British Journal of Educational Psychology (2014), 84, 502–519

The British Psychological Society

www.wileyonlinelibrary.com

An alternative time for telling: When conceptual instruction prior to problem solving improves mathematical knowledge

Emily R. Fyfe[1]*, Marci S. DeCaro[2] and Bethany Rittle-Johnson[1]

[1]Department of Psychology and Human Development, Vanderbilt University, Nashville, Tennessee, USA

[2]Department of Psychological and Brain Sciences, University of Louisville, Kentucky, USA

Background. The sequencing of learning materials greatly influences the knowledge that learners construct. Recently, learning theorists have focused on the sequencing of instruction in relation to solving related problems. The general consensus suggests explicit instruction should be provided; however, when to provide instruction remains unclear.

Aims. We tested the impact of conceptual instruction preceding or following mathematics problem solving to determine when conceptual instruction should or should not be delayed. We also examined the learning processes supported to inform theories of learning more broadly.

Sample. We worked with 122 second- and third-grade children.

Method. In a randomized experiment, children received instruction on the concept of math equivalence either before or after being asked to solve and explain challenging equivalence problems with feedback.

Results. Providing conceptual instruction first resulted in greater procedural knowledge and conceptual knowledge of equation structures than delaying instruction until after problem solving. Prior conceptual instruction enhanced problem solving by increasing the quality of explanations and attempted procedures.

Conclusions. Providing conceptual instruction prior to problem solving was the more effective sequencing of activities than the reverse. We compare these results with previous, contrasting findings to outline a potential framework for understanding when instruction should or should not be delayed.

Research in psychology and education indicates that the sequencing of learning material can be just as important as the content itself (e.g., McNeil *et al.*, 2012; Murdock, 1962; Saffran, Aslin, & Newport, 1996). For example, research on the use of multiple representations suggests that learners exhibit better transfer when concrete examples precede abstract examples, rather than the reverse (e.g., Goldstone & Son, 2005; McNeil & Fyfe, 2012). Further, research on practice effects indicates that learners benefit when different problem types are interleaved, rather than sequenced in a blocked order (e.g., Rohrer, 2012). Clearly, learners exposed to the same material can gain varying levels of knowledge based solely on how it is sequenced.

*Correspondence should be addressed to Emily R. Fyfe, Department of Psychology and Human Development, 230 Appleton Place, Peabody #552, Vanderbilt University, Nashville, TN 37203, USA (email: emily.r.fyfe@vanderbilt.edu).

DOI:10.1111/bjep.12035

Recent work has focused on the sequencing of instruction and problem solving. Several researchers suggest delaying instruction until learners have been prepared to attend by participating in a preceding problem solving task (e.g., Schwartz & Bransford, 1998). Others suggest providing instruction first to ensure learners quickly hone in on key information during subsequent problem solving (e.g., Kirschner, Sweller, & Clark, 2006). We contrasted these alternative sequences in the domain of math equivalence. Specifically, we tested the impact of providing conceptual instruction before or after mathematics problem solving.

Several researchers in psychology and education recommend delaying instruction until after an exploration phase with relevant problems (e.g., Kapur & Bielaczyc, 2012; Schwartz, Lindgren, & Lewis, 2009). Schwartz *et al.* (2009) have proposed a *preparation for future learning* account, suggesting that problem exploration facilitates knowledge of problem structure that helps learners understand subsequent instruction at a deeper level. Similarly, Kapur and Bielaczyc (2012) endorse delaying instruction to increase *productive failure*, suggesting that struggling with novel problems first can help learners make sense of instruction later. Having learners struggle to solve problems prior to instruction is also a recommended practice in mathematics education (e.g., Dewey, 1910; Hiebert & Grouws, 2007). None of these accounts predicts that learners will successfully solve the initial problems, but rather that learners gain key experiences from problem solving that augments learning from subsequent instruction.

Growing evidence has accumulated in support of delaying instruction (e.g., DeCaro & Rittle-Johnson, 2012; Kapur, 2011, 2012; Schwartz, Chase, Oppezzo, & Chin, 2011; Schwartz & Martin, 2004). For example, middle-school students who explored density problems prior to instruction demonstrated greater transfer than students who received instruction before solving the problems (Schwartz *et al.*, 2011). Similarly, middle-school students learned a math concept better when they solved problems over several class periods followed by a culminating lecture, than when they received a lecture followed by practice during each class period (Kapur, 2011).

Despite evidence supporting this solve–instruct approach, it seems unlikely that it will optimize learning under all conditions. Rather, the most effective order may depend on features of both instruction and problem solving. For example, most studies that have manipulated the order of instruction and problem solving have given instruction on the concepts and procedures (e.g., Kapur, 2011; Schwartz & Martin, 2004; Schwartz *et al.*, 2011). *Conceptual instruction* focuses on domain principles, whereas *procedural instruction* focuses on step-by-step procedures (Hiebert & Lefevre, 1986). Including procedural information may render subsequent problem solving a rote practice activity (e.g., Schwartz *et al.*, 2011). Thus, when instruction includes procedures, it may be best to delay instruction to give learners a chance to generate procedures on their own. In contrast, conceptual instruction does not transmit ready-made solutions and may be more effective before problem solving.

Several lines of research support this hypothesis. For example, providing conceptual instruction first can familiarize learners with relevant principles, allow them to integrate the principles with the problem solving activity, and engage in productive cognitive processing (e.g., Berthold & Renkl, 2010; Wittwer & Renkl, 2008). Prior conceptual instruction may also help reduce the demands of problem solving (e.g., Sweller, van Merrienboer, & Paas, 1998), by narrowing the problem space and limiting subsequent search (Bonawitz *et al.*, 2011). Finally, learners solving problems without instruction often fail to generate correct procedures (e.g., Klahr & Nigam, 2004; Mayer, 2004).

