Gesturing May Not Always Make Learning Last

Caroline E. Byrd (cbyrd@nd.edu)
Nicole M. McNeil (nmcneil@nd.edu)
Sidney K. D'Mello (sdmello@nd.edu)
Department of Psychology, University of Notre Dame

Susan Wagner Cook (susan-cook@uiowa.edu)
Department of Psychology, University of Iowa

Abstract

Studies suggest that mimicking specific gestures prior to math instruction facilitates learning. However, benefits could be due to the eye movements that accompany gesture, rather than to gesture per se. Children (*M* age = 8 yrs, 9 mos) who solved pretest equations incorrectly were taught a correct strategy for solving equations. They were randomly assigned to mimic gestures instantiating the strategy, the eye movements that accompany those gestures, or speech only prior to and during instruction. Children completed an immediate posttest and a 4-week follow-up test. We hypothesized that children in the eye movement and gesture conditions would retain more from instruction when compared to children in the speech only condition. Posttest performance was similar across conditions. Contrary to hypotheses, children in the gesture condition retained less from instruction when compared to children in the other conditions. Results suggest that there may not always be benefits of gesture during instruction.

Keywords: cognitive development; embodied cognition; mathematics learning; gesture; problem solving

Algebra is a "disaster" for most students in the United States (National Research Council, 1998, p.1). Lack of readiness for algebra can be traced back to misunderstandings of pre-algebra concepts in elementary school (Carpenter, Franke, & Levi, 2003; Knuth, Stephens, McNeil, & Alibali, 2006). For example, most children (ages 7 to 11) in the U.S. do not understand how to solve *math equivalence problems*, which are equations that have operations on both sides of the equal sign (e.g., 3 + 4 + 5 = 3 + __; Alibali, 1999; McNeil & Alibali, 2005b; Perry, 1991; Perry, Church, & Goldin-Meadow, 1988, Rittle-Johnson & Alibali, 1999). Difficulties with these problems are not easily "fixed" by instruction, as children often revert back to old, incorrect strategies a few weeks after being taught a correct strategy (Cook, Mitchell, & Goldin-Meadow, 2008). Given the importance of understanding math equivalence to future success and the apparent difficulties in helping children achieve this understanding, it is important to investigate the mechanisms that facilitate learning of this fundamental concept.

Research suggests that gesture may be a particularly powerful mechanism for creating new knowledge in the domain of mathematics (Alibali & Goldin-Meadow, 1993; Cook & Goldin-Meadow, 2006). For example, Cook and Goldin-Meadow (2006) successfully increased the rate at which children gestured when explaining their solutions to math equivalence problems by exposing children to teachers' gestures during a lesson on math equivalence. The researchers then examined the relation between children's gesture production and learning. Children who produced gestures of their own after viewing the teachers' gestures were more likely than those who did not produce gestures to both retain and generalize the knowledge gained during instruction. These results suggest gesturing, particularly gesturing a correct strategy as demonstrated by the instructor, can help children benefit from a lesson. However, because not all of the children who observed the teachers' gestures actually produced those same gestures themselves, the gestures children produced could have possibly been a reflection of their "readiness to learn," rather than a causal factor in the learning process. To address this concern, Broaders, Cook, Mitchell, and Goldin-Meadow (2007) performed a more direct experimental manipulation of gesture. They found that children who were simply told to gesture during their explanations of their solutions to math equivalence problems added more new strategies to their repertoires through gesturing and showed a greater benefit from instruction compared to children told not to gesture.

