

Developmental Psychology

Relations Among Conceptual Knowledge, Procedural Knowledge, and Procedural Flexibility in Two Samples Differing in Prior Knowledge

Michael Schneider, Bethany Rittle-Johnson, and Jon R. Star

Online First Publication, August 8, 2011. doi: 10.1037/a0024997

CITATION

Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011, August 8). Relations Among Conceptual Knowledge, Procedural Knowledge, and Procedural Flexibility in Two Samples Differing in Prior Knowledge. *Developmental Psychology*. Advance online publication. doi: 10.1037/a0024997

Relations Among Conceptual Knowledge, Procedural Knowledge, and Procedural Flexibility in Two Samples Differing in Prior Knowledge

Michael Schneider
ETH Zurich

Bethany Rittle-Johnson
Vanderbilt University

Jon R. Star
Harvard University

Competence in many domains rests on children developing conceptual and procedural knowledge, as well as procedural flexibility. However, research on the developmental relations between these different types of knowledge has yielded unclear results, in part because little attention has been paid to the validity of the measures or to the effects of prior knowledge on the relations. To overcome these problems, we modeled the three constructs in the domain of equation solving as latent factors and tested (a) whether the predictive relations between conceptual and procedural knowledge were bidirectional, (b) whether these interrelations were moderated by prior knowledge, and (c) how both constructs contributed to procedural flexibility. We analyzed data from 2 measurement points each from two samples ($N_s = 228$ and 304) of middle school students who differed in prior knowledge. Conceptual and procedural knowledge had stable bidirectional relations that were not moderated by prior knowledge. Both kinds of knowledge contributed independently to procedural flexibility. The results demonstrate how changes in complex knowledge structures contribute to competence development.

Keywords: conceptual and procedural knowledge, procedural flexibility, mathematics learning, structural equation models

When children practice solving problems, does this also enhance their understanding of the underlying concepts? Under what circumstances do abstract concepts help children invent or implement correct procedures? How do knowledge of concepts and procedures each contribute to flexible problem solving in a domain?

These questions tap a central research topic in the field of cognitive development: the relations between conceptual and procedural knowledge. *Conceptual knowledge* can be defined as knowledge of the concepts of a domain and their interrelations, whereas *procedural knowledge* can be defined as the ability to execute action sequences to solve problems (e.g., Canobi, Reeve, & Pattison, 2003; Rittle-Johnson, Sielger, & Alibali, 2001). Both kinds of knowledge have been hypothesized to contribute to the ability to solve a range of problems flexibly and efficiently, so-called *procedural flexibility* (Blöte, van der Burg, & Klein, 2001; Kilpatrick, Swafford, & Findell, 2001; Star & Seifert, 2006).

The goals of the current study were threefold: (a) test whether the predictive relations between conceptual and procedural knowledge were bidirectional, (b) evaluate whether these interrelations

were moderated by prior knowledge, and (c) explore how both constructs contributed to procedural flexibility. The introduction is organized around these three goals.

Possible Relations Between Conceptual and Procedural Knowledge

A primary goal of the current study was to empirically test the longitudinal relationship between conceptual and procedural knowledge using a more rigorous methodology than has been applied in past research. There are four different theoretical viewpoints on the causal interrelations of conceptual and procedural knowledge, each one supported by some empirical evidence (cf. Haapasalo & Kadjevich, 2000; Rittle-Johnson & Siegler, 1998). *Concepts-first theories* posit that children initially acquire conceptual knowledge, for example, through parent explanations or by innate constraints, and then derive and build procedural knowledge from it through repeated practice solving problems (e.g., Gelman & Williams, 1998; Halford, 1993). *Procedures-first theories* posit that children first learn procedures, for example, by means of explorative behavior and then gradually derive conceptual knowledge from them by abstraction processes, such as representational redescription (e.g., Karmiloff-Smith, 1992; Siegler & Stern, 1998). A third possibility, sometimes labeled *inactivation view* (e.g., by Haapasalo & Kadjevich, 2000), is that conceptual and procedural knowledge are mutually independent (e.g., Resnick, 1982; Resnick & Omanson, 1987). On the basis of the plausibility of both the concepts-first view and the procedures-first view, Rittle-Johnson et al. (2001) proposed a fourth possibility with their *iterative model*.

Michael Schneider, Institute for Behavioral Sciences, ETH Zurich, Zurich, Switzerland; Bethany Rittle-Johnson, Psychology & Human Development Department, Vanderbilt University; Jon R. Star, Harvard Graduate School of Education, Harvard University.

Correspondence concerning this article should be addressed to Michael Schneider, Institute for Behavioral Sciences, ETH Zurich, Universitätstrasse 41, Zurich 8092, Switzerland. E-mail: schneider@ifv.gess.ethz.ch

The causal relations may be bidirectional, with increases in conceptual knowledge leading to subsequent increases in procedural knowledge and vice versa (Canobi & Bethune, 2008; Rittle-Johnson & Alibali, 1999).

Unfortunately, after decades of research, it is still not clear which of the four viewpoints on the longitudinal relations between conceptual and procedural knowledge is adequate. Low validity of the measures is one likely source of this problem (Schneider & Stern, 2010). Schneider and Stern used four measures of each kind of knowledge at three time points in a longitudinal design. Depending on which pair of measures they used for their analyses, the results supported the concepts-first view, the procedures-first view, the iterative model, or the inactivation view.

Further, if the four manifest measures for a particular knowledge type assessed the same kind of knowledge with high validity, they should be strongly related to a commonly underlying latent factor. This was not the case empirically. For each kind of knowledge and each measurement point, the latent factor explained less than 50% of the pooled variance of its indicators. More than half of the variance of each measure did not validly indicate conceptual or procedural knowledge but instead reflected unsystematic measurement error or task-specific competencies (e.g., verbal abilities on an explanation task or knowledge about diagrams in a diagram task). Schneider and Stern suggested that subsequent studies should pay greater attention to questions of measurement and that latent variable analyses are a useful tool for this. Therefore, a main goal of the current study was to investigate the predictive relations between conceptual and procedural knowledge by means of latent variable analyses. We expected our results to be in line with earlier studies that supported the assumption of bidirectional relations between the two kinds of knowledge over time (e.g., Canobi & Bethune, 2008; Rittle-Johnson & Alibali, 1999; Rittle-Johnson et al., 2001). Two further aims of our study were to test whether prior knowledge moderated the relations between conceptual and procedural knowledge and how both kinds of knowledge contributed to procedural flexibility.

Possible Moderating Influences of Prior Knowledge

The iterative model and the three competing viewpoints are all based on the assumption that the longitudinal relations between conceptual and procedural knowledge are always the same and, thus, stable over different situations. This could, for example, be the case if the relations between kinds of knowledge are determined by the stable architecture of the human information processing system (e.g., Anderson et al., 2004; Karmiloff-Smith, 1992; Sun, Merrill, & Peterson, 2001). This assumption has never been tested empirically, so the second goal of this study was to evaluate whether these interrelations between conceptual and procedural knowledge were moderated by prior knowledge.

Empirical results indicate that there is variability in the longitudinal relations between conceptual and procedural knowledge. In a review of the mathematics learning literature, Rittle-Johnson and Siegler (1998) found that the relations differed between domains, studies, and even participants within studies. This may explain why the four conflicting theoretical viewpoints have coexisted for many years: Each viewpoint is supported by some empirical studies but countered by others. In this study, we analyzed a possible source of this heterogeneity—a moderating effect of prior knowl-

edge on the assessments and the predictive relations between kinds of knowledge. Differences in prior knowledge are among the strongest sources of individual differences in learning processes (Ackerman & Cianciolo, 2000; Smith, diSessa, & Roschelle, 1994).

First consider children with little prior knowledge of the target content. Karmiloff-Smith (1992) analyzed knowledge acquisition processes in young children, who had little prior knowledge in their fields of learning. She concluded that, most likely, procedural knowledge developmentally precedes and facilitates a later conceptual understanding in many domains, because children first need to explore a domain on a practical level. For example, children typically learn counting procedures before they understand most of the underlying concepts (e.g., Frye, Braisby, Lowe, Maroudas, & Nicholls, 1989; LeFevre et al., 2006). Subsequently, they abstract the underlying concepts from concrete impressions and procedures gained by exploring. This suggests that for learners with little prior knowledge, the influence of procedural knowledge on the subsequent acquisition of conceptual knowledge might be stronger than vice versa.

