How do people evaluate problem-solving strategies? Efficiency and intuitiveness matter

Sarah A. Brown (sbrown23@wisc.edu)
David Menendez (dmenendez@wisc.edu)
Martha W. Alibali (mwalibali@wisc.edu)
Department of Psychology, University of Wisconsin – Madison
1202 W. Johnson Street, Madison, WI 53706 USA

Abstract

What factors affect whether learners adopt a new problem-solving strategy? Potential factors include learners' evaluations of alternative strategies and the degree of similarity between their existing strategy and the alternatives. A first step in answering this question is investigating how people evaluate strategies. This exploratory study investigated how people evaluate strategies for solving algebraic word problems, and how these evaluations vary as function of individual differences. Undergraduates rated three strategies on six dimensions and judged each pair of strategies for similarity. Factor analysis showed that evaluations could be reduced to two constructs: efficiency and intuitiveness. We calculated factor scores for each participant for each strategy. Efficiency did not predict similarity ratings on its own, but it did interact with Need for Cognition. These results suggest stable learner characteristics and moment-to-moment evaluations of strategies influence judgments about strategy similarity.

Keywords: strategy change; problem solving; instruction

Introduction

Why do people shift from using one problem-solving strategy to using another? Explaining why people change their minds is fundamental for theories of learning (Siegler, 2000). One factor that has been shown to promote strategy change is exposure to alternative strategies (e.g., Booth, Lange, Koedinger & Newton, 2013; Brown & Alibali, 2018; Fyfe & Rittle-Johnson, 2016), a common occurrence in classrooms. Thus, understanding how exposure to alternatives leads to strategy change is also relevant for educational practice.

Past research has focused on two broad classes of factors in explaining patterns of strategy change. Some studies have focused on characteristics of the learners, such as their level of ability or achievement within the target domain (e.g., Torbeyns, DeSmedt, Ghesquiere & Verschaffel, 2009), their encoding of problem features (e.g., Siegler, 1976), or their confidence in their existing strategies (e.g., Brown & Alibali, 2018). Other studies have focused on features of the context in which a new strategy is presented, such as whether the context is more abstract or more situated (e.g., Bassok & Holyoak, 1989), or whether the strategy is associated with a specific person. For example, Riggs, Alibali, and Kalish (2015; 2017) presented middle-school students and adults with a novel strategy for solving a specific type of algebraic problems. For some of the participants, the novel strategy was introduced as tied to a specific person (e.g. "Molly's strategy"), and for others, it was not. Both middle-school students and adults were more likely to use the novel strategy when it was not tied to a specific person.

Past work demonstrates that the likelihood of strategy change depends on characteristics of the learners and on the way in which alternative strategies are presented. In this research, we consider an additional factor: how learners perceive the novel strategies to which they are exposed. We suggest that learners' evaluations of particular strategies also influence their likelihood of strategy change. More specifically, we expect that the degree to which learners view a novel strategy as a good alternative to their current approach affects how likely they are to adopt that novel strategy. From this perspective, it is important to understand how learners evaluate different strategies, and to determine the features or dimensions that people use to distinguish "good" strategies from "bad" strategies.

Past research provides some limited information about the dimensions that people consider when evaluating strategies. For example, in a previous pilot study, we asked child participants to explain their ratings of how "smart" various strategies were (for a partial analysis of these data, see Alibali & Prather, 2007). We found that participants invoked a range of dimensions, including the accuracy of the strategy, the difficulty of applying the strategy, and the overall "goodness" or appropriateness of the strategy. Other past research has shown that adults tend to prefer strategies that they view as efficient (e.g., Walsh & Anderson, 2009). In the present study, we considered a range of dimensions along which people could evaluate strategies.

It is also of interest to understand whether participants view strategies as "similar" to one another. Participants may be more willing to adopt a strategy that they view as largely similar to one they are currently using, and less willing to adopt a strategy that they view as "far out" or unusual. Thus, in addition to evaluations of individual strategies, we also sought information about how similar participants viewed pairs of strategies.