Prior conceptual instruction can facilitate the generation of correct procedures (e.g., Perry, 1991).

Although these studies suggest that providing conceptual instruction prior to problem solving is beneficial, DeCaro and Rittle-Johnson (2012) found the opposite. Children who solved and explained problems prior to conceptual instruction solved fewer problems correctly during the intervention, but demonstrated better understanding of the math concept on a post-test than children who received conceptual instruction first. To our knowledge, this is the only experiment that manipulated the timing of instruction and problem solving using conceptual instruction only.

Given the potential benefits of providing conceptual instruction first and the paucity of research on the topic, it is important to explore the boundary conditions for when conceptual instruction should be delayed. For example, in domains with misconceptions, the misconceptions should be activated and addressed to optimize learning (e.g., Vosniadou & Vamvakoussi, 2006). However, previous studies on the timing of instruction have not involved tasks with common misconceptions or have not activated misconceptions during problem solving. When problem solving does activate challenging misconceptions, it seems likely that prior conceptual instruction will be beneficial. In the current study, we employed two techniques thought to activate and engage misconceptions: inclusion of familiar problem types in line with a common misconception, and side-by-side contrast of the familiar problem with a novel problem type (Vosniadou & Vamvakoussi, 2006).

The use and nature of self-explanation prompts during problem solving may also matter. Most studies in the area have not included explanation prompts. Self-explanation prompts can help learners integrate information (e.g., Chi, 2000), and conceptual instruction can improve the quality of self-explanations (e.g., Matthews & Rittle-Johnson, 2009). Thus, their inclusion during problem solving may particularly benefit learners who receive conceptual instruction first. Although DeCaro and Rittle-Johnson (2012) had some children self-explain, children did not often use the information from the instruction to explain subsequent problems. That is, providing conceptual instruction first did not reliably increase the conceptual content of children's explanations. Different explanation prompts can trigger different cognitive processes and lead to different learning outcomes (e.g., Nokes, Hausmann, VanLehn, & Gershman, 2011). Thus, in the current study, we used conceptual self-explanation prompts to facilitate knowledge integration.

In sum, prior conceptual instruction is thought to support key learning processes including knowledge integration and procedure generation, suggesting that conceptual instruction may be better *prior* to problem solving. However, research testing the sequence of conceptual instruction and problem solving found benefits for the solve–instruct approach (DeCaro & Rittle-Johnson, 2012). Here, we adopted the general design from DeCaro and Rittle-Johnson (2012) to further examine when conceptual instruction should precede or follow mathematics problem solving. We made changes to facilitate integration and generation by activating misconceptions and using conceptual self-explanation prompts. Thus, in contrast to DeCaro and Rittle-Johnson (2012), conceptual instruction prior to problem solving may support greater learning than the reverse.

We examined children's learning in the domain of *math equivalence* – the idea that two sides of an equation represent the same quantity. Math equivalence is foundational for arithmetic and algebra and requires knowledge of concepts (e.g., the meaning of the equal sign) and procedures (e.g., for solving problems; Kieran, 1981). Yet, elementary curricula do not typically include definitions of the equal sign or math equivalence problems – problems with operations on both sides of the equal sign (e.g., 3 + 4 + 5 = 3 + __; Powell, 2012). Children in Western countries often interpret the equal sign as an operator, meaning 'get the answer', rather than as a relational symbol that indicates two equal amounts (e.g., Baroody & Ginsburg, 1983; McNeil & Alibali, 2005). Further, this operational view often leads to poor performance on math equivalence problems (e.g., McNeil & Alibali, 2005). In the current study, children received instruction on the concept of math equivalence either before or after solving math equivalence problems.

Method

Participants

Consent was obtained from 183 second- and third-grade children from 12 classrooms in two public schools serving a middle-class population. A pre-test assessing both procedural and conceptual knowledge of math equivalence (see *assessment*) was administered to identify children who demonstrated a high level of prior knowledge. Forty-seven children were excluded for scoring 75% or higher on either the conceptual or procedural knowledge pre-test measure. We used this exclusion criterion to retain a sample of children who had a range of prior knowledge, but still had room to learn from the intervention. Data from 14 additional children were excluded due to experimenter error (n = 2), missing assessments (n = 3), or diagnosed learning disabilities (n = 9). The final sample (N = 122, Mage = 8.2 years, 57% female, 31% ethnic minorities) consisted of 81 second-graders from six classrooms and 41 third-graders from six classrooms.

Design

The experiment had a pre-test–intervention–post-test design with a 2-week delayed retention test. For the intervention, children were randomly assigned to the instruct–solve (n = 60) or solve–instruct (n = 62) conditions.

Procedure

Children completed the pre-test in their classrooms in one 20-min session. Those who met our inclusion criterion then completed a one-on-one intervention and immediate post-test in a 50-min session at the school. The intervention consisted of a conceptual instruction phase and a problem solving phase. The only difference between conditions was the sequencing of the phases. Two weeks after the intervention, children completed the retention test.

Instruction

During the conceptual instruction phase, children were taught the relational meaning of the equal sign in the context of five non-standard number sentences (e.g., 3 + 4 = 3 + 4). For each example, the experimenter identified the two sides of the number sentence, defined the equal sign as meaning the same amount as, and explained how the two sides of the number sentence were equal. No procedures were discussed.

