Wait For It... Delaying Instruction Improves Mathematics Problem Solving: A Classroom Study

Abbey Marie Loehr,[1] Emily R. Fyfe,[1] Bethany Rittle-Johnson[1]

[1] Vanderbilt University

Engaging learners in exploratory problem-solving activities prior to receiving instruction (i.e., explore-instruct approach) has been endorsed as an effective learning approach. However, it remains unclear whether this approach is feasible for elementary-school children in a classroom context. In two experiments, second-graders solved mathematical equivalence problems either before or after receiving brief conceptual instruction. In Experiment 1 (n = 41), the explore-instruct approach was less effective at supporting learning than an instruct-solve approach. However, it did not include a common, but often overlooked feature of an explore-instruct approach, which is provision of a knowledge-application activity after instruction. In Experiment 2 (n = 47), we included a knowledge-application activity by having all children check their answers on previously solved problems. The explore-instruct approach in this experiment led to superior learning than an instruct-solve approach. Findings suggest promise for an explore-instruct approach, provided learners have the opportunity to apply knowledge from instruction.

Correspondence:

Abbey Marie Loehr, 230 Appleton Place, Peabody #552, Vanderbilt University, Nashville, TN 37203; Phone: (615) 343-7149. Email: abbey.loehr@vanderbilt.edu

Keywords:

mathematics learning, problem solving, exploratory learning, direct instruction

Some researchers and educators uphold the importance of attempting to solve problems even if the learner doesn't know how to solve them (Chi, 2009; Dewey, 1902/1956; Piaget, 1973; Schwartz, Lindgren, & Lewis, 2009; Schwartz & Martin, 2004). Learners are encouraged to forge ahead and figure them out using what they already know to get started. Other researchers and educators feel that learners should be armed with a sufficient tool set (e.g., instruction on a procedure or a worked example) before productively attempting novel problems (Hiebert et al., 2003; Kirschner, Sweller, & Clark, 2006; Roelofs, Visser, & Terwel, 2003; Sweller & Cooper, 1985).

Both approaches to problem solving have potential benefits and can be fruitful for different reasons. For example, some theories of learning focus on how people learn through exploration and self-discovery of their environment without explicit instruction (Hirsh-Pasek, Golinkoff, Berk, & Singer, 2009; Piaget, 1973; Schulz & Bonawitz, 2007; Sylva, Bruner, & Genova, 1976). Exploration of an unfamiliar topic or problem is thought to support learning by increasing motivation, encouraging broad hypothesis testing, and improving depth of understanding (e.g., Bonawitz et al., 2011; Piaget, 1973; Sylva et al., 1976; Wise & O'Neill, 2009). Other theories of learning turn to how learning is supported through guidance and instruction provided by a more knowledgeable other (Csibra & Gergely, 2009; Kirschner et al., 2006; Tomasello, Carpenter, Call, Behne, & Moll, 2005; Vygotsky, 1978). Explicit instruction is thought to increase available cognitive resources and support the development of accurate knowledge (Kirschner et al., 2006; Klahr & Nigam, 2004; Koenig & Harris, 2005; Sweller, van Merrienboer, & Paas, 1998; Tomasello et al., 2005).

Although historically theorists have debated the superiority of exploration or instruction, many contemporary learning theorists now agree that both are important. Indeed, researchers have begun to investigate ways exploration and explicit instruction might be combined, instead of contrasted, to maximize the benefits of each (Lorch et al., 2010; Mayer, 2004; Schwartz & Bransford, 1998).

One promising combination that has been endorsed by researchers in psychology and education is to provide opportunities for problem exploration prior to explicit instruction (i.e., an explore-instruct approach; Hiebert & Grouws, 2007; Kapur, 2012; Schwartz, Chase, Chin, & Oppezzo, 2011). For example, prior exploration can be used to prepare learners for future instruction (Schwartz et al., 2009). Explicit instruction often presupposes some level of prior knowledge that novices lack. Problem exploration activates and builds prior knowledge, which allows learners to extract more from subsequent instruction (Schwartz & Bransford, 1998; Schwartz et al., 2009). Prior exploration can also create opportunities for productive failure (Kapur, 2011, 2012). That is, most learners will struggle to solve the problems correctly, but this struggle will ultimately lead to deeper processing of subsequent instruction.

A growing body of evidence supports the potential of an explore-instruct approach for a wide range of topics, including elementary mathematics (DeCaro & Rittle-Johnson, 2012), middle-school science (Schwartz et al., 2011), high-school statistics (Kapur, 2012), analogical reasoning (Needham & Begg, 1991), cell functioning (Taylor, Smith, van Stolk, & Spiegelman, 2010) and psychology (Schwartz & Bransford, 1998). Experimental studies on the timing of instruction relative to solving unfamiliar problems provide the most direct evidence for the effectiveness of an explore-instruct approach. For example, middle-school students who learned about density by solving problems prior to instruction learned more than students who received instruction first and then solved problems as practice (Schwartz et al., 2011). Similarly, engaging students in novel problem-solving tasks prior to instruction on the concept of average speed was a more effective approach than a traditional instruct-then-practice sequence (Kapur, 2011). Kapur (2012) recently replicated those findings with high-school students learning statistics. Those in the explore-instruct condition demonstrated higher conceptual understanding and transfer on a posttest relative to students who received instruction first.

Most evidence in support of an explore-instruct approach comes from research with adolescents and adults using complex problem-solving tasks. Given the high cognitive demands of problem exploration (Kirschner et al., 2006), the benefits of an explore-instruct approach may not generalize to younger children who have more limited cognitive resources (Gathercole, Pickering, Ambridge, & Wearing, 2004). Indeed, delaying instruction has been criticized for failing to direct attention to critical information, leaving learners to use available resources on inefficient trial and error strategies (Clark, 2009; Sweller et al., 1998). Further, the instruction used in previous research included a step-by-step solution procedure that may render subsequent problem-solving a rote form of practice in the instruct-solve condition. The advantages of an explore-instruct sequence may have arisen because the instruction on a procedure interfered with meaningful problem solving in the instruct-solve condition (see Perry, 1991), not because problem exploration better prepared people to learn from the instruction.

Given these limitations in past research, we previously conducted a series of studies with elementary-school children and provided instruction focused exclusively on concepts. This initial research shed light on when and why an explore-instruct approach is effective with elementary-school children learning mathematical equivalence. In two previous studies, elementary-school children solved novel math problems and received immediate trial-by-trial accuracy feedback either before or after receiving conceptual instruction on the meaning of the equal sign (DeCaro & Rittle-Johnson, 2012; Fyfe, DeCaro, & Rittle-Johnson, 2014). Both studies occurred in a one-on-one tutoring context. In the first study, children in the explore-instruct condition had greater conceptual knowledge at both posttest and a two-week delayed retention test than children in the instruct-solve condition (DeCaro & Rittle-Johnson, 2012). During the intervention, children in the explore-instruct condition tried a wider variety of solution procedures, encoded key problem features more often and better gauged their level of understanding.