Constant Change Problems

We address these issues in the domain of constant change problems, which are algebraic problems in which a rate changes constantly from the beginning to the end of an interval of time or space, as in the following example:

> Fuel is pumped into the fuel tank of an airplane for a period of 13 minutes. The rate at which it is pumped increases steadily over the interval from 30 gallons per minute to 186 gallons per minute. How many gallons are pumped into the tank over the 13-minute interval?

These problems are well-suited for studying strategy change because there are multiple strategies that can be used to find the correct answer. In this study we focused on three strategies, the Discrete, Gauss, and Area strategies.

For the problem shown above, the *Discrete* strategy involves finding the constant by which the rate changes minute to minute, using this constant to calculate the number of gallons pumped into the tank in each minute, and adding these numbers together. We refer to this strategy as the *Discrete* strategy because it treats the change in rate as occurring in discrete increments. The Discrete strategy is the most common strategy used to solve constant change problems by college undergraduates (Riggs et al., 2015). Although it yields the correct answer when implemented correctly, it is error prone and time consuming.

The *Gauss* strategy involves adding the initial and final rate and multiplying the sum by half of the time period (e.g., $(\mathbf{30}+\mathbf{186}) \times \frac{\mathbf{13}}{\mathbf{2}}$). This strategy is less common than the discrete strategy, but it is an efficient and accurate way to solve constant change problems.

The *Area* strategy involves creating a geometric model of the problem (see Figure 1) and finding its area using either an algebraic formula (i.e., for the area of trapezoid) or using integration. Spontaneous use of this strategy is rare among college students. Although this strategy yields the correct answer, it may be perceived as complex because it involves drawing a model of the problem. The Area strategy may also be viewed as less intuitive because the link between the area of a diagram and the amount of fuel pumped into the tank may not be readily apparent to all learners.

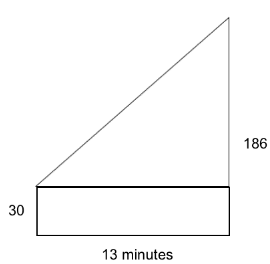

Figure 1. Sample drawing for the Area strategy.

Individual Differences

Past research has suggested that individual differences in ability and experience influence people's strategy choices. These factors may influence strategy choice by affecting strategy evaluations. In this research, we examine whether patterns of strategy evaluations and similarity judgements vary as a function of participants' mathematics ability and their inclination to engage in effortful cognitive activities, termed their Need for Cognition (hereafter, NfC, Cacioppo & Petty, 1982). We expected that mathematics ability might influence participants' understanding of the target strategies, and NfC might influence their evaluations of features of the strategies such as complexity. Some research also suggests that NfC is related to participants' willingness to adopt new strategies (Menendez, Brown & Alibali, 2017).

Current Study

In brief, the goal of this exploratory study was to investigate how people evaluate different strategies. This study is a first step in determining how people's perceptions of strategies influence strategy change. We presented participants with constant change problems and introduced each of the strategies described above. Participants were asked to rate each strategy along six dimensions: goodness, commonness, complexity, length, easiness to remember, and intuitiveness. Finally, participants were asked to judge how similar or different the target strategies were from one another.

Our first goal was to determine whether these six dimensions could be consolidated into a smaller number of factors. If this were the case, the resulting factor structure could give insight into the dimensions that participants find relevant when evaluating strategies. A second goal was to investigate potential relations between these factor scores and similarity ratings. Are strategies that differ more in their factor scores rated as more different from one another than strategies that are more similar in factor scores? Finally, we examined whether this relation changed as a function of individual differences in NfC and mathematics ability.

Method

Participants

Participants were 32 undergraduate students (20 female, 12 male) from a large Midwestern university; they received extra credit in Introductory Psychology in exchange for participation. The sample was 66% Caucasian, 25% Asian or Asian-American, 6% Native American or Pacific Islander, and 3% Hispanic or Latinx. Undergraduates completed the study individually in a laboratory setting and provided written consent at the outset of the session.

