Preservice Teachers' Beliefs about Students' Mathematical Knowledge Structure as a Foundation for Formative Assessments

Stephanie Herppich and Jörg Wittwer

University of Freiburg

Author Note

Stephanie Herppich, Department of Educational Science, University of Freiburg, 79085 Freiburg, Germany. Jörg Wittwer, Department of Educational Science, University of Freiburg, 79085 Freiburg, Germany.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Correspondence concerning this article should be addressed to Stephanie Herppich, Department of Educational Science, University of Freiburg, 79085 Freiburg, Germany.

E-mail: stephanie.herppich@ezw.uni-freiburg.de

Acknowledgements: We thank Bethany Rittle-Johnson and John Star for kindly providing us with the instruments used in their research. Moreover, we thank Jessica Hassler, Ruth Flierl, Mara Niebuhr, Julia Pflumm, Manuel Rischer, and Julia Tritschler for helping us with many practical aspects of the project.

Manuscript version of:
Herppich, S., & Wittwer, J. (in press). Preservice teachers' beliefs about students' mathematical knowledge structure as a foundation for formative assessments. *Journal of Teaching and Teacher Education.*
The final publication is available at ScienceDirect via https://doi.org/10.1016/j.tate.2018.06.011

Abstract

In an experiment, we studied beliefs about the structure of students' mathematical knowledge that may affect teachers' formative assessments. Using a novel approach that simulated an assessment situation, we measured the beliefs of N = 42 preservice mathematics teachers. Teachers' responses revealed that they predominantly thought of students' conceptual and procedural knowledge as being symmetrically related but not identical, which is in line with recent findings about students' knowledge. Their assessments may, thus, not be biased by incorrect beliefs. Teachers, however, did not believe conceptual and procedural knowledge to be more interrelated as a result of students' increased expertise.

Keywords: Preservice teacher beliefs; Formative assessment; Students' mathematical knowledge; Conceptual knowledge; Procedural knowledge

Preservice Teachers' Beliefs about Students' Mathematical Knowledge Structure as a Foundation for Formative Assessments

Introduction

Imagine that a mathematics teacher administers a test on the area of rectangles. She finds that most students were able to correctly calculate the areas. The teacher might use this result to formatively assess her students' knowledge (Bennett, 2011). Did the students learn how to carry out various multiplication procedures errorlessly and are, therefore, able to calculate the areas or do they also understand how the area of a rectangle is determined? What about those students who did not succeed in the test? Informed by this assessment, the teacher then could make an instructional decision (Bishop & Whitfield, 1972; Shavelson & Stern, 1981; cf. Author, this issue). For instance, the teacher could decide to proceed to different problems and re-teach the procedure to the students who failed. The teacher could also decide to choose problems that probe the students' knowledge first.

The correctness of these assessments would exert a strong influence on what students will learn in class. Often, however, the assessment of students' knowledge to decide next steps in teaching may be one of many professional tasks a teacher has to complete and may be carried out rather informally and based on incomplete information (Hascher, 2008; Kelley & Jacoby, 1996; Shavelson, 2006). In these situations, teachers are assumed to look for cues that reduce uncertainty about the students' knowledge, for example information about a student's prior performance (Brunswik, 1956; Koriat, 1997; Oudman, van de Pol, Bakker, Moerbeek, & van Gog, this issue; Thiede et al., 2015). These cues in turn elicit beliefs that guide the teacher's judgment (Mueller & Dunlosky, 2017; Witherby & Tauber, 2017). In situations, such as the one sketched above, teachers' beliefs about fundamental aspects of students' mathematical knowledge may impact assessments, for example, teachers' beliefs about students' conceptual and procedural knowledge and how these knowledge types are related with each other (Nickerson, 1999). For instance, a teacher might believe that a student who is able to calculate the area of a rectangle (i.e., procedural knowledge) also knows how this area is determined conceptually (i.e., conceptual knowledge). In this study, we were interested in preservice teachers' beliefs about the relationship between students' conceptual knowledge and procedural knowledge. More specifically, we examined those teacher beliefs about students' knowledge that become visible when teachers judge students' performance in concrete situations based on little available information.

Relationship Between Conceptual and Procedural Mathematical Knowledge

There is agreement in psychological, instructional design, and educational research on (learning) mathematics that procedural and conceptual knowledge are fundamental constituents of mathematical proficiency (e.g., Crooks & Alibali, 2014; Hatano & Inagaki, 1986; Rittle-Johnson, Fyfe, & Loehr, 2016; Schoenfeld, 2007; see Gray & Tall, 1994, for a more integrated view). Procedural knowledge is usually defined as "the ability to execute action sequences to solve problems[1]" (Schneider, Rittle-Johnson, & Star, 2011, p. 1525; cf. Crooks & Alibali, 2014; Hatano & Inagaki, 1986). There is more variation in the definition of conceptual knowledge (e.g., Crooks & Alibali, 2014; de Jong & Ferguson-Hessler, 1996; Smith & Ragan, 2005). In this article, we use the definition provided by Schneider et al. (2011, p. 1525) and refer to conceptual knowledge as "knowledge of the concepts of a domain and their interrelation". More specifically, in drawing on taxonomies of learning outcomes or knowledge components (Gagné, 1985; Koedinger, Corbett, & Perfetti, 2012), we focused on one aspect of conceptual knowledge, that is, knowledge of a theorem and its application in varying contexts in our study.

In their study, Schneider et al. (2011) found that conceptual and procedural knowledge are positively and symmetrically related with each other over time. That is, when a student's conceptual knowledge increases this will also entail an increase in procedural knowledge and

[1] We use the term *problem* with the meaning *mathematical task*.

vice versa to a similar extent (cf. Rittle-Johnson, Schneider, & Star, 2015). Moreover, Schneider et al. (2011) found that the correlational relationship between conceptual and procedural knowledge became stronger as a result of a three-day intervention. They cautiously interpret this result as a sign for knowledge integration during the development of higher expertise in the domain. The assumption of (possible) knowledge integration with increasing expertise is supported by theoretical and empirical findings from mathematics (Gray & Tall, 1994) and from other domains (e.g., Ericsson, 2006; Hatano & Inagaki, 1986; Schmidt & Rikers, 2007).

The relationship between students' conceptual and procedural knowledge in mathematics that has been observed in research may also be assumed by (preservice) teachers and, therefore, part of their beliefs. Thus, in line with the findings of Schneider et al. (2011), teachers might believe conceptual and procedural knowledge to be distinct but symmetrically related types of knowledge that become increasingly related and integrated as more knowledge is acquired.

Teachers' Knowledge and Practices Related to Conceptual and Procedural Knowledge

So far, teachers' beliefs about the interrelation between students' conceptual and procedural knowledge have not been examined. Indirect evidence about potential beliefs, however, comes from various fields of teacher research. First, Depaepe, Verschaffel, and Kelchtermans (2013) report in their review on pedagogical content knowledge (PCK) in mathematics education "that (pre-service) teachers' PCK is more procedural than conceptual…." (p. 21). To substantiate this claim, they cite work conducted by Cankoy (2010), who established that high school teachers predominantly proposed memorization-oriented procedural instructional strategies for teaching the zero-concept (cf. Kinach, 2002, for a similar example on integer addition and subtraction). In addition, Depaepe et al. (2013) report on the findings of Tirosh (2000), who observed that most participants did not know about students' naïve conceptions about fractions and, consequently, often attributed students' mistakes in dealing with fractions to forgotten algorithms and not to misconceptions.

Second, Thiede et al. (2015) found that teachers more accurately judged students' procedural computation skills than their conceptual knowledge of number sense. Drawing on the cue-utilization theory (Koriat, 1997), the researchers concluded that teachers either used evidence (i.e., cues) for the students' learning which was not informative (i.e., diagnostic) for judgments of conceptual knowledge. Alternatively, the teachers might have ignored diagnostic cues. This could have happened because they believed other cues to be more diagnostic (e.g., cues for students' procedural knowledge).

Third, research has found that teachers predominantly use instructional strategies that focus on procedures instead of concepts. For example, drawing on data from the TIMSS video studies, Richland, Stigler, and Holyoak (2012) showed that US teachers use mathematical problems to teach procedures most of the time, even when the problems were designed to teach the concepts underlying a procedure (cf. Schoenfeld, 1988). In addition, Lachner and Nückles (2016) revealed that German high school teachers tended to explain how to solve a word problem about an extremum problem with a focus on the algorithm to carry out and not with a focus on the underlying concepts. Similarly, Kinach (2002) found that she had to teach preservice teachers how to explain with a focus on conceptual understanding *against* their beliefs that procedural explanations were good explanations.

Overall, the findings show that mathematics teachers (still – cf. Schoenfeld, 1988) seem to be biased towards a focus on procedural knowledge in their teaching practice across mathematical content domains and across countries (e.g., Northern Cyprus: Cankoy, 2010; Germany: Lachner & Nückles, 2016; Israel: Tirosh, 2000; USA: Kinach, 2002; Richland et al., 2012; Schoenfeld, 1988). Based on these findings and in line with Kinach (2002), it is plausible to assume that teachers' beliefs about students' knowledge are biased towards procedural knowledge, too. As teachers' knowledge about (students') procedural knowledge seems to be more differentiated than their knowledge about (students') conceptual knowledge, they might, at least in part, equate conceptual understanding with procedural skill (Richland et al., 2012). Thus, a teacher might believe that a student who has correctly solved a procedural problem has understood the associated concepts as well.