A second study identified a boundary condition for when an explore-instruct approach supports greater learning (Fyfe et al., 2014). In this study, children again received trial-by-trial accuracy feedback (e.g., the correct answer). They also received high-quality self-explanation prompts during the solve phase, in which they were asked to explain the conceptual rationale of the correct solutions (e.g., "why does [the correct answer] make this a true number sentence"). Receiving conceptual instruction first (instruct-solve condition) improved the quality of children's self-explanations. In turn, the children in the instruct-solve condition had greater knowledge at posttest and retention test, and the quality of children's verbal self-explanations partially mediated the impact of condition on the outcomes. Clearly, the nature of the problem exploration task is important.

Current study

In the current studies, we evaluated the effectiveness of an explore-instruct approach compared to an instruct-solve approach in elementary-school classrooms rather than in one-on-one tutoring settings. Several features of our previous one-on-one tutoring studies were not feasible in a classroom context with elementary-aged children, including the provision of immediate, trial-by-trial accuracy feedback and the prompts for individual self-explanations. For example, many second-graders are still learning to write and providing accurate, written explanations can be difficult. Thus, in the current studies children solved math equivalence problems on worksheets without feedback or prompts to self-explain, allowing children to engage in problem solving independently. Immediate trial-by-trial feedback during problem solving is beneficial for low-knowledge children in an explore-instruct approach (Fyfe, Rittle-Johnson, & DeCaro, 2012), so it is important to evaluate the effectiveness of the approach without feedback. Our previous study suggests that without high-quality self-explanation prompts, an explore-instruct approach should be more effective (DeCaro & Rittle-Johnson, 2012). The current studies will help evaluate this claim in an elementary classroom context.

As in our previous work, we focused on children's understanding of mathematical equivalence. Mathematical equivalence is the idea that two sides of an equation represent the same quantity and it is often symbolized by the equal sign (=). Knowledge of mathematical equivalence is a critical prerequisite for understanding higher-level algebra (e.g., Falkner, Levi, & Carpenter, 1999). For example, students' definitions of the equal sign are associated with their performance on algebraic equations (Knuth, Stephens, McNeil, & Alibali, 2006), and errors that reflect misunderstanding of equality are predictive of difficulties in Algebra 1 (Booth, Barbieri, Eyer, & Pare-Blagoev, 2014). Yet, elementary curricula do not typically include definitions of the equal sign or mathematical equivalence problems (e.g., problems with operations on both sides of the equal sign such as 3 + 4 + 5 = 3 + ☐; Powell, 2012; Rittle-Johnson, Matthews, Taylor, & McEldoon, 2011). As a result, elementary-school children in the U.S. often struggle to understand mathematical equivalence. For example, they often exhibit misconceptions about the meaning of the equal sign, viewing it as an operator signal that means "adds up to" or "get the answer" (e.g., Matthews, Rittle-Johnson, McEldoon, & Taylor, 2012; McNeil & Alibali, 2005; Rittle-Johnson et al., 2011). Further, they often solve mathematical equivalence problems incorrectly by adding up all of the numbers in the problem or only the numbers before the equal sign (e.g., answering 15 or 12 for 3 + 4 + 5 = 3 + ☐, rather than 9; McNeil & Alibali, 2005). Unfortunately, children's difficulties with mathematical equivalence are often robust, persisting into middle school, high school, and even adulthood (e.g., Chesney & McNeil, 2014; Knuth et al., 2006; McNeil & Alibali, 2005; McNeil, Rittle-Johnson, Hattikudur, & Petersen, 2010). In sum, evidence from past research stresses the importance of establishing understanding of mathematical equivalence early in elementary school.

Experiment 1

Method

Participants

All children from two second-grade classrooms in a parochial school serving a predominately white, middle- to upper-class population were eligible to participate. Children in the second grade are typically 7–8 years of age. Data from two children were excluded because they were absent from school on either the day of the intervention or posttest. The final sample included 41 children.

Design

The experiment had a pretest–intervention–posttest design. The intervention occurred in children's classrooms during their normal mathematics instruction. Children were randomly assigned to the instruct-solve (n = 22) or the explore-instruct (n = 19) condition.

Assessment and Coding

The mathematical equivalence assessment was a shortened version of the assessment used in our one-on-one tutoring studies (DeCaro & Rittle-Johnson, 2012; Fyfe et al., 2014) to meet district standards for limiting classroom time spent on testing. The assessment included both procedural and conceptual knowledge scales. The procedural knowledge scale assessed children's ability to solve mathematical equivalence problems. Some procedural items had the same structure as the problems presented during the intervention (i.e., familiar items) and some items contained a new problem feature, such as the unknown on the left side of the equal sign (i.e., transfer items). As in prior work (McNeil, 2007) children's responses were coded as correct if they were within one of the correct answer. The conceptual knowledge scale assessed children's understanding of two key concepts of equivalence: (a) the relational meaning of the equal sign and (b) the structure of equations including equations with operations on both sides of the equal sign. Four conceptual knowledge items required an explanation or reconstruction of an equation. Two raters independently coded 20% of these responses, and agreement was very high (kappas = 0.90 – 0.93). Table 1 contains example items and scoring criteria for both the procedural and conceptual knowledge scales. Different versions of the assessment were used at pretest, midtest (during the intervention), and posttest.

Pretest. The pretest was a brief version of the assessment that only contained the conceptual knowledge scale. It included five conceptual items (four were equation structure items and one was an equal sign item). One child did not complete two of the items at pretest, so his/her score was calculated based on the number of items completed.

Midtest. The midtest was also a brief version of the assessment that only contained the conceptual knowledge scale. It included four items, all of which were equation structure items.

Posttest. The posttest was a more comprehensive assessment that contained both the procedural and conceptual knowledge scales. The procedural knowledge scale included eight items (four were familiar items and four were transfer items). The conceptual knowledge scale included eight items (three were equal sign items and five were equation structure items).

Table 1.

Example items from the procedural and conceptual knowledge scales on the mathematical equivalence assessment.