Researchers have begun to investigate the role of gesture in learning by asking children to mimic gestures instantiating particular strategies for solving math equivalence problems prior to instruction. Cook, Mitchell, and Goldin-Meadow (2008) tested the hypothesis that gestures play a role in the creation and retention of knowledge by comparing posttest performance among children told to mimic a gesture instantiating a correct, "equalize" strategy for solving math equivalence problems, children told to mimic speech describing that strategy, and children told to mimic both speech and gesture. Prior to instruction, children mimicked their assigned behavior three times (either in speech, gesture, or both), and children also mimicked that behavior before and after solving problems on their own during instruction on the problems. Children in all three groups performed similarly on an immediate posttest; however, the mimicked behavior affected how well children retained the knowledge gained from instruction. Children who mimicked an equalize gesture (i.e., moving the L hand from L to R under the L side of the problem, pausing, and then moving the R hand from L to R under the R side of the problem) performed better on a delayed follow-up test than those who did not mimic the gesture. Interestingly, there were no differences between the gesture and "both" groups, suggesting that there was something about mimicking the gesture per se—with or without speech—that led to more robust learning or consolidation of what was learned.

The results from Cook et al. (2008) suggest that gesture plays a role in conceptual change by "making learning last." Goldin-Meadow, Cook, and Mitchell (2009) extended these findings by showing that mimicking specific gestures prior to instruction on math equivalence problems not only helps children maintain a correct, learned strategy, but also helps them generate a correct strategy on their own. Children were taught to mimic gestures that instantiated a different strategy than the strategy teachers taught in the lesson, in order to examine whether children's gestures alone prior to instruction can create new ideas. Children who were told to mimic a grouping gesture before and during instruction on the equalize strategy performed better on a posttest than children who were not told to mimic the grouping gesture.

Although these findings support the idea that gesturing facilitates learning, the gestures children have mimicked in these studies have all been *relational* gestures that move children's attention back and forth across the equal sign. It is, therefore, unclear how the gestures facilitate learning. We posited that the benefits of these gestures could be due to the relational eye movements that accompany the gestures, rather than to the gestures themselves.

Many researchers have investigated associations between eye movements and cognition. Grant and Spivey (2003) showed that participants' eye movements predict correct problem solving. In their study, they used animation (visual pulsing) to induce problem solvers to fixate on the critical feature of a problem (as previously revealed in successful participants' eye movements). Results indicated that drawing problem solvers' attention in this way can help solvers develop problem-solving insights. Thus, participants' eye movements may serve as an embodied physical mechanism that stimulates new ways of thinking about a problem (Grant & Spivey, 2003; but see van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009 for an alternative view). Thomas and Lleras (2007) provided additional evidence for the link between eye movements and cognition in a study that manipulated participants' eye movements. They showed that directing eye movements on a tracking task in a pattern that embodies a correct solution leads to successful problem solving (Thomas & Lleras, 2007). Additionally, research conducted with adults indicates that looking back and forth across the equal sign is correlated with correct strategies to solve math equivalence problems (Chesney, McNeil, Brockmole, & Kelley, 2013). Thus, we theorized that the beneficial effects of gesture on learning of math equivalence could be driven, in part, by eye movements that embody relational thinking.

The present study was designed to directly compare the effects of mimicking gestures to the effects of mimicking the eye movements that accompany those gestures. It built off of Cook et al.'s (2008) design by using both "speech only" and "gesture" conditions and comparing them to an eye movement condition. We hypothesized that children in the eye movement condition would perform better than children in the speech only condition and similarly to children in the gesture condition, thus demonstrating that the beneficial effects of gesture may not depend on the hand movements themselves, but rather on a more general attentional-guidance mechanism that co-occurs with gesture. We also hypothesized that the number of times children's eyes moved back and forth across the equal sign (coded from video) would be an important predictor of learning.

Method

Participants

Participants were 70 children (34 boys, 36 girls; *M* age = 8 years, 9 months). The race/ethnicity of the sample was 17% African-American or black, 4% Asian, 7% Hispanic or Latino, 6% Other, and 66% white. Sessions were conducted in a quiet room in a research lab, a local school, and a local afterschool program.

Design

The study was a pretest-intervention-posttest design, with a 4-week follow-up, akin to the design used by Cook and colleagues (2008). The first session consisted of a pretest, pre-instruction, instruction, and a posttest. The second session consisted of a follow-up test and a brief lesson on math equivalence tailored specifically to the child's needs. Both sessions were videotaped, so that we could study the strategies children used when solving the problems.