Materials

In the first part of the experiment, each participant was presented with two constant change problems. For each problem, participants saw worked examples of three different strategies, one at a time. For each strategy, participants made six different ratings. Thus, this part of the computerized survey included a total of 36 rating questions.

Each strategy was presented on its own page. For each strategy, participants were asked to rate how good, common, complicated, and easy to remember the strategy was. In addition, participants rated how long it would take to implement the strategy and the degree to which the strategy made sense. All ratings were made on a continuous slider scale from 1 (e.g., not at all) to 5 (e.g., very much). No other anchors were given.

Table 1. Mean ratings (and SD) for each strategy. For each column values with different subscripts indicate significant differences. Across rows, subscripts have no meaning.

	Goodness	Makes sense	Commonness	Complexity	Easiness to remember	Length
Discrete	3.79 *(0.88)* ^{a}	3.94 *(0.75)* ^{a}	3.55 *(0.88)* ^{a}	2.89 *(0.89)* ^{a}	3.41 *(0.83)* ^{a}	3.25 *(0.88)* ^{a}
Gauss	2.70 *(1.10)* ^{b}	2.80 *(1.09)* ^{b}	2.75 *(1.00)* ^{b}	2.27 *(0.86)* ^{b}	3.54 *(1.03)* ^{a}	2.19 *(0.91)* ^{b}
Area	2.94 *(1.04)* ^{b}	3.08 *(1.03)* ^{b}	2.28 *(1.02)* ^{c}	2.60 *(1.00)* ^{a,b}	3.37 *(0.94)* ^{a}	2.67 *(0.79)* ^{c}

To minimize demands, participants first completed the ratings for all three strategies for the first constant change problem, before moving to the next problem. The order in which participants saw the first and the second problem was counterbalanced. Within problems, the order in which the strategies were presented was randomized for each participant and was independent of the order on the previous problem. The order of the rating questions was fixed for all problems, strategies and participants.

After rating each strategy for each of the two problems, participants were asked to explicitly compare the strategies. We used the same two problems and strategy descriptions as in the strategy rating section of the survey. For this section, one strategy was presented on the left and one on the right. Participants rated similarity on a continuous slider scale that ranged from 1 (not similar at all) to 5 (very similar). Each comparison was presented on its own page and participants were not allowed to go back after leaving a page.

For the similarity ratings, the order in which the problems were presented was counterbalanced and participants had to complete all comparisons before moving to the next problem. The three comparisons were: Discrete vs. Gauss, Area vs. Discrete, and Gauss vs. Area. The order in which the strategy comparisons were presented was randomized and the order for the second problem did not depend on the first.

Participants also reported demographic information, including SAT/ACT math scores, and completed a NfC scale (Cacioppo & Petty, 1982). We transformed SAT/ACT scores into percentile scores for analysis, using records from 2015 (when most participants would have taken the tests).

Procedure

Students participated in this study individually. They first completed the computer-based survey that requested strategy ratings and comparisons, and they then reported demographic information and competed the NfC scale. The study took approximately 30 minutes to complete.

Results

Strategy Ratings

Each participant was asked to rate each of three strategies on six dimensions: how good the strategy is, how much sense the strategy makes, how common the strategy is, how easy the strategy is to remember, and how long it takes to use the strategy. We conducted a series of linear mixed effects models to examine differences in strategy ratings. We used non-orthogonal contrasts for strategy type and re-centered appropriately to evaluate all strategy comparisons. Average ratings for each of the strategies on each of the dimensions can be seen in Table 1.

We first considered participants' ratings of how *good* each strategy was. On average, participants rated the Discrete strategy as better than both the Gauss strategy, $B = -1.09$, $F(1, 30.59) = 12.62$, $p < .01$ and the Area strategy, $B = -0.82$, $F(1, 32.05) = 8.95$, $p < .01$. Participants rated the Gauss strategy and the Area strategy similarly, $B = 0.26$, $F(1, 30.66) = 0.86$, $p = .36$. Thus, participants viewed Area and Gauss as equally good, but viewed both as significantly worse than Discrete.