Problem solving

During the problem solving phase, children solved 12 problems on a computer, presented in four sets. Each set contained three problems with similar addends. The first problem in each set was a standard arithmetic problem with operations on the left side of the equal sign and the blank on the right. Standard arithmetic problems activate an 'operational' misconception, in which children focus on adding all the numbers together (McNeil & Alibali, 2005). The two remaining problems in each set were math equivalence problems with operations on both sides of the equal sign. The three problems in a set appeared one at a time, but remained visible on the same screen after they were solved.

After solving each problem, children received correct-answer feedback. On the math equivalence problems, children were also prompted to self-explain. Specifically, after the correct answer was given, children received one of two prompts: 'Why does x make this a true number sentence?' or 'Why does it make sense to put x in the box and not some other number?' (x denotes the correct answer). Finally, on the last problem in each set, children were asked to describe their problem solving procedure before receiving feedback. Figure 1 provides a screen shot of a completed set of problems.

This problem solving phase was similar to that used in DeCaro and Rittle-Johnson (2012), but with several differences. First, we presented problems in sets rather than individually to facilitate spontaneous comparison. Second, we began each set with one standard arithmetic problem to activate prior knowledge of standard equation formats. Third, our self-explanation prompts were more conceptual in nature and focused only on correct solutions.

Measures and coding

Assessment

The math equivalence assessment was adapted from past work (Matthews, Rittle-Johnson, McEldoon, & Taylor, 2012; Rittle-Johnson, Matthews, Taylor, & McEldoon, 2011). Two parallel forms were used: Form 1 at pre-test and Form 2 at post-test and retention test. The assessment included procedural (eight items) and conceptual (10 items) knowledge scales (see Table 1). The procedural scale assessed children's use of correct strategies to solve learning and transfer problems. The conceptual scale assessed two key concepts of equivalence: The meaning of the equal sign and the structure of equations. Two previous studies used a construct-modelling approach to develop and validate the assessment (Matthews *et al.*, 2012; Rittle-Johnson *et al.*, 2011). Results from this work indicate that items vary in difficulty and that children with higher scores know more about the construct.

5 + 3 + 9 = [17]

5 + 3 + 9 = [10] + 7

5 = 3 + 9 = 5 + [12]

How did you solve this problem?

Your answer is 12. The correct answer is 12.

Why does 12 make this a true number sentence?

Figure 1. Screen shot of a completed problem set from the intervention.

Table 1. Example items from the math equivalence assessment

Item type	Example items	Scoring criteria
Procedural	Four items per subscale	
Learning items (α = .87)	Solve problem with operation on right side (8 = 6 + □)	Use correct procedure (if ambiguous, must be ±1 of correct answer)
	Solve problem with operations on both sides, blank on right (7 + 6 + 4 = 7 + □)	Same as above
Transfer items (α = .89)	Solve problem with operations on both sides, blank on left (□ + 6 = 8 + 6 + 5)	Same as above
	Solve problem with operations on both sides, includes subtraction (5 − 2 + 4 = □ + 4)	Same as above
Conceptual	Five items per subscale	
Equal sign items (α = .66)	Define equal sign	Provide relational definition (e.g., the same amount as)
	Rate definitions of equal sign as good, not good, or don't know	Rate 'two amounts are the same' as a good definition
Structure items (α = .73)	Reproduce math equivalence problems from memory	Reconstruct numerals, operators, equal sign, and blank in correct location
	Indicate whether equations such as 3 = 3 are true or false	Correctly recognize non-standard equations as true or false

Note. Cronbach's alphas are for retention test. Alphas were similar at post-test, but somewhat lower at pre-test largely due to floor effects on some items.

Equation structure mid-test

Three items from the assessment were also used to assess knowledge of structure mid-way through the intervention. These items were administered after the first phase of the intervention (i.e., problem solving or instruction).

Cognitive ability measures

We administered two measures of cognitive ability to serve as control variables. First, we measured working memory capacity using the backward digit-span task (Wechsler, 2003) and the backward letter-span task. One task was administered at the beginning and the other at the end of the intervention session, in counterbalanced order.

Children were read a series of numbers or letters and asked to repeat the numbers/ letters in reverse order. Children received one point for every correct series. Scores from the two tasks were averaged to form a working memory score.

We also measured retrieval fluency (Gaddes & Crockett, 1975 as cited in Brocki & Bohlin, 2004) – the controlled search and retrieval of information from long-term memory – at the end of the intervention session. Children were asked to name as many items from a category (i.e., 'animals' and 'things to eat') as possible within a 1-min span. Children received one point for every distinct item named in a category. Scores from each category were averaged to form a fluency score. Five children were missing fluency scores.

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 (see Schafer & Graham, 2002).

Additional measures

Several measures were administered for exploratory purposes (e.g., achievement goal measures), but were not relevant for the primary research question and thus are not considered further.

Coding

On the conceptual knowledge assessment, several items required a written response. Two raters independently coded 20% of the responses, and agreement was high (κs = .96–.98). On the procedural knowledge assessment, problem solving procedures were inferred from children's written work. Inter-rater agreement on whether a procedure was correct was high (κ = .98). Children's procedure reports and self-explanations during the intervention were audio-recorded and coded by two trained raters (see Tables 2 and 3 for codes). Inter-rater reliability on children's procedures (κ = .72) and self-explanation types (κ = .96) yielded substantial agreement.