Experimental Conditions

Children were randomly assigned to one of three conditions: speech only (n = 23), gesture (n = 24), or eye movement (n = 23). Each child received the same instruction on math equivalence and the same assessments; the only aspect that varied was the behavior children were asked to mimic during pre-instruction and instruction. One experimenter served as the lesson facilitator during pre-instruction and instruction, and a different experimenter, who was blind to the child's condition, served as the tester, administering the pretest, posttest, and follow-up test. The lesson facilitator also taught the brief lesson at the end of the second session.

In the *speech only condition*, children were shown a video of a teacher standing in front of a problem saying the phrase "I want to make one side equal to the other side."

In the *gesture condition*, children were shown a video of a teacher standing in front of a problem saying the same phrase while simultaneously producing a relational, equalize gesture (moving the L hand from L to R under the L side of the problem, pausing, and then moving the R hand from L to R under the R side of the problem).

In the *eye movement condition*, children were shown a video of a teacher standing in front of a problem saying the same phrase while simultaneously moving their eyes across the problem in a way that simulated the eye movements that would co-occur with gesture. To encourage eye movements,

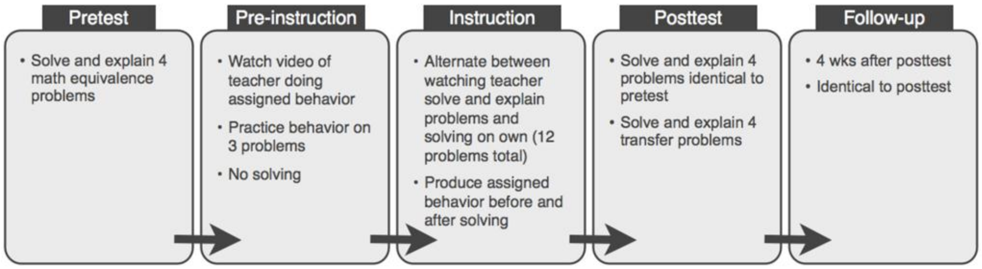

Figure 1: Overview of Procedure.

an arrow moved underneath each side of the problem during the video (from L to R under the L side of the problem, disappearing briefly, and then from L to R under the R side of the problem). Thus, this condition emulated all aspects of the gesture condition except the actual hand gestures.

Procedure

Pretest Children solved four equations with equivalent addends on each side of the equal sign (5 + 4 + 6 = __ + 6, 3 + 5 + 9 = __ + 9, 8 + 4 + 3 = 8 + __, 7 + 5 + 8 = 7 + __) and explained how they solved each problem (cf. Alibali, 1999). Because our goal was to examine how different instructional conditions affect children's learning from a lesson on math equivalence, analyses were limited to children who solved all pretest problems incorrectly ($N = 70$).

Pre-instruction The lesson facilitator showed children a video of a teacher demonstrating a behavior and asked the children to mimic that behavior for three problems of the format $a + b + c =$ __ $+ c$. The behavior a given child was asked to mimic depended on which condition he or she was in (as described above). In all conditions, the facilitator showed the video twice to ensure that all children understood the procedure. Children were shown the videos on a laptop, and then children were presented with two additional math equivalence problems alone on the laptop screen for them to practice doing the behavior on their own.

Videos of children's faces and laptop screens in all conditions were recorded during pre-instruction and instruction using customized software and the laptop's built-in camera. The videos of children's faces (including eye movements) and what they were seeing on the laptop (e.g., a particular problem) were temporally aligned so that eye movements (coded from videos) could be connected to every point during pre-instruction and instruction.