Next consider children who have some prior knowledge of the target content. Many studies have shown that existing conceptual knowledge about the target content is one of the most important determinants of subsequent learning processes, including the acquisition of new procedures (Hecht, Close, & Santisi, 2003; Schneider, Grabner, & Paetsch, 2009). Conceptual knowledge is general and abstract, and thus can be generalized to new problem types. By contrast, procedural knowledge is more tied to routine problems familiar from practice (Rittle-Johnson et al., 2001). Therefore, when a learner needs to solve a problem he has never encountered before, prior conceptual knowledge should support the generation of new procedures (Gelman & Williams, 1998). This suggests that for learners with some prior knowledge of the target content, conceptual knowledge might have a stronger influence on the subsequent acquisition of procedural knowledge than vice versa. From a methodological point of view, this would mean that prior knowledge is a *moderator* of the predictive relations between conceptual and procedural knowledge (cf. Baron & Kenny, 1986).

The amount of prior knowledge a person has in a domain might not only moderate the predictive relations between conceptual and procedural knowledge, but also the validities of potential measures of the two kinds of knowledge. Researchers often assess individuals' procedural knowledge by routine tasks familiar from practice, because procedural knowledge is thought to be tied to these problems. On the other hand, conceptual knowledge is often measured by new problems, where people have to resort to their knowledge of domain concepts to construct new solution approaches (Bisanz & LeFevre, 1992; Gelman & Williams, 1998; Halford, 1993). However, what is a routine problem and what is a new problem changes as a result of prior knowledge and experience (Schneider & Stern, 2010). As a consequence, the same task might be a new problem that assesses conceptual knowledge for a person with little prior knowledge but a familiar problem that assesses procedural knowledge for a person with higher prior knowledge. This is especially a problem in pretest–posttest designs or longitudinal studies in which the same measure is used multiple times with the intention to assess one construct in the same way each time.

In the current study, we evaluated whether the interrelations of conceptual knowledge, procedural knowledge, and procedural flexibility were stable over two samples of students differing in their prior knowledge.

Relations to Procedural Flexibility

Conceptual and procedural knowledge are important sources of competence in a domain but certainly not the only sources. Another source of competence is procedural flexibility, where learners know multiple procedures and apply them adaptively to a range of situations (Baroody & Dowker, 2003; Rittle-Johnson & Star, 2007; Star & Seifert, 2006; Verschaffel, Luwel, Torbeyns, & Van Dooren, 2009). For example, expert mathematicians know and use more procedures than novices, even choosing to use different procedures when attempting identical problems on different occasions (Dowker, 1992).

Procedural flexibility is typically assessed both as (a) ability to solve problems in more than one way (often with prompting) and (b) success in choosing the most appropriate procedure to solve a given problem on the basis of problem features and situational demands (see Verschaffel et al., 2009, for a review). It is considered important because people who develop procedural flexibility are more likely to use or adapt existing procedures when faced with unfamiliar transfer problems and to have a greater understanding of domain concepts (e.g., Blöte et al., 2001; Hiebert & Wearne, 1996). For example, knowledge of multiple procedures for multidigit arithmetic calculations was related to greater accuracy on transfer problems and greater conceptual knowledge of arithmetic (Carpenter, Franke, Jacobs, Fennema, & Empson, 1998). Because of this, flexibility is discussed as an important component of mathematical proficiency in several recent mathematics education policy documents (Kilpatrick et al., 2001; U.S. Department of Education, 2008).

Although flexibility is assessed independently of conceptual and procedural knowledge in research studies, it is a less familiar construct to the education community. In particular, it is unclear how flexibility should be situated within the conceptual/procedural knowledge framework. As noted earlier, flexibility is related to procedural knowledge (e.g., Star, 2005) and to conceptual knowledge (e.g., Baroody, Feil, & Johnson, 2007). However, past research on flexibility has not included evidence for the validity of the measures nor evaluated whether development of flexibility is predicted by prior conceptual and/or procedural knowledge in the domain. Thus, our third goal was to provide much needed evidence for these two issues.

The Present Studies

Conceptual Knowledge, Procedural Knowledge, and Procedural Flexibility as Latent Variables

We had three research questions. First, can latent variable analyses replicate and validate earlier findings with manifest measures that indicate bidirectional, predictive relations between conceptual and procedural knowledge (Canobi & Bethune, 2008; Rittle-Johnson & Alibali, 1999; Rittle-Johnson et al., 2001)? Second, are the predictive relations between conceptual and procedural knowledge the same in a sample with low prior knowledge and in a

sample with higher prior knowledge? Third, how do conceptual and procedural knowledge each contribute to developing procedural flexibility, modeled as a latent variable?

To investigate these research questions, we analyzed data from two empirical investigations of the acquisition of conceptual knowledge, procedural knowledge, and procedural flexibility. Each study tested more than 200 middle school students' knowledge about linear equation solving before and after several lessons on the topic. The two samples differed in prior algebra instruction and, thus, in prior knowledge at the first measurement point. The students in Study 1 were tested near the beginning of the school year and had received very limited prior instruction on equation solving. The participants of Study 2 were tested toward the end of the school year and had participated in a prealgebra curriculum during that year, so that they already had some knowledge about the content of the study. In addition to differences in knowledge across studies, all students received instruction on the topic between the first and second measurement points, leading to knowledge increases from pretest to posttest.

In both studies, we modeled conceptual knowledge, procedural knowledge, and procedural flexibility as latent factors underlying our manifest (i.e., actually assessed) measures, as advocated for by Schneider and Stern (2010). Manifest measures that are used to estimate a latent factor are also referred to as *factor indicators* in the literature. Compared with manifest measures, latent factors have the advantage that they only take on variance that is common to all of their indicators, and thus, the variance does not reflect random noise in the data. For this reason, a latent factor usually reflects a construct with a reliability that is greater than the reliability of any of the manifest measures used to estimate the factor (Bollen, 2002; Ullman, 2007). Latent variable analyses allow for explicit tests about whether two latent factors, which could stand, for example, for two kinds of knowledge, are significantly different from each other. Latent variable analyses can also be used to test whether the relations between the manifest measures and the underlying latent factors change over time or groups of persons. The stability of factor loadings is referred to as *factorial measurement invariance* (Bollen, 2002; Vandenberg & Lance, 2000) and is desirable in many contexts because it aids the comparability of results across studies and across measurement points within studies. To our knowledge, our study is the first successful attempt to model the longitudinal relations between conceptual knowledge and procedural knowledge by means of latent variable analyses (cf. Schneider & Stern, 2010).

Equation Solving

We investigated our research questions in the domain of linear equation solving. Because equation solving draws on multiple principles and allows for a variety of solution procedures, it is an ideal domain for studying issues of learning and transfer (VanLehn, 1996). It is also an important topic for students to learn; it is considered a basic skill by many (e.g., U.S. Department of Education, 2008) and is recommended as a curriculum focal point in middle-school mathematics by the National Council of Teachers of Mathematics (2006).

Typically, conceptual knowledge measures for equation solving focus on understanding of equivalent expressions by asking students to decide and justify if two expressions or equations are

equivalent (e.g., Alibali, Knuth, Hattikudur, McNeil, & Stephens, 2007). In contrast, procedural knowledge for equation solving is typically measured by asking students to solve algebraic equations. In the current study, we focused on multistep linear equations such as $3 * (x + 1) + 6 = 5 * (x + 1)$, as they require some of the longest procedures that students must implement by eighth grade and because they can be solved in multiple ways (cf. Rittle-Johnson & Star, 2007). Finally, procedural flexibility is measured by asking students to solve problems in more than one way and to recognize and evaluate alternative solution procedures (Blöte et al., 2001; Rittle-Johnson & Star, 2007; Star & Rittle-Johnson, 2008).

Study 1

Rationale

In Study 1, we investigated the relations between conceptual and procedural knowledge and procedural flexibility with students who had had very limited instruction about equation solving prior to the first measurement point. Between the first and the second measurement point, each student participated in one of three experimental lessons on multistep equation solving. All students studied worked examples of different procedures for solving equations, discussed the examples with a partner, and solved practice problems. The lessons differed in whether and what types of comparisons students were asked to make when studying the worked examples. The interventions were described and empirically compared by Rittle-Johnson, Star, and Durkin (2009) using the manifest measures. In Study 1, we controlled for differences between intervention groups and instead reanalyzed the data from Rittle-Johnson et al. (2009) with a focus on the use of latent variables, the quality of the measures, and the longitudinal relations between the different types of knowledge over time, none of which have been reported previously.

Method

Participants. Participants were drawn from 11 classrooms at a low-performing, urban middle school in Massachusetts. In these classes, 72% of the students were White, 9% African American, 9% Hispanic, and 9% Asian American. Teachers identified classes they felt were prepared to learn about multistep equation solving; nine of the classes were eighth-grade classes (five regular- and four honors-level classes) and two of the classes were seventh-grade classes (both honors level). Students were using the Connected Mathematics 2 Curriculum (Lappan, Fey, Fitzgerald, Friel, & Phillips, 2009). Most teachers reported only spending a few days on linear equation solving. In five of the classes, teachers reported briefly introducing some of their students to multistep equations but provided little opportunity to practice solving them.