Item Type	Task	Scoring Criteria
Procedural		(α = .92 in Exp. 1; α = .68 in Exp. 2)
Familiar problems	Solve problem with operation on right side (8 = 6 + ☐) or operations on both sides, blank on right (e.g., 3 + 4 = ☐ + 6)	Response must be within 1 of correct answer

Transfer problems	Solve problems with operations on both sides, blank on left or includes subtraction (e.g., ☐ + 2 = 6 + 5)	Same as above
Conceptual		(α = .75 in Exp. 1; α = .74 in Exp. 2)
Meaning of equal sign	Define equal sign	1 point for relational definition (e.g., the same amount)
	Rate definitions of equal sign as good, not good, or don't know	1 point for choosing "two amounts are the same" as a best, over "add" and "the answer to the problem"
Structure of equations	Reproduce 4 + 3 + 9 = 4 + ☐ from memory after viewing for 5 seconds	1 point for correctly reconstructing numerals, operators, equal sign and blank in correct location
	Judge 3 = 3 and 7 = 3 + 4 as true or false	1 point for judging both equations as true

Note. Cronbach alphas are for posttest. Alphas were somewhat lower for the brief assessment at pretest, largely due to floor effects on some items.

Procedure

All sessions occurred in children's classrooms during their mathematics class period. Children completed the pretest in one 20-minute session. A few days later, children participated in the intervention for approximately 60 minutes, with one condition meeting in one classroom and the other condition meeting in a second classroom. Research assistants administered the intervention to the entire class. The intervention consisted of an instruction phase and a solve phase, with the order of the phases depending on condition. In the *instruct-solve* condition, children completed the instruction phase first, followed by the solve phase. In the *explore-instruct* condition, the order of the phases was simply reversed and children completed the solve phase first, followed by the instruction phase. The following day, children completed the posttest in one 30-minute session.

Instruction phase. Children received conceptual instruction on the relational meaning of the equal sign that was nearly identical to instruction from our tutoring studies (DeCaro & Rittle-Johnson, 2012; Fyfe et al., 2014). Specifically, four closed, non-standard equations (e.g., 3 + 4 = 3 + 4; 4 + 4 = 3 + 5; 7 = 3 + 4) were printed on laminated cardstock. For each equation, the experimenter identified the two sides of the equation, defined the equal sign as meaning the same amount as, and explained how the two sides of the equation were equal. A single math equivalence problem (i.e., 5 + 4 + 3 = 5 + ☐) was presented at the end of the instruction and the experimenter reviewed how to identify the sides of the equation as well as the meaning of the equal sign. Children were prompted with simple questions throughout the instruction to encourage attendance and engagement (e.g., "What is on the first side of the problem?"). No solution procedures were discussed, as children tend to ignore conceptual instruction when procedural instruction is also presented (Perry, 1991).

Solve phase. During the solve phase, children individually completed a problem-solving workbook with 13 problems to solve. The first 12 problems were presented in sets of four problems with similar addends (presented on the same page). The first two problem sets began with two easier problems meant to activate prior knowledge (e.g., 3 + ☐ = 10 and 8 = ☐ + 5). The remaining 8 problems were four- and five-addend math equivalence problems with operations on both sides of the equal sign (e.g., 3 + 5 = 6 + ☐). On the last page of the workbook, there was a final math equivalence problem to solve as well as a prompt to define the equal sign ("What does the equal sign mean?"). The types and number of problems were very similar to those in our one-on-one tutoring studies. Because children in the explore-instruct condition were required to solve unfamiliar problems without receiving instruction, a few additional hints (e.g., think about what the equal sign means) were provided to guide attention to important problem features and promote invention of procedures. All children were told the correct answer to the last math equivalence problem at the end of the solve phase. This was intended to motivate children in the explore-instruct condition to attend to instruction and to alert children in the instruct-solve condition of their performance. No additional feedback was provided.

Midtest. We administered a brief conceptual knowledge measure after the first phase of the intervention across conditions (i.e., problem solving or instruction; see *Assessment and Coding*).

Results

Pretest

At pretest, children exhibited moderate conceptual knowledge (see Table 2). There were no significant differences between conditions at pretest for conceptual knowledge, $F(1, 39) = 0.29$, $p = .60$, $\eta_p^2 = .01$.

Table 2.

Performance on outcome measures by condition.

	Explore-Instruct		Instruct-Solve	
Outcome	M	SD	M	SD
Pretest				
Conceptual Knowledge (%)	40	27	44	20
Posttest				
Procedural Knowledge (%)	76	38	87	22
Conceptual Knowledge (%)	66	27	88	17
Intervention				
Problem-Solving Accuracy (%)	78	32	92	18
Midtest Scores (%)	57	34	74	21

Posttest

We examined procedural and conceptual knowledge using two separate ANCOVAs. Each model included condition as the between-subject factor and pretest scores as a covariate. In addition, the data were examined by estimating a Bayes factor using Bayesian Information Criteria

(Wagenmakers, 2007), comparing the fit of the data under the null hypothesis and the alternative hypothesis. Table 2 provides the means and standard deviations on primary outcome measures by condition.

Procedural knowledge. For procedural knowledge, there was no effect of condition, $F(1, 38) = 0.93$, $p = .34$, $\eta_p^2 = .02$. An estimated Bayes factor (null/alternative) suggested that the data were only 3.91:1 in favor of the alternative hypothesis, or rather, .26 times more likely to occur under a model including an effect for order of instruction, rather than a model without it. Children in the explore-instruct condition exhibited similar procedural knowledge as children in the instruct-solve condition (see Table 2). Errors were relatively rare, but generally reflected common misconceptions in this domain. Specifically, adding only the numbers before the equal sign or adding all the numbers in the problem accounted for 68% of errors (12% of all trials).

Conceptual knowledge. For conceptual knowledge, there was a significant effect of condition, $F(1, 38) = 11.70$, $p = .002$, $\eta_p^2 = .24$. Contrary to our hypothesis, children in the instruct-solve condition exhibited higher conceptual knowledge than children in the explore-instruct condition (see Table 2). An estimated Bayes factor (null/alternative) suggested that the data were .027:1 in favor of the alternative hypothesis, or rather, 37.04 times more likely to occur under a model including an effect for order of instruction, rather than a model without it.

However, the scores were not normally distributed, with over half of the children solving 7 or 8 (out of 8) of the items correctly. Thus, we also used binomial logistic regression to predict the odds of scoring above 80% to ensure the effects did not depend on our method of analysis. We included condition and pretest scores in the model. Results were consistent with the ANCOVA. Children in the instruct-solve condition were significantly more likely than children in the explore-instruct condition to solve over 80% of the conceptual items correctly (17 of 22 [77%] vs. 7 of 19 [37%], $\beta = 1.87$, $z = 2.43$, $Wald\ (1, N = 41) = 5.94$, $p = .02$.

Intervention Activities

To explore the effect of condition during learning, we examined performance during the intervention. Pretest scores were included as covariates in analyses of continuous measures.