Instruction Even though conditions differed in terms of what they saw during pre-instruction, all children saw the same instruction (cf. Cook et al., 2008). Children watched a video of a teacher explaining how to use the equalize strategy to solve six more problems of the same type ($a + b + c =$ __ $+ c$). For each problem, the teacher described the strategy "I want to make one side equal to the other side" both before and after solving the problem. Each time the teacher said that phrase, an arrow moved underneath each side of the problem while the teacher made the relational, "equalize" gesture (moving the L hand from L to R under the L side of the problem, pausing, and then moving the R hand from L to R under the R side of the problem). Thus, children in all conditions were exposed to the equalize strategy 12 times in speech, gesture, and eye movements (encouraged by the arrows). This ensured that all children were exposed to the same representations of equivalence.

After each of the teacher videos, children saw another problem (presented alone on the laptop screen). First, children were asked to reproduce the behavior they practiced during pre-instruction. Next, the lesson facilitator placed a transparency sheet over the laptop screen and asked children to solve the problem using a transparency marker. Children were not given any feedback about correctness. Finally, children were asked to reproduce their behavior they practiced during pre-instruction again. Children who produced behaviors other than what they had practiced during pre-instruction were reminded to only produce the behaviors they were instructed to mimic.

Posttest Immediately after instruction, children completed a posttest administered by the tester that included the pretest equations (see above) along with transfer equations that differed in surface features (7 + 4 + 6 = __ + 3, 6 + 2 + 8 = 5 + __, 1 + 5 = __ + 2, 6 – 1 = 3 + __). Problem-solving strategies were coded as correct or incorrect based on a system used in previous research (e.g., McNeil & Alibali, 2004; Perry et al., 1988). For most problems, correctness could be inferred from the solution itself (e.g., for 7 + 5 + 8 = 7 + __, a solution of 27 indicated an incorrect "add all" strategy and a solution of 13 indicated a correct strategy). If the solution was ambiguous, then strategy correctness was coded based on children's verbal explanation (e.g., for 7 + 5 + 8 = 7 + __, the explanation "I added 7 plus 5" indicated an incorrect strategy and the explanation "I added 5 plus 8" indicated a correct strategy). Agreement between coders for a randomly selected 20% subsample was 100%.

Follow-up test Approximately four weeks after the first session, children completed a follow-up test identical to the posttest. Agreement between coders was 99%.

Results

The learning rate was high, with a majority (60%) solving at least one of the first four posttest equations correctly. This learning rate was comparable to the 52% learning rate in Cook et al.'s (2008) study. As in Cook et al., there was no evidence of significant differences in performance solving equations among conditions during instruction, $F(2, 67) = 0.30$, $p = .74$, or on the immediate posttest, $F(2, 67) = 0.02$, $p = .98$.

Following Cook et al. (2008), we tested if children across conditions differed in how well they maintained the knowledge gained during instruction over the 4-week delay. We conducted an ANCOVA with condition as the independent variable, number of posttest equations correct (out of 8) as the covariate, and number of follow-up equations correct (out of 8) as the dependent variable. Not surprisingly, posttest performance significantly predicted follow-up performance, $F(1, 66) = 52.93$, $p < .001$, with higher posttest equation solving performance associated with higher follow-up equation solving performance. The effect of condition was also significant, $F(2, 66) = 3.90$, $p = .025$, $\eta_p^2 = .11$. Contrary to our hypothesis, it was children in the gesture condition who did not retain the knowledge they had gained during instruction (see Figure 2). Simple contrasts indicated that children in the gesture condition had significantly lower retention scores than children in both the eye movement condition, $p = .035$, Cohen's $d = .63$, and the speech only condition, $p = .011$, Cohen's $d = .76$. Children in the eye movement condition did not differ from children in the speech only condition, $p = .66$.

Findings were robust, even when we varied aspects of the analysis. For example, conclusions were unchanged when we limited the analysis to the four equations that matched the pretest equations, when we limited the analysis to only the four transfer equations, when we excluded the children who did not show evidence of learning from instruction (i.e., children who did not solve at least one posttest equation correctly), and when we excluded the children who demonstrated a correct strategy in gesture at the pretest.