All 239 students from these classes participated. Three students were excluded from the analyses because they were absent from two of the three intervention sessions or did not complete the pretest and the posttest. Eight additional students were excluded because they had been the third member of a triad during the intervention phase, and our analysis model only allowed for data from dyads of students (see later). All subsequent results are reported for the remaining 228 students. Among these students

were 44 seventh-grade students and 184 eighth-grade students, and 130 were girls. The average age was 13.3 years (range 11.9–15.7 years). The school used the Measures of Academic Progress (MAP) as a norm-referenced test to measure mathematics achievement and growth. Students' average score was in the 74th percentile, but there was great variability, with scores ranging from the 12th to the 99th percentile.

Materials.

Assessments. The measures assessed conceptual knowledge, procedural knowledge, and procedural flexibility, as reported in Rittle-Johnson et al. (2009). Details of the assessment, including scoring criteria, are in the Appendix. We used the following three assessments. (a) There were 13 conceptual knowledge items designed to tap students' verbal and nonverbal knowledge of algebra concepts, such as maintaining equivalence and the meaning of variables. (b) The nine procedural knowledge items assessed students' ability to solve equations, with three familiar problems and six novel problems. Familiar problems had the same problem features (but not numbers) as problems presented during the intervention, whereas novel problems included a new problem feature. We also coded students' solution procedure on each item as using a correct algebraic procedure (i.e., included a simplified equation based on a valid transformation of the equation, such as distributing across the parentheses), an incorrect algebraic procedure (i.e., included a simplified equation based on an invalid transformation of the equation, such as distributing incorrectly), or an informal procedure such as guess-and-test. (c) The 20 flexibility knowledge items tapped students' knowledge of multiple procedures for solving equations and their ability to recognize and evaluate unfamiliar solution steps for accuracy and efficiency. The same assessments were administered at Times 1 and 2.

Two independent coders categorized the open responses across the assessment for 20% of the sample. Kappa scores ranged from .64 to 1.00, with a mean of .84. Discrepancies were discussed, and codes were altered when deemed appropriate by the primary coder.

Instruction packets. Students completed three 1-day lessons on multistep equations such as $3(x + 5) = 12$, $4(y + 2) + 6(y + 2) = 20$, and $7(n + 5) = 4(n + 5) + 9$. For a majority of each lesson, students studied packets of worked examples and answered reflection questions about the examples with a partner. The worked examples illustrated two different solution methods and three different versions of the packets were created that varied in whether and how the worked examples were paired. Packets either presented (a) the same problem solved with two different procedures side by side (i.e., compare methods), (b) two different problems solved with the same procedure side-by-side (i.e., compare problems), or (c) a single problem solved with one procedure on each page (i.e., sequential). As reported in Rittle-Johnson et al. (2009), for students who did not attempt to use algebra at pretest, comparing problems or sequential study of examples aided learning the most. For students who attempted algebra at pretest, comparing methods was most beneficial.

Procedure. The assessment was administered by a researcher in a group setting during one of students' regular mathematics classes (Time 1). Students were given 13 min to solve the equations in order to encourage use of efficient solution methods. They had about 30 min to finish the remainder of the assessment. For the next 3 days, students completed the three lessons. They were randomly paired with another student in their class for the partner

work. In eight cases, triads were created instead of dyads because of an uneven number of students. On the fifth day, the students completed the assessment again (Time 2).

Statistical analyses. In order to have the required minimum number of two indicators for each latent factor, for each of the three constructs in our study (i.e., conceptual knowledge, procedural knowledge, and procedural flexibility), we split the respective items into two groups (cf. Little, Cunningham, Shahar, & Widaman, 2002). We assigned items with odd item numbers to an Item Group A, items with even numbers to an Item Group B, and computed the sum score for each group. In the following, we refer to these sum scores as *manifest or measured variables*. When a latent factor stands alone, at least three indicators are needed for that factor to be identified. However, in more complex models where the factor is allowed to correlate with other constructs, two indicators for each factor are enough (Anderson & Gerbing, 1988, p. 415; Huizinga, Dolan, & van der Molen, 2006, p. 2030).

We specified the structural equation model (SEM) displayed in Figure 1. For each kind of knowledge and each measurement point, the respective Manifest Measures A and B served as indicators of a latent variable reflecting the amount of knowledge. At both measurement points, all latent factors could intercorrelate. At Time 2, also the residuals of the latent factors were allowed to intercorrelate. Procedural flexibility was modeled only for Time 2, because students typically exhibit little or no flexibility prior to experience solving problems in a domain (cf. Torbevens, Verschaffel, & Ghesquière, 2005), and in accordance with this, performance was near floor on most of our flexibility items at Time 1. Finally, we specified regression paths from all factors at Time 1 to all factors at Time 2, thus, creating a cross-lagged panel model (Burkholder & Harlow, 2003).

In our model, the factor loadings indicate the reliability of our measures. The lower the error variance of each manifest measure, the greater is the measures' loading on the underlying latent factor, which stands for the assessed construct (Ullman, 2007). The correlations between the latent factors indicated the divergent validity of the respective assessments. The higher these correlations are the higher is the degree of overlap between the measures of the two constructs and the lower is their divergent validity (Eid & Diener, 2006). Finally, the regression paths between the latent factors at Time 1 and Time 2 indicated the strength of the predictive relations between the constructs the factors stand for (Burkholder & Harlow, 2003).

In the following, we refer to this model, which is depicted in Figure 1, as our *basic model*. We derived a series of models from the basic model by introducing constraints on some of the model parameters, and we investigated how this affected the model fit to the empirical data.

The data were collected in different classrooms and for three treatment conditions within each classroom. If not accounted for, this multilevel structure of the data can lead to an underestimation of standard errors. We solved this problem by z standardizing each of our manifest measures separately for Time 1 and Time 2, first within each treatment condition and then within each classroom. Treatment condition and classroom were orthogonal to each other, because the students in each classroom were randomly assigned to the treatment conditions. Therefore, the two consecutive standardizations led to a data set with neither significant mean differences (all $p > .9$, all $\eta^2 < .002$) nor significant variance differences (all $p > .7$) between treatment conditions or classrooms, as confirmed by analyses of variance and Levene's tests carried out with each of our 10 manifest measures. Z-standardized measures

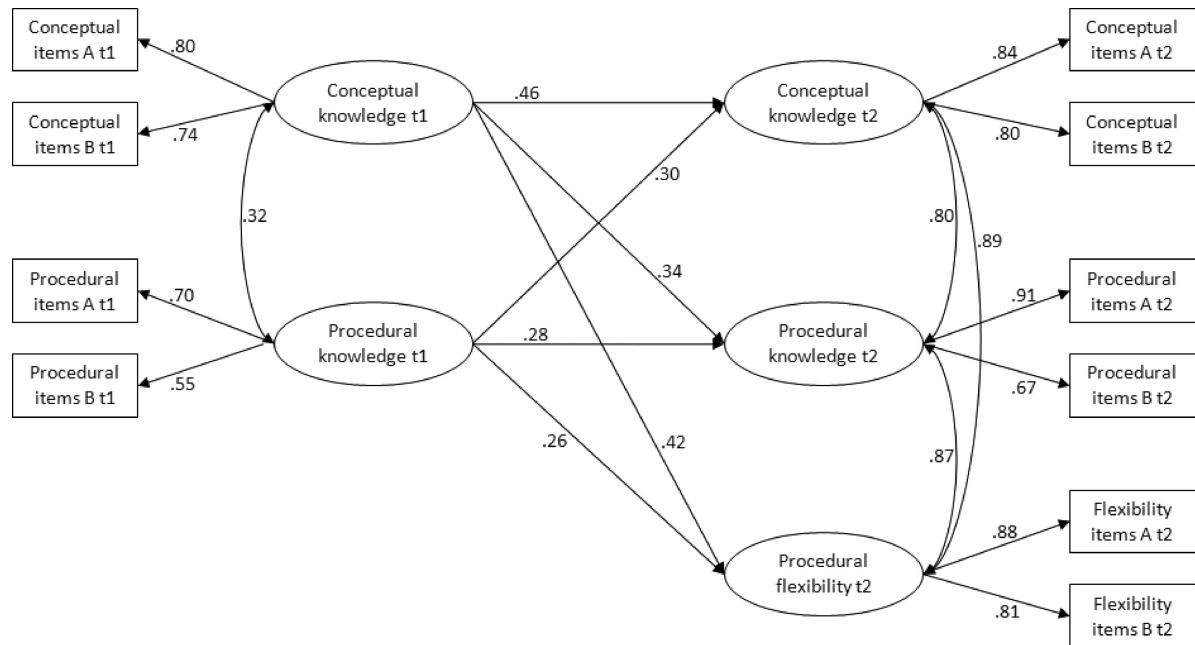


Figure 1. Factor loadings, factor intercorrelations, and regression paths of the best fitting structural equation model (Model 1k) of the relations among conceptual knowledge, procedural knowledge, and procedural flexibility in Study 1. A = Item Group A (sum of items with odd numbers); B = Item Group B (sum of items with even numbers); t1 = Time 1; t2 = Time 2. All estimated coefficients are significant with $p < .01$.

have a mean of 0 and a variance of 1. Accordingly, the intraclass correlations were $< .01$ for all measures, because the standardization left no mean differences between classes or treatment groups.