Problem-solving accuracy. Because children in the explore-instruct group explored problems prior to instruction, we expected them to solve fewer problems correctly than children in the instruct–solve group. On average, children solved 11 out of the 13 intervention problems correctly in the problem-solving packet (SD = 3.4). There was a marginal effect of condition, $F(1, 38) = 2.80$, $p = .10$, $\eta_p^2 = .07$. Although accuracy was quite high in both groups, children in the instruct-solve group solved somewhat more problems correctly than children in the explore-instruct group (see Table 2). Problem-solving success in the explore-instruct group likely reflects the fact that children were not excluded from the study for already knowing how to solve the problems, unlike in our tutoring studies, because all children in the classroom received instruction. Additionally, we did not have a measure of procedural knowledge at pretest.

Equal sign definition. We had all children provide a written definition of the equal sign at the end of the problem-solving packet. For children in the instruct-solve conditions, instruction occurred before they defined the equal sign. The number of children who provided a relational definition of the equal sign was significantly higher in the instruct-solve condition (82%) than in the explore-instruct condition (42%), χ^2 (1, N = 41) = 6.93, $p = .008$. Thus, conceptual instruction did impact children's knowledge of the equal sign.

Midtest. Children received a brief midtest during the intervention, which assessed children's conceptual knowledge of equation structure after the first phase. There was a marginal effect of condition, $F(1, 38) = 4.06$, $p = .051$, $\eta_p^2 = .10$. Children in the instruct-solve condition had higher midtest scores than children in the explore-instruct condition (see Table 2). The same was true when we considered just the items that assessed children's encoding of the problem structure, $F(1, 38) = 2.91$, $p = .096$, $\eta_p^2 = .07$ (instruct-solve M = 70%, explore-instruct M = 50%).

Discussion

Contrary to our hypothesis, results from Experiment 1 did not support the effectiveness of an explore-instruct approach in elementary-school classrooms. Thus, findings from our previous tutoring study (DeCaro & Rittle-Johnson, 2012) did not generalize to a classroom setting without feedback. Rather, children in the instruct-solve condition had greater conceptual knowledge and similar procedural knowledge at posttest, compared to children who solved problems first. Intervention performance indicated that children in the instruct-solve condition also had somewhat greater problem-solving success, although problem-solving accuracy was quite high for both groups. Children who received instruction first also provided more accurate definitions of the equal sign at the end of the solve phase and had somewhat greater knowledge of equation structures at midtest, suggesting conceptual instruction impacted intervention performance during the solve phase.

There are several potential reasons why the explore-instruct approach was less effective at supporting learning in a classroom context compared to an instruct-solve sequence. One possibility is the lack of guidance provided during the exploratory solve phase. Previous tutoring studies have included features of problem exploration that require a level of guidance that is not feasible in a classroom setting (i.e., self-explanation prompts and immediate trial-by-trial feedback). Unfortunately, if individual guidance is required for the explore-instruct approach to be effective with elementary-school children, it seems unlikely to have practical use for typical classroom lessons.

However, a second possibility is that the explore-instruct condition was less effective because of the lack of a *knowledge-application* activity following instruction. One assumed, but often overlooked feature of an explore-instruct approach is the inclusion of an additional problem-solving task after instruction. The task allows learners to apply knowledge from the instruction to a relevant activity. In this way, the taught information is used immediately, and hence, integrated with prior knowledge. One common knowledge-application task is to provide learners with worksheets containing problems that can be solved using information from the instruction (Kapur, 2011, 2012; Schwartz et al., 2011). For example, in Schwartz et al. (2011) after students in the explore-instruct condition received explicit instruction on the concepts and procedures of density and speed, they completed a worksheet of word problems. In our previous tutoring studies, an immediate posttest likely functioned as an knowledge-application activity, as children had the opportunity to use knowledge from the instruction immediately after receiving it (DeCaro & Rittle-Johnson, 2012; Fyfe et al., 2014). However, in Experiment 1, the posttest occurred on the following day. Thus, children did not have an opportunity to immediately use knowledge gained from the instruction.

Experiment 2

The goal of Experiment 2 was to evaluate an explore-instruct approach that included an explicit knowledge application activity in a classroom context. The knowledge application activity was to revisit the problems previously solved during the explore phase and to potentially re-solve and change any incorrect answers. In that way, children could apply the information from the instruction to relevant, previously-solved problems. Revisiting problems is more time efficient and may be more beneficial for addressing misconceptions than solving a new problem set. For example, revisiting problems that were previously solved incorrectly and correcting them using relevant knowledge gained from instruction may help students overcome misconceptions. Indeed, students process problems more deeply upon encountering impasses and detecting errors (VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003). In Experiment 1, an explicit knowledge-application activity immediately following instruction was not implemented and may have revealed a boundary condition for when an explore-instruct approach supports greater learning. With the inclusion of a knowledge-application activity, we predicted that children in the explore-instruct condition would exhibit greater knowledge of mathematical equivalence than children in the instruct-solve condition.

Method

Participants

All children from three second-grade classrooms in one public school serving a working- to middle-class population and one predominately White parochial school serving a middle- to upper-class population were eligible to participate. Demographic information was not available for the parochial school, but at the public school 48% of students were eligible for free or reduced price lunch and approximately 36% of students were ethnic minorities (28% African-American, 5% Hispanic, 3% Asian). Children in the second grade are typically 7–8 years of age. Data from ten children were excluded because they were absent from school the day of the intervention (n = 3), the experimenter did not implement the knowledge application activity (n = 3) or indicators suggested that they had a disability (n = 4). The final sample included 47 children.

Design

The experiment had the same pretest–intervention–posttest design and procedure as Experiment 1 with a few exceptions. Within each classroom, children were randomly assigned to one of two conditions: instruct-solve (n = 24) or explore-instruct (n = 23).

Assessment and Coding

The pretest, midtest, and posttest assessments of mathematical equivalence knowledge were similar to those used in Experiment 1. The pretest included the same 5-item conceptual knowledge scale as Experiment 1. It also included a brief procedural knowledge scale (three familiar items) to provide a more informative prior knowledge measure. The midtest was identical to the midtest in Experiment 1. The posttest was nearly identical to Experiment 1. We included an additional equal sign item on the conceptual knowledge scale so that our assessment of equal sign items and equation structure items was more balanced. It now included four equal sign items and five equation structure items. To compensate for the additional time necessary to complete the conceptual knowledge scale, we omitted two items from the procedural knowledge scale so that it now contained six items (three familiar and three transfer).

As in Experiment 1, two raters independently coded 20% of relevant responses, and agreement was very high (kappas = 0.96). A few children did not complete two of the conceptual items (one child at pretest and two at posttest), so their conceptual knowledge scores were calculated based on the items they did complete.