To further probe these unexpected effects of the gesture condition, we coded children's level of adherence to the modeled equalize gestures during pre-instruction and instruction. Equalize gestures were coded using a system established in previous work (cf. Alibali & Goldin-Meadow, 1993). Children were given a score of "1" on each equation if they ever completed the full equalize gesture with two different hands as demonstrated, "0.5" if they made a different equalize gesture (equalize gestures are gestures that distinguish the two sides of the equation, for example changing hand shape in between the left and right sides of the equation), and "0" if their gesture was not an equalize gesture. Children's scores across all nine problems were added together for a total level of adherence score. Children's adherence in the gesture condition was far from perfect, with a mean level of adherence of 4.88 (*SD* = 3.16). However, 54% of children made an equalize gesture on at least half of the problems, and only 8% of children never made an equalize gesture. There was no evidence that the degree of adherence was associated with retention. We conducted a multiple regression with level of adherence (out of 9) as the independent variable, number of posttest equations correct (out of 8) as the covariate, and number of

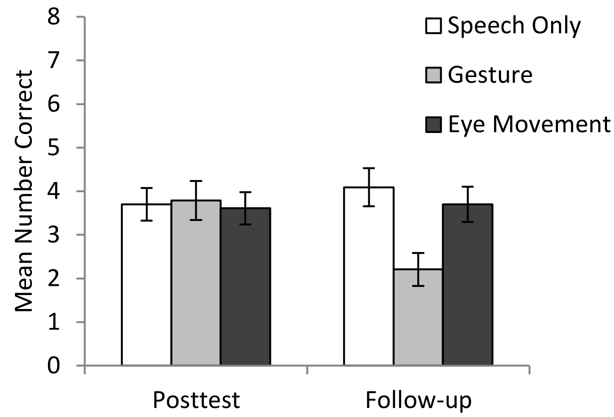

Figure 2: Equation Solving Performance at Posttest and Follow-up test by Condition.

follow-up equations correct (out of 8) as the dependent variable. Level of adherence was not a significant predictor of follow-up performance, after controlling for posttest performance, $b = .22$, $t(23) = 1.08$, $p = .29$.

Adherence in the speech only and eye movement conditions was coded in a similar manner as the gesture condition for each equation seen during pre-instruction and instruction. For the speech only condition, children were given a score of "1" on each equation if their phrase ever indicated one side "equal to" another side, "0.5" if they used "equal," "equals," or "equal as" in the phrase instead of "equal to," and "0" if they did not indicate one side and another. For the eye movement condition, children were given a score of "1" on each equation if they ever directly switched between looking to one side and looking to the other side, "0.5" if they switched directions while looking above the problem, and "0" if they did not switch directions. Agreement between coders for a 20% random subsample was 98% for speech only, 100% for gesture, and 91% for the eye movement condition. Children's scores across all nine problems were added together for a total level of adherence score. One child in the eye movement condition was excluded from this analysis because his eye movements were not properly recorded; thus, the final sample for this analysis included 69 children.

We first conducted an ANOVA with condition as the independent variable and level of adherence (out of 9) as the dependent variable. The effect of condition was significant, $F(2, 66) = 16.26$, $p < .001$, $\eta_p^2 = .33$. Simple contrasts indicated that children in the gesture condition had significantly lower level of adherence scores ($M = 4.88$, $SD = 3.16$) than children in both the speech only condition, $p < .001$ ($M = 7.91$, $SD = 1.79$), and the eye movement condition, $p < .001$ ($M = 8.34$, $SD = 1.36$). Children in the speech only condition did not differ from children in the eye movement condition, $p = .84$. Next, we conducted a multiple regression analysis with level of adherence (out of 9) as the independent variable, condition as the covariate, and number of posttest equations correct (out of 8) as the dependent variable. After controlling for condition, level of adherence significantly predicted posttest performance, $b = .38$, $t(65) = 2.17$, $p = .033$. Thus, level of adherence to the exact behavior predicted immediate learning.