Standards in Mathematics Teaching and Teacher Education

Internationally, there has been an increasing emphasis on teaching conceptual knowledge in mathematics for some decades. Specifically, students should acquire conceptual understanding to be able to understand procedures, to adapt them if necessary, and to apply them flexibly to a variety of situations (Crooks & Alibali, 2014; National Council of Teachers of Mathematics (NCTM), 2014; cf. Hatano & Inagaki, 1986). This emphasis on teaching and learning conceptual knowledge has been anchored in mathematics learning and teaching standards (e.g., Australia: The Australian Association of Mathematics Teachers (AAMT), 2006; USA: NCTM, 2000; 2014). In Germany, where this study was conducted, The Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (KMK) issued national standards for learning mathematics in 2003 to 2012 (for different types of schools, e.g., KMK, 2003). These standards emphasize mathematics as a means for flexible problem solving and focus on learning the conceptual underpinnings of procedures. In addition, the KMK released standards for mathematics teacher education in 2008. They require teachers to apply findings and suggestions from educational research (on mathematics). In this regard, many suggestions emphasize teaching of conceptual understanding (e.g., DMV, GDM, & MNU, 2008; Leuders, 2007; Prediger, 2009; cf. Crooks & Alibali, 2014; Rittle-Johnson et al., 2015). (German) younger teachers and preservice teachers in advanced phases of teacher education, thus, should have repeatedly learned that conceptual knowledge in mathematics is important. When we assume that a persons' beliefs develop during education (Stahl, 2011), such younger teachers' beliefs might be biased towards a focus on conceptual knowledge. Hence, a teacher might believe that a student who has understood the concepts associated with a procedure can correctly solve a related procedural problem, too.

How Teachers' Beliefs About Students' Knowledge May Influence Formative Assessments

In this study, we were interested in those teachers' beliefs about the structure of students' mathematical knowledge that are particularly relevant for formative assessments. Assessments of students are formative when they are meant to help learning (e.g., Black & Wiliam, 2009; Harlen, 2005). Shavelson (2006) has organized types of formative assessment along a continuum from informal/ unplanned to formal/ planned. At one extreme he puts on-the-fly assessments, which are teachers' reactions to unexpected teachable moments in class. An example is a teacher who 1) realizes that a student has just uttered a misconception and 2) spontaneously decides to challenge this misconception (cf. van Es & Sherin, 2002). At the other extreme he puts embedded-in-the-curriculum assessments. This term refers, for example, to tests that are carefully created when a series of lessons is planned to explicitly surface and analyze the students' understanding of critical aspects of the topic at hand at transitions from one aspect of the topic to the next. The assessments we are interested in, tend towards the informal/ unplanned end of the continuum. As in the opening example, teachers may judge students' understanding based on information about their students' knowledge not unpreparedly. They even may have time to think about the students' understanding without severe time pressure (Ostermann, Leuders, & Nückles, 2015). At the same time, evaluating such information is often one of many competing tasks. Mostly, it may not be part of a formal formative assessment procedure but carried out rather informally and under uncertainty (Hascher, 2008; Kelley & Jacoby, 1996). To arrive at judgments in such situations, according to cue-utilization theory, teachers search for cues that help them to reduce uncertainty. Depending on whether these cues are diagnostic for the students' understanding, the judgment becomes more or less accurate (Brunswik, 1956; Koriat, 1997; Oudman et al., this issue;

Thiede et al., 2015). Recent research on how people judge their own and others' learning, however, has suggested that it is not the cues per se that influence judgments but people's beliefs associated with the cues (Mueller & Dunlosky, 2017; Witherby & Tauber, 2017; cf. Kelley & Jacoby, 1996). For example, when people were made believe that words printed in blue are more easily processed than words printed in green, their belief that ease of processing leads to better memory made them assume that they would recall more blue than green words. Memory, however, did not differ depending on the color (Mueller & Dunlosky, 2017). In the same vein, teachers' judgments about students' mathematical knowledge can be influenced by their beliefs about the relationships between conceptual and procedural knowledge (Nickerson, 1999).

These beliefs could be elicited by information about a student's conceptual or procedural knowledge and measured via teachers' reaction to such cues. Nevertheless, to capture these beliefs, it may not be advisable to rely on self-reports (e.g., ask teachers: "Do you think that a student who has correctly solved a procedural problem understands the associated mathematical concepts?"). This is because the cognitions that influence behavior may not be identical with the reasons for behavior that people give or the beliefs that they state (Chinn, Buckland, & Samarapungavan, 2011; Nisbett & Wilson, 1977). Instead, these cognitions should be inferred from behavior (Chinn et al., 2011). Moreover, to consistently predict behavior, captured beliefs should be relatively specific to the behavior of interest (Rott & Leuders, 2016; Stahl, 2011; cf. the correspondence principle described by Ajzen & Fishbein, 1977, for the predictive value of attitudes). Thus, when we are interested in beliefs that are relevant to teachers' assessments of students' knowledge, we should not ask teachers about their general assumptions about the relationships between conceptual and procedural knowledge. Instead, to capture teachers' beliefs in line with these requirements, we asked participants to, actually, *assess* students' knowledge (structure) using a case scenario approach, which allowed us to infer the cognitions of interest (Kaiser, Südkamp, & Möller, 2017; Klug, Bruder, Kelava, Spiel, & Schmitz, 2013).

Present Study, Research Questions, and Hypotheses

We examined preservice mathematics teachers' beliefs about the relationship between students' conceptual and students' procedural knowledge as fundamental constituents of their mathematical proficiency. In particular, we were interested in those beliefs that most probably inform relatively informal formative assessments. We measured these beliefs behavior-related via judgments of students' performance in concrete situations. We examined the following research questions and hypotheses:

1. *Structure-of-Knowledge Research Question:* Based on the findings about the relationships between students' conceptual and procedural mathematical knowledge, on teachers' knowledge and instructional practices, and based on developments in teacher education, we wondered how preservice teachers conceive of the relationship between students' conceptual and procedural knowledge. We tested three mutually exclusive hypotheses: Preservice teachers conceive of students' conceptual and procedural knowledge as either…
    1) distinct types of knowledge that are symmetrically related to some extent (*symmetrically-related-but-not-identical-types-of-knowledge hypothesis*) or
    2) partially related types of knowledge where information about procedural knowledge is more relevant for judging conceptual and procedural knowledge than is information about conceptual knowledge (*procedural-bias hypothesis*) or
    3) partially related types of knowledge where information about conceptual knowledge is more relevant for judging conceptual and procedural knowledge than is information about procedural knowledge (*conceptual-bias hypothesis*).
2. *Prior-Knowledge Hypothesis:* As is the case in other knowledge domains, there is evidence that students' conceptual knowledge and procedural knowledge become more

related and integrated with increasing expertise. Accordingly, we hypothesized that preservice teachers assume a stronger relationship between conceptual and procedural knowledge for more knowledgeable students than for less knowledgeable students.

Methodology and Methods

Participants and Design

Participants were *N* = 42 mathematics preservice teachers in the second part of German teacher education who received training to teach at the secondary level in middle-track and vocational schools (Realschulen and Berufliche Schulen). Whereas in the first part of teacher education there are only some school placement phases that are framed by predominantly theoretical university studies, preservice teachers teach regularly during this second part. The participants had a mean age of 27.84 years (*SD* = 5.41, four participants did not indicate their age) and had been teaching for 1.62 years on average (*SD* = 4.04; including teaching experiences outside school; not indicated by seven participants). All were in the first year of one and a half years in this training phase. Their mean school leaving grade in mathematics was 1.91 (*SD* = 0.85; with German grades ranging from 1 = *very good* to 6 = *insufficient*; not indicated by one participant). Of the participants, 23 were female and 17 were male, 2 did not report gender.

We examined our research questions with a 2 x 2 x 2 x 2 mixed-design experiment with one between-subjects factor and three within-subjects factors (cf. Table 1). All participants read a case scenario that introduced to them a fictitious class of junior high school students. To half of the participants (*n* = 22), the students were introduced as having *high relevant prior knowledge* and, to half of the participants (*n* = 20), the students were introduced as having *low relevant prior knowledge*. This distinction was made to simulate more knowledgeable students versus less knowledgeable students. The participants were randomly assigned to the two groups. The groups did neither differ with regard to age, *t*(36) =

0.28, p = .80, nor with regard to teaching experience, $t(33) = -0.89$, p = .38, nor school leaving mathematics grade, $t(38) = -0.03$, p = .98, nor gender, $\chi^2(1, N = 40) = 0.05$, p = .82.

Afterwards, all participants read four case descriptions that varied within participants whether a fictitious student had worked on a conceptual problem or a procedural problem (= *type of start problem*) and whether the respective problem was solved correctly or not (= *success at start problem*). In each of the case descriptions, we then presented to the participants four additional problems the student would work on. There were two procedural additional problems and two conceptual additional problems (= *type of additional problems*). We counterbalanced the order of presenting the four case descriptions. The dependent variables were 1) the participants' ratings as to how likely each fictitious student would be able to solve each of the additional problems and 2) the number of arguments expressing independence of the problems in the participants' justifications for each rating.