The students learned in dyads during the intervention phase. Therefore, persons were not independent units of analysis in our study. If not accounted for, this dyadic structure of the data would lead to an underestimation of the standard errors and an overestimation of the statistical significances in our analyses. Two approaches are frequently used for modeling dyadic data: multilevel modeling and treating dyads as unit of analysis (Kenny, in press). In the former approach, differences between persons within each dyad and differences between dyads are modeled simultaneously. Thus, multilevel regression models can include predictor variables on both levels, individual persons and dyads. A drawback of multilevel models are the large sample sizes needed to obtain valid results, in particular with complex structural equation models like the ones in our study (cf. Meuleman & Billiet, 2009). Indeed, when we tried modeling the multilevel structure of our data, we ran into convergence problems. For this reason, we used the alternative approach and treated dyads as units of analysis (i.e., each line in our data set corresponded to one dyad), as recommended by Kenny, Kashy, and Cook (2006, p. 100). This approach does not decompose overall effects into person-level effects and dyad-level effects and, thus, is statistically less demanding.

We treated dyads ($n = 114$) as the unit of analysis, with each unit consisting of two persons with two (at Time 1) or three (at Time 2) latent factors, respectively (see also Newsom, 2002). Thus, we specified all latent factors and paths displayed in Figure 1 twice, once for Person 1 and once for Person 2. For example, the model comprised two latent factors for conceptual knowledge at Time 1, one for Person 1 and one for Person 2 of the dyad. The model also comprised two correlations between conceptual and procedural knowledge at Time 1, one for Person 1 and one for Person 2.

It was arbitrary who was Person 1 and who was Person 2 of each dyad. We did not expect to find different model parameters for Persons 1 and Persons 2, because both belong to the same population. For example, there is no reason to expect that the correlation between conceptual and procedural knowledge would be different for all Persons 1 in the sample than for all Persons 2 in the sample. We, thus, constrained all factor loadings and paths coefficients to be equal for Person 1 and Person 2, which is why Figure 1 displays all relevant outcomes of our model. For further details on modeling dyadic data in SEM, see Kenny et al. (2006).

We analyzed the covariance matrix of our measures in the program Mplus (Muthén & Muthén, 1998–2007) by means of maximum-likelihood estimation. Missing data were handled by the full-information maximum likelihood (FIML) procedure implemented in Mplus. Data were missing because students were occasionally absent from class or did not finish the assessments in the available time, with data missing for only 6% of students. We assumed the data were missing at random, a requirement for the FIML procedure. We set the factor metrics by fixing the loading of the first indicator of each latent factor to one. Some of our measures did not follow a normal distribution (see Table 1). We accounted for this by bootstrapping the standard errors of our model parameters using 500 draws (cf. Nevitt & Hancock, 2001).

Results

As expected, students were low in prior knowledge. At Time 1, their accuracy was only 24% on the conceptual knowledge items and 20% on the procedural knowledge items. On the procedural knowledge items, only 23% of students used a correct algebraic solution method to solve at least one equation, and correct algebraic methods were only used on 9% of items.

Table 1 displays the solution rates together with Cronbach's alpha for each scale, the number of items on that scale, and the number of valid cases. Where the number of valid cases is smaller than the total sample size, students were absent from a data collection session or did not complete all tests. Cronbach's alphas range from .39 to .74. Although some are lower than is advisable for psychometric tests of homogeneous constructs, the obtained values are still acceptable in the context of this article because persons' knowledge in a domain usually has multiple facets (e.g., different procedures, different concepts). Students can know one facet without necessarily knowing all other facets (Schneider & Stern, 2009). The low alpha coefficients indicate this partly fragmented nature of knowledge and conform to previous findings from similar analyses (Schneider & Stern, 2010).

Before fitting the full model, we tested, separately for each measurement point and each pair of constructs (conceptual knowledge, procedural knowledge, and procedural flexibility), whether they were better fit by a two-factors model (with the two factors standing for the two constructs) than by a more parsimonious one-factor model (Research Question 2). The former case would indicate good divergent validities of our measures. In particular, evidence that all three con-

Table 1
Study 1 Performance Summary: Percentage Correct, Standard Deviation, Cronbach's Alpha, Number of Items, Number of Valid Cases, Skewness, and Kurtosis for Each Manifest Measure

Measure	Items A							Items B						
	<i>M</i>	<i>SD</i>	α	<i>n</i> items	<i>n</i> valid cases	Skewness	Kurtosis	<i>M</i>	<i>SD</i>	α	<i>n</i> items	<i>n</i> valid cases	Skewness	Kurtosis
Pretest														
Conceptual knowledge	30	25	.69	7	221	0.91	0.05	18	19	.40	6	217	0.92	0.14
Procedural knowledge	17	21	.57	5	225	1.42	1.92	23	23	.39	4	225	0.67	-0.39
Posttest														
Conceptual knowledge	54	26	.62	7	220	-0.03	-0.96	42	26	.58	6	218	0.25	-0.76
Procedural knowledge	47	34	.75	5	220	0.18	-1.30	35	31	.57	4	220	0.50	-0.81
Procedural flexibility	61	22	.68	10	216	0.03	-0.94	54	22	.71	10	215	0.17	-0.88

structs are distinct from each other in pairwise comparisons supports a model with three latent factors (at Time 2).

The fits of the estimated models are displayed in the upper part of Table 2 (Models 1a–1h). A comparative fit index (CFI) greater than .95, a root-mean-square error of approximation (RMSEA) less than .05, and a standardized root-mean-square residual (SRMR) less than .08 indicate a good absolute fit of a model. When two models are compared, the model with the smaller Akaike information criterion (AIC) or Bayesian information criterion (BIC) should be chosen. These information criteria combine the absolute fit of a model with a correction function penalizing for less parsimonious model assumptions. The model fit indices differ in their advantages and disadvantages. Thus, some indices can indicate a good fit of a model while others indicate a bad fit (Hu & Bentler, 1999; Ullman, 2007).

At the first measurement point, the two-factors model (Model 1b) has a better fit than the one-factor model as indicated by all indices. At the second measurement point, the picture is more mixed. For each pair of constructs, some fit indices are better for the two-factors models (Models 1d, 1f, and 1h), and other fit indices are better for the one-factor models (Models 1c, 1e, and 1g). In particular, the AIC is lower for the two-factors models, and the BIC is lower for the one-factor models. Thus, the divergent validities of our measures are low. We modeled the constructs by separate latent factors in the subsequent longitudinal analyses in spite of this for three reasons. First, conceptual knowledge and procedural knowledge are clearly distinct at Time 1. Second, the distinction of conceptual knowledge, procedural knowledge, and procedural flexibility is of a high theoretical importance and cannot be investigated empirically when the constructs are not modeled as separate factors. Third, as expected and as we show empirically in a later section, conceptual knowledge and procedural knowledge influence each other over time. Therefore, the very close relation of the latent factors at Time 2 is in line with our theory.

In a next step, we estimated the fit of three different versions of our basic model to investigate the longitudinal relations between the

different types of knowledge (see Method section). Their fits are displayed in the lower part of Table 2. Model 1i was the basic model itself. In Model 1j, we constrained the factor loadings to be equal at Time 1 and Time 2 (i.e., we assumed weak factorial measurement invariance; Research Question 3). Model 1j is more parsimonious than Model 1i, because a smaller number of parameters (i.e., in this case, different factor loadings) have to be estimated. We derived Model 1k from Model 1j by restraining the predictive relations from conceptual knowledge at Time 1 to procedural knowledge at Time 2 and from procedural knowledge at Time 1 to conceptual knowledge at Time 2 to be equally strong. All three models have good CFI, RMSEA, and SRMR coefficients. In addition, Model 1k has the lowest AIC (together with Model 1j) and the lowest BIC and is more parsimonious than the other two models. In structural equation modeling, when alternative models have about the same fit, one chooses the most parsimonious of the models, because it is more simple and, thus, more economical than its competitors (Ullman, 2007). Therefore, Model 1k, with invariant factor loadings and equally strong relations between conceptual and procedural knowledge, describes the relations between the variables obtained in Study 1 best.

