Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Conceptual and Procedural Knowledge of a Mathematics Problem: Their Measurement and Their Causal Interrelations

Permalink

https://escholarship.org/uc/item/63d3w5p7

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

ISSN

1069-7977

Authors

Schoner, Gregor
Wilimzig, Caludia

Publication Date

2005

Conceptual and Procedural Knowledge of a Mathematics Problem: Their Measurement and Their Causal Interrelations

Michael Schneider (mschneider@mpib-berlin.mpg.de)
Elsbeth Stern (stern@mpib-berlin.mpg.de)
Max Planck Institute for Human Development
Lentzeallee 94, 14195 Berlin, Germany

Abstract

Some learning theories see conceptual knowledge as a source of children's procedural knowledge. Others assume the opposite to be true or posit bi-directional causal relations. Empirical tests of these assumptions are hampered by the lack of knowledge on how to obtain valid measures of these two constructs. We assessed four different measures of both constructs before and after an intervention and modelled the two kinds of knowledge as underlying latent factors in SEM. This enabled us to test the quality of our measures as well as the adequacy of the above-mentioned assumptions. Conceptual knowledge was a source of children's procedural knowledge, but not vice versa. In contrast to procedural knowledge, conceptual knowledge could be assessed with high internal consistency.

Keywords: learning theories; conceptual knowledge; procedural knowledge; declarative knowledge; SEM.

Introduction

Several theories of learning and cognition posit that our behaviour is shaped by at least two different kinds of knowledge: one providing an abstract understanding of the principles and relations between pieces of knowledge in a certain domain, and another one enabling us to quickly and efficiently solve problems. In recent empirical research on mathematics learning the former is frequently named *conceptual knowledge*, while the latter is labelled *procedural knowledge* (e.g., Baroody, 2003).

Cognitive models of the relations between these different kinds of knowledge can facilitate our general understanding of the human mind, but may also be helpful for designing the contexts in which knowledge is to be conveyed.

However, different theories make different predictions as to the interrelations between conceptual and procedural knowledge. One major difference is described by Rittle-Johnson, Siegler, and Alibali (2001) who distinguish between *concepts-first* and *procedures-first theories*. According to concepts-first theories, children will initially acquire conceptual knowledge, for example by listening to verbal explanations, and will then, by practice, derive procedural knowledge from it. Procedures-first theories, on the contrary, posit that children initially acquire procedural knowledge in a specific domain, for example by trial-and-error learning, and then gradually abstract conceptual knowledge from it by reflection.

Based on the fact that there is empirical evidence for both kinds of theories, Rittle-Johnson et al. propose a third possibility, i.e. their *Iterative Model*: there may be bi-directional causal links between conceptual and procedural knowledge. Increase in one kind of knowledge will, then, prompt increase in the other one as well.

The Measurement Problem

Despite these controversies and the importance of the field, there are comparatively few empirical studies addressing the relationships between the two kinds of knowledge. As a review by Rittle-Johnson and Siegler (1998) shows, these studies yielded partly inconsistent results and had serious methodological limitations.

While some of these limitations, such as the use of correlational designs and non-gradual, dichotomous measures, can easily be overcome, one problem was of a more general nature: it is as yet unclear how conceptual and procedural knowledge can be measured independently of each other and with a sufficient degree of validity. Since it would seem that some cognitive procedure is always needed to derive actions from (static) conceptual knowledge representations, how can we find out whether any given action, for example a subject's response to a test item, is rooted in conceptual or in procedural knowledge or, to different degrees, in both?

While this question is largely ignored in the literature, Rittle-Johnson et al. (2001) gave a well-founded answer in the context of a study they conducted to test the Iterative Model. They measured children's conceptual and procedural knowledge before and after an intervention that was designed to increase both kinds of knowledge, and showed that children's initial conceptual knowledge predicted gains in procedural knowledge, and gains in procedural knowledge predicted improvements in conceptual knowledge. They distinguished between assessments of conceptual and procedural knowledge on the basis of the novelty of the tasks at posttest, the assumption being that children will apply acquired procedural knowledge when solving the routine tasks they already know from the intervention, but will resort to conceptual understanding when challenged to produce new solutions to hitherto unknown transfer tasks.

But even this elaborated answer fails to account for certain parts of the problem because we cannot tell with any certainty that children do not use conceptual knowledge to generate answers to routine tasks. Furthermore, if there are routine tasks only in the posttest, conceptual and procedural knowledge can only be independently assessed for the posttest, and no evaluation of the increase of both kinds during an intervention is possible.