----- Insert Table 1 about here ----

Stimuli

Case scenarios. To manipulate the assumed prior knowledge of the fictitious students, an introductory text informed the participants in the high prior knowledge condition that they had been teaching a class for several weeks on the same topic, the students had already had ample opportunity to practice and seemed to be doing well. In the low prior knowledge condition, the participants were informed that they had only started a new topic some days ago, there had barely been time to practice so far and the students only had a first understanding.

Subsequently, all participants were informed that they had assessed their students' current knowledge with a short test. Then, we asked them to judge how some of these students who had or had not correctly solved particular problems in this test would do on additional similar problems. In the subsequently presented case descriptions, the participants learned whether a student had correctly solved a problem from the test (i.e., the start problem) and were asked to rate, for each of the four additional problems, how likely this student would correctly solve the additional problem. Additionally, they were asked to justify their rating for each additional problem. Gender of the students in the case descriptions was held constant (female) and their names were chosen to be the most popular in the junior high school age group at the time of the study. The participants did not receive more details about the fictitious class to focus their attention exclusively on the information about the students' performance on the different problem types.

Conceptual and procedural problems. All problems were developed to fit with a common theme (i.e., the area of a rectangle) and to represent either a common aspect of conceptual knowledge or of procedural knowledge according to the definitions by Schneider et al. (2011, p. 1525; cf. Figure 1). We transformed these definitions to Gagné's (1985) cognitively based taxonomy of learning outcomes as used by Smith and Ragan (2005) in their book on instructional design to provide concrete guidelines for the construction of learning and assessment problems. Along these guidelines, we constructed the problems. More specifically, we assumed Schneider et al.'s (2011) conceptual knowledge to comprise, first, knowledge about concepts, that is, knowledge about categories and their attributes, including the ability to classify objects as members or non-members of a category. Second, we assumed conceptual knowledge to comprise knowledge about principles, that is, *if-then* rules that describe the relationships between concepts and "help us to predict, explain or control circumstances in our environment…" (Smith & Ragan, 2005, p. 81). Our conceptual problems all represented applications of the rule (i.e., the theorem): If a geometric shape is a rectangle, its area can be determined by length times width. Our operational definition in this study focused on one aspect of conceptual knowledge, that is, knowledge of the theorem that relates the concept area of a rectangle to the operation of multiplication and its application to varying contexts (cf. Koedinger et al., 2012; Smith & Ragan, 2005). Aspects such as the justification of the theorem or a broader meaning of the concept of area were not focused. Accordingly, we assumed procedural knowledge to comprise knowledge about procedures, that is, a series of strictly defined unambiguous steps that need to be carried out to complete a task. In our case, the procedure is to carry out the necessary steps to complete a multiplication problem[2]. To evoke a larger intra-individual variance in the participants' ratings, we, moreover, varied the difficulty of the problems. To do so, we changed aspects or mathematical *concepts* in the problems that did not alter the common conceptual or procedural aspect of the problems. For example, the easy additional conceptual problem was more abstract than the conceptual start problem and, in addition, to solve the more difficult additional conceptual problem, the student had to recognize that she needs to add m and 2b and to use parentheses to provide the formula. Practically, our mathematical problems were based on instruments used in the studies of Rittle-Johnson et al. (2009; cf. Schneider et al., 2011). We adapted items from these instruments to fit them with the requirements of our study (e.g., common theme). To do so, we used common mathematics textbooks for junior high school grades (Baum, 2012; Lergenmüller, 2006).

---- Insert Figure 1 about here ----

Measures

Ratings of likelihood for solving and justifications thereof. We borrowed a measure of judgments of learning from research on metacomprehension (Dunlosky & Rawson, 2005; see also Haase, Renkewitz, & Betsch, 2013) to determine preservice teachers' beliefs about

---

[2] To do so, a person has to categorize a problem as multiplication problem. Thus, according to Gagné, procedural knowledge always comprises conceptual knowledge as a prerequisite as well.

the relationship between students' conceptual and procedural knowledge in mathematics. After presenting each additional problem, we asked the participants to judge how likely the student would be able to solve this problem, in the participant's opinion, on a scale ranging from 0% (anchored as: "[Students' name] will certainly not solve the problem correctly.") to 100% (anchored as: "[Students' name] will certainly solve the problem correctly.") in 10% steps. In addition to 0% and 100%, 50% was anchored: "[Students' name] will solve the problem correctly with a probability of 50%." That is, the participants actually *assessed* (fictitious) students in a situation with limited information. This setting focuses on the variables that are relevant to our research question and, accordingly, allowed us to infer teachers' beliefs about students' knowledge structure from judgments that approximate assessments in class without interference of confounding variables (cf. Kaiser et al., 2017).

In addition, we asked the participants to briefly justify each rating. We coded whether these justifications comprised arguments that expressed that a rating was very tentative, uncertain or not possible due to the start problem and the additional problem being independent from each other with regard to their conceptual and procedural characteristics. For example, when a participant justified the likelihood rating for the easier conceptual problem after the procedural problem had not been solved correctly by the student with: "There isn't any information whether Sophie [the fictitious student] has a sound understanding of A [the area], problem 2 [the start problem] is a calculus problem.", this was coded as argument expressing independence. For each statement in the justification, one argument was coded. A second rater double-coded the justifications of the entire sample. Interrater agreement was acceptable for a high-inferential rating: 93%, $\kappa = .69$. After calculating interrater agreement both coders discussed and resolved remaining differences.

Ratings of problem difficulty. We used a measure introduced by Nathan and Petrosino (2003) to assess how difficult the preservice teachers assumed the problems to be. That is, we asked the participants to rank the six problems according to the difficulty they assumed the respective problem to entail for students. We explicitly allowed for tied ranks. For example, if a teacher had assumed the conceptual start problems to be the easiest to solve for junior high school students, he or she could have assigned this problem rank 1. If this teacher had assumed the second procedural additional problem and the second conceptual additional problem to be the most difficult he or she could have assigned them rank 6. And if this teacher had perceived the remaining problems to be on a similar level of difficulty somewhere in between he or she could have assigned them rank 3.

Procedure

See Table 1 for the experimental procedure.

Hypothetical Data Patterns

To examine our research questions, we applied the repeated-measures contrasts approach developed by Furr and Rosenthal (2003b; for more applications of this approach see, e.g., Tong, 2015; Tubaldi, Ansunini, Dematté, Tirindelli, & Castiello, 2008; van der Weiden, Ruys, & Aarts, 2013). In the logic of this approach, one derives an expected pattern of results based on a hypothesis. Afterwards, the expected data points are transformed into orthogonal contrast weights, which are used to compare the expected pattern with a given sample's data. This way, hypotheses about whether given data conforms with expected data patterns can be tested more powerfully and more directly than with omnibus-methods such as ANOVA (Furr & Rosenthal, 2003a, b). In this study, we derived expected data patterns that represent the three different hypotheses on teachers' beliefs about the relationships between students' conceptual and procedural knowledge (cf. our *Structure-of-Knowledge Research Question*) and about the extent to which conceptual and procedural knowledge are integrated with each other (cf. our *Prior-Knowledge Hypothesis*). To do so, we, first, determined potential likelihood ratings for the assumption that conceptual knowledge and procedural knowledge were entirely *unrelated* (see first row in Table 2). We set the hypothetical likelihood for solving a same type additional problem after the start problem had been solved at 80%. We chose this likelihood because we constructed the same type problems to be similar with regard to conceptual or procedural aspects but not completely isomorphic (cf. the stimuli section). Similarly, we set the hypothetical likelihood for solving a same type additional problem after the start problem had not been solved correctly for both types of start problem at 20%[3]. If we assume conceptual and procedural knowledge to be completely unrelated, evidence that a student can, for example, correctly solve a conceptual start problem would not tell us whether the student can solve a procedural additional problem and vice versa. Thus, for unrelated knowledge types we assumed the participants' ratings to tend towards the mean of the scale (cf. Schmidt-Atzert, 2010) when a conceptual start problem was followed by a procedural additional problem or vice versa. We did not set the likelihood for solving to exactly 50% because of two reasons. First, 50% ratings would result in contrast weights that equal zero. Zero contrast weights and pertaining observed data points, however, are dropped from the comparison of given and hypothetical data patterns. As the hypothetical data patterns only differ in the ratings for other type additional problems, it is essential that these ratings are represented in the comparison. Second, it is plausible to assume that a likelihood rating deviates towards the positive end of the scale after a correctly solved problem and towards the negative end of the scale after an incorrectly solved problem as a result of anchoring effects (Tversky & Kahnemann, 1974).