The coefficients of Model 1k are displayed in Figure 1. All factor loadings were greater than .5 and were significant with $p < .01$, demonstrating acceptable reliabilities of our measures. The factor loadings were constrained to be equal at Time 1 and Time 2, but they could vary slightly among the four cases due to technical details of their standardization in Mplus (i.e., only unstandardized coefficients can be constrained, but we report standardized coefficients). The factor intercorrelations are low at Time 1 and very high at Time 2, indicating an increasing overlap and a decreasing divergent validity of the measures of the three kinds of knowledge. The bidirectional predictive relations between conceptual and procedural knowledge (Research Question 1) had standardized regressions coefficients of about .3, were significantly greater than 0, and were in the same range as the findings reported by Rittle-Johnson et al. (2001). Both conceptual and procedural knowledge at Time 1 contributed significantly to procedural flex-

Table 2
Fit Indices of the Models Estimated for Study 1

Model type, measurement point, and model no.	Model description	χ^2	df	p	CFI	RMSEA	SRMR	AIC	BIC
<i>Divergent validities</i>									
Time 1									
1a	Conceptual and procedural knowledge, 1 factor	112	23	<.001	.576	.184	.140	10632	10689
1b	Conceptual and procedural knowledge, 2 factors	14	16	.573	1.000	.000	.056	10548	10625
<i>Time 2</i>									
1c	Conceptual and procedural knowledge, 1 factor	46	22	.002	.903	.098	.061	10238	10299
1d	Conceptual and procedural knowledge, 2 factors	29	17	.039	.953	.077	.077	10231	10305
1e	Conceptual knowledge and procedural flexibility, 1 factor	29	22	.155	.981	.052	.045	10046	10107
1f	Conceptual knowledge and procedural flexibility, 2 factors	9	16	.924	1.000	.000	.033	10038	10115
1g	Procedural knowledge and procedural flexibility, 1 factor	31	22	.093	.974	.060	.060	10078	10139
1h	Procedural knowledge and procedural flexibility, 2 factors	14	16	.579	1.000	.000	.047	10074	10150
<i>Longitudinal relations</i>									
1i	Basic model	166	156	.271	.990	.024	.074	25552	25755
1j	Invariant factor loadings assumed	167	158	.299	.991	.022	.074	25549	25745
1k	Invariant factor loadings and symmetrical predictive relations assumed	169	159	.276	.990	.024	.078	25549	25743

Note. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.

ibility at Time 2, with standardized regression coefficients of .42 and .26, respectively (Research Question 3).

Discussion

In all, the findings from Study 1 indicate that conceptual knowledge, procedural knowledge, and procedural flexibility were assessed reliably and at least partly independently of each other. At Time 1, the two-factors model clearly had a better fit than the one-factor model. However, at Time 2, some fit indices were better for the two-factors models while others were better for the one-factor models. This supports the notion that conceptual and procedural knowledge can sometimes partly overlap and are hard to measure independently of each other (Schneider & Stern, 2010). It should be noted, though, that this problem seems large or negligibly small depending on the measurement point.

The three longitudinal models that modeled conceptual knowledge, procedural knowledge, and procedural flexibility as three separate entities had excellent fits to the data and suggest that in the overall context of a longitudinal study, these three constructs can, indeed, be modeled as interrelated but separate latent factors. The relatively high factor loadings that changed only modestly across measurement points demonstrate that our constructs were assessed with acceptable reliabilities. The comparison of the three alternative longitudinal models demonstrated the adequacy of the iterative model, which assumes bidirectional predictive relations. In addition, conceptual and procedural knowledge at Time 1 each predicted students' procedural flexibility at Time 2, supporting the importance of both types of knowledge for gaining flexibility.

Study 2

Rationale

In Study 1, students had received limited instruction in algebra and very little instruction on equation solving. In Study 2, we investigated whether the relations among conceptual knowledge, procedural knowledge, and procedural flexibility were different when students had already received classroom instruction on equation solving and thus had greater domain knowledge. The method and analyses were almost exactly the same as in Study 1. The data from Study 2 have not been analyzed and published before.

Method

Participants and procedure. Participants were drawn from two urban public middle schools from the same school district; we had worked with one of the schools in Study 1, but during a different school year. The students were diverse, in terms of ethnicity (55% White, 20% Hispanic, 15% Asian, 9% African American, and 1% Native American), language spoken at home (35% spoke a language other than English at home), and socio-economic background (55% received free- or reduced-price lunch). The head of the math department at each school identified 2 seventh-grade advanced-level classes, 10 eighth-grade advanced-level classes, and 2 eighth-grade regular-level classes they felt were prepared for solving multistep equations. Teachers used a mixture of the Connected Mathematics 2 curriculum and algebra I textbooks. Most teachers reported spending a few months on linear

equation solving, all reported discussing and practicing multistep equations, and a majority (8/14) reported working on linear equation solving again within a month of the study.

All 325 students from these classes participated. Fourteen students were excluded from analyses because they were absent for both assessments or for two of the three intervention sessions. Seven additional students were excluded because they were the third member of a triad during the intervention. Of the remaining 304 students, 245 were in advanced mathematics classes, 154 were girls, and the average age was 14.0 years (range 12.0–16.3 years). On average, they scored in the 64th percentile on a standardized math assessment, but there was a wide range (from 9th to the 99th percentile).

The procedure and analyses were identical to Study 1. We analyzed data from 157 dyads (i.e., 147 dyads with two members and 10 dyads with only one member, due to the exclusion of some participants from the analyses as described earlier).

Materials. The assessment was the same as the one used in Study 1 except that two equations on the procedural flexibility assessment were modified without changing the nature of these tasks. Interrater reliability on open-response items for 20% of the sample, measured by kappa, ranged from .62 to 1.00, with a mean of .88.

The intervention packets were very similar to those used in Study 1, with three different conditions varying in how the worked examples were paired. Two of the conditions were the same as in Study 1 (compare methods and compare problems), and the third condition had students compare equivalent equations—for example, $3(x + 1) = 12$ and $4(y + 2) = 16$ —solved with the same procedure. In all packets, there were four fewer worked examples presented in each lesson than in Study 1. There was no main effect of condition in this study.

Results and Discussion

Differences between samples in Studies 1 and 2 at pretest. The defining difference between the samples was that students in Study 2 had received a significant amount of instruction related to the topic under study, while the sample in Study 1 had received very limited prior instruction on equation solving. We thus assumed that the students in Study 2 had more relevant prior knowledge than the students in Study 1. If this was the case, the students in Study 2 should have more conceptual and procedural knowledge for equation solving than the students in Study 1. We used tests for the differences on our measures as an implementation check, although our measures were not intended to capture the full range of algebra knowledge that students in Study 2 were expected to have.

As expected, students in Study 2 had greater prior knowledge at Time 1 than students in Study 1. Their accuracy was higher on the conceptual knowledge items ($M = 33\%$, $SD = 27$, in Study 2 vs. $M = 25\%$, $SD = 20$, in Study 1), Mann-Whitney $U = 25892$, $p = .005$, and on the procedural knowledge items ($M = 32\%$, $SD = 32$, in Study 2 vs. $M = 19\%$, $SD = 20$, in Study 1), Mann-Whitney $U = 25878$, $p < .001$. The largest difference between the two samples was their use of algebra to solve equations on the procedural knowledge assessment. Students in Study 2 solved 45% of the equations using a correct algebraic procedure at pretest, whereas students in Study 1 only solved 9% of the equations using

a correct algebraic procedure, Mann–Whitney $U = 17222$, $p < .001$. This greater use of algebraic procedures in Study 2 was due in large part to many more of the students in this study using a correct algebraic procedure at least once: 61% vs. 22% of students, $\chi^2(1) = 75.148$, $p < .001$. Most students in Study 2 had spent several months studying linear equation solving, and this added instructional time had the greatest impact on how students solved the equations.

Study 2 analyses. Table 3 shows the percentage correct, Cronbach's alpha, number of items, and number of valid cases for each manifest measure. Cronbach's alphas were relatively good. Percentage correct was higher at Time 2 than at Time 1, indicating that the students gained additional knowledge of equation solving by participating in the interventions between Time 1 and Time 2.

As in Study 1, we compared the fit of a two-factors model and the fit of a one-factor model for each pair of constructs and each measurement point (see the upper part of Table 4; Research Question 2). The two-factors models (Models 2b, 2d, 2f, and 2h) had excellent fits to the data based on all indices. In contrast, none of the one-factor models fit the data well. Both the AIC and BIC values indicated that the two-factors models were a better fit than the one-factor models, providing evidence for the divergent validity of the measures in Study 2.

In a second step of our analyses, we fit a series of longitudinal models to the data. These three models, Models 2i, 2j, and 2k, were the same as Models 1i, 1j, and 1k in Study 1. CFI, RMSEA, and SRMR indicated good fits of all three models. The values of AIC and BIC were best for Model 2k, which assumed invariant factor loadings and symmetrical longitudinal relations between conceptual and procedural knowledge.