To overcome these problems and provide further evidence for the Iterative Model in the current study, we used a design that is very similar to the one reported in Rittle-Johnson et al. (2001), but we assessed four different measures of conceptual knowledge and four different measures of procedural knowledge before and after the intervention. We then tried to model children's conceptual and procedural knowledge at both points in time, respectively, as two latent factors that to various degrees influence children's behaviour in the eight measures.

The success of this method depends on an adequate choice of measures. We will therefore give a short overview of the different characteristics of conceptual and procedural knowledge and then explain our choice of measures on these grounds.

Characteristics of conceptual and procedural knowledge

In the context of the above-mentioned studies, conceptual knowledge is seen as the knowledge of the core concepts and principles and their interrelations in a certain domain. Accordingly, it is assumed to be stored in some form of relational representation, like schemas, semantic networks or hierarchies (e.g. Byrnes & Wasik, 1991). Because of its abstract nature and the fact that it can be consciously accessed, it can be largely verbalized and flexibly transformed through processes of inference and reflection. It is, therefore, not bound up with specific problems but can in principle be generalized for a variety of problem types in a domain (e.g. Baroody, 2003).

Procedural knowledge, in contrast, is seen as the knowledge of operators and the conditions under which these can be used to reach certain goals (e.g. Byrnes & Wasik, 1991). Further, it allows people to solve problems quickly and efficiently because it is to some degree automated. Automatization is accomplished through practice and allows for a quick activation and execution of procedural knowledge, since its application, as compared to the application of conceptual knowledge, involves minimal conscious attention and few cognitive resources (see Johnson, 2003, for an overview). Its automated nature, however, implies that procedural knowledge is not or only partly open to conscious inspection and can, thus, be hardly verbalized or transformed by higher mental processes. As a consequence, it is tied to specific problem types (e.g. Baroody, 2003).

The distinction between conceptual and procedural knowledge, as it is understood here, is similar to the well-known distinction between declarative and procedural knowledge. We see conceptual knowledge as one kind of declarative knowledge among others, e.g. knowledge of examples and memories of specific situations.

The present study

As described above, it is still unclear how conceptual and procedural knowledge can be assessed with sufficient degrees of validity and independently of each other. Previous studies on conceptual and procedural knowledge can thus be criticized for trying to empirically examine the relations between conceptual and procedural knowledge before it is known how the constructs can be measured.

To provide more valid evidence on the relations between conceptual and procedural knowledge in the present study we assessed children's conceptual and procedural knowledge each by four different measures before and after an intervention. A three-step strategy was then used for data analysis.

First, separate confirmatory factor analyses (CFA) were done for the pretest measures of conceptual knowledge, the pretest measures of procedural knowledge, the posttest measures of conceptual knowledge and the posttest measures of procedural knowledge. For each quadruple of measures, the convergent validity, i.e. whether all four assessments really measure the same construct and thus load high on an underlying factor, was evaluated. We expect this to be the case, since all measures are closely related to the different characteristics of conceptual or procedural knowledge, as will be discussed below, and have been used before in a similar form to measure one of the two kinds of knowledge, as reported in published studies.

Second, if these factors were found, the divergent validity of the measures could be separately evaluated for the pretest and the posttest data, i.e. it could be ascertained whether the measures of conceptual and procedural knowledge really assess two significantly different constructs rather than only one. This would imply that the values of the two knowledge factors will, in part, vary independently of each other. Only then can they be used to investigate the relations between conceptual and procedural knowledge.

Finally, those factors which, during the previous steps, had proved to be adequate were to be used to test different hypotheses concerning the causal relations between conceptual and procedural knowledgeby means of structural equation modelling (SEM).

As far as we know, conceptual and procedural knowledge have never before been estimated by factor values in a study. This proceeding should, however, lead to more valid results than the dominant strategy of using sum scores because the measure-specific error variances are factored out and do not distort the estimates.

Since there is empirical evidence supporting the concepts-first view as well as evidence supporting the procedures first-view, we expect the results of the SEM analyses to confirm the Iterative Model of Rittle-Johnson et al. (2001).

Method

Participants and procedure

230 fifth-grade and sixth-grade volunteers from 10 primary schools in Berlin, Germany, participated (median age: 11

years). They were tested in small groups on two consecutive days at our research institute, each student working individually, and without seeing the others, at a computer in a quiet environment. On the first day, the students did the pretest and received the first part of the intervention. On the second day, they received the second part of the intervention and did the posttest.