Data analysis

We worked with children from 12 classrooms with an average of 10 children per classroom (min = 3, max = 17). To examine the variability due to classroom, we calculated intraclass correlations on the outcome measures, controlling for predictor variables outlined below (see Table 4; Kenny, Kashy, & Cook, 2006). The intraclass correlations were low for all but one of the outcomes. We used ANCOVA models to analyse the effect of condition, but we checked whether classroom clustering affected

Table 2. Children's procedures for solving math equivalence problems during the intervention

		% Trials		
Procedure	Sample report (5 + 3 + 9 = 5 + __)	Instruct–solve	Solve–instruct	Total
Correct procedures				
Equalize	9 plus 3 plus 5 is 17 and 5 plus 12 is 17	46*	24	35
Add-Subtract	I added the first three then I took away 5	2	3	2
Grouping	There was already a 5, so you add 3 and 9	1	1	1
Ambiguous	I started at 5 and counted up	6	7	7
Incorrect procedures				
Add All	I added all those numbers up and it made 22	16	15	16
Add-to-Equal	I added up the 5, the 3, and the 9	8	13	11
Carry	I copied the 5 from over there	4	5	4
Add Two	Because 5 plus 5 is 10	3	4	4
Guess	I just guessed	4	7	6
Ambiguous	I put 15 in my head and counted on	10*	19	14

Note. *$p < .05$.

Table 3. Children's explanations (e.g., 'Why does *x* make this a true number sentence?')

		% Trials		
Explanation type	Sample explanation	Instruct–solve	Solve–instruct	Total
Equality concept	The two sides are supposed to be equal	46*	26	36
Correct procedure	Because adding 4 to 6 is the only one that makes 10, and 7 + 3 is 10	14	15	14
Incorrect procedure	You add it to all of them	12	15	13
Answer	Because it's the right answer	8^{τ}	14	11
Don't know	I don't know	8	12	10
Other/random	It's just a fact	13^{τ}	18	16

Note. $^{*}p < .05$; $^{\tau}p < .10$.

Table 4. Intraclass correlations for primary outcomes, controlling for predictor variables

Outcome	ICC
Procedural knowledge	
Post-test	.061
Retention test	.054
Conceptual knowledge	
Post-test	.066
Retention test	.174
Intervention	
Accuracy	.029
Conceptual explanation frequency	.043

the results because clustering in the data can lead to inflation of alpha values. Specifically, we ran all of the analyses reported below with a set of 11 dummy variables for the 12 classrooms in addition to the other predictors. Condition effects remained the same. Thus, we report results without the dummy variables for classroom.

Results

Pre-test and cognitive factors

At pre-test, conceptual knowledge was similar in the instruct–solve and solve–instruct conditions (M = 30%, SD = 21% vs. M = 32%, SD = 18%), $F < 1$, as was procedural knowledge (M = 18%, SD = 15% vs. M = 18%, SD = 13%), $F < 1$. Working memory span scores were also similar in the instruct–solve and solve–instruct conditions (M = 3.8, SD = 1.1 vs. M = 3.9, SD = 1.1), $F < 1$, as were retrieval fluency scores (M = 13.4, SD = 3.5 vs. M = 13.9, SD = 3.4), $F < 1$. Finally, there were no significant differences between conditions in terms of age, grade, gender, or ethnic minority status, ps > .20. Conditions were well matched at the onset of the study.

Post-test and retention test

To analyse children's performance, we examined procedural and conceptual knowledge using ANCOVAs. For procedural knowledge, the ANCOVA included time (post-test and retention test) and subscale (learning and transfer) as within-subject factors and condition (instruct–solve and solve–instruct) as the between-subject factor. For conceptual knowledge, the model was the same, except the two subscales were equal sign and equation structure. We also included pre-test scores, age, working memory, and fluency as covariates. Table 5 presents the main effects in each model for the primary outcomes.

Procedural knowledge

As shown in Figure 2, children's procedural knowledge was similar across learning and transfer subscales and also stable across time. Further, children in the instruct–solve condition exhibited higher procedural knowledge than children in the solve–instruct condition, and this was robust across subscale and time. Indeed, there was a main effect of condition (see Table 5), with children in the instruct–solve condition solving more procedural items correctly ($M = 62\%$, $SE = 4\%$) than children in the solve–instruct condition ($M = 49\%$, $SE = 4\%$). Further, there were no main effects of time or subscale, nor did they interact with condition or each other, Fs < 2.3.

Conceptual knowledge

As shown in Figure 3, children's conceptual knowledge was higher on the equal sign subscale than on the structure subscale, and their scores were relatively stable across time. Further, children in the instruct–solve condition exhibited higher conceptual knowledge than children in the solve–instruct condition, but only on the structure subscale. Indeed, there was a main effect of subscale, but not of time or condition (see Table 5). However, there was a marginal interaction between condition and subscale, $F(1, 115) = 3.53$, $p = .06$, $\eta_p^2 = .03$, but not with time, $F < 1$. To examine the condition by subscale interaction, we tested the effect of condition for each subscale. For the equal sign subscale, there was no effect of condition, $F < 1$. For the structure subscale, there was a significant effect of condition, $F(1, 115) = 4.66$, $p = .03$, $\eta_p^2 = .04$. Children in the instruct–solve condition ($M = 52\%$, $SE = 3\%$) exhibited higher knowledge of structure than children in the solve–instruct condition ($M = 43\%$, $SE = 3\%$).

Table 5. ANCOVA main effects for primary outcomes

	Procedural knowledge		Conceptual knowledge		Intervention accuracy		Explanation frequency	
	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2
Pre-test procedural	2.21	.02	0.19	.00	3.25	.03	1.89	.08
Pre-test conceptual	10.85**	.09	69.31***	.38	6.13*	.05	3.58**	.14
Age	2.04	.02	2.58	.02	4.95*	.04	2.59*	.11
Working memory	1.97	.02	2.83	.02	0.95	.01	1.14	.05
Retrieval fluency	8.57**	.07	9.49**	.08	2.73	.02	1.59	.07
Condition	4.72*	.04	1.75	.02	16.95***	.13	3.22**	.13
Time	0.68	.01	2.85	.02	–	–	–	–
Subscale	0.11	.00	16.59***	.13	–	–	–	–

Note. *$p < .05$; **$p < .01$; ***$p < .001$.