Next, we considered participants' ratings of how much *sense* each strategy made. On average, participants rated the Discrete strategy higher than the Gauss strategy, $B = -1.13$, $F(1, 31.03) = 15.30$, $p < .001$, and higher than the Area strategy on this dimension, $B = -0.82$, $F(1, 38.02) = 16.00$, $p < .001$. Participants' ratings of Gauss and Area were comparable, $B = 0.31$, $F(1, 32.45) = 1.37$, $p = .25$. Thus, participants thought that the Discrete strategy made more sense than either of the other strategies.

Third, we considered participants' ratings of how *common* the strategies were. On average, participants rated the Discrete strategy as more common than both Gauss, $B = -0.81$, $F(1, 32.18) = 9.04$, $p < .01$, and Area, $B = -1.26$, $F(1, 34.36) = 23.87$, $p < .001$. Participants also rated Gauss as more common than Area, $B = -0.5$, $F(1, 30.84) = 4.25$, $p < .05$. In sum, participants viewed the Discrete strategy as the most common, the Area strategy as the least common, and the Gauss strategy as in between.

Next, we considered participants' ratings of *complexity.* Participants rated the Discrete strategy as more complex than Gauss, $B = -0.59$, $F(1, 30.93) = 6.03$, $p = .02$, but no different in complexity than Area, $B = -0.31$, $F(1, 32.27) = 1.67$, $p = .21$. Participants rated Area and Gauss as similarly complex, $B = 0.28$, $F(1, 32.79) = 2.07$, $p = .16$. In sum, participants viewed Gauss as less complex than Discrete, but no other comparisons were significant.

Next, we considered participants' ratings of how *easy to remember* each strategy was. Overall, participants did not differentiate among the strategies on this dimension. Participants rated the Discrete strategy similarly to both the Gauss strategy, $B = 0.12$, $F(1, 31.68) = 0.23$, $p =.63$, and the Area strategy, $B = 0.003$, $F(1, 33.16) = 0.0002$, $p = .99$. Participants also rated the Area strategy as comparable to the

Gauss strategy, *B* = -0.11, *F*(1, 30.29) = 0.29, *p* = .60.

Finally, we examined participants' ratings of *how long* each strategy would take to implement. Participants believed that the Discrete strategy would take longer than Gauss, *B* = -1.05, *F*(1, 30.11) = 17.01, $p < .001$ and longer than Area, *B* = -0.62, *F*(1, 33.28) = 8.66, $p < .01$. Participants also thought the Area strategy would take longer than Gauss strategy, *B* = 0.44, *F*(1, 32.70) = 4.86, *p* = .03. In sum, participants viewed the Gauss strategy as taking the least amount of time to implement, and the Discrete strategy as taking the longest.

Factor Analysis

We averaged participants' ratings for each dimension across problems and strategies. We conducted an exploratory factor analysis on these six composite scores in order to determine the underlying factor structure of participants' ratings. An examination of the scree plot suggested a two-factor solution. We extracted two factors using a maximum likelihood estimator. We first allowed the factors to correlate using a promax rotation. We found that the factors were correlated (*r* = -.42). However, allowing the factors to correlate led to a Heywood case. Using a varimax rotation, which forces the factors to be orthogonal, eliminated this issue, so we proceeded with this model.

Table 2. Factor loadings of strategy ratings.

	Efficiency	Intuitiveness
Complexity	**1**	-0.05
Easiness to remember	**-0.75**	0.26
Length	**0.37**	-0.09
Makes Sense	**-0.36**	**0.68**
Goodness	-0.22	**0.97**
Commonness	-0.02	**0.56**

As seen in Table 2, five of the six rating categories loaded on only one of the two factors. The only rating category that loaded onto more than one factor was whether the strategy made sense. We did not consider this to be an issue because the factor loading was much higher for one factor than the other. Complexity, easiness to remember, and length all loaded onto the first factor. Goodness, commonness, and the degree to which the strategy made sense all loaded onto the second factor. The two-factor solution suggests that participants consider two dimensions when evaluating mathematical strategies. The first factor reflects the *efficiency* of the strategies, including how complex and easy to remember the strategy seems and the amount of time participants believe would be needed to implement the strategy. The second dimension reflects the *intuitiveness* of the strategies; it included the degree to which the strategy makes sense, is common, and is "good." These two factors explained 61% of the variance in participants' ratings.