Since in Berlin the general mathematical properties of decimal fractions are usually not taught before the end of sixth grade, our participants had no recent school instruction on the topic of our study. They could, however, have some relevant prior conceptual and procedural knowledge due to the usual first-grade to fourth-grade lessons on diagrams or on how to compute distances and prices.

Intervention

The intervention was the same for all children and is adopted from Rittle-Johnson et al. (2001). In each of our 160 tasks, a decimal fraction was presented to the children on the computer screen together with a number line. The children had to click on the position on the number line that corresponded to the value of the decimal fraction. Afterwards they received feedback as to the correct position. In 16 tasks, the children additionally had to give a written explanation for the correct answer after the feedback.

The intervention was designed to activate and increase children's conceptual and procedural knowledge. We therefore expected increases from pretest to posttest to occur in all of our measures.

Assessments

As a starting point for selecting our measures, we used the idea, as discussed above, that assessments of conceptual and procedural knowledge can be distinguished on the basis of the novelty of the task in the posttest. Children are more likely to use their procedural knowledge when solving routine tasks, but tend to rely on their conceptual knowledge for solving hitherto unknown transfer tasks.

A task is considered a routine task if it requires participants to locate a given decimal fraction on a number line, because this is the problem type used in the intervention. A task is considered a transfer task if it involves knowledge about decimal fractions, but does not require participants to actually place a decimal fraction on a number line.

Accordingly, conceptual knowledge was measured by having the children solve four different types of transfer tasks and by scoring the correctness of their answers. Procedural knowledge was assessed by having the children solve four types of routine problems and by measuring four different aspects of their behaviour. The order of the tasks was block-randomised for each participant.

We will now introduce our four measures of conceptual knowledge. Of course we cannot yet tell if they really measure a construct like conceptual knowledge because this still has to be examined empirically. We will, therefore, use the term "measures of conceptual knowledge" as a short form for the more correct expression "measures that are hypothesized to measure conceptual knowledge". The same applies to similar expressions concerning the other measures or factors throughout this paper.

Evaluation of procedures (*ce*) Different verbal descriptions of problem solving procedures for the routine problems were successively presented to the children. The children had to evaluate the quality of each procedure as *rather good* or *rather bad*. This evaluation of procedures has been used to access conceptual knowledge by Canobi, Reeve, and Pattison (1998) and many others. It requires children to rely on their conceptual knowledge base for deriving some kind of measure for the quality of procedures.

Translation into diagrams (*ct*) In each task a decimal fraction was presented on the screen together with four pie charts. Each pie chart consisted of a white and a grey area. The children had to click on the pie chart in which the ratio between the grey area and the whole area corresponded to the decimal fraction.

This measure and the following (*cs*) have been used by, for example, Byrnes and Wasik (1991) to assess conceptual knowledge of fractions. Task solution requires the ability to translate knowledge of the relations between different numbers into knowledge of the relations between geometrical areas.

Size comparisons (*cs*) Pairs of decimal fractions were presented on the screen, and for each pair, the children had to click on the number with the higher value. This assessment reflects children's understanding of the ordinal relations between decimal fractions.

Written explanations (*cw*) The children were asked four different questions about the general properties of decimal fractions, such as: "In which real-life situations whole numbers would be used rather than decimal fractions?", and had to write down their answers. The correctness of each answer was coded by two independent raters.

We used the following four measures of procedural knowledge.

Problem solving correctness (*pc*) The children had to indicate the position of a given decimal fraction on a number line by using the mouse to move a lever to this position. They were instructed to maximize the correctness of their answers and not to think about time. The percentage of correct answers was scored by the computer.

As described above, problem solving correctness, especially on routine problems, is a frequent measure of procedural knowledge.

Problem solving duration (*pd*) The participants had to locate the positions of given decimal fractions on a number line by freely clicking at it as fast as they could. The time they needed was measured in milliseconds, log-transformed, and multiplied with 10 by the computer. This measure has been used by Canobi et al. (1998) and in numerous studies on the acquisition of procedural knowledge.

Asymmetry of access (*pa*) For the measurement of *pa* the children had to solve two different types of tasks (*A* and *B*) in an ABBA design. Their problem solving times were measured. The value of *pa* was then computed as the difference between the mean problem solving time pa_B for *Type B* tasks and the mean problem solving time pa_A for *Type A* tasks.