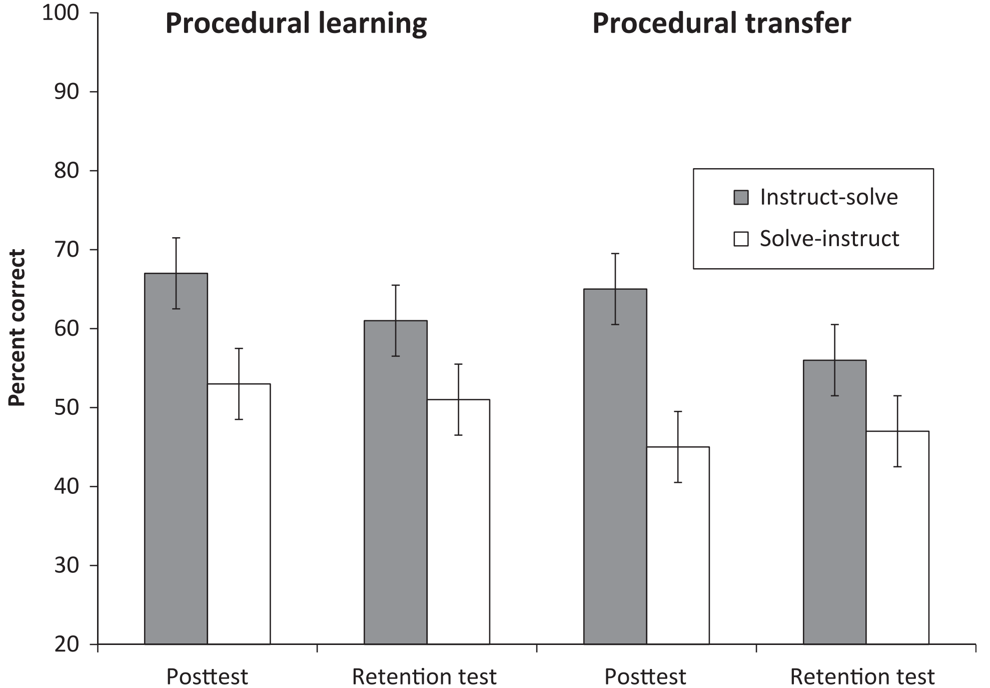

Figure 2. Procedural learning and transfer by condition at post-test and retention test. *Note*. Scores are estimated marginal means. Error bars represent standard errors.

Intervention activities

To better understand the superiority of the instruct–solve condition, we performed secondary analyses on intervention measures. We examined accuracy and explanations using a similar ANCOVA model as above, but without the time and subscale factors.

Accuracy

We examined children's accuracy during the intervention problem solving phase, focusing on the eight math equivalence problems. There was a main effect of condition (see Table 5). Children in the instruct–solve condition solved more intervention problems correctly (M = 4.4, SE = 0.3) than children in the solve–instruct condition (M = 2.7, SE = 0.3). Children were asked to describe how they solved four of the equivalence problems. As shown in Table 2, children reported correct procedures on close to half of those trials, which is consistent with their accuracy scores. The percentage of children using at least one correct procedure was significantly higher in the instruct–solve condition (67%) than in the solve–instruct condition (44%), $\chi^2(1, N = 121) = 6.14$, $p = .01$. Thus, prior instruction facilitated the generation of correct procedures.

Explanations

Children were prompted to self-explain on the eight equivalence problems. We examined the percentage of trials on which children used each explanation type in a MANCOVA and found a main effect of condition (see Table 5). Children in the instruct–solve condition

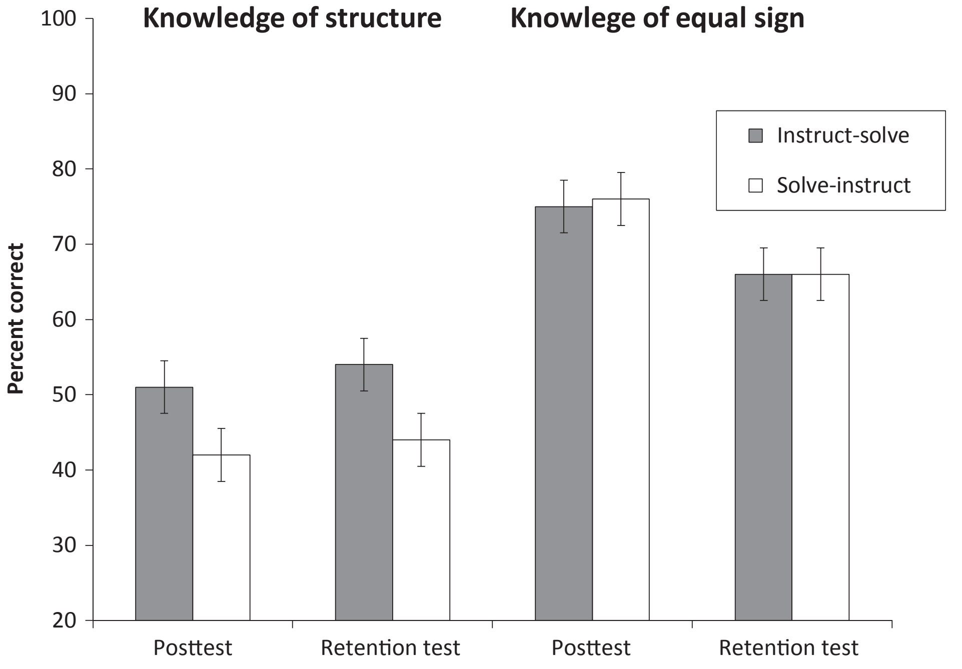

Figure 3. Conceptual knowledge of structure and equal sign by condition at post-test and retention test. *Note*. Scores are estimated marginal means. Error bars represent standard errors.

provided more equality-concept explanations than children in the solve–instruct condition, $F(1,\ 114) = 15.80$, $p < .001$, $\eta_p^2 = .13$. In contrast, children in the solve–instruct condition provided more answer-focused explanations, $F(1,\ 114) = 2.95$, $p = .09$, $\eta_p^2 = .03$, or random explanations that did not fall into an easily identifiable category, $F(1,\ 114) = 2.90$, $p = .09$, $\eta_p^2 = .03$. Thus, prior instruction enhanced the quality of children's explanations.