The estimated coefficients of Model 2k are displayed in Figure 2. All factor loadings were greater than or equal to .7, indicating good reliabilities of our measures. Conceptual knowledge, procedural knowledge, and procedural flexibility were intercorrelated with $rs = .6$ at both measurement points, but could still be assessed with good divergent validities as established before by Models 2b, 2d, 2f, and 2h. The results indicate bidirectional relations between conceptual and procedural knowledge over time (Research Question 1), and these relations were equally strong in both directions. Conceptual and procedural knowledge at Time 1 also contributed to procedural flexibility at Time 2 (Research Question 3).

Comparison of the data from Study 1 and Study 2. In order to explicitly test whether relations in our structural equation models differed between the two studies, we combined the data sets

from Study 1 and Study 2 into a single data set and fit a series of multigroup models (Muthén & Muthén, 1998–2007) with the samples of the two studies being the two groups compared. Model coefficients were estimated independently for the two groups.

The fits of the three multigroup models are displayed in Table 5. Model 3a has the same structure as the basic Models 1a and 2a. However, it was specified as a multigroup model. All parameters were allowed to vary between the two groups. Separately for each group, constant factor loadings over time and symmetrical predictive interrelations of conceptual and procedural knowledge were assumed, because this had been established by the previous analyses (Model 1k and Model 2k).

Model 3b was derived from Model 3a by additionally constraining the patterns of factors loadings to be equal for the two groups. Model 3c was derived from Model 3b by additionally specifying the predictive relations between conceptual and procedural knowledge to be equally strong in both samples. CFI, RMSEA, and SRMR indicate good fits of all three models to the data. AIC and BIC are best for Model 3c, which is most parsimonious. In this model, the standardized coefficients of the regression paths between conceptual knowledge and procedural knowledge at Time 1 and Time 2 all lie between .29 and .35.

In all, the predictive relations between conceptual and procedural knowledge were bidirectional (Research Question 1), equally strong in both directions, and constant across the two samples (Research Question 2). Conceptual and procedural knowledge both contributed to procedural flexibility (Research Question 3).

General Discussion

Longitudinal Relations Between Conceptual and Procedural Knowledge

The relations between conceptual and procedural knowledge during learning and development have been hotly debated for decades. However, many of these publications ignored the problem of measuring the two kinds of knowledge validly and partly independently of each other. In the current studies, we modeled conceptual and procedural knowledge as latent variables. This allowed us to better account for the indirect relation between overt behavior and the underlying knowledge structures than was possible in previous research. A cross-lagged panel design allowed us to directly test and compare the predictive relations from conceptual knowledge to procedural knowledge and vice versa.

Table 3
Study 2 Performance: Percentage Correct, Standard Deviation, Cronbach's Alpha, Number of Items, Number of Valid Cases, Skewness, and Kurtosis for Each Manifest Measure

Measure	Items A							Items B						
	M	SD	α	n items	n valid cases	Skewness	Kurtosis	M	SD	α	n items	n valid cases	Skewness	Kurtosis
Pretest														
Conceptual knowledge	38	31	.77	7	285	0.43	-0.96	27	26	.64	6	285	0.77	-0.21
Procedural knowledge	35	35	.81	5	285	0.60	-1.02	30	32	.67	4	285	0.77	-0.61
Posttest														
Conceptual knowledge	55	30	.73	7	286	-0.19	-1.10	48	32	.74	6	286	0.08	-1.18
Procedural knowledge	56	36	.81	5	286	-0.23	-1.38	47	37	.73	4	286	0.05	-1.40
Flexibility knowledge	60	26	.78	10	286	-0.31	-0.63	59	27	.82	10	286	-0.24	-0.92

Table 4
Fit Indices of the Models Estimated for Study 2

Model type, measurement point, and model no.	Model description	χ^2	df	p	CFI	RMSEA	SRMR	AIC	BIC
Divergent validities									
Time 1									
2a	Conceptual and procedural knowledge, 1 factor	62	23	<.001	.809	.104	.071	13509	13573
2b	Conceptual and procedural knowledge, 2 factors	13	14	.526	1.000	.000	.031	13479	13570
Time 2									
2c	Conceptual and procedural knowledge, 1 factor	50	22	.001	.926	.090	.063	13537	13604
2d	Conceptual and procedural knowledge, 2 factors	18	17	.381	.997	.021	.049	13515	13597
2e	Conceptual knowledge and procedural flexibility, 1 factor	58	22	<.001	.927	.103	.066	13424	13491
2f	Conceptual knowledge and procedural flexibility, 2 factors	20	16	.209	.991	.042	.034	13398	13483
2g	Procedural knowledge and procedural flexibility, 1 factor	69	22	<.001	.906	.118	.078	13429	13496
2h	Procedural knowledge and procedural flexibility, 2 factors	14	16	.578	1.000	.000	.034	13386	13471
Longitudinal relations									
2i	Basic model	180	156	.092	.978	.031	.065	33398	33624
2j	Invariant loadings assumed	180	158	.112	.980	.030	.065	33394	33614
2k	Invariant loadings and symmetrical predictive relations assumed	181	159	.109	.980	.030	.065	33393	33610

Note. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.

Our empirical results strongly support an iterative model, which poses bidirectional relations between the two kinds of knowledge over time. The predictive relations between conceptual and procedural knowledge from Time 1 to Time 2 were significant and lay in the range from .29 to .35 in the most comprehensive model (Model 3c). The relations were not only bidirectional but even symmetrical.

These findings are in line with a number of recent studies that have found indirect evidence for bidirectional relations between conceptual and procedural knowledge in various mathematical domains and by means of different methods. In particular, conceptual sequencing of practice problems supported improvements in 7- and 8-year-old children's procedural and conceptual knowledge about arithmetic (Canobi, 2009), direct instruction on one type of knowledge led to improvements in the other type of knowledge in fourth and fifth graders' learning about equivalence problems (Rittle-Johnson & Alibali, 1999), and iterating between lessons on concepts and procedures on decimals supported greater procedural knowledge than presenting concept lessons before procedure lessons in a sample of sixth graders (Rittle-Johnson & Koedinger, 2009). In our study, we confirmed these findings using a more adequate method and extended the findings by looking at older student learning, a more complex topic with multistep procedures.

Overall, converging empirical evidence from different content areas, age groups, and research methods strongly supports an iterative model of the development of conceptual and procedural knowledge. Hence, instruction focusing on only one of the two kinds of knowledge is not desirable. Conceptual knowledge may help with the construction, selection, and appropriate execution of

problem-solving procedures. At the same time, practice using procedures may help students develop and deepen understanding of concepts. Both kinds of knowledge are intertwined and can strengthen each other over time.

A Lack of Moderating Influences of Prior Knowledge

Students' prior knowledge did not moderate the predictive relations between conceptual and procedural knowledge. We compared two samples differing in their amount of instruction and, thus, in their amount of prior conceptual and procedural knowledge for equation solving. A multigroup structural equation model allowed us to analyze the data from the two samples simultaneously and to explicitly test for differences between the learning processes in the two groups. Despite the statistical power that comes with our sample of 532 participants, we found no evidence for a moderating effect of prior knowledge on the predictive relations between conceptual and procedural knowledge. Future research is needed to test for moderating effects for larger differences in prior knowledge or in other content domains.

One explanation for the coherence of these findings and for the lack of a moderating effect of prior knowledge is that the bidirectional relations between conceptual and procedural knowledge are a basic property of the architecture of the human information processing system. Unfortunately, so far, there is virtually no connection between studies of conceptual and procedural knowledge and studies of the architecture of cognition (e.g., Anderson et al., 2004). Making these connections would be a breakthrough for future research on conceptual and procedural knowledge, because it would enable researchers not only to describe the relations