In *Type A* tasks, the children were presented with a decimal fraction together with several number lines. On each number line, a position was marked by an arrow. The children had to click on the number line that showed the correct position of the decimal fraction. *Type B* tasks worked the other way round, i.e. children had to choose a decimal fraction that corresponded to a position indicated on a number line.

This measure is rarely used but is subject to empirical validation in a study by Anderson, Fincham and Douglass (1997). It is based on the assumption that changes in children's conceptual knowledge should affect their problem solving times for *none* or for *both* kinds of tasks because this knowledge is undirected. An increase in procedural knowledge for one type of task, in contrast, should have little effect on the other task type because procedural knowledge is goal-directed and, thus, asymmetric. So a growing asymmetry between the solution times for both types of tasks can be attributed to an increase in procedural knowledge.

Dual-task costs (*pu*) For the measurement of *pu* the children had to solve two different types of tasks, again in an ABBA design. In the single-task condition, the children in each task saw a decimal fraction on the screen and had to click on one of four arrows that indicated its potential positions on a number line. Their problem solving time, pu_{single}, was measured. In the dual-task condition, children again received the same tasks, but had to do a second task simultaneously, i.e. counting names they heard on a headphone. Their problem solving time, pu_{dual}, was measured. *Pu* was computed as pu_{dual} minus pu_{single}.

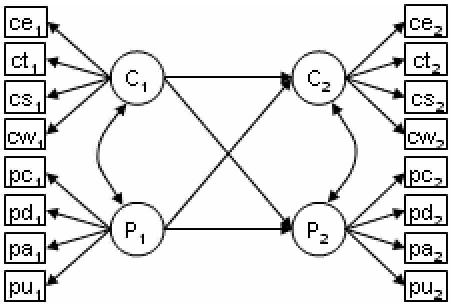
Dual-task costs have been shown to be negatively related to children's procedural knowledge. A possible explanation for this could be that individuals with higher procedural knowledge need less cognitive resources for solving a task (see Johnson, 2003).

Analysis

The program MPlus was used to test our hypotheses. Because of their robustness to non-normal distributions, we chose the estimator MLM and the Satorra-Bentler scaled χ^2-statistic (see Satorra & Bentler, 1999) for the analyses.

The basic SEM model used for analysis is shown in Figure 1. Subscript 1 indicates a pretest variable, subscript 2 a posttest variable. Squares represent observed, i.e. measured, variables; circles stand for hypothetical latent factors.

To evaluate the convergent validity of the pretest measures of conceptual knowledge, we specified a *Model* C_1+ describing the four measures as loading on one underlying factor C_1, and an alternative *Model* C_1- describing the four measures as varying independently. The same was done for the posttest measures of conceptual knowledge (*Model* C_2+ and *Model* C_2-), the pretest measures of procedural knowledge (*Model* P_1+ and *Model* P_1-), and the posttest measures of procedural knowledge (*Model* P_2+ and *Model* P_2-). The model fit indices of these models were compared. The lower the χ^2 value and the higher the probability *p* of finding the obtained data under the assumption that the specified model holds for the population, the better is the relative fit of a model. A *Comparative Fit Index* (CFA) above .95, a *Weighted Root Mean Square of Residuals* (WRMR) below 1.0, and a *Root Mean Square Error of Approximation* (RMSEA) below .06 indicate an acceptable absolute model fit.



C: conceptual knowledge; P: procedural knowledge; ce: evaluation of procedures; ct: translation into diagrams; cs: size comparisons; cw: written explanations; pc: problem solving correctness; pd: problem solving duration; pa: asymmetry of access; pu: dual-task costs.

Figure 1: Basic SEM model used for the analyses.

The internal consistency of each factor was further estimated by Cronbach's α, which can be computed if the factor is taken as a scale with its indicators, i.e. the measured variables, as the items of the scale.

To test our hypotheses concerning the divergent validities of the pretest measures, we specified a *Model* C_1P_1 that describes the four measures of conceptual knowledge as loading on one latent factor and the four measures of procedural knowledge as loading on another one. An alternative model, *Model* K_1, was specified that describes all eight measures as loading on only one latent factor. Similar models were specified for the posttest data (*Model* C_2P_2 and *Model* K_2).

To test the assumption of bi-directional causal relations between the kinds of knowledge, we had to test whether the coefficients of the paths leading from *C1* and *P1* to *C2* and from *C1* and *P1* to *P2* in Figure 1 significantly differ from zero.

Eight of the initial 230 participants were excluded from the analyses because they obviously did not comply with the instructions or had severe language problems due to their recent migration to Germany. There were no missing data.