If preservice teachers assume conceptual knowledge and procedural knowledge to be *strongly symmetrically related* (see the second row in Table 2), evidence on one type of knowledge would be informative for expectations about the other type of knowledge. In this case, the likelihood ratings for correctly solving an additional problem should be similar or identical for the same and the other type of problem. Based on the two hypothetical data

[3] To test the robustness of our findings, we repeated all analyses with the hypothetical likelihoods 70% versus 30% and 90% versus 10%. As results remained virtually the same, we only report the analyses pertaining to the 80% versus 20% likelihoods.

patterns described for unrelated types of knowledge and for strongly symmetrically related types of knowledge, we derived the potential data pattern for the *symmetrically-related-but-not-identical hypothesis* (see third row in Table 2) as the average of the two patterns.

If preservice teachers assume evidence about conceptual knowledge to be informative for students' conceptual *and* procedural knowledge, as stated by the *conceptual-bias hypothesis*, the pattern of likelihood ratings should be asymmetrical (see fourth row in Table 2). More precisely, when the start problem is conceptual there should be similar positive or negative likelihood ratings respectively for all additional problems. For example, if a student was able to solve the conceptual start problem, a teacher with a conceptual bias might expect that this student can quite likely solve not only the additional conceptual problems (e.g., 80% likelihood) but also the additional procedural problems (e.g., 80% likelihood). However, when the start problem is procedural this teacher's likelihood ratings should only be clearly positive or negative, respectively, for procedural additional problems (e.g., 80% when the start problem is solved correctly and 20% when it is not solved correctly). The ratings for conceptual additional problems should, by contrast, tend towards the mean (e.g., 55% and 45%) because the teacher would not extrapolate from information about a student's procedural knowledge to a student's conceptual knowledge. If a teacher has a procedural bias as postulated by the *procedural-bias hypothesis* (see fifth row in Table 2), on the other hand, his or her ratings should be similar when the start problem is procedural but differ between conceptual and procedural additional problems when the start problem is conceptual.

We proceeded similarly for the number of arguments that expressed independence in preservice teachers' justifications of the likelihood ratings. If preservice teachers assumed conceptual and procedural knowledge to be unrelated and evidence on one type of knowledge to be not informative for judgments about the other type of knowledge, the participants should refer to the independence of the problems more often when they justify ratings for other type additional problems than when they justify ratings for same type problems (see Table 2). On the contrary, participants should seldom refer to the independence of the problems overall in their justifications if they assumed conceptual and procedural knowledge to be closely related. Furthermore, when teachers assumed conceptual knowledge to be important for solving conceptual and procedural tasks, they should only refer to independence of the problems if a conceptual problem followed a procedural start problem. Again, this pattern should emerge vice versa when preservice teachers perceived procedural knowledge to be informative for procedural and conceptual problems.

---- Insert Table 2 about here ----

Results

We used an α-level of 0.05. All analyses were performed using Excel 2013 and SPSS 24.0.0. Some participants did not respond to all items. Therefore, the *N* per analysis slightly fluctuates. As fluctuations remained in the range of only n = 1-2 persons, we excluded missing values from the analyses. None of the demographic variables showed systematic relationships with the dependent variables in the study. Pertaining analyses are not reported here in detail for the sake of conciseness but can be requested from the authors.

To test whether the teachers assumed corresponding conceptual problems and procedural problems to be equally difficult to solve for junior high school students, we analyzed the results of the difficulty ranking task (Nathan & Petrosino, 2003). As Table 3 shows, the participants perceived conceptual and procedural problems as approximately equally difficult. Moreover, they perceived the levels of difficulty, overall, as intended by us. To ease interpretations, we performed all subsequent analyses with easier and more difficult problems of the same type collapsed into one variable.

---- Insert Table 3 about here ----

Structure of knowledge research question.

We analyzed the preservice teachers' likelihood ratings to examine which of the three hypotheses on beliefs about the relationships between students' conceptual and procedural knowledge describes the data best (cf. Furr &Rosenthal, 2003b). First, we tested to what extent the data generally conformed with each of the three data patterns suggested by the three hypotheses (test of three one-pattern one-group hypotheses; see Figure 2 for descriptive results). To derive the necessary contrast weights, we *z*-standardized the likelihood ratings that represented the respective hypothesis (see the hypothetical data patterns section). The contrast weights for the pattern that pertains to the *symmetrically-related-but-not-identical-types-of-knowledge hypothesis* were 1.14, 0.67, -1.14, -0.67, 1.14, 0.67, -1.14, -0.67, ordered along the pattern displayed in Table 2. Generally, our data manifested the pattern suggested by the symmetrically-related-but-not-identical-types-of-knowledge hypothesis, as shown by a positive and significant test statistic, $t(38) = 14.61$, $p < .001$, $r_{contrast} = .92$ (large effect). The contrast weights for the *conceptual-bias hypothesis* were 1.08, 1.08, -1.08, -1.08, 1.08, 0.18, -1.08, -0.18. This pattern also conformed with our data, $t(38) = 13.81$, $p < .001$, $r_{contrast} = .91$ (large effect). Finally, the contrast weights for the *procedural-bias hypothesis* were 1.08, 0.18, -1.08, -0.18, 1.08, 1.08, -1.08, -1.08. And this pattern could also potentially explain our data, $t(38) = 14.80$, $p < .001$, $r_{contrast} = .92$ (large effect). All contrasts remained significant after Bonferroni correction for family-wise error rate, which lowered the alpha-level to $p = .017$ (Furr & Rosenthal, 2003b).

To test whether one hypothesis represented the data better than the other two, we, second, calculated three three-pattern one-group contrasts (Furr & Rosenthal, 2003b). We used the contrast weights reported above for the three hypotheses. The analyses showed that the data manifested the *symmetrically-related-but-not-identical-types-of-knowledge* pattern significantly more strongly than the averaged *conceptual-bias* pattern and *procedural-bias* pattern, $t(38) = 14.61$, $p < .001$, $r_{contrast} = .92$ (large effect). The *conceptual-bias* pattern represented the data significantly worse than the other two patterns did, $t(38) = -2.08$, $p = .04$, $r_{contrast} = .32$ (medium effect). Finally, there was no significant difference between the *procedural-bias* pattern and the other two patterns, $t(38) = -.79$, $p = .43$, $r_{contrast} = .13$ (small effect). Only the comparison of the *symmetrically-related-but-not-identical* pattern with the other two patterns remained significant after applying Bonferroni correction that reduced the alpha value to $p = .017$. Overall, the sample, thus, manifested the *symmetrically-related-but-not-identical-types-of-knowledge* pattern more strongly than the two bias patterns.

---- Insert Figure 2 about here ----

To compare the evidence for the three hypotheses on beliefs about the relationships between students' conceptual and procedural knowledge we, moreover, analyzed whether the preservice teachers' justifications for their likelihood ratings comprised arguments that indicated independence of the problems regarding their conceptual and procedural characteristics. We, first, tested whether the arguments were evenly distributed over conceptual start problems and procedural start problems. To do so, we used a Wilcoxon signed rank test. An asymmetric distribution would indicate a conceptual bias or a procedural bias depending on the direction of the skew, whereas an even distribution would indicate no bias. The descriptive values suggested that there were more arguments indicating independence of the problems after the procedural start problem than after the conceptual start problem (see Figure 2). The pertaining test was not significant but showed a trend, $Z = -1.83$, $p = .07$, $r = -.20$ (small effect). Thus, it provided weak evidence for a conceptual bias as the teachers marginally more often referred to the independence of start problem and additional problem in their justifications following a procedural start problem than following a conceptual start problem. Second, we compared the number of arguments indicating independence of problems between same and different types of additional problems for each level of success at the start problem at each of the two start problems. The first comparison between same type problems and other type problem for the correctly solved conceptual start problem resulted in $Z = -0.39$, $p = .71$, $r = -.04$ (small effect). The second comparison for the conceptual start problem not being correctly solved resulted in $Z = -0.72$, $p = .47$, $r = -.08$ (small effect), the third comparison for the correctly solved procedural problem in $Z = -3.35$, $p < .01$, $r = -.37$ (medium effect), and the fourth comparison for the procedural problem not being correctly solved in, $Z = -2.38$, $p = .02$, $r = -.26$ (medium effect; see Figure 2 for the descriptive data). These results provide some evidence that preservice teachers did not perceive the information about the student's performance at the procedural start problem to be relevant for judging her performance at the conceptual additional problems and considered data on conceptual knowledge as more informative for judging how likely a student would solve an additional problem even of the other knowledge type. This is because the teachers expressed independence of problems for other type problems more often than for same type problems only when the start problem was procedural. They did not do so when the start problem was conceptual. After Bonferroni correction for family-wise error rate ($p = .0125$), the interpretations of the *p*-values remained the same for justification of ratings following conceptual start problems and for justification of ratings following the correctly solved procedural problem. The difference in independence arguments that accompanied ratings for same type additional problems and other type additional problems when the procedural start problem had not been solved correctly reduced to a trend. Overall, the pattern of our results provides strong evidence for the hypothesis that preservice teachers conceive of students' conceptual and procedural knowledge as symmetrically related but not identical types of knowledge if we look at the likelihood ratings. At the same time, there is some evidence that teachers assume data about students' procedural knowledge to be less informative for judgments about their mathematical skills than data about students' conceptual knowledge, indicating a conceptual bias, if we look at the justifications.