A final step-wise regression analysis was conducted with number of follow-up equations correct (out of 8) as the dependent variable. In the first step, condition and number of posttest equations correct (out of 8) were entered as the covariates. In the second step, level of adherence (out of 9) was entered as an independent variable. In the third step, the interaction between condition and level of adherence was entered as another independent variable to test for a moderation effect. The R^2 change for the second model including level of adherence (R^2 $\Delta= .003$) was not significant, $p = .52$, and the R^2 change for the third model including the level of adherence by condition interaction (R^2 $\Delta = .011$) was also not significant, $p = .52$. Thus, although level of adherence was a significant predictor of posttest performance, level of adherence was not a significant predictor of maintenance of learning in the long-term, and neither was the level of adherence by condition interaction.

Eye Movement Analyses

The primary rationale for conducting the present study was related to participants' eye movements. We hypothesized that participants' eye movements back and forth across the equal sign during instruction would be an important predictor of learning. Thus, we calculated the number of times children looked back and forth across the equal sign on each problem (i.e., *related* the two sides), and then averaged across all problems seen during instruction.

We first tested if eye movements during instruction differed across conditions. We conducted an ANOVA with condition as the between-subjects factor and average number of relational eye movements as the dependent variable. There was a significant effect of condition, $F(2, 66) = 26.75$, $p < .001$, $\eta_p^2 = .45$. Simple contrasts indicated that, as predicted, children in the speech only condition made significantly fewer relational eye movements than did children in the eye movement condition, $p < .001$, and the gesture condition, $p < .001$. Children in the eye movement condition did not differ from children in the gesture condition, $p = .78$. There were no significant differences across conditions for relational eye movements while solving problems during instruction.

Next, we considered whether children's relational eye movements while solving problems predicted their performance. The average number of eye movements back and forth across the equal sign during instruction was a marginally significant predictor of performance on the instruction problems, $b = .40$, $t(67) = 1.96$, $p = .054$, but it did not predict performance at posttest or follow-up.

Discussion

We hypothesized that children in the eye movement and gesture conditions would learn and retain more from instruction on math equivalence when compared to children in the speech only condition. However, contrary to our expectations, children in the gesture condition actually retained less of the knowledge they had gained during instruction when compared to children in the other two conditions. Overall, these results suggest that there may be some limits to the benefits of gesture during instruction. At the same time, however, results provide some support for the hypothesis that relational eye movements back and forth across the equal sign are associated with learning. Specifically, children who produced more relational eye movements while solving problems during instruction, on average, solved more problems correctly during instruction.

Previous research has detailed the benefits of gesture in various contexts – spontaneous gesture during a lesson (Cook & Goldin-Meadow, 2006), being told to gesture (Broaders et al., 2007), and self-producing gestures of correct strategies (Cook et al., 2008, Goldin-Meadow et al., 2009). Thus, it is important to determine what it was about the present study that reversed these benefits.

There are several potential reasons why the expected benefits of gesture (on both learning and retention) were not found in this study, and each provides fodder for future research. First, the physical presence of the teacher may moderate the effects of gesture on learning and retention. Perhaps watching a video of a teacher gesturing and mimicking that gesture may be experienced differently than watching a teacher gesture in real life and mimicking that gesture. Children may have felt uncomfortable mimicking the gestures of someone who was not there to see them.

Second, redundancy of information during the instruction video may moderate the effects of gesture on learning and retention. In the present study, aspects of each condition were included in the instruction (i.e., an arrow moved under the problem while the teacher spoke and gestured). Cook et al.'s (2008) study did not include an eye movement condition, so the instruction only included speech and gesture, without an arrow. Perhaps the arrow that appeared underneath the problem attracted attention away from the teacher's gestures (which were lower on the screen, below the arrow) and interfered with children's mimicry of the behavior in the gesture condition. Redundancy during the instruction video may have resulted in a weaker instantiation of the physical gesture in memory for children in the gesture condition, and thus, a weaker connection to and embodiment of the strategy when performing the behavior themselves (Kalyuga, Chandler, & Sweller, 1999). Beyond being a detriment to retention in the gesture condition, the arrow during instruction may have provided additional benefit for children in the speech only condition. Indeed, children in the speech only condition in this study solved a greater number of problems correctly on average during instruction (3.4 out of 6) than did children in the speech condition in Cook et al.'s (2008) study (1.8 out of 6).