Procedure

The procedure was very similar to Experiment 1 with the addition of a concluding knowledge application activity in both conditions: the check phase. The intervention was administered in small groups within intact classrooms to accommodate school scheduling constraints. Research assistants worked with small groups of 4–6 children assigned to the same condition, allowing for children within the same classroom to be randomly assigned to different conditions. In the *instruct-solve condition*, children completed the instruction phase first, followed by the solve phase and then the checking phase. In the *explore-instruct condition*, children completed the solve phase first, followed by the instruction phase and then the checking phase. The instruct and solve phases were identical to Experiment 1. Instruction was administered at the small group level, and children completed the solve phase individually.

During the checking phase, all children were asked to independently check their answers from the solve phase without feedback or guidance. Experimenters provided purple pens and instructed children to either place a check mark by answers they deemed correct or to write their new answer above their old answer. This allowed us to track original and changed answers. For children in the explore-instruct condition, the checking phase encouraged children to use information from instruction to check their previous problem-solving efforts. For children in the instruct-solve condition, the checking phase allowed children to go back over their work and check for mistakes.

Data Analysis

One child was absent the day of the pretest. Imputing missing independent variables leads to more precise and unbiased conclusions than omitting participants with missing data (Peugh & Enders, 2004). We used the expectation-maximization algorithm for maximum likelihood estimation via the missing value analysis in SPSS (Schafer & Graham, 2002) to impute the missing pretest score.

Unlike Experiment 1, children in Experiment 2 sat in small groups within the classroom. Specifically, we worked with children in 11 small groups of 4–6 children each. To test for nonindependence at the small group level, we calculated unconditional intraclass correlations on the outcomes, using an approach that allows for negative non-independence recommended by Kenny, Kashy, Mannetti, Pierro, and Livi (2002). The intraclass correlation was .01 for procedural knowledge and –.13 for conceptual knowledge. Because traditional analysis of variance models assume independence in the data and this assumption was violated for conceptual knowledge, we used multilevel modeling to account for nesting within group. We specified the use of restricted maximum-likelihood estimation and compound symmetry for the variance-covariance structure in the models (Kenny, Kashy, & Cook, 2006). The significance tests used the Satterthwaite (1946) approximation to estimate the degrees of freedom. Because students worked in small groups, our model had two levels: (1) the individual level and (2) the small-group level, with condition at level 2.

Pretest scores were grand mean centered and condition was dummy coded, with the instruct-solve condition coded as 0.

Results

Pretest

At pretest, children exhibited moderate procedural knowledge and low conceptual knowledge (see Table 3). There were no significant differences between conditions at pretest for procedural knowledge, β = –7.95, *p* = .22, or conceptual knowledge, β = –1.19, *p* = .86.

Table 3.

Performance on outcome measures by condition.

	Explore-Instruct		Instruct-Solve	
Outcome	M	SD	M	SD
Pretest				
Procedural Knowledge (%)	46	25	54	29
Conceptual Knowledge (%)	33	22	34	24
Posttest				
Procedural Knowledge (%)	92	14	78	25
Conceptual Knowledge (%)	69	27	72	26
Intervention				
Problem-Solving Accuracy (%)	60	31	88	22
Midtest Scores (%)	53	29	63	25
Answer Changes (Freq.)	2.8	3.5	0.5	1.2
Incorrect-to-Correct Changes (Freq.)	2.4	3.4	0.2	0.5

Posttest

Table 3 provides the means and standard deviations on primary outcome measures by condition. Table 4 shows the results of the two-level modeling analyses predicting procedural knowledge and conceptual knowledge at posttest. Effects of pretest knowledge (procedural and conceptual knowledge scores) and condition were included in the model.

Table 4.

Parameter estimates for posttest outcome measures.

Variable	Procedural Knowledge	Conceptual Knowledge
Intercept	79.51 (3.49)***	71.87 (4.23)***
Condition	12.88 (4.86)*	-1.61 (5.90)
Pretest Procedural	-5.32 (12.55)	-1.67 (15.09)
Pretest Conceptual	6.78 (14.23)	39.81 (17.13)*

Note. Condition is coded 1 for explore-instruct and 0 for instruct-solve. Unstandardized coefficients are shown with standard errors in parentheses. Pretest scores were mean-centered.

* $p < .05$, *** $p < .001$

Procedural knowledge. For procedural knowledge, there was a significant effect of condition, β = 12.9, $p = .04$ (see Table 3). Consistent with our prediction, children in the explore-instruct condition exhibited higher procedural knowledge than children in the instruct-solve condition (see Table 3). An estimated Bayes factor (null/alternative) suggested that the data were .49:1 in favor of the alternative hypothesis, or rather, 2.09 times more likely to occur under a model including an effect for order of instruction than a model without it.

However, the scores were not normally distributed, with over half of the children solving all of the items correctly. Thus, we also used binomial logistic regression to predict the odds of scoring 100% to ensure the effects did not depend on our method of analysis. We included condition and pretest scores in the model. Results were consistent with the multi-level model. Children in the explore-instruct condition were significantly more likely than children in the instruct-solve condition to solve 100% of the procedural items correctly (16 of 23 [70%] vs. 9 of 24 [38%], β =

1.30, $z = 2.08$, *Wald* $(1, N = 47) = 4.33$, $p = .04$. Errors were rare, but generally reflected reflect common misconceptions in this domain. Specifically, adding only the numbers before the equal sign or adding all the numbers in the problem regardless of the placement of the equal sign accounted for 45% of errors (7% of all trials).

Conceptual knowledge. For conceptual knowledge, there was no effect of condition, $\beta = -1.61$, $p = .79$ (see Table 4). An estimated Bayes factor (null/alternative) suggested that the data were only 6.54:1 in favor of the alternative hypothesis, or rather, .15 times more likely to occur under a model including an effect for order of instruction than a model without it. Children in the explore-instruct condition exhibited similar conceptual knowledge as children in the instruct-solve condition (see Table 3).

Intervention Activities

To explore the effect of condition during learning, we examined performance during the intervention using the same multilevel model as for the posttest. Pretest scores were included as covariates in analyses of continuous measures.

Problem-solving accuracy. On average, children solved 9.6 out of the 13 intervention problems correctly in the problem-solving packet ($SD = 3.9$). There was a significant effect of condition, $\beta = -24.5$, $p = .007$. As expected, children in the instruct-solve group solved more problems correctly than children in the explore-instruct group (see Table 3).

Equal sign definition. As expected, the number of children who provided a relational definition of the equal sign was significantly higher in the instruct-solve condition (50%) than in the explore-instruct condition (13%), χ^2 $(1, N = 47) = 7.38$, $p = .007$. Thus, conceptual instruction did impact children's knowledge of the equal sign.

Midtest. There was no effect of condition on the midtest, which assessed children's conceptual knowledge of equation structure after the first phase, $\beta = -6.58$, $p = .29$. Children in the explore-instruct condition had similar midtest scores as children in the instruct-solve condition (see Table 4). The same was true when we considered just the items that assessed children's encoding of the problem structures from memory, $\beta = -1.04$, $p = .93$ (explore-instruct M = 65%, instruct-solve M = 67%).