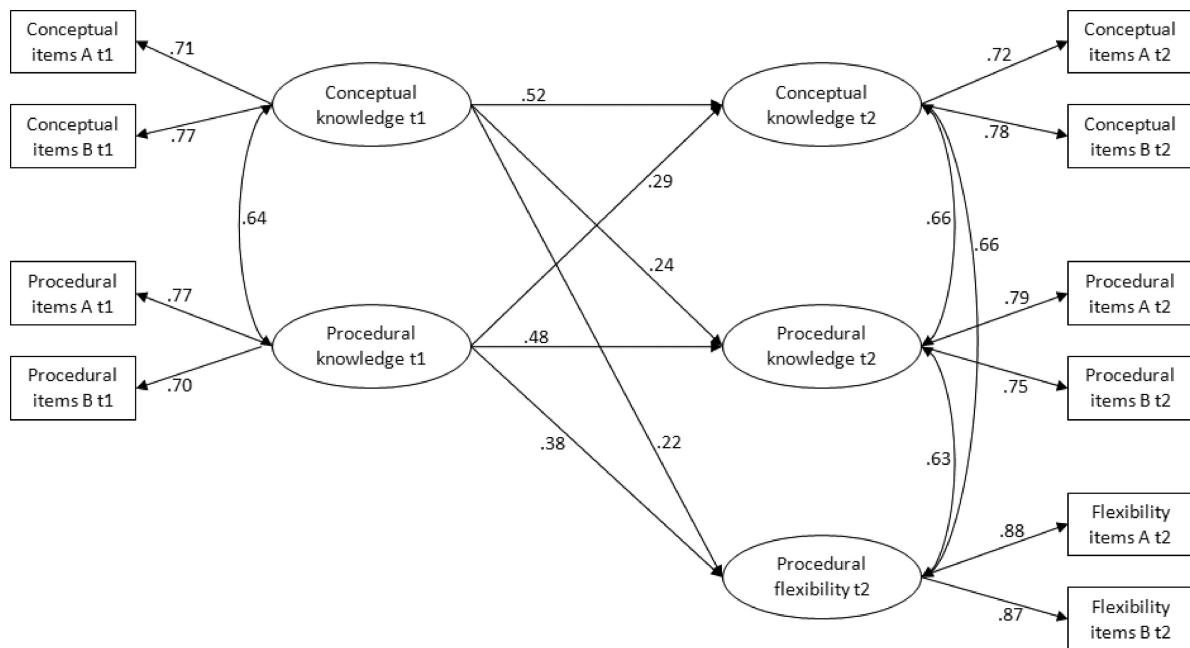


Figure 2. Factor loadings, factor intercorrelations, and regression paths of the best fitting structural equation model (Model 2k) of the relations between conceptual knowledge, procedural knowledge, and procedural flexibility in Study 2. A = Item Group A (sum of items with odd numbers); B = Item Group B (sum of items with even numbers); t1 = Time 1; t2 = Time 2. All estimated coefficients are significant with $p < .05$.

between the two kinds of knowledge but also to explain why their interrelations are the way they are in terms of the human information processing architecture.

Relations to Procedural Flexibility

We also investigated how conceptual and procedural knowledge contribute to procedural flexibility, that is, knowledge of how to solve problems flexibly and efficiently. Conceptual and procedural knowledge were both important for supporting procedural flexibility. In our two samples, conceptual and procedural knowledge at Time 1 both contributed individually to procedural flexibility at Time 2. Given the importance of procedural flexibility for mathematical proficiency, it is important to recognize the benefits of both types of prior knowledge. Flexibility does not seem to come from conceptual or procedural knowledge alone—children may gain flexibility by using both types of knowledge (Baroody & Dowker, 2003).

The literature suggests several mechanisms by which conceptual or procedural knowledge can strengthen procedural flexibility. Conceptual knowledge can increase flexibility by guiding attention to important problem features and aid choice of the most appropriate procedures, adapting the choice to the specific problem and context at hand. At the same time, procedural flexibility rests on knowing more than one procedure. A sufficiently large repertoire of problem-solving procedures is necessary for flexibly adapting the behavior to the problem at hand (Siegler & Lemaire, 1997). Background knowledge about the effectiveness, strengths, and weaknesses of each procedure helps people to flexibly choose the most adequate procedure in a situation (Star, 2005). Our results suggest that these mechanisms are not mutually exclusive. Concepts and procedures can simultaneously strengthen procedural flexibility by a multitude of mechanisms. Instruction addressing conceptual or procedural knowledge may, thus, also be expected to have positive indirect effects on procedural flexibility.

Table 5
Fit Indices of the Multigroup Models of Longitudinal Relations Estimated With the Data From Study 1 and Study 2

Model no.	Model description	χ^2	df	p	CFI	RMSEA	SRMR	AIC	BIC
3a	Basic model	360	328	.106	.985	.027	.072	58932	59407
3b	Identical loadings in both groups assumed	378	341	.083	.982	.028	.074	58923	59352
3c	Identical loadings and identical predictive relations in both groups assumed	383	351	.118	.985	.026	.075	58908	59300

Note. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.

Measuring Conceptual and Procedural Knowledge

When new intelligence or personality tests are developed, researchers use a rigorous set of psychometric methods to evaluate the reliability and validity of the new test. In contrast, research on conceptual and procedural knowledge so far has not used this approach (Schneider & Stern, 2010). In the current study, a latent variable approach helped us to investigate several aspects of the divergent validity and the reliability of our measures. All of our measures had acceptable reliabilities, since their loadings on the latent factors standing for the different kinds of knowledge are all higher than .5 and are statistically significant (cf. Bollen, 2002).

Conceptual knowledge, procedural knowledge, and procedural flexibility could be assessed partly independently of each other. In most cases, the relations between our measures were fit better or at least equally well by a factor for each construct than by a single latent factor underlying pairs of measures. In addition, the longitudinal models with a separate latent factor for each of our constructs had excellent fits to the data.

Latent variable analyses also allowed us to test whether our measures functioned the same at different measurement points and in samples differing in prior knowledge. The pattern of factor loadings was the same for both measurement points and both studies. The only aspects of the factor structure that changed over time and samples were the intercorrelations between conceptual and procedural knowledge that lay at $r = .3$ for Time 1 in Study 1, where the participants' knowledge was lowest, and were greater .6 in all other cases. Since the assessment tasks and test instructions were always the same, the changing correlations likely indicate changes in the assessed knowledge structures. Some authors have suggested that a person with low expertise in a domain has fragmented knowledge and does not see how different pieces of knowledge in a domain, for examples, concepts and procedures, relate to each other. A higher expertise in a domain enables learners to integrate more and more pieces of knowledge into a coherent knowledge structure (Baroody & Dowker, 2003; Linn, 2006; Schneider & Stern, 2009). Our findings are in line with this interpretation. However, high correlations between latent factors assessing knowledge structures only indicate that these knowledge structures frequently appear together. This does not necessarily imply a high level of cognitive integration of the two knowledge structures. Further research with more explicit measures of knowledge integration or fragmentation is needed.

Overall, the structural equation models of the final models (see Figures 1 and 2) had excellent fits to the data. This is all the more remarkable as our data sets had quite complex structures. The data came from different samples, interventional groups, classrooms, dyads, and measurement points. We carefully choose the adequate modeling strategies to address each of these points. The excellent model fits indicate that this strategy was successful and that our models adequately reflect the relations between the assessed constructs in our samples.

Future studies will have to test whether our results generalize over different measures. As explained earlier, the choice of measures can sometimes influence the quality of the obtained results (Schneider & Stern, 2010). For logistical reasons, we had only one type of measure for each kind of knowledge and could not control for these effects here. Future studies on knowledge acquisition in dyads should be conducted with larger sample sizes so that mul-

tilevel modeling can be used to simultaneously investigate effects on the level of individual persons and effects on the dyad level.

In summary, we showed that latent variable modeling can be used to improve the reliability of measures of conceptual knowledge, procedural knowledge, and procedural flexibility. Our analyses yielded clear evidence for bidirectional and even symmetric predictive relations between conceptual and procedural knowledge. These relations were stable over two large samples differing in their prior knowledge. Conceptual and procedural knowledge at Time 1 independently predicted students' procedural flexibility at Time 2. These findings add to a growing body of evidence that conceptual and procedural knowledge develop in an iterative fashion and both support the development of procedural flexibility.

References

- Ackerman, P. L., & Caciolo, A. T. (2000). Cognitive, perceptual speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: Applied*, *6*, 259–290. doi:10.1037/1076-898X.6.4.259
- Alibali, M. W., Knuth, E. J., Hattikudur, S., McNeil, N. M., & Stephens, A. C. (2007). A longitudinal examination of middle school students' understanding of the equal sign and equivalent equations. *Mathematical Thinking and Learning*, *9*, 221–247. doi:10.1080/10986060701360902
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*, 411–423. doi:10.1037/0033-2909.103.3.411
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglas, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*, 1036–1060. doi:10.1037/0033-295X.111.4.1036
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. doi:10.1037/0022-3514.51.6.1173
- Baroody, A. J., & Dowker, A. (Eds.). (2003). *The development of arithmetic concepts and skills: Constructing adaptive expertise*. Mahwah, NJ: Erlbaum.
- Baroody, A. J., Feil, Y., & Johnson, A. R. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education*, *38*, 115–131.
- Bisanz, J., & LeFevre, J. A. (1992). Understanding elementary mathematics. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 113–136). Mahwah, NJ: Elsevier. doi:10.1016/S0166-4115(08)60885-7
- Blöte, A. W., van der Burg, E., & Klein, A. S. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instruction effects. *Journal of Educational Psychology*, *93*, 627–638. doi:10.1037/0022-0663.93.3.627
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*, 605–634. doi:10.1146/annurev.psych.53.100901.135239
- Burkholder, G. J., & Harlow, L. L. (2003). An illustration of a longitudinal cross-lagged panel design for larger structural equation models. *Structural Equation Modeling*, *10*, 465–486. doi:10.1207/S15328007SEM1003_8
- 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
- Canobi, K. H., & Bethune, N. E. (2008). Number words in young children's conceptual and procedural knowledge of addition, subtraction and inversion. *Cognition*, *108*, 675–686. doi:10.1016/j.cognition.2008.05.011
- Canobi, K. H., Reeve, R. A., & Pattison, P. E. (2003). Patterns of knowl-