Mid-test

Children received a brief mid-test during the intervention that assessed their knowledge of structure. There was a main effect of condition, $F(1,\ 115) = 11.83$, $p = .001$, $\eta_p^2 = .09$. Children in the instruct–solve condition exhibited higher knowledge of equation structure (M = 48%, SE = 4%) than children in the solve–instruct condition (M = 31%, SE = 4%).

Exploratory mediation analyses

Given that performance during the intervention differed by condition, we explored whether these differences helped explain learning outcomes. Specifically, we examined whether problem solving accuracy and equality-concept explanations mediated the relationship between condition and learning outcomes. We used a bootstrapping technique recommended by Preacher and Hayes (2008), in which we obtained estimates for the indirect effect of condition. Bootstrapping involved the extraction of 5,000 samples from the data and an estimation of the indirect effect in each extracted data set.

This produced a 95% bias-corrected confidence interval for each indirect effect, which is significant if it excludes zero. In our models, we included condition as the between-subject factor, intervention accuracy and frequency of equality-concept explanations as the mediators, and the covariates from the primary analyses (pre-test scores, age, working memory, and fluency).

For procedural knowledge at post-test, problem solving accuracy (CI: 6.1, 25.6) and equality-concept explanations (CI: 0.8, 9.5) were significant mediators. Indeed, including both mediators rendered the direct effect of condition non-significant (p = .85). However, problem solving accuracy was a significantly stronger mediator than equality-concept explanations (CI: 0.7, 18.6). At retention, including both mediators reduced the effect of condition (p = .08), but only problem solving accuracy was a significant mediator (CI: 8.6, 27.4). A similar pattern emerged for conceptual knowledge of structure. At post-test, problem solving accuracy (CI: 3.1, 13.3) and equality-concept explanations (CI: 1.2, 8.8) were significant mediators, with no difference in the strength of these effects (CI: −3.1, 10.8). Indeed, including both mediators rendered the direct effect of condition non-significant (p = .72). At retention, including both mediators eliminated the effect of condition (p = .49), but only problem solving accuracy was a significant mediator (CI: 4.6, 17.4).

Discussion

Research in psychology and education indicates that the sequencing of learning material impacts the knowledge that learners construct (e.g., McNeil *et al.*, 2012; Rohrer, 2012). We evaluated the sequencing of conceptual instruction and problem solving for children learning math equivalence. Children received instruction on the concept of equivalence before or after solving and explaining challenging mathematics problems with feedback. In contrast to previous research (DeCaro & Rittle-Johnson, 2012), providing conceptual instruction first resulted in better procedural knowledge and conceptual knowledge of structure than delaying instruction. Children in the instruct–solve condition also solved more problems correctly and provided more conceptual explanations during the intervention. We discuss these key learning processes and then reflect on the results in the light of previous findings.

Potential explanatory mechanisms

Children in the instruct–solve condition outperformed children in the solve–instruct condition on most outcome measures. Given reports to the contrary, it is necessary to identify the mechanisms underlying these findings. Our results point to the roles of conceptual explanations and the acquisition of correct problem solving procedures.

Self-explanation involves generating explanations for oneself in an attempt to make sense of new information (Chi, 2000). Although the act of explaining can aid learning, researchers suggest that the content of the explanations matters (Chi, De Leeuw, Chiu, & LaVancher, 1994; Renkl, 1997). Importantly, instruction improved the quality of children's explanations in the current study. Children in the instruct–solve condition provided more equality-concept explanations than children in the solve–instruct condition. Consistent with prior research (Matthews & Rittle-Johnson, 2009), these conceptually oriented explanations were associated with greater learning outcomes.

Although the conceptual self-explanations played a positive role, our results highlight two caveats. First, self-explanation prompts that focus on why information is correct may be more effective after conceptual instruction. Here, the prompts allowed children to meaningfully integrate the instruction into the problem solving task (e.g., Wittwer & Renkl, 2008). Without the prior instruction, however, children often provided shallow explanations. Second, the positive role of conceptual self-explanations may diminish over time. Equality-concept explanations no longer mediated the effect of condition after a 2-week delay.

Acquiring correct problem solving procedures is another key source of cognitive change and can be a strong predictor of subsequent performance (e.g., Rittle-Johnson, 2006; Siegler & Shipley, 1995). Here, children who received conceptual instruction first were more likely to generate and use correct procedures than children who solved the problems first. Indeed, prior instruction often facilitates the generation of other domain knowledge (e.g., Chen & Klahr, 1999; Perry, 1991). For example, in Perry (1991), instruction on the concept of equivalence facilitated the generation and transfer of correct procedures, and this was more effective than directly telling children a correct procedure. Thus, prior conceptual instruction may facilitate problem solving during learning and greater knowledge gains overall.