Third, the space in which children learn and produce gestures may moderate the effects of gesture on learning and retention. In the present study, children gestured to a problem presented on a laptop after watching a video of a teacher gesturing. In Cook et al.'s (2008) study, the teacher's gestures, the children's gestures, and the problems children solved during instruction were all in the exact same space (at a board). Perhaps children in Cook et al.'s (2008) study anticipated being imitated by the teacher again after their turn, which may have resulted in deeper encoding of the strategy in memory. Also, children in the present study were sitting down and making fairly small hand movements compared to the Cook et al. (2008) study in which children were standing and making larger gestures.

Finally, the act of gesturing towards a laptop screen may have been unnatural or awkward, thus adding extraneous cognitive load to the learning task. Extraneous cognitive load makes processing information during learning more difficult. Students who are burdened by this extraneous load are not able to construct the depth of knowledge that other students may be able to because they cannot devote all of their cognitive resources to the learning process (Sweller, van Merrienboer, & Paas, 1998). Children found it difficult to mimic the exact gesture (recall that children in the gesture condition had significantly lower scores on the level of adherence scale than did children in the other conditions), and some voiced these difficulties (e.g., "This is hard."). Producing unnatural gestures in instructional settings may increase cognitive load and have a negative effect on children with low expertise in a content area (like in the present study) (e.g., Post, van Gog, Paas, & Zwaan, 2013).

Overall, results provide important data regarding potential limits to the benefits of gesture during instruction. These findings not only advance theory and provide future avenues of study, but also provide educators with an important caveat when designing lessons and learning materials for teaching children the concept of mathematical equivalence.

Acknowledgments

This paper is based on a master's thesis conducted by Byrd under the direction of McNeil, funded by a National Science Foundation Graduate Research Fellowship. Thanks to Jill Lany, James Brockmole, and members of the Cognition Learning and Development Lab at Notre Dame. Thanks also to the students, parents, teachers, and administrators at the participating school and afterschool programs.

References

Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology, 35*, 127-145.

Alibali, M. W., & Goldin-Meadow, S. (1993). Transitions in learning: What the hands reveal about a child's state of mind. *Cognitive Psychology*, *25*, 468-523.

Broaders, S., Cook, S. W., Mitchell, Z., Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, *136*, 539-550.

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school*. Portsmouth, NH: Heinemann.

Chesney, D. L., McNeil, N. M, Brockmole, J. R., & Kelley, K. (2013). An eye for relations: Eye tracking indicates long-term negative effects of operational thinking on understanding of math equivalence. *Memory & Cognition*, *41*, 1079-1095.

Cook, S. W., & Goldin-Meadow, S. (2006). The role of gesture in learning: Do children use their hands to change their minds? *Journal of Cognition and Development*, *7*, 211-232.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*, 1047-1058.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, *20*, 267-272.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, *14*, 462-466.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, *13*, 351-371.

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education*, *37*, 297-312.

McNeil, N. M., & Alibali, M. W. (2004). You'll see what you mean: Students encode equations based on their knowledge of arithmetic. *Cognitive Science*, *28*, 451-466.

McNeil, N. M., & Alibali, M. W. (2005b). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development, 76*, 883-899.

National Research Council. (1998). *The nature and role of algebra in the K-14 curriculum*. Washington, DC: National Academy Press.

Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development, 6*, 449-468.

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

Post, L. S., van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of simultaneously observing and making gestures while studying grammar animations on cognitive load and learning. *Computers in Human Behavior*, *29*, 1450-1455.

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*, 175-189.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251-296.

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin & Review*, *14*, 663-668.