- edge in children's addition. *Developmental Psychology, 39*, 521–534. doi:10.1037/0012-1649.39.3.521
- Carpenter, T. P., Franke, M. L., Jacobs, V. R., Fennema, E., & Empson, S. B. (1998). A longitudinal study of invention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education, 29*, 3–20. doi:10.2307/749715
- Dowker, A. (1992). Computational estimation strategies of professional mathematicians. *Journal for Research in Mathematics Education, 23*, 45–55. doi:10.2307/749163
- Eid, M., & Diener, E. (Eds.). (2006). *Handbook of multimethod measurement in psychology*. Washington, DC: American Psychological Association. doi:10.1037/11383-000
- Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls, J. (1989). Young children's understanding of counting and cardinality. *Child Development, 60*, 1158–1171. doi:10.2307/1130790
- Gelman, R., & Williams, E. M. (1998). Enabling constraints for cognitive development and learning: Domain specificity and epigenesis. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (5th ed., pp. 575–630). New York, NY: Wiley.
- Haapasalo, L., & Kadjevich, D. (2000). Two types of mathematical knowledge and their relation. *Journal für Mathematik-Didaktik, 21*, 139–157.
- Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Erlbaum.
- Hecht, S. A., Close, L., & Santisi, M. (2003). Sources of individual differences in fraction skills. *Journal of Experimental Child Psychology, 86*, 277–302. doi:10.1016/j.jecp.2003.08.003
- Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and Instruction, 14*, 251–283. doi:10.1207/s1532690xci1403_1
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. doi:10.1080/1070551990540118
- Huijzinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*, 2017–2036. doi:10.1016/j.neuropsychologia.2006.01.010
- Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.
- Kenny, D. A. (in press). Dyadic analyses of family data [Commentary]. *Journal of Pediatric Psychology, 36*, 630–633. doi:10.1093/jpepsy/jsq124
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford.
- Kilpatrick, J., Swafford, J. O., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academies Press.
- Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S. N., & Phillips, E. D. (2009). *Connected mathematics 2*. Upper Saddle River, NJ: Pearson Education.
- LeFevre, J.-A., Smith-Chant, B. L., Fast, L., Skwarchuk, S.-L., Sargla, E., Arnup, J. S., ... Kamawar, D. (2006). What counts as knowing? The development of conceptual and procedural knowledge of counting from kindergarten through Grade 2. *Journal of Experimental Child Psychology, 93*, 285–303. doi:10.1016/j.jecp.2005.11.002
- Linn, M. C. (2006). The knowledge integration perspective on learning and instruction. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 243–264). New York, NY: Cambridge University Press.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*, 151–173. doi:10.1207/S15328007SEM0902_1
- Meuleman, B., & Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Survey Research Methods, 3*, 45–58.
- Muthén, L. K., & Muthén, B. O. (1998–2007). *MPlus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.
- National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics*. Reston, VA: Author.
- Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling, 8*, 353–377. doi:10.1207/S15328007SEM0803_2
- Newsom, J. T. (2002). A multilevel structural equation model for dyadic data. *Structural Equation Modeling, 9*, 431–447. doi:10.1207/S15328007SEM0903_7
- Resnick, L. B. (1982). Syntax and semantics in learning to subtract. In T. P. Carpenter, J. M. Moser, & T. A. Romburg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 136–155). Hillsdale, NJ: Erlbaum.
- Resnick, L. B., & Omanson, S. F. (1987). Learning to understand arithmetic. In R. Glaser (Ed.), *Advances in instructional psychology: Vol. 3. Reading/arithmetic/verbal comprehension/classroom behavior* (pp. 41–95). Hillsdale, NJ: Erlbaum.
- Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*, 175–189. doi:10.1037/0022-0663.91.1.175
- Rittle-Johnson, B., & Koedinger, K. R. (2009). Iterating between lessons concepts and procedures can improve mathematics knowledge. *British Journal of Educational Psychology, 79*, 483–500. doi:10.1348/000709908X398106
- Rittle-Johnson, B., & Siegler, R. S. (1998). The relation between conceptual and procedural knowledge in learning mathematics: A review. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 75–110). Hove, United Kingdom: Psychology Press.
- Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*, 346–362. doi:10.1037/0022-0663.93.2.346
- Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*, 561–574. doi:10.1037/0022-0663.99.3.561
- Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology, 101*, 836–852. doi:10.1037/a0016026
- Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: Their interrelations in Grades 5 and 6. *Journal of Educational Psychology, 101*, 359–372. doi:10.1037/a0013840
- Schneider, M., & Stern, E. (2009). The inverse relation of addition and subtraction: A knowledge integration perspective. *Mathematical Thinking and Learning, 11*, 92–101. doi:10.1080/10986060802584012
- Schneider, M., & Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: A multimethod approach. *Developmental Psychology, 46*, 178–192. doi:10.1037/a0016701
- Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM via the choice/no choice method. *Journal of Experimental Psychology: General, 126*, 71–92. doi:10.1037/0096-3445.126.1.71
- Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General, 127*, 377–397. doi:10.1037/0096-3445.127.4.377
- Smith, J. P., diSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Jour-*

- nal of the Learning Sciences*, 3, 115–163. doi:10.1207/s15327809jls0302_1
- Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, 36, 404–411.
- Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, 18, 565–579. doi: 10.1016/j.learninstruc.2007.09.018
- Star, J. R., & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology*, 31, 280–300. doi: 10.1016/j.cedpsych.2005.08.001
- Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skill to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science*, 25, 203–244. doi:10.1207/s15516709cog2502_2
- Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2005). Simple addition strategies in a first-grade class with multiple strategy instruction. *Cognition and Instruction*, 23, 1–21. doi:10.1207/s1532690xci2301_1
- Ullman, J. B. (2007). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (5th ed., pp. 676–780). Boston, MA: Pearson.
- U.S. Department of Education. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: Author.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–69. doi:10.1177/109442810031002
- VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, 47, 513–539. doi:10.1146/annurev.psych.47.1.513
- Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education*, 24, 335–359. doi:10.1007/BF03174765

Appendix

Sample Items for Assessing Conceptual, Procedural, and Flexibility Knowledge

Problem type	Sample items	Scoring
Procedural knowledge		
Familiar ($n = 3$)	$\frac{1}{2}(x + 1) = 10$ $3(h + 2) + 4(h + 2) = 35$ $3(m - 2)/5 = 33/5$ $3(2x + 3x - 4) + 5(2x + 3x - 4) = 48$	1 pt for each correct answer
Novel ($n = 6$)		1 pt for each correct answer
Procedural flexibility knowledge		
Generate multiple methods ($n = 6$, from 4 question stems)	Solve this equation in two different ways: $3(y + 1) = 4(y + 1) + 2(y + 1)$ Which of your ways do you think is easiest and fastest?	1 pt for two correct unique solutions
Recognize multiple methods ($n = 8$, from 2 question stems)	For the equation $2(x + 1) + 4 = 12$, identify all possible steps that could be done next. (4 choices) $5(x + 3) + 6 = 5(x + 3) + 2x$ $6 = 2x$ What step did the student use to get from the first line to the second line?	1 pt for choosing solution with fewest steps 1 pt for each correct choice
Evaluate nonconventional methods ($n = 6$, from 2 question stems)	Do you think that this is a good way to start this problem? (a) a very good way, (b) OK to do, but not a very good way, (c) not OK to do Explain your reasoning.	2 pt for Choice a, 1 pt for Choice b, 0 pts for Choice c
Conceptual knowledge ($n = 13$)	Which of the following is a like term to (could be combined with) $7(j + 4)$? (a) $7(j + 10)$, (b) $7(p + 4)$, (c) j , (d) $2(j + 4)$, (e) a and d Here are two equations: $98 = 21x$ $98 + 2(x + 1) = 21x + 2(x + 1)$ Look at this pair of equations. Without solving the equations, decide if these equations are equivalent (have the same answer). Explain your reasoning.	2 pt if accurately evaluates efficiency or justifies why OK to do; 1 pt if simply states that step is OK to do. 1 pt for choosing d 1 pt for selecting “yes (they have the same answer)” 1 pt for mentioning equivalence of equations

Received June 12, 2010

Revision received April 20, 2011

Accepted April 27, 2011 ■