Prior-Knowledge Hypothesis

We also hypothesized that preservice teachers would assume students' knowledge types to be more integrated when these students had more prior knowledge than when they had less prior knowledge. To test this hypothesis, we performed a one-pattern two-groups contrast (Furr & Rosenthal, 200b3) that compared the extent to which both groups represented the symmetrically related pattern in the likelihood ratings (see Figure 3 for descriptive results). The contrast weights derived from the predicted pattern were 0.94, 0.94, -0.94, -0.94, 0.94, 0.94, -0.94, -0.94. There was no significant difference between the groups, $t(37) = 1.11$, $p = .27$, $r_{contrast} = .18$ (small effect). Accordingly, the high prior knowledge group did not manifest the symmetrically related pattern more substantially than the low prior knowledge group. As the two groups did not differ significantly in their beliefs from each other, we were interested whether the entire sample was more representative for the *unrelated* pattern or for the *symmetrically related* pattern. A posteriori, we, therefore, conducted a two patterns one group contrast (Furr & Rosenthal, 2003b). Again, we employed the contrast weights 0.94, 0.94, -0.94, -0.94, 0.94, 0.94, -0.94, -0.94 for the *symmetrically related* pattern. In addition, we used the contrast weights 1.30, 0.22, -1.30, -0.22, 1.30, 0.22, -1.30, -0.22 for the *unrelated* pattern and tested whether one of the hypothetical patterns represented our data significantly better than the other (see Figure 2 for descriptive results). We found that the participants believed conceptual knowledge and procedural knowledge to be only non-significantly more unrelated than related, $t(38) = 1.43$, $p = .16$ $r_{contrast} = .22$ (small effect). Thus, the belief *symmetrically-related-but-not-identical* seems, indeed, to be a good description for the empirical data pattern.

Second, we analyzed whether preservice teachers in the high prior knowledge group expressed fewer arguments referring to the independence of problems for other type additional problems than preservice teachers in the low prior knowledge group (see Figure 3 for means and standard deviations). This was not the case as shown by a non-significant Mann-Whitney-U test, $Z = -0.22$, $p = .83$, $r = -.03$ (small effect). Accordingly, our prior knowledge hypothesis was neither supported by the likelihood ratings nor by their justifications.

---- Insert Figure 3 about here ----

Discussion

This study examined preservice teachers' beliefs about the relationship between students' conceptual and procedural mathematical knowledge. We were particularly interested in those beliefs that may influence teachers' relatively informal formative assessments and instructional decisions under uncertainty. Therefore, we measured these beliefs based on a case-scenario approach with a judgment task that requested preservice teachers to assess fictitious students' performance based on limited information. We found, first, that preservice teachers rated the relationships between students' answers to conceptual problems and students' answers to procedural problems in a way that was most consistent with the belief that conceptual and procedural knowledge are symmetrically related but not identical types of knowledge.

Literature on the structure of students' mathematical knowledge, on teachers' knowledge and instructional practices, and on standards in mathematics teaching and their impact on teacher education suggested three different hypotheses about preservice mathematics teachers' beliefs regarding the relationships between the knowledge types. Our data provides evidence that teachers' beliefs, measured by judgments of fictitious students' performance in an approximation to formative assessment situations (Chinn et al., 2011; cf. Kaiser et al., 2017; Klug et al., 2013), predominantly reflect those relationships that have recently been found by research on students' knowledge (Rittle-Johnson et al., 2015; Schneider et al., 2011). These beliefs entail that preservice teachers do not automatically impute conceptual knowledge to a student who has correctly solved a procedural problem, neither do they believe that a student, who has good conceptual knowledge, can automatically solve a related procedural problem. To come back to the situation sketched in the introduction section, a teacher who finds that most students in class were able to correctly calculate areas of rectangles and who judges the students' knowledge based on this information would probably not assume that the students have understood how the area of rectangles is determined as well. This outcome draws a more positive picture of teachers' beliefs than has been done by previous research. For example, preservice teachers who assume procedural explanations generally to be good explanations, as in the study of Kinach (2002), might not differentiate between procedural and conceptual knowledge but show a procedural bias and impute conceptual understanding to someone who is able to carry out a procedure. If we work on the premise that beliefs develop biographically (Feldon, 2007; Stahl, 2011), we may speculate, accordingly, that the education experienced by the preservice teachers in our study facilitated the development of differentiated beliefs.

In addition to capturing teachers' judgments of fictitious students' performance, we also asked the participants to justify their ratings after they had indicated how likely they assumed a fictitious student to solve a conceptual or a procedural problem. We analyzed whether the participants expressed that start problem and additional problem were independent from each other with regard to their conceptual and procedural characteristics. These justifications show that the teachers trusted information about students' conceptual knowledge as a basis for judging conceptual *and* procedural knowledge more than they trusted information about students' procedural knowledge as a basis for judging conceptual knowledge. This slight conceptual bias may mean that preservice teachers tend to impute procedural skills when students demonstrate conceptual understanding. Given the teachers' judgments, however, it might be more likely that the participants explicitly devaluate information about students' procedural knowledge as a source for judging conceptual knowledge instead of overestimating the value of information about conceptual knowledge. Remember that the hypothetical data pattern that was associated with the *conceptual-bias hypothesis* represented the judgment data worst as compared with the other hypothetical patterns. In learning about the importance of conceptual knowledge in teacher education (DMV, GDM, & MNU, 2008; KMK, 2003, 2008; Leuders, 2007; Prediger, 2009), preservice teachers may also learn that a focus on procedural knowledge in teaching is criticized (Kinach, 2002; Prediger, 2009; Richland, 2012). Based on these results, future research should examine in which ways and to what extent these beliefs actually influence teachers' assessments and instructional actions. Our study suggests that, particularly, preservice teachers and perhaps younger teachers may not regularly extrapolate from information about one type of knowledge to the other type of knowledge and that this effect should be particularly visible if information about procedural knowledge is given. Accordingly, teachers might, for example, probe students' understanding if this student can calculate the area of a rectangle as an effect of their beliefs about the relationship between conceptual and procedural knowledge. Of course, we do not assume that these beliefs are the only determinants of teachers' assessments but rather one of many (Author et al., this issue; Südkamp, Kaiser, & Möller, 2012). Similarly, we do not assume necessarily direct influences on assessments and actions but rather indirect influences (cf. Ajzen, 1991). Thus, there is need for additional research. Given that little is known about teachers' characteristics that are relevant for their assessments (Machts, Kaiser, Schmidt, & Möller, 2016; Südkamp et al., 2012), however, our study constitutes an important first step.

Our second research question asked whether preservice teachers assumed students' conceptual and procedural knowledge to be more strongly related for students with high prior knowledge (i.e., more expertise) than for students with low prior knowledge. If teachers believed the types of knowledge to become more related and integrated with increasing knowledge, these beliefs would probably yield different judgments for case-scenario students with low prior knowledge and for case-scenario students with high prior knowledge. However, the results did not suggest that teachers believed students' knowledge to be more closely related as a result of increased prior knowledge. Instead, the believed knowledge structure could best be described as conceptual and procedural knowledge being symmetrically related but not identical, independent of the fictitious students' level of prior knowledge. In terms of the example provided in the introduction section, a teacher might not assume that the students have understood how the area of rectangles is determined when they can correctly calculate areas of rectangles, if these students have low prior knowledge. Consequently, the teacher might want to probe the students' understanding. If, however, the students have high prior knowledge, the teacher could assume that they, most probably, have acquired the conceptual knowledge when they are able to carry out the procedures (given appropriate instruction, cf. Hatano & Inagaki, 1986). Hence, a belief in knowledge integration could reduce the perceived need for formative assessment. Teachers who do not believe in knowledge integration with expertise development, however, cannot become more confident about one type of knowledge when they predominantly have information about the other type of knowledge with increasing students' knowledge.

A limitation of our study is, that we manipulated the information about students' prior knowledge (high vs. low) as a between-subjects factor. Consequently, the participants could not compare high prior knowledge students with low prior knowledge students in their judgments. This might have masked differences because measures of likelihood are more suitable for capturing relative differences than for capturing absolute judgments (Haase et al., 2013). Moreover, whereas the power for within-subject comparisons was adequate to detect at least medium size effects, the number of participants was only adequate to detect large between-subject effects (Faul, Erdfelder, Lang, & Buchner, 2007). However, between-group effects in the current designs were that small that a change of design as proposed above seems more promising to get clearer results than (only) an increase in sample size.

Another limitation to the generalizability of our results is that we operationalized conceptual and procedural knowledge along only one definition (Schneider et al., 2011) and one specific operationalization of the definition (Gagné, 1985; Smith & Ragan, 2005), for one content, and one type of problems. In operationalizing conceptual knowledge, we focused on different applications of the principle or theorem: If a geometric shape is a rectangle, its area can be determined by length times width (cf. Gagné, 1985; Smith & Ragan, 2005). Despite this potentially narrow focus, the participants' justifications of their likelihood ratings imply that they often generalized the given operationalization to other aspects of conceptual knowledge, such as the broader meaning of the concept of area. This was evident from justifications, such as "She has understood the principle", "If she has understood problem 1 [the start problem], she will be able to solve this problem, because she has a concept of it", "has a concept of area calculations", "in problem 1 [the start problem] one shows general understanding of areas". Hence, the results may be generalizable beyond our specific operationalization. Nevertheless, there exist several definitions and operationalizations of conceptual knowledge in the literature (cf. Crooks & Alibali, 2014; Chi; Siler, Jeong, Yamauchi, & Hausmann, 2001; de Jong & Ferguson-Hessler, 1996; Gagné, 1985; Koedinger et al., 2012). To validate our results, thus, future research should refer to other definitions and operationalizations, including more explicit references to other aspects of conceptual knowledge, such as the justifications of theorems. According definitions and operationalization have been collated by Crooks and Alibali (2014) for three mathematical domains.