Results

Descriptives and reliabilities

The means and standard deviations for the measures and *Cohen's d* as the effect size of the pretest-posttest changes

Table 1: Descriptives for the measures and their change.

Measure	*mean*		*SD*		*d*
	pre	post	pre	post	
ce	59	68	21	22	0.4
ct	51	70	17	19	1.1
cs	75	84	16	13	0.6
cw	17	22	21	24	0.2
pc	57	88	22	16	1.6
pd	80	78	4	3	0.6
pa	-5	-4	8	5	0.2
pu	9	2	10	2	1.0

ce: evaluation of procedures [%]; ct: translation into diagrams [%]; cs: size comparisons [%]; cw: written explanations [%]; pc: problem solving correctness [%]; pd: problem solving duration [10×ln(ms)]; pa: asymmetry of access [Δs]; pu: dual-task costs [Δs].

Table 2: Indicators of reliability.

Measure	*n* of tasks	Cronbach's α		*Stability*
		pre	post	
ce	8	.44	.53	.40
ct	20	.67	.79	.47
cs	20	.74	.72	.62
cw	4	.54	.60	.75
pc	20	.82	.76	.41
pd	20	.93	.94	.48
pa_A	20	.78	.69	.20
pa_B	20	.73	.74	.17
pa	-	-	-	.19
pu_{single}	40	.92	.92	.58
pu_{dual}	40	.93	.94	.51
pu	-	-	-	.44

ce: evaluation of procedures; ct: translation into diagrams; cs: size comparisons; cw: written explanations; pc: problem solving correctness; pd: problem solving duration; pa_A: asymmetry of access (Type A); pa_B: asymmetry of access (Type B); pa: asymmetry of access; pu_{single}: dual-task costs (single condition), pu_{dual}: dual-task costs (dual condition); pu: dual-task costs.

of the means are shown in Table 1. As *t*-tests revealed, all pretest-posttest differences were significant, with *p*s < 0.05.

Table 2 gives an overview of the number of tasks and Cronbach's α for each sum score and the stability of the measures, that is the correlations between pretest and posttest scores.

Convergent validities

Table 3 gives an overview of the model fit indices for the eight different models. They suggest that the assumption of the latent factors *C1*, *C2*, and *P2* is well justified by the data, since the χ^2 values are lower for the one-factor models than for the independence models. Moreover, these one-factor models show very good CFI values and good WRMR indices, although the RMSEA of *Model* C_2+ is suboptimal.

Contrary to expectations, the estimation of Model P_1+ did not converge. As the good fit of Model P_1- suggests, this might be due to the fact that the four measured variables are mutually independent. In this case, no factor values can be estimated and no convergence is possible. This is further indicated by the Cronbach's α values: for the pretest measures of conceptual knowledge, α is .68, for the pretest measures of procedural knowledge, it is -.13, and for the posttest measures it is .69 and .12, respectively. The negative value indicates that the four measures of procedural knowledge in the pretest, in contrast to the other measures, did not with any internal consistency assess a single construct. As a consequence, the divergent validity could only be tested for the posttest measures.

Table 3: Model fit indices (convergent validities).

	χ^2	*df*	*p*	*CFI*	*WRMR*	*RMSEA*
Model C_1+	1	2	0.73	1.00	0.18	0.00
Model C_1-	160	6	0.00	- [2]	5.10	0.34
Model P_1+	- [1]	- [1]	- [1]	- [1]	- [1]	- [1]
Model P_1-	7	6	0.34	- [2]	0.99	0.02
Model C_2+	8	2	0.02	0.97	0.92	0.12
Model C_2-	194	6	0.00	- [2]	5.45	0.38
Model P_2+	3	2	0.21	0.96	0.49	0.05
Model P_2-	32	6	0.00	- [2]	2.25	0.14

[1] No estimates due to lack of convergence.

[2] CFI is not defined for this model.

Divergent validities

The model fit indices of the one-factor model (*Model* K_2) and the two-factor model (*Model* C_2P_2) are shown in Table 4. Since the indices of the two-factor model are only slightly better than those of the one-factor model, we used the scaled difference chi-square test (Satorra & Bentler, 1999), which is adequate for Satorra-Bentler scaled test statistics, for testing whether the difference in the chi-square values is significant.

Table 5 shows the coefficients that were included in the test. The resulting scaling correction factor is $c = 0.871$. The corrected chi-square difference value is statistically significant, $\chi^2_{diff}(1, N = 222) = 5.211 / 0.871 = 5.983$, $p = .01$. Thus the two-factor solution fits the data significantly better than the one-factor solution. The two

Table 4: Model fit indices (divergent validity).