Importantly, the mediating effect of problem solving accuracy was not limited to post-intervention measures of problem solving (i.e., procedural learning and transfer), but also generalized to measures of conceptual knowledge. Theoretically, this is consistent with the perspective that conceptual and procedural knowledge develop iteratively, with bi-directional relations between the two types of knowledge (e.g., Canobi, 2009; Cowan *et al.*, 2011; Dowker, 1998; Schneider, Rittle-Johnson, & Star, 2011). Here, we show that conceptual instruction facilitates the use of correct problem solving procedures, and, in turn, these improvements in problem solving are related to subsequent improvements in conceptual knowledge of structure.

These results support arguments that it is the processing and performance during learning that matters most (e.g., Klahr & Nigam, 2004; Rittle-Johnson, 2006). For example, in Klahr and Nigam (2004), children who mastered a scientific procedure during learning performed well on a transfer task, whether they generated the procedure or were told the procedure. Similarly, here, children who solved intervention problems correctly or explained them conceptually performed well on subsequent tests, regardless of condition. However, children in the instruct–solve condition were more likely to engage in these activities than children in the solve–instruct condition. Thus, 'what is learned is more important than how it is taught' (Klahr & Nigam, 2004, p. 662), but how it is taught still matters because of the processes that are supported.

There are likely additional processes at work that may help explain why children in the instruct–solve condition outperformed those in the solve–instruct condition in this study. For example, an instruct–solve approach may have been more consistent with the instructional practices used in their classrooms.

The timing of instruction

The current results contrast with a growing literature suggesting that instruction should be delayed. For example, researchers find that learners gain key experiences from problem solving that augments learning from subsequent instruction (e.g., DeCaro & Rittle-Johnson, 2012; Kapur, 2011; Schwartz *et al.*, 2011). Importantly, our results do not contradict the conclusions from these studies, but suggest a need to identify boundary conditions. Based on comparisons across these studies, we propose three factors that may influence the timing of instruction.

Type of instruction

First, the *type of instruction* provided might matter. In most studies in which the sequencing of instruction was manipulated, the instruction focused on concepts and procedures. For example, Schwartz *et al.* (2011) taught students the concept of density and the formula for computing it. When procedural instruction is provided first, students often apply the step-by-step procedure rather than explore the problems more broadly (e.g., Schwartz *et al.*, 2011). But procedural instruction may be beneficial after problem solving, as students often fail to generate correct procedures on their own (e.g., Kapur & Bielaczyc, 2012). In contrast, when instruction is solely conceptual in nature, it can be useful to provide the instruction first. Rather than transmitting ready-made solutions, conceptual instruction can guide the problem solving activity, thus enhancing its generative nature (e.g., Wittwer & Renkl, 2008). Solving subsequent problems can then ‘elicit an active processing of the [instructional] explanation’, resulting in the construction and integration of new knowledge (Berthold & Renkl, 2010, p. 36).

Self-explanation prompts

The type of instruction likely matters, but cannot fully explain the contrasting results. DeCaro and Rittle-Johnson (2012) also employed conceptual instruction, but found that delaying it was more effective. Another feature common to these two studies was the inclusion of self-explanation prompts during problem solving; however, the nature of the prompts varied. Thus, the *type of self-explanation prompts* may also influence the timing of instruction. The current prompts focused on knowing why a number sentence is true, which seemed to help children integrate the instruction with the problem solving activity. This supports suggestions that conceptually focused prompts are beneficial for knowledge integration (Berthold, Eysink, & Renkl, 2009). In DeCaro and Rittle-Johnson (2012), the prompts focused on correct and incorrect answers and thus were less conceptual. As a result, they seemed less likely to support the integration of information. Indeed, in DeCaro and Rittle-Johnson (2012), children rarely provided conceptual explanations. Overall, providing conceptual instruction first may be particularly effective if learners are supported in using this information via the use of conceptual self-explanation prompts.

Activation of misconceptions

Finally, the *activation of misconceptions* during problem solving may impact the timing of instruction. Previous studies on the timing of instruction have not involved tasks with common misconceptions or have not activated relevant misconceptions. In the current study, we employed two knowledge-activation techniques: inclusion of familiar problem types in line with a common misconception and side-by-side contrast with a novel problem type (e.g., Vosniadou & Vamvakoussi, 2006). Standard arithmetic problems can activate an ‘operational’ perspective, in which children focus on adding the numbers together and finding the total (McNeil & Alibali, 2005). This operational understanding conflicts with a relational understanding needed to solve math equivalence problems (McNeil, 2008).

If problem solving is difficult due to the activation of misconceptions, prior instruction is likely necessary to make children's effort and struggle more productive. Because children in the instruct–solve condition were taught the concept of equivalence beforehand, the instruction may have helped them make sense of the cognitive conflict during problem solving. In contrast, children exploring problems first may have experienced a counterproductive level of struggle.

The type of instruction, self-explanation prompts and activation of misconceptions likely influence the timing of instruction. Certainly, other factors matter as well, including characteristics of the learner and of the domain. Future research should attempt to determine the optimal sequencing of tasks across a variety of settings and populations.

Conclusion

The current results suggest that minor differences in the sequencing of learning materials can alter the knowledge that learners construct. By comparing the current findings on the sequencing of instruction with previous, contrasting findings, we can better understand when instruction should or should not be delayed. Despite arguments in favour of delaying instruction, the results suggest there is a time for providing conceptual instruction before problem solving. Providing only conceptual instruction, promoting conceptually based explanations and activating misconceptions during problem solving is one promising time.