Individual Differences

We next created factor scores for each strategy for each participant. This meant that each participant had 6 factor scores (one for each strategy for intuitiveness and efficiency). We then examined correlations between factor scores for each strategy and Need for Cognition. Participants with higher NfC scores had lower efficiency scores for the Discrete strategy (*r* = -0.38, $p < .05$) and the Gauss strategy (*r* = -0.41, $p < .05$), but not for the Area strategy (*r* = -0.12, *ns*). None of the correlations of NfC with intuitiveness scores were significant. We also examined correlations between factor scores and participants' math percentile scores. None of these correlations were significant.

Predicting Similarity Ratings

Participants rated the Discrete and Gauss strategies as the most similar (M = 2.82, SD = 1.10), followed by the Discrete and Area strategies (M = 2.79, SD = 1.07). Participants rated the Gauss and Area strategies as the least similar (M = 2.44, SD = 1.10). We wanted to explore whether participants' similarity ratings were related to how efficient and intuitive they perceived each strategy to be. We were also interested in how individual differences might moderate these relations. To address these issues, we first calculated difference scores for each pair of strategies from the factor scores (e.g., Efficiency (Discrete) – Efficiency (Area)), yielding an efficiency difference score and an intuitiveness difference score for each strategy comparison. We took the absolute value of these difference scores, so that higher values indicate that participants thought that the two strategies differed more in that dimension. We then used these difference scores to predict participants' similarity ratings for each pair of strategies.

We analyzed these data using a linear mixed effects model. We predicted participants' similarity ratings for each comparison on the basis of their efficiency difference, their intuitiveness difference, NfC, math percentile, and the specific strategy comparison (Discrete vs. Area, Discrete vs. Gauss, or Gauss vs. Area). We allowed NfC and math percentile to interact with the efficiency difference and intuitiveness difference. We also included by-subject random intercepts and by-subject random slopes for strategy comparison, efficiency difference, and intuitiveness difference. The model specification was as follows:

*Strategy Similarity ~ Strategy comparison + Efficiency Difference * NfC + Intuitiveness Difference * NfC + Efficiency Difference * math percentile + Intuitiveness Difference * math percentile + (1 + Strategy comparison + Efficiency Difference + Intuitiveness Difference| Subject)*

This analysis revealed an interaction between NfC and efficiency difference on similarity ratings, $\chi^2(1, N = 32) = 5.53$, *p* = .0187. Although the related simple effects were not significant, as seen in Figure 2, participants with low NfC tended to rate strategies as less similar if they perceived them to be different in efficiency. In contrast, participants with high NfC tended to rate strategies as more similar if they perceived them to be different in efficiency. No other effects were significant, all *p*'s > .05.