Checking behavior. As expected, checking behavior differed by condition (see Table 4). Children in the explore-instruct condition made significantly more changes than children in the instruct-solve condition, $\beta = 2.17$, $p = .01$, and this effect remained unchanged when we considered just the frequency of incorrect-to-correct changes, $\beta = 2.13$, $p = .01$. Further, more children in the explore-instruct condition made an incorrect-to-correct change at least once (48%) compared to children in the instruct-solve condition (17%), χ^2 (1, N = 47) = 5.25, $p = .02$. In the explore-instruct condition, over half of the children (52%) changed at least one answer, and of these children 92% made an incorrect-to-correct change at least once.

Given that children's checking behavior differed by condition, we re-examined children's accuracy on the problem-solving packet after the checking had occurred. Recall that problem-solving accuracy was initially higher in the instruct-solve condition relative to the explore-instruct condition. After answer-checking, children's accuracy on the intervention problems increased in the explore-instruct condition (M = 60% vs. M = 78%), $t(22) = -3.27$, $p = .003$, but not in the instruct-solve condition (M = 88% vs. M = 88%), $t(23) = -0.78$, $p = .45$. Indeed, after the checking activity, problem-solving accuracy was statistically similar in the explore-instruct (M = 78%, SD = 27%) and instruct-solve conditions (M = 88%, SD =22%), $\beta = -9.49$, $p = .31$. Answer checking clearly benefitted children in the explore-instruct condition.

Discussion

Delaying conceptual instruction until after problem exploration led to similar conceptual knowledge and greater procedural knowledge than the reverse sequence (instruct-solve). These results provide additional support for the effectiveness of an explore-instruct approach and identify an important feature that has been commonly assumed and overlooked in past research. Specifically, a concluding activity that allows students to apply knowledge from instruction to related problems seems to be necessary to achieve the advantages of an explore-instruct approach. In Experiment 1, children did not complete this activity and learned less in the explore-instruct condition compared to the instruct-solve condition. In Experiment 2, all children checked their work from the solve phase. For children in the explore-instruct condition, this activity occurred immediately after receiving instruction and seemed to play an important role for integrating knowledge from instruction with problem solving. Children in the explore-instruct condition detected their errors and revised knowledge during the checking phase.

Providing conceptual instruction first did have an initially positive impact on children's performance. For example, children in the instruct-solve condition exhibited higher problem-solving performance and knowledge of the equal sign during the intervention. However, this positive impact was not long-lasting. Indeed, following the checking phase, children in the explore-instruct condition exhibited similar problem-solving accuracy as children in the instruct-solve condition. Further, children in the explore-instruct condition exhibited greater procedural knowledge at posttest.

General Discussion

A growing body of evidence has documented the benefits of delaying instruction until after an opportunity for exploratory problem solving (DeCaro & Rittle-Johnson, 2012; Kapur, 2010, 2011, 2012; Kapur & Bielaczyc, 2012; Needham & Begg, 1991; Schwartz & Bransford, 1998; Schwartz et al., 2011; Schwartz & Martin, 2004; Taylor et al., 2010). However, the majority of past research comparing an explore-instruct approach to a conventional instruct-solve approach has used complex problem-solving tasks with adolescents or adults and instruction that included information on solution procedures (Kapur, 2010, 2011, 2012; Schwartz et al., 2011; Schwartz et al., 2009; Schwartz & Martin, 2004). In an effort to generalize findings to younger learners with limited cognitive resources, our recent tutoring studies targeted elementary-school children learning about mathematical equivalence (DeCaro & Rittle-Johnson, 2012; Fyfe et al., 2014). Additionally, we provided only conceptual instruction, which required children in both conditions to invent solution procedures, given past research indicating that including procedural instruction with conceptual instruction can lead children to ignore the conceptual instruction (Perry, 1991). The current experiments sought to replicate the findings of DeCaro and Rittle-Johnson (2012) in a classroom context that did not include immediate trial-by-trial feedback or prompts to self-explain.

In Experiment 1, we failed to provide evidence in favor of an explore-instruct approach. Indeed, children who received instruction first exhibited higher conceptual knowledge than children in the explore-instruct condition. However, children in the explore-instruct approach did not have an opportunity to apply what they learned from instruction. Thus, in Experiment 2, we employed the same design but included a knowledge-application phase in which children used information from the instruction to check their previous problem-solving performance. In this study, we found support for an explore-instruct approach. Specifically, children in the explore-instruct condition exhibited higher procedural knowledge and similar conceptual knowledge as children in the instruct-solve condition.

The current study is the first to provide evidence for the benefits of an explore-instruct approach relative to an instruct-solve sequence with elementary-school children in a classroom context. There are a number of proposed mechanisms that are theorized to support learning when instruction is delayed until after an opportunity to engage in problem exploration. For example, problem exploration activates prior knowledge and promotes attention to important problem features (DeCaro & Rittle-Johnson, 2012; Schwartz et al., 2011). Children often fail to notice key information in the learning environment, and learning what information to attend to is a prominent process underlying learning and development (Case, Harris, & Graham, 1992; Siegler, 1989). Problem exploration creates an opportunity for productive failure (e.g., Bjork, 1994; Kapur & Bielaczyc, 2012), which may motivate and prepare students to learn from subsequent instruction. Relatedly, engaging in challenging problem exploration may reduce an illusion of knowing (Glenberg, Wilkinson, & Epstein, 1982). Indeed, in DeCaro and Rittle-Johnson (2012), children's ratings of understanding during the intervention were correlated with their knowledge retention in the explore-instruct condition, but not in the instruct-practice condition. Overall, prior problem exploration is thought to promote deeper processing of instruction (Schwartz et al., 2011; Schwartz & Martin, 2004).

In addition to providing support for an explore-instruct approach more generally, the current experiments also provide evidence for the importance of a concluding knowledge-application activity, during which learners can apply information from instruction to related problems. This activity allows learners in an explore-instruct approach to use what they just learned in an integrative and productive way. Previous studies that provided instruction on concepts and procedures have included a final problem-solving worksheet for students to practice applying the instruction (Kapur, 2011; Schwartz et al., 2011). In the current studies, we provided instruction on the concepts only and children were required to apply instruction to problems to figure out how to solve them. Children's answer-checking behaviors during the intervention in Experiment 2 suggest the check phase promoted application of the instruction, including generating correct solution procedures. Approximately half of the children in the explore-instruct phase detected and corrected at least one of their errors during the solve phase. Indeed, absence of the checking phase in Experiment 1 seemed to eliminate potential benefits of an explore-instruct approach. In Experiment 1, children's problem solving accuracy in the explore-instruct condition was similar during the intervention and on the posttest, suggesting the instruction phase had little impact on procedural knowledge. Their lower conceptual knowledge at posttest suggests that they also suffered from not reflecting on and using the instructed concepts immediately after instruction.