Acknowledgements

Research supported by National Science Foundation Grant DRL-0746565 to Bethany Rittle-Johnson and Institute of Education Sciences, U.S. Department of Education, training grants R305B080008 and R305B080025. Fyfe is supported by a Graduate Research Fellowship from the National Science Foundation. The authors thank Abbey Loehr, Maryphyllis Crean, Polly Colgan, Rachel Ross, and Maddie Black for their assistance with data collection and coding as well as the staff, teachers, and children at Lockeland Elementary School and Sylvan Park Paideia Design Center.

References

Baroody, A. J., & Ginsburg, H. P. (1983). The effects of instruction on children's understanding of the "equals" sign. *Elementary School Journal*, *84*, 199–212. Retrieved from http://www.jstor.org/stable/1001311.

Berthold, K., Eysink, T. H. S., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science*, *37*, 345–363. doi:10.1007/s11251-008-9051-z

Berthold, K., & Renkl, A. (2010). How to foster active processing of explanations in instructional communication. *Educational Psychology Review*, *22*, 25–40. doi:10.1007/s10648-010-9124-9

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*, 322–330. doi:10.1016/j.cognition.2010.10.001

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology*, *26*, 571–593. doi:10.1207/s15326942dn2602_3

Canobi, K. H. (2009). Concept-procedure interactions in children's addition and subtraction. *Journal of Experimental Child Psychology*, *102*, 131–149. doi:10.1016/j.jecp.2008.07.008

Chen, Z., & Klahr, D. (1999). All other things being equal: Children's acquisition of the control of variables strategy. *Child Development*, *70*, 1098–1120. doi:10.1111/1467-8624.00081

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439–477. doi:10.1207/s15516709cog1803_3

Cowan, R., Donlan, C., Shepherd, D.-L., Cole-Fletcher, R., Saxton, M., & Hurry, J. (2011). Basic calculation proficiency and mathematics achievement in elementary school children. *Journal of Educational Psychology*, *103*, 786–803. doi:10.1037/a0024556

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

Dewey, J. (1910). *How we think*. Boston, MA: Heath.

Dowker, A. (1998). Individual differences in normal arithmetical development. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 275–301). Hove, England: Psychology Press.

Gaddes, W. H., & Crockett, D. J. (1975). The Spreen-Benton Aphasia Tests: Normative data as a measure of normal language development. *Brain and Language*, *2*, 257–280. doi:10.1016/S0093-934X(75)80070-8

Goldstone, R., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *Journal of the Learning Sciences*, *14*(1), 69–110. doi:10.1207/s15327809jls1401_4

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

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–27). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

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

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

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

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, *12*, 317–326. doi:10.1007.BF003110621

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*, 75–86. doi: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*, 661–667. doi:10.1111/j.0956-7976.2004.00737.x

Matthews, P., & Rittle-Johnson, B. (2009). In pursuit of knowledge: Comparing self-explanations, concepts, and procedures as pedagogical tools. *Journal of Experimental Child Psychology*, *104*, 1–21. doi:10.1016/j.jecp.2008.08.004

Matthews, P. G., 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*, 220–254. Retrieved from http://www.jstor.org/stable/10.5951/jresematheduc.43.3.0316

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, *59*, 14–19. doi:10.1037/0003-066X.59.1.14

McNeil, N. M. (2008). Limitations to teaching children 2 + 2 = 4: Typical arithmetic problems can hinder learning of mathematical equivalence. *Child Development*, *79*, 1524–1537. doi:10.1111/j.1467-8624.2008.01203.x

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*, 883–899. doi:10.1111/j.1467-8624.2005.00884.x

McNeil, N. M., Chesney, D. L., Matthews, P. G., Fyfe, E. R., Petersen, L. A., Dunwiddie, A. E., & Wheeler, M. C. (2012). It pays to be organized: Organizing arithmetic practice around equivalent values facilitates understanding of math equivalence. *Journal of Educational Psychology*, *104*, 1109–1121. doi:10.1037/a0028997

McNeil, N. M., & Fyfe, E. R. (2012). "Concreteness fading" promotes transfer of mathematical knowledge. *Learning and Instruction*, *22*, 440–448. doi:10.1016/j.learninstruc.2012.05.001.

Murdock, B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*, 482–488. doi:10.1037/h0045106.

Nokes, T. J., Hausmann, R. G. M., VanLehn, K., & Gershman, S. (2011). Testing the instructional fit hypothesis: The case of self-explanation prompts. *Instructional Science*, *39*, 645–666. doi:10.1007/s11251-010-9151-4

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

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*, 525–556. doi:10.3102/00346543074004525

Powell, S. R. (2012). Equations and the equal sign in elementary mathematics textbooks. *The Elementary School Journal*, *112*, 627–648. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3374577/

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891. doi:10.3758/BRM.40.3.879

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*, 1–29.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, *77*, 1–15. doi:10.1111/j.1467-8624.2006.00852.x

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

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, *24*, 355–367. doi:10.1007/s10648-012-9201-3

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928. doi:10.1126/science.274.5294.1926

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177. doi:10.1037/1082-989X.7.2.147

Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology*, *47*, 1525–1538. doi:10.1037/a0024997

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*, 475–522. doi:10.1027/s1532690xci1604_4

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*, 759–775. doi: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 efficacy of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*, 129–184. doi:10.1027/s1532690xci2202_1

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. Simon & G. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31–76). Hillsdale, NJ: Erlbaum.

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

Vosniadou, S., & Vamvakoussi, X. (2006). Examining mathematics learning from a conceptual change point of view. In L. Verschaffel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds.), *Instructional psychology: Past, present and future trends. Fifteen essays in honor of Eric de Corte* (pp. 55–70). Oxford, UK: Elsevier.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). San Antonio, TX: The Psychological Corporation.

Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, *43*, 49–64. doi:10.1080/00461520701756420

Received 8 October 2013; revised version received 26 December 2013