As with other case scenario and simulation based studies (e.g., Kaiser et al., 2017; Klug et al., 2013; Krolak-Schwerdt et al., 2013), finally, our approach uses a simplified model of reality to infer teachers' beliefs. This approach allowed us to capture the outcomes of interest from over 40 participants while preventing interference of many confounding variables (cf., Kaiser et al., 2017; Wang, Treat, & Brownell, 2008). Thus, the efficiency of our method and the internal validity of our study are very high, although at the expense of external validity. Nevertheless, similar approaches have proven to capture outcomes that are relevant to behavior in reality (Bledow & Frese, 2009)

Notwithstanding the limitations, our study showed that (German) preservice teachers have relatively unbiased beliefs about the relationships between students' conceptual and procedural knowledge, making an important step towards the examination of teachers' characteristics that may impact teachers' assessment of students (cf. Machts et al., 2016). The structure of these beliefs is in line with recent findings about the structure of students' knowledge and allows for cautious optimism about (teacher) education. A task for future research is now to relate such judgments to teachers' formative assessments and instructional decisions.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179-211. https://doi.org/10.1016/0749-5978(91)90020-T

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*, 888-918.

Author (this issue)

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Principles, Policy & Practice, 18,* 5–25. https://doi.org/10.1080/0969594x.2010.513678

Baum, M. (2004). *Lambacher-Schweizer: Mathematik für Gymnasien* [Lambacher-Schweizer: Mathematics for academic track secondary schools]. Leipzig, Germany, Klett.

Bishop, A. J., & Whitfield, R. C. (1972). *Situations in teaching.* Maidenhead Berkshire, UK: McGraw-Hill.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment Evaluation and Accountability, 2*1, 5-31. https://doi.org/10.1007/s11092-008-9068-5.

Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative and its relationship to performance. *Personnel Psychology, 62*, 229-258. https://doi.org/10.1111/j.1744-6570.2009.01137.x

Borko, H., Roberts, S., & Shavelson, R. J. (2008). Teachers' decision making: from Alan J. Bishop to today. In P. Clarkson, & N. Presmeg (Eds.), *Critical issues in mathematics education*, (pp. 37-67). Berlin, Germany: Springer.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.

Cankoy, O. (2010). Mathematics teachers' topic-specific pedagogical content knowledge in the context of teaching a 0, 0! and /0. *Kuram ve Uygulamada Egitim Bilimleri [Educational Sciences: Theory and Practice], 10*, 749-769.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. (2001). Learning from human tutoring. *Cognitive Science, 25*, 471–533. https://doi.org/10.1207/s15516709cog2504_1

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. L. A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist, 46*, 141-167. https://doi.org/10.1080/00461520.2011.587722

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review, 34,* 344-377. https://doi.org/10.1016/j.dr.2014.10.001

de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*, 105–113.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education, 34*, 12-25. https://doi.org/10.1016/j.tate.2013.03.001

DMV, GDM, & MNU (2008). Standards für die Lehrerbildung im Fach Mathematik. Empfehlungen von DMV, GDM, MNU [Standards for teacher education in mathematics. Suggestions of DMV, GDM, MNU]. Retrieved from http://madipedia.de/wiki/Stellungnahmen

Dunlosky, J., & Rawson, K. A. (2005). Why does rereading improve metacomprehension accuracy? Evaluating the levels-of-disruption hypothesis for the rereading effect. *Discourse Processes, 40*, 37-55. https://doi.org/10.1207/s15326950dp4001_2

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683-704). Cambridge: Cambridge University Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist, 42*, 123-137. https://doi.org/10.1080/00461520701416173

Furr, R. M., & Rosenthal, R. (2003a). Evaluating theories efficiently: The nuts and bolts of contrast analysis. *Understanding Statistics, 2*, 45-67. https://doi.org/10.1207/S15328031US0201_03

Furr, R. M., & Rosenthal, R. (2003b). Repeated-measures contrasts for "multiple-pattern" hypotheses. *Psychological Methods, 8*, 275-293. https://doi.org/10.1037/1082-989X.8.3.275

Gagné, R. M. (1985). *The conditions of learning* (4th ed.). New York, NY: Holt, Rinehart & Winston.

Gray, E. M., & Tall, D. O. (1994). Duality, ambiguity, and flexibility: A "proceptual" view of simple arithmetic. *Journal for Research in Mathematics Education, 25*, 116-140. https://doi.org/10.2307/749505

Haase, N., Renkewitz, F., & Betsch, C. (2013). The measurement of subjective probability: Evaluating the sensitivity and accuracy of various scales. *Risk Analysis, 33*, 1812-1828. https://doi.org/10.1111/risa.12025

Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal*, 16, 207-223. https://doi.org/10.1080/09585170500136093.

Hascher, T. (2008). Diagnostische Kompetenz im Lehrerberuf [Diagnostic competences in the teaching profession]. In C. Kraler, & M. Schratz (Eds.), *Wissen erwerben, Kompetenzen entwickeln. Modelle zur Kompetenzorientierten Lehrerbildung* [Gaining knowledge, developing competences. Models for competence-oriented teacher education] (pp. 71–86). Münster, Germany: Waxmann

Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japa*n (pp. 262–272). New York, NY: Freeman.

Kaiser, J., Südkamp, A., & Möller, J. (2017). The effects of student characteristics on teachers' judgment accuracy: Disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology*, *109*, 871-888. https://doi.org/10.1037/edu0000156

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language*, *35*, 157-175. https://doi.org/10.1006/jmla.1996.0009

Kinach, B. M. (2002). Understanding and learning-to-explain by representing mathematics: Epistemological dilemmas facing teacher educators in the secondary mathematics "methods"' course. *Journal of Mathematics Teacher Education*, *5*, 153-186. https://doi.org/10.1023/A:1015822104536

Klug, J., Bruder, S., Kelava, A., Spiel, C., & Schmitz, B. (2013). Diagnostic competence of teachers: A process model that accounts for diagnosing learning behavior tested by means of a case scenario. *Teaching and Teacher Education, 30*, 38-46. https://doi.org/10.1016/j.tate.2012.10.004

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction Framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36*, 757-798. https://doi.org/10.1111/j.1551-6709.2012.01245.x

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349. https://doi.org/10.1037/0096-3445.126.4.349

Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2013). The impact of accountability on teachers' assessments of student performance: a social cognitive analysis. *Social Psychology of Education, 16*, 215-239. https://doi.org/10.1007/s11218-013-9215-9

Lachner, A., & Nückles, M. (2016). Tell me why! Content knowledge predicts process-orientation of math researchers' and math teachers' explanations. *Instructional Science, 44*, 221-242. https://doi.org/10.1007/s11251-015-9365-6

Lergenmüller, A. (Ed.) (2006). *Mathematik neue Wege* [Mathematics new ways]. Hannover, Germany, Schroedel.

Leuders, T. (2007). Fachdidaktik und Unterrichtsqualität im Bereich Mathematik [Subject didactics and instructional quality in mathematics]. In K.-H. Arnold (Ed.), *Unterrichtsqualität und Fachdidaktik* [Instructional quality and subject didactics] (pp. 205–234). Bad Heilbrunn, Germany: Klinkhardt.

Machts, N., Kaiser, J., Schmidt, F. T. C., & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review, 19*, 85-103. https://doi.org/10.1016/j.edurev.2016.06.003

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language, 93*, 245-258. https://doi.org/10.1016/j.jml.2016.10.008

Nathan, M. J., & Petrosino, A. (2003). Expert blind spot among preservice teachers. *American Educational Research Journal, 40*, 905-928. https://doi.org/10.3102/00028312040004905

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics (NCTM). (2014). *Principles to actions: Ensuring mathematical success for all*. Reston, VA: Author.

Nickerson, R. S. (1999). How we know -- and sometimes misjudge -- what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*, 737-759. https://doi.org.10.1037//0033-2909.125.6.737

Nisbett R. E., & Wilson T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–95.

