concrete elements in addition to simple transcription. To evaluate this possibility, we next consider a more sophisticated case of equation construction, which requires the reasoner to select, remember, and apply an appropriate equation. This task cannot be accomplished without some modeling, so we can use it to evaluate whether mapping occurs in modeling.

Experiment 3

Method

Participants Participants were 32 undergraduates at the University of Illinois who had recently completed introductory physics participating in exchange for monetary compensation.

Procedure Participants completed a short test containing several elementary mechanics problems, and other distracters. Some problems were parts of other experiments, which will not be discussed here. The target problem was the fifth problem in a set of sixteen**.** In this problem, participants were told a story about several asteroids (with masses m_1, m_2, and m_3) and a single asteroid (mass m). For each pairing of the single asteroid and the other asteroid, the participants were to construct the Newtonian gravitation equation,

$$F = G\frac{m_1 m}{r^2}.$$

Participants were not reminded of the gravitation equation, and had to retrieve it from memory. Thus, transcription is not possible in this case, and some modeling was required to select and construct an appropriate equation.

Constructing this equation requires deciding which mass to place on the left, and which on the right. If the alignment of concrete features is a core component of equation construction, then in a situation with few semantic factors, participants should tend to place terms in the order in which they were introduced in the problem. If concrete alignment is a shortcut separate from modeling, there is no particular reason why one term should be placed to the left. The order in which the terms were introduced was counterbalanced: in one condition, the single asteroid was described first, and was described as the agent (it moved from asteroid to asteroid). The other condition introduced the asteroids first, which were described as moving past the asteroid.

Results 27 participants responded either correctly, or made only minor errors not relevant for our purposes (such as neglecting to square the distance in the denominator). Of these, 24 (12 in each counterbalancing condition) placed the term that was introduced in the narrative first in the equation ($p<.001$ by Fisher's exact test). Two participants in the asteroids-first condition and one in the asteroid-first condition placed the masses in the inverted order.

Discussion The effect of the order in which terms appear in a problem statement on how reasoners order equation terms is not limited to cases in which it provides a suitable shortcut. This suggests that relational equation reversals are ubiquitous because the processes involved in modeling an equation involve relational mapping, and the attended relations include concrete aspects of algebraic notation.

General Discussion

Three experiments matched predictions made by the concrete alignment account of equation use: First, participants constructing division equations were affected by relation binding in problem statements, just as are people writing the more common multiplication equations. Second, participants freely constructing addition or subtraction equations did so more successfully when the problem statement afforded maintaining both concrete structure and semantic features. Third, participants proved more likely to match low-level physical structure when doing so led to correct answers than when it did not, suggesting that the use of physical structure occurs after or along with semantic processing. Finally, participants matched concrete details even when the problem did not afford any simple heuristic solution.

Beyond simplifying existing accounts of empirical phenomena and providing new testable predictions, by making the alignment of concrete notations a central component of correct equation construction, the current proposal suggests approaches to teaching students how to read and understand equations. In particular, it suggests that rather than trying to instruct students that physical structure is irrelevant, or exclusively focusing on the intra-mathematical articulation of implications, it may be possible to help students understand equations as sensible utterances by providing interpretation routes (i.e., mappings onto natural-language descriptions or imagistic models) that are both interpretable and maintain concrete relational structure. That is, rather than seeing mappings like this as a shortcut to be averted, we can see them as a route to potential understanding.

Finally, this research mirrors suggestions that our ability for wholly abstract relational reasoning may be strikingly limited. Algebra and its accompanying notation are a paradigmatic case of purely symbolic thought. That experienced users of algebra rely on concrete physical structure suggests the interpretation that purely symbolic thought is itself largely achieved not though complex abstract internal resources, but through the cooption—in this case via external formal notations—of resources typically devoted to representing concrete relations and features.

Acknowledgments

We gratefully acknowledge funding from a research grant from the University of Richmond, and by Department of Education, Institute of Education Sciences grants R305H050116 and R305A110060.

References

Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science*, *29*(3), 313–341.

Clement, J. (1982). Algebra word problem solutions: Thought processes underlying a common misconception. *Journal for Research in Mathematics Education*, *13*(1), 16-30.

Clement, J., Lochhead, J., & Monk, G. S. (1981). Translation difficulties in learning mathematics. *American Mathematical Monthly*, *88*(4), 286-290.

Fisher, K. J., Borchert, K., & Bassok, M. (in press). Following the standard form: Effects of equation format on algebraic modeling. *Memory & Cognition*.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of educational psychology*, *87*, 18–32.

Kirshner, D., & Awtry, T. (2004). Visual salience of algebraic transformations. *Journal for Research in Mathematics Education*, *35*(4), 224–257.

Koedinger, K. R., Alibali, M. W., & Nathan, M. J. (2008). Trade-offs between grounded and abstract representations: Evidence from algebra problem solving. *Cognitive Science*, *32*(2), 366-397.

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*(2), 129-164.

Landy, D., & Goldstone, R. L. (2007). How abstract is symbolic thought? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 720-733.

Mestre, J. P. and Lochhead, J. (1983). The Variable-Reversal Error among Five Cultural Groups. In Bergeron, J. and Herscovics, N., editors, *Proceedings of the Fifth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, pages 180-188, Montreal, CA. International Group for the Psychology of Mathematics Education.Education, pp. 180 - 188.

MacGregor, M., & Stacey, K. (1993). Cognitive models underlying students' formulation of simple linear equations. *Journal for Research in Mathematics Education*, *24*(3), 217–232.

Martin, S. A., & Bassok, M. (2005). Effects of semantic cues on mathematical modeling: Evidence from word-problem solving and equation construction tasks. *Memory & Cognition*, *33*(3), 471-478.