	χ^2	*df*	*p*	CFI	WRMR	RMSEA
Model K_2	47	20	0.00	.93	1.19	0.08
Model C_2P_2	41	19	0.00	.94	1.17	0.07

Table 5: Parameters for the χ^2 difference test.

	unscaled χ^2	scaled χ^2	*df*	scaling factor
Model K_2	46.657	46.478	20	1.004
Model C_2P_2	41.446	41.001	19	1.011
difference	5.211		1	

factors are correlated with $r = .84$. The suboptimal descriptive fits of the two-factor model are probably due to the low internal consistency of the procedural factor P_2.

Knowledge influences over time

The standardized coefficient for the regression of the posttest conceptual latent factor (*C2*) on the pretest conceptual latent factor (*C1*) is .92; the one for the regression of the posttest procedural latent factor (*P2*) on the pretest conceptual latent factor is .51. Both are statistically significant, with *p*s < .05.

Discussion

The present study was conducted to examine two interrelated questions: (1) Can conceptual and procedural knowledge be measured with sufficient degrees of validity and independently of each other, and if so, (2) are the causal relations between the two kinds bi-directional, as suggested by the Iterative Model, or uni-directional?

Concerning the first question, we can say that conceptual knowledge was successfully measured by the assessments we used. The convergent validity of our measures of conceptual knowledge was high for, both, the pretest and the posttest data. Our results thus demonstrate that the construct of conceptual knowledge is quite useful to explain children's patterns of behaviour over a variety of very different tasks.

The construct of procedural knowledge was found to be useful to explain patterns of different aspects of children's problem solving behaviour for the posttest, but not for the pretest, as confirmatory factor analyses revealed.

The most plausible explanation for this lack of a pretest procedural latent factor is that children, at the pretest, may have lacked procedural knowledge itself. In this case the between-persons variances of the four pretest measures would reflect children's measure-specific baselines rather than differences in their procedural knowledge.

Our second question can only partly be answered. Since we did not find a pretest procedural latent factor, we were not able to completely analyse the causal relations between children's conceptual and procedural knowledge before and after the intervention. We could, however, examine how their conceptual pretest knowledge influences their conceptual and procedural knowledge after the intervention. Here the strong influence of the conceptual pretest knowledge on the conceptual posttest knowledge is not surprising, since conceptual knowledge has been described, e.g. in studies on conceptual change, as being quite stable. What is more interesting is the finding that children's conceptual pretest knowledge fairly well predicts their procedural posttest knowledge. Obviously, conceptual knowledge is a valuable source for children's procedural knowledge.

The over-all picture of our results is not consistent with the procedures-first view. Clearly, the children do not have initial task-specific procedural knowledge that helps them to acquire the relevant concepts. There may be bi-directional causal relations after the children have already acquired both kinds of knowledge, but contrary to our expectations, no evidence for these processes was found in our study. For children that are relatively new to a mathematical domain, like those in our sample, the concepts-first view, rather than the procedures-first view or the Iterative Model, seems to be adequate: children start out with some prior conceptual knowledge that, then, serves as a source for new conceptual and procedural knowledge.

The low internal consistency of the posttest procedural latent factor might suggest that procedural knowledge is a hierarchical rather than a one-dimensional construct. This would be in accordance with several studies where the construct of automaticity was found to be multi-dimensional (see Besner, Stolz, & Boutilier, 1997).

References

Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(4), 932-945.

Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 1-33). Mahwah, NJ: Erlbaum.

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review, 4*(2), 221-225.

Byrnes, J. P., & Wasik, B. A. (1991). Role of conceptual knowledge in mathematical procedural learning. *Developmental Psychology, 27*(5), 777-786.

Canobi, K. H., Reeve, R. A., & Pattison, P. E. (1998). The role of conceptual understanding in children's addition problem solving. *Developmental Psychology, 34*(5), 882-891.

Johnson, A. (2003). Procedural memory and skill acquisition. In I. B. Weiner (Ed.), *Handbook of Psychology* (pp. 499-523). Hoboken, NJ: Wiley.

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. 75-328). Hove, UK: 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*(2), 346-362.

Satorra, A., & Bentler, P. M. (1999, August 3). *A scaled difference chi-square test statistic for moment structure analysis* [working paper]. University Pompeu Fabra, Department of Economics. Retrieved February 1, 2005, from the World Wide Web: http://ssrn.com/abstract=199064