Ostermann, A., Leuders, T., & Nückles, M. (2015). Wissen, was Schülerinnen und Schülern schwer fällt. Welche Faktoren beeinflussen die Schwierigkeitseinschätzung von Mathematikaufgaben? [Knowing what is hard for students. Which factors influence difficulty judgments of mathematical problems?] *Journal für Mathematik-Didaktik, 36*, 45-76. https://doi.org.10.1007/s13138-015-0073-1

Oudman, S., van de Pol, J., Bakker, A., Moerbeek, M., & van Gog, T. (this issue). Effects of different cue types on the accuracy of primary school teachers' judgments of students' mathematical understanding. *Teaching and Teacher Education*. Advance online publication. https://doi.org/10.1016/j.tate.2018.02.007

Prediger, S. (2009). Inhaltliches Denken vor Kalkül – Ein didaktisches Prinzip zur Vorbeugung und Förderung bei Rechenschwierigkeiten [Thinking through content before calculations – an educational principle for prevention and promotion in the case of arithmetic difficulties]. In A. Fritz, & S. Schmidt (Eds.). *Fördernder Mathematikunterricht in der Sek. I. Rechenschwierigkeiten erkennen und überwinden*

[Promoting mathematics instruction in lower secondary grades. Recognizing and overcoming arithmetic difficulties] (pp. 213-234). Weinheim, Germany: Beltz.

Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the Conceptual Structure of Mathematics. *Educational Psychologist, 47*, 189-203. https://doi.org/10.1080/00461520.2012.667065

Rittle-Johnson, B., Fyfe, E. R., & Loehr, A. M. (2016). Improving conceptual and procedural knowledge: The impact of instructional content within a mathematics lesson. *British Journal of Educational Psychology, 86*, 576-591. https://doi.org/10.1111/bjep.12124

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*, 587–597. https://doi.org/10.1007/s10648-015-9302-x

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. https://doi.org/10.1037/a0016026

Rott, B., & Leuders, T. (2016). Inductive and deductive justification of knowledge: Flexible judgments underneath stable beliefs in teacher education. *Mathematical Thinking and Learning, 18*, 271-286. https://doi.org/10.1080/10986065.2016.1219933

Schmidt, H. G., & Rikers, R. M. J. P. (2007). How expertise develops in medicine: knowledge encapsulation and illness script formation. *Medical Education, 41*, 1133-1139. https://doi.org/10.1111/j.1365-2923.2007.02915.x

Schmidt-Atzert, L. (2010). Beobachtungsfehler und Beobachtungsverzerrungen [Observational mistakes and biases]. In K. Westhoff, Hagemeister, K., Kersting, M., Lang, F., Mosbrugger, H., Reimann, G., & Stemmler, G. (Eds.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* [Basic knowledge for vocational suitability assessments according to DIN 33430] (3rd ed., pp. 74-77). Lengerich, Germany: Pabst.

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

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of "well-taught" mathematics courses. *Educational Psychologist, 23*, 145-166.

Schoenfeld, A. H. (2007). What is mathematical proficiency (and how can it be assessed)? In A. H. Schoenfeld (Ed.), *Assessing mathematical proficiency* (pp. 59-73). Cambridge, UK: Cambridge University Press.

Shavelson, R. J. (2006). *On the integration of formative assessment in teaching and learning with implications for teacher education.* Paper prepared for the Stanford Education Assessment Laboratory and the University of Hawaii Curriculum Research and Development Group. Retrieved from http://www.stanford.edu/dept/SUSE/SEAL/Reports_Papers/Paper.htm

Shavelson, R. J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research, 51*, 455–498.

Smith, P. L., & Ragan, T. J. (2005). *Instructional design*. Hoboken, NJ: Wiley.

Stahl, E. (2011). The generative nature of epistemological judgments. In J. Elen, E. Stahl, R. Bromme, & G. Clarebout (Eds.), *Links between beliefs and cognitive flexibility – Lessons learned* (pp. 37–60). Dordrecht, The Netherlands: Springer.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*, 743-762. https://doi.org/10.1037/a0027627

The Australian Association of Mathematics Teachers (AAMT) (2006). *Standards for excellence in teaching mathematics in Australian schools.* Retrieved from http://www.aamt.edu.au/Better-teaching/Standards

The Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (KMK). (2003). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss* (Beschluss vom 4.12.2003) [Educational standards in mathematics for the intermediate-level school leaving qualification (decision of December 4, 2003)]. Retrieved from https://www.kmk.org/themen/qualitaetssicherung-in-schulen/bildungsstandards.html

The Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (KMK). (2008). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung* (Beschluss der Kultusministerkonferenz vom 16.10.2008 i. d. F. vom 06.10.2016) [Content-related requirements of all Länder for subjects and subject didactics in teacher education (decision of October, 16, 2008 in the version of October 6, 2016)]. Retrieved from https://www.kmk.org/themen/allgemeinbildende-schulen/lehrkraefte/lehrerbildung.html

Thiede, K. W., Brendefur, J. L., Osguthorpe, R. D., Carney, M. B., Bremner, A., Strother, S., . . . Jesse, D. (2015). Can teachers accurately predict student performance? *Teaching and Teacher Education, 49,* 36-44. https://doi.org/10.1016/j.tate.2015.01.012

Tirosh, D. (2000). Enhancing prospective teachers' knowledge of children's conceptions: The case of division of fractions. *Journal for Research in Mathematics Education, 31,* 5-25.

Tong, E. M. W. (2015). Differentiation of 13 positive emotions by appraisals. *Cognition and Emotion, 29,* 484-503. https://doi.org/10.1080/02699931.2014.922056

Tubaldi, F., Ansuini, C., Demattè, M. L., Tirindelli, R., & Castiello, U. (2008). Effects of olfactory stimuli on arm-reaching duration. *Chemical Senses, 33*, 433-440. https://doi.org/10.1093/chemse/bjn010

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131. https://doi.org/10.1126/science.185.4157.1124

van der Weiden, A., Ruys, K. I., & Aarts, H. (2013). A matter of matching: How goals and primes affect self-agency experiences. *Journal of Experimental Psychology: General*, *142*, 954-966. https://doi.org/10.1037/a0030079

van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, *10*, 571-596.

Wang, S. S., Treat, T. A., & Brownell, K. D. (2008). Cognitive processing about classroom-relevant contexts: Teachers' attention to and utilization of girls' body size, ethnicity, attractiveness, and facial affect. *Journal of Educational Psychology*, *100*, 473–489. https://doi.org/10.1037/0022- 0663.100.2.473

Witherby, A. E., & Tauber, S. K. (2017). The concreteness effect on judgments of learning: Evaluating the contributions of fluency and beliefs. *Memory & Cognition*, *45*, 639-650. https://doi.org/10.3758/s13421-016-0681-0

Table 1

Research Design and Procedure

<table>
<tr><th>Case scenario with between-subjects manipulation</th><th>Case description 1</th><th>Case description 2</th><th>Case description 3</th><th>Case description 4</th><th>Additional data</th></tr>
<tr><td>Fictitious student introduced as having high prior knowledge (n = 22)</td><td rowspan="2">Conceptual start problem, solved by student
Followed by
2 procedural problems
2 conceptual problems</td><td rowspan="2">Conceptual start problem, not solved by student
Followed by
2 procedural problems 2 conceptual problems</td><td rowspan="2">Procedural start problem, solved by student
Followed by
2 procedural problems
2 conceptual problems</td><td rowspan="2">Procedural start problem, not solved by student
Followed by
2 procedural problems
2 conceptual problems</td><td rowspan="2">Participants…
- Rate problem difficulty
- Provide demogra-phical data</td></tr>
<tr><td>Fictitious students introduced as having low prior knowledge (n = 20)</td></tr>
<tr><td></td><td colspan="4">For each additional problem: Participants rate how likely the fictitious student would solve the problem and justify their rating</td><td></td></tr>
</table>

Note. The first column shows the between-subjects conditions, the participants were assigned to in the beginning of the study. The following four columns show how we used four case descriptions to vary our three within-subjects factors. The last column shows the additional instruments we used. All participants were first assigned to one of the between-subjects conditions, then responded to all case descriptions (in counterbalanced order) and then to all additional instruments.

Table 2

Expected Data Patterns for Each of the Hypotheses on Preservice Mathematics Teachers' Conceptions About the Relationships Between Students' Conceptual Knowledge and Procedural Knowledge

	Hypothesis	Conc+Conc	Conc+Proc	Conc-Conc	Conc-Proc	Proc+Proc	Proc+Conc	Proc-Proc	Proc-Conc
Likelihood ratings in %	Unrelated	80	55	20	45	80	55	20	45
	Bi-dir. related	80	80	20	20	80	80	20	20
	bi-dir. related but not identical	80	67,5	20	32,5	80	67,5	20	32,5
	Conceptual knowledge bias	80	80	20	20	80	55	20	45
	Procedural knowledge bias	80	55	20	45	80	80	20	20
Arguments expressing	Unrelated	-	+	-	+	-	+	-	+
	Bi-dir. related	-	-	-	-	-	-	-	-

indepen-dence of problems	Conceptual knowledge bias	-	-	-	-	-	+	-	+
	Procedural knowledge bias	-	+	-	+	-	-	-	-

Notes. Conc+Conc = conceptual start problem, correctly solved by the student, additional conceptual problems; Conc+Proc = conceptual start problem, correctly solved by the student, additional procedural problems; Conc-Conc = conceptual start problem, not correctly solved by the student, additional conceptual problems; Conc-Proc = conceptual start problem, not correctly solved by the student, additional procedural problems; Proc+Proc = procedural start problem, correctly solved by the student, additional procedural problems; Proc+Conc = procedural start problem, correctly solved by the student, additional conceptual problems; Proc-Proc = procedural start problem, not correctly solved by the student, additional procedural problems; Proc-Conc = procedural start problem, not correctly solved by the student, additional conceptual problems. Numbers indicate likelihood ratings in accordance with each of the theoretically expected relationships. + = argument that expresses independence of problems present in explanation for likelihood rating; - = argument that expresses independence of problems *not* present in explanation for likelihood rating.