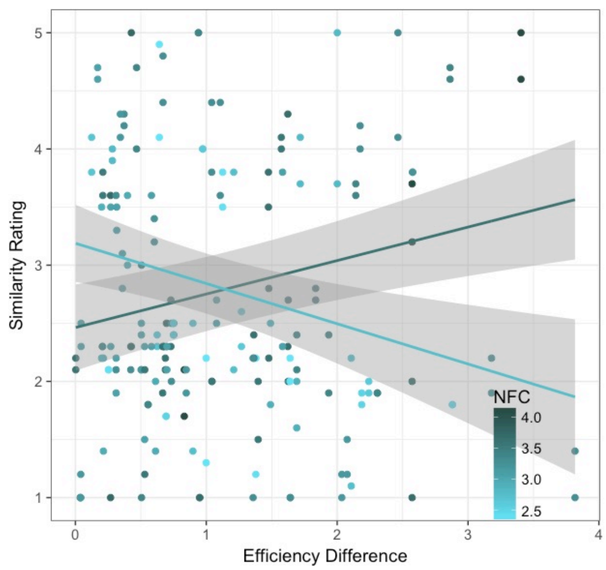

Figure 2. Similarity ratings as a function of absolute differences in efficiency factor scores and Need for Cognition. Higher values on the y-axis indicate greater similarity. Dots are colored based on Need for Cognition. The dark line indicates predicted values for participants who scored high on the Need for Cognition scale (1 SD above the mean). The light line indicates predicted values for participants who scored low on the Need for Cognition scale (1 SD below the mean). Shaded regions represent 95% CIs. A 0 on the x-axis indicates that participants thought of the two strategies as equal in efficiency.

Discussion

This research examined how participants evaluated different strategies for solving constant change problems. We found that participants viewed the Discrete strategy as the most common, as making the most sense, and as better than the Gauss and Area strategies. Additionally, we found that participants rated strategies along two primary dimensions, which we characterized as intuitiveness and efficiency.

Participants in this study judged the Discrete strategy to be the most common, which accurately reflects the "ground truth", as the Discrete strategy was the most commonly used strategy in previous studies of constant change problems (e.g., Riggs et al., 2015). However, participants also saw limitations of this strategy, noting that it was complex and took a long time to implement, relative to the other strategy options. The Discrete strategy was rated highly on all of the intuitiveness variables and relatively low on efficiency. If participants notice the drawbacks of this strategy, then why is it used so frequently?

One possibility is that participants may value intuitiveness *over* efficiency. This is a rational choice in situations in which it is important to solve problems correctly. In such situations, a strategy that seems likely to result in a correct answer but that may be laborious may be preferable to a strategy that is easy but unlikely to result in a correct answer.

However, valuing intuitiveness over efficiency might also deter learners from adopting new ways to solve problems. If someone does not care about efficiency, they might choose to shift only to a more intuitive strategy, and not to a more efficient one. Valuing intuitive strategies over all else can be problematic when the most intuitive strategies are not correct. In the domain of mathematics, this issue has been studied in children solving equivalence problems (e.g., 3 + 4 + 6 = 3 + __), a type of math problem with which children often struggle (McNeil, 2007; Perry, Church, & Goldin-Meadow, 1988). Children often solve equivalence problems using a strategy that is intuitive given their past experience with arithmetic (adding all the numbers), but that does not yield a correct answer on these particular problems (McNeil, 2008). The current study suggests that it may be especially difficult for children to abandon such incorrect but intuitive strategies. Empirical findings support this view. Abandoning such strategies is a complicated process, often requiring feedback or extensive practice (McNeil, Fyfe, Petersen, Dunwiddie, & Brletic-Shipley, 2011).

We found that participants considered efficiency in evaluating the similarity of pairs of strategies, but the relation varied depending on participants' level of Need for Cognition. For participants with low Need for Cognition, similar efficiency seemed to be a basis for overall judgements of similarity, but this was not the case for participants with high Need for Cognition. More generally, the findings suggest that individual differences may influence the dimensions that people view as important when evaluating strategies and judging their similarity.

The implications of this study extend beyond the domain of mathematics. Some research suggests that people's intuitive theories about the world are often wrong (see Shtulman, 2017). Recent studies have shown that even experts continue to endorse intuitive, but incorrect, theories under certain circumstances (Coley, & Tanner, 2015; Shtulman, & Harrington, 2016; Shtulman, & Valcarcel, 2012). The present work study suggests that may be precisely the intuitiveness of these theories that may make conceptual change difficult.

The contributions of this study should be understood in the context of its limitations. First and most important, our sample size was small, so our conclusions are necessarily tentative.

Second, we did not measure participants' own strategy use before or after their ratings. Our aim was to assess how people think about different strategies to see if this could inform research on strategy change. We view this work a first step towards understanding how perceptions of strategies influence strategy use and strategy change. In future work, we plan to test whether learners' own perceptions of various strategies are related to their willingness to adopt new strategies.