Differences in Experiment 2 and DeCaro and Rittle-Johnson (2012) suggest additional considerations for the knowledge application activity. In Experiment 2, we found that an explore-instruct approach supported greater procedural knowledge, whereas in DeCaro & Rittle-Johnson (2012), it supported greater conceptual knowledge. One potential explanation for this discrepancy may be due to the nature of the concluding knowledge application activities. While children in the previous tutoring study completed the immediate posttest after receiving instruction, children in Experiment 2 revisited previously-solved problems from the solve phase. This knowledge application included more explicit guidance to apply instruction to problems and notice errors, which may have better supported procedural knowledge compared to simply completing a posttest. Indeed, procedural knowledge on the posttest across conditions in the previous study (72%) and the instruct-solve condition in the current study (78%) was very similar. In contrast, procedural knowledge for the explore-instruct condition in the current study (92%) was much higher. Additionally, the knowledge checking activity in Experiment 2 did not include items targeting conceptual knowledge, whereas the posttest did in DeCaro and Rittle-Johnson (2012). This may help explain why differences in conceptual knowledge were not apparent in Experiment 2, but were apparent in DeCaro and Rittle-Johnson. In an explore-instruct approach, learners need opportunities to apply instructed content, and the demands of the task will influence what types of knowledge are better processed and developed.

Despite the demonstrated benefits of an explore-instruct approach in Experiment 2, the results from Experiment 1 suggest it does not always lead to superior learning. Our series of studies have begun to identify boundary conditions for the effectiveness of an explore-instruct approach. First, in a tutoring setting that includes effective self-explanation prompts, an instruct-solve approach can be more effective when it facilitates high-quality explanations during the solve phase that draw on content from conceptual instruction (Fyfe et al., 2014). Second, when exploratory problem solving activates prior misconceptions, instruction prior to problem solving may be necessary to make problem exploration more productive (Fyfe et al., 2014). Third, a concluding knowledge application activity seems critical for learning in an explore-instruct approach.

While there are several positive contributions of the current study, limitations remain. For example, we have posited that the primary reason for the difference in results across Experiment 1 and Experiment 2 is the inclusion of an explicit knowledge-application activity in Experiment 2. However, an alternative explanation for the difference in results across experiments is the size of the instruction group. In Experiment 1, the lesson was delivered to the entire class, whereas in Experiment 2, the lesson was delivered to students in small groups. One possibility is that students learn more effectively from smaller groups when using the explore-instruct approach. Indeed, past research has shown that undergraduates collaborating in small groups with at least one high knowledge student benefitted from learning-by-invention (Wiedmann, Leach, Rummel, & Wiley, 2012). Further, many previous studies showing benefits of an explore-instruct approach have had students working in pairs or small groups (Kapur, 2011; Kapur & Kinzer, 2009; Schwartz et al., 2011).

Additional limitations stem from the constrained, small-scale nature of the current study. For example, research assistants administered a single scripted lesson in a classroom with constrained exploratory activities. While this is a step in the right direction and extends past work from tutoring studies, future research should include teacher-implemented lessons with more extensive problem exploration activities. Also, the current study was implemented in traditional classrooms that spent little time engaging in exploratory problem-solving as part of their regular activities. Thus, an explore-instruct approach should be tested in classrooms whose students are accustomed to engaging in exploratory activities. It is also important to evaluate the impact of an explore-instruct approach with elementary-school children learning other content. For example, a majority of children are able to invent correct solution procedures given conceptual instruction on mathematical equivalence, and this is not the case for other topics (e.g., fraction division; Sharp & Adams, 2002). Children also have persistent misconceptions about mathematical equivalence, and an explore-instruct approach may have different effects and boundary conditions for learning about topics without common misconceptions. Overall, further research is needed to determine potential boundary conditions that impact the optimal sequencing of learning activities, including the features of the target topic.

In conclusion, results from the current studies support the effectiveness of an explore-instruct approach for learning in elementary-school classrooms. One essential feature of an explore-instruct approach was providing a knowledge application activity following instruction. The benefits of delaying instruction until after problem exploration seem to be evident even in the context of a classroom setting without extensive problem-solving guidance.

Acknowledgements

Research supported by National Science Foundation grant DRL-0746565 to Bethany Rittle-Johnson. Fyfe is supported by a Graduate Research Fellowship from the National Science Foundation. The authors thank Katherine L. McEldoon for her assistance with data collection as well as the staff, teachers, and children at St. Henry School and Sylvan Park Paideia Design Center.

References

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition, 120*(3), 322–330. http://dx.doi.org/10.1016/j.cognition.2010.10.001

Booth, J. L., Barbieri, C., Eyer, F., & Pare-Blagoev, E. J. (2014). Persistent and pernicious errors in algebraic problem solving. *The Journal of Problem Solving, 7*(1), Article 3. Retrieved from http://docs.lib.purdue.edu/jps/vol7/iss1/3

Case, L. P., Harris, K. R., & Graham, S. (1992). Improving the mathematical problem-solving skills of students with learning disabilities: Self-regulated strategy development. *The Journal of Special Education, 26*(1), 1–19. http://dx.doi.org/10.1177/002246699202600101

Chesney, D. L., & McNeil, N. M. (2014). Activation of operational thinking during arithmetic practice hinders learning and transfer. *The Journal of Problem Solving, 7*(1), Article 4. Retrieved from http://docs.lib.purdue.edu/jps/vol7/iss1/4.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73–105. http://dx.doi.org/10.1111/j.1756-8765.2008.01005.x

Clark, R. E. (2009). How much and what type of guidance is optimal for learning from instruction? In S. T. T. M. Duffy (Ed.), *Constructivist instruction: Success or failure* (pp. 158–183). New York, NY: Routledge.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153. http://dx.doi.org/10.1016/j.tics.2009.01.005

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology, 113*(4), 552–568. http://dx.doi.org/10.1016/j.jecp.2012.06.009

Dewey, J. (1902/1956). *The child and the curriculum and the school and society*. Chicago, IL: University of Chicago Press.

Falkner, K. P., Levi, L., & Carpenter, T. P. (1999). Children's understanding of equality: A foundation for algebra. *Teaching Children Mathematics, 6*(4), 232–236.

Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2014). An alternative time for telling: When conceptual instruction prior to problem solving improves mathematical knowledge. *British Journal of Educational Psychology*, Early View. http://dx.doi.org/10.1111/bjep.12035

Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology, 104*(4), 1094–1108. http://dx.doi.org/10.1037/a0028389

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177–190. http://dx.doi.org/10.1037/0012-1649.40.2.177

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory and Cognition, 10*(6), 597–602. http://dx.doi.org/10.3758/BF03202442

Hiebert, J., Gallimore, R., Garnier, H., Bogard Givvin, K., Hollingsworth, H., Jacobs, J., . . . Stigler, J. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study*. Washington D.C.: U.S. Department of Education, National Center for Education Statistics.

Hiebert, J., & Grouws, D. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 371–404). Charlotte, NC: Information Age Publishing.

Hirsh-Pasek, K., Golinkoff, R. M., Berk, L. E., & Singer, D. G. (2009). *A mandate for playful learning in preschool: Presenting the evidence*. New York, NY: Oxford University Press.

Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science, 38*(6), 523–550. http://dx.doi.org/10.1007/s11251-009-9093-x

Kapur, M. (2011). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science, 39*(4), 561–579. http://dx.doi.org/10.1007/s11251-010-9144-3

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*(4), 651–672. http://dx.doi.org/10.1007/s11251-012-9209-6

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45–83. http://dx.doi.org/10.1080/10508406.2011.591717

Kapur, M., & Kinzer, C. K. (2009). Productive failure in CSCL groups. *International Journal of Computer-Supported Collaborative Learning, 4*(1), 21–46.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford Press.

Kenny, D. A., Kashy, D. A., Mannetti, L., Pierro, A., & Livi, S. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology, 83*(1), 126–137. http://dx.doi.org/10.1037//0022-3514.83.1.126

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. http://dx.doi.org/10.1207/s15326985ep4102_1

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*(10), 661–667. http://dx.doi.org/10.1111/j.0956-7976.2004.00737.x

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*(4), 297–312.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261–1277. http://dx.doi.org/10.1111/j.1467-8624.2005.00849.x

Lorch, R. F., Jr., Lorch, E. P., Calderhead, W. J., Dunlap, E. E., Hodell, E. C., & Freer, B. D. (2010). Learning the control of variables strategy in higher and lower achieving classrooms: Contributions of explicit instruction and experimentation. *Journal of Educational Psychology, 102*(1), 90–101. http://dx.doi.org/10.1037/a0017972

Matthews, P., Rittle-Johnson, B., McEldoon, K., & Taylor, R. (2012). Measure for measure: What combining diverse measures reveals about children's understanding of the equal sign as an indicator of mathematical equality. *Journal for Research in Mathematics Education, 43*(3), 316–350.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14–19. http://dx.doi.org/10.1037/0003-066X.59.1.14

McNeil, N. M. (2007). U-shaped development in math: 7-year-olds outperform 9-year-olds on equivalence problems. *Developmental Psychology, 43*(3), 687–695. http://dx.doi.org/10.1037/0012-1649.43.3.687

McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development, 76*(4), 883–899. http://dx.doi.org/10.1111/j.1467-8624.2005.00884.x

McNeil, N. M., Rittle-Johnson, B., Hattikudur, S., & Petersen, L. A. (2010). Continuity in representation between children and adults: Arithmetic knowledge hinders undergraduates' algebraic problem solving. *Journal of Cognition and Development, 11*, 437–457. http://dx.doi.org/10.1080/15248372.2010.516421

Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory & Cognition, 19*(6), 543–557. http://dx.doi.org/10.3758/BF03197150

Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development, 6*(4), 449–468. http://dx.doi.org/10.1016/0885-2014(91)90049-J

Piaget, J. (1973). *To understand is to invent*. New York, NY: Grossman Publishers.

Powell, S. R. (2012). Equations and the equal sign in elementary mathematics textbooks. *Elementary School Journal, 112*(4), 627–648. http://dx.doi.org/10.1086/665009

Rittle-Johnson, B., Matthews, P. G., Taylor, R. S., & McEldoon, K. L. (2011). Assessing knowledge of mathematical equivalence: A construct-modeling approach. *Journal of Educational Psychology, 103*(1), 85–104. http://dx.doi.org/10.1037/a0021334

Roelofs, E., Visser, J., & Terwel, J. (2003). Preferences for various learning environments: Teachers' and parents' perceptions. *Learning Environments Research, 6*(1), 77–110. http://dx.doi.org/10.1023/A:1022915910198

Satterthwaite, F. E. (1946). An approximate distribution of estimation of variance components. *Biometrics Bulletin, 2*, 110–114. http://dx.doi.org/10.2307/3002019

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology, 43*(4), 1045–1050. http://dx.doi.org/10.1037/0012-1649.43.4.1045

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475–522. http://dx.doi.org/10.1207/s1532690xci1604_4

Schwartz, D. L., Chase, C. C., Chin, D. B., & Oppezzo, M. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*, 759–775. http://dx.doi.org/10.1037/a0025140

Schwartz, D. L., Lindgren, R., & Lewis, S. (2009). Constructivist in an age of non-constructivist assessments. In S. Tobias & T. Duffy (Eds.), *Constructivist instruction: Success or failure?* (pp. 34–61). New York, NY: Routledge.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22*(2), 129–184. http://dx.doi.org/10.1207/s1532690xci2202_1

Sharp, J., & Adams, B. (2002). Children's constructions of knowledge for fraction division after solving realistic problems. *The Journal of Educational Research, 95*(6), 333–347. http://dx.doi.org/10.1080/00220670209596608

Siegler, R. S. (1989). Mechanisms of cognitive development. *Annual Review of Psychology, 40*, 353–379. http://dx.doi.org/10.1146/annurev.psych.40.1.353

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59–89. http://dx.doi.org/10.1207/s1532690xci0201_3

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. http://dx.doi.org/10.1023/A:1022193728205

Sylva, K., Bruner, J. S., & Genova, P. (1976). The role of play in the problem-solving of children 3–5 years old. In J. S. Bruner, A. Jolly & K. Sylva (Eds.), *Play - Its role in development and evolution* (pp. 244–260). New York: Basic Books.

Taylor, J. L., Smith, K. M., van Stolk, A. P., & Spiegelman, G. B. (2010). Using invention to change how students tackle problems. *CBE Life Sci Educ, 9*(4), 504–512. http://dx.doi.org/10.1187/cbe.10-02-0012

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*(5), 675–735. http://dx.doi.org/10.1017/s0140525x05000129

VanLehn, K. A., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction, 21*(3), 209–249. http://dx.doi.org/10.1207/S1532690XCI2103_01

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wiedmann, M., Leach, R., Rummel, N., & Wiley, J. (2012). Does group composition affect learning by invention? *Instructional Science, 40*, 711-730. http://dx.doi.org/10.1007/s11251-012-9204-y

Wise, A. F., & O'Neill, K. (2009). Beyond more versus less: A reframing of the debate on instructional guidance. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure* (pp. 82–105). New York, NY: Routledge.