Third, we included only correct strategies in the current study. This was an intentional choice in order to identify differences in factors other than correctness along which people evaluate strategies. Future research should investigate whether the same factor structure is found for incorrect strategies. It is possible that when rating incorrect strategies, learners base their judgments on other factors. In future research, it would also be of interest to examine differences in evaluations of intuitive but incorrect strategies and counterintuitive but correct strategies. Such work would help establish whether people value correctness or intuitiveness more highly.

Finally, there was a great deal of individual variation in people's perceptions of strategies. Although some individual difference characteristics, such as Need for Cognition, were related to participants' evaluations of strategies, others, such as performance on standardized math tests, were not. Future research should further investigate what types of individual differences—such as prior knowledge, confidence, attitudes towards mathematics, or participants' own strategy use when solving these problems—influence strategy evaluations.

In sum, we found that participants judged strategies along two primary dimensions: intuitiveness and efficiency. The strategy that scored highest on the intuitiveness ratings (the Discrete strategy) is the same strategy that prior research had identified as the most commonly used. This study has potential theoretical implications for research on strategy use and conceptual change, as well as practical implications for teachers and others who may wish to encourage learners to change the ways that they solve problems.

Acknowledgments

We thank Cher Yang and Feier Yang for their help in data collection. This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Awards #R305B150003 and #R305B090009 to the University of Wisconsin- Madison. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.

References

Alibali, M. W. & Prather, R. W. (2007, March). *Improvements in problem encoding lead to changes in strategy evaluations*. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Boston, Massachusetts.

Bassok, M. & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 153–166.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, *25*, 24-34.

Brown, S. A. and Alibali, M. W. (2018). Promoting strategy change: Mere exposure to alternative strategies helps, but feedback can hurt. *Journal of Cognition & Development*, in press.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116-131.

Coley, J. D., & Tanner, K. (2015). Relations between intuitive biological thinking and biological misconceptions in biology majors and nonmajors. *CBE-Life Sciences Education*, *14*, 1-19.

Fyfe, E. R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology*, *108*, 1-16.

McNeil, N. M. (2007). U-shaped development in math: 7-year-olds outperform 9-year-olds on equivalence problems. *Developmental Psychology*, *43*, 687-695.

McNeil, N. M. (2008). Limitations to teaching children 2+ 2= 4: Typical arithmetic problems can hinder learning of mathematical equivalence. *Child Development*, *79*, 1524-1537.

McNeil, N. M., Fyfe, E. R., Petersen, L. A., Dunwiddie, A. E., & Brletic-Shipley, H. (2011). Benefits of practicing 4 = 2 + 2: Nontraditional problem formats facilitate children's understanding of mathematical equivalence. *Child Development*, *82*, 1620-1633.

Menendez, D., Brown, S. A., & Alibali, M. W. (2017, July). *I'll do what she's doing: Promoting the adoption of new problem-solving strategies in undergraduates*. Poster presented at the Midwest Meeting on Mathematical Thinking, Minneapolis, Minnesota.

Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, *3*, 359-400.

Riggs, A. E., Alibali, M. W., & Kalish, C. W. (2015). Leave her out of it: Person-presentation of strategies is harmful for transfer. *Cognitive Science*, *39*, 1965-1978.

Riggs, A. E., Alibali, M. W., & Kalish, C. W. (2017). Does it matter how Molly does it? Person-presentation of strategies and transfer in mathematics. *Contemporary Educational Psychology*, *51*, 315-320.

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. Hachette UK: Basic books

Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science*, *8*, 118-137.

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*, 209-215.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*, 481-520

Siegler, R. S. (2000). The rebirth of children's learning. *Child Development*, *71*, 26-35.

Torbeyns, J., De Smedt, B., Ghesquiere, P., & Verschaffel, L. (2009). Acquisition and use of shortcut strategies by traditionally schooled children. *Educational Studies in Mathematics, 71*, 1-17.

Walsh, M. & Anderson, J. R. (2009). The strategic nature of changing your mind. *Cognitive Psychology, 58*, 416–440.