Table 3

Participants' Rankings of Problem Difficulty

Problem	Median Rank	Range	Test of difference between ratings for corresponding problems
Conceptual start problem	2	1-6	
Procedural start problem	2	1-4	Z = -1.35, p > .05, r = -.21 (small effect)
Easy additional conceptual problem	3	1-6	
Easy additional procedural problem	3	1-5	Z = -0.18, p > .05, r = -.03 (small effect)
More difficult additional conceptual problem	6	2-6	
More difficult additional procedural problem	5	3-6	Z = -1.94, p > .05, r = -.31 (medium effect)

Figure 1. Conceptual and procedural problems used as stimuli in the study. The start problems were presented to participants as either correctly solved or not correctly solved by a fictitious student. The participants were then ask to judge how well this respective student would do on each of the four additional problems.

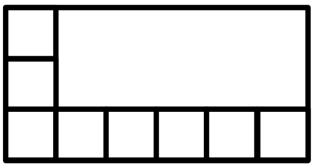
Figure 2. The bars show preservice teachers' average ratings of how likely fictitious students would be able to solve additional conceptual problems and additional procedural problems when these students had or had not correctly solved a conceptual start problem or a procedural start problem. Ratings are averaged over easier and more difficult additional tasks. Error bars represent standard deviations of average ratings. The line represents the number of arguments expressing independence of problems in the preservice teachers' justifications of their ratings. Abbreviations translate as follows: Conc+Conc = conceptual start problem, correctly solved by the student, additional conceptual problems; Conc+Proc = conceptual start problem, correctly solved by the student, additional procedural problems; Conc-Conc = conceptual start problem, not correctly solved by the student, additional conceptual problems; Conc-Proc = conceptual start problem, not correctly solved by the student, additional procedural problems; Proc+Proc = procedural start problem, correctly solved by the student, additional procedural problems; Proc+Conc = procedural start problem, correctly solved by the student, additional conceptual problems; Proc-Proc = procedural start problem, not correctly solved by the student, additional procedural problems; Proc-Conc = procedural start problem, not correctly solved by the student, additional conceptual problems.

Figure 3. Preservice teachers' average ratings of how likely fictitious students would be able to solve additional conceptual problems and additional procedural problems when these students had or had not correctly solved a conceptual start problem or a procedural start problem. Ratings are averaged over easier and more difficult additional tasks. They are shown separately for participants, who were informed that the fictitious students had low prior knowledge and for participants, who were informed that the fictitious students had high prior knowledge. Error bars represent standard deviations of average ratings.

Abbreviations translate as follows: Conc+Conc = conceptual start problem, correctly solved by the student, additional conceptual problems; Conc+Proc = conceptual start problem, correctly solved by the student, additional procedural problems; Conc-Conc = conceptual start problem, not correctly solved by the student, additional conceptual problems; Conc-Proc = conceptual start problem, not correctly solved by the student, additional procedural problems; Proc+Proc = procedural start problem, correctly solved by the student, additional procedural problems; Proc+Conc = procedural start problem, correctly solved by the student, additional conceptual problems; Proc-Proc = = procedural start problem, not correctly solved by the student, additional procedural problems; Proc-Conc = procedural start problem, not correctly solved by the student, additional conceptual problems.

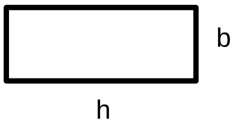
Conceptual start problem

State a formula to calculate the area of the figure below in the unit „little squares“.



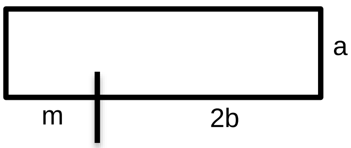
Easy additional conceptual problem

State a formula to calculate the area of the figure below.



More difficult additional conceptual problem

State a formula to calculate the area of the figure below.



Procedural start problem

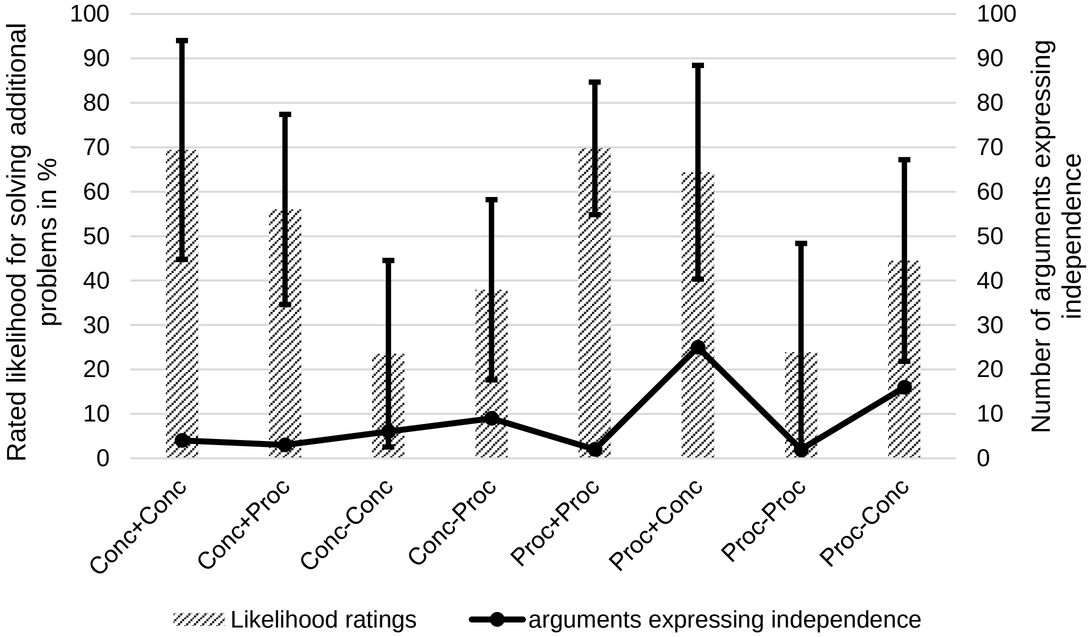
Calculate A = 3m · 5,5m

Easy additional procedural problem

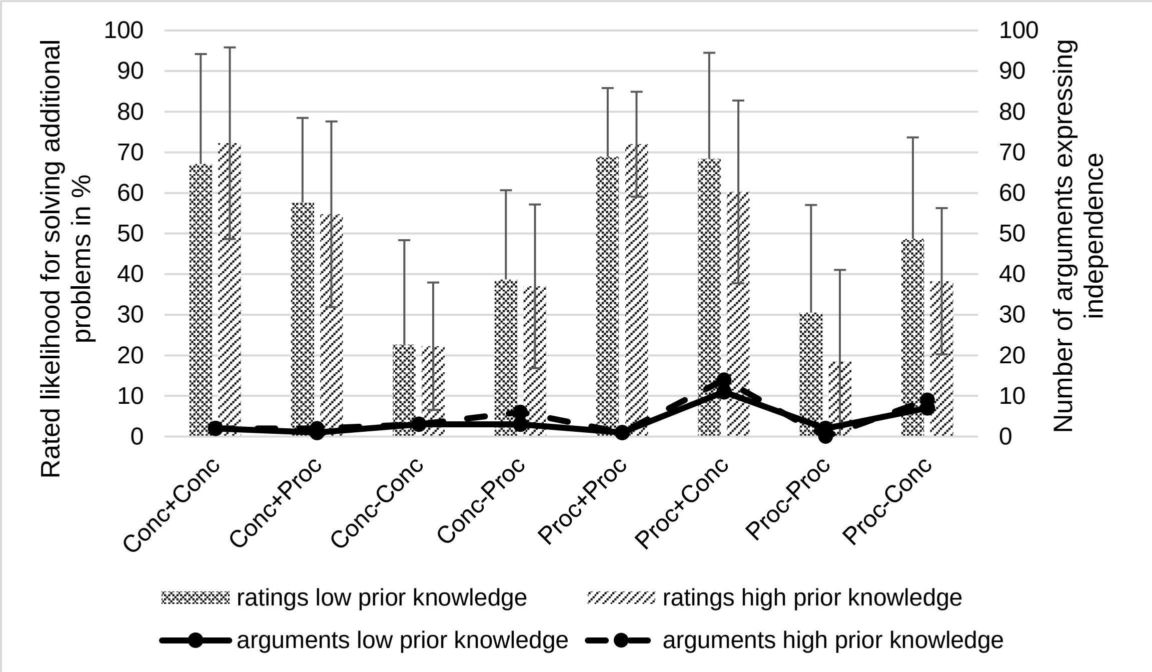
Calculate $A = \frac{3}{4}\,m \cdot \frac{1}{2}\,m$

More difficult additional procedural problem

Calculate $A = x \cdot y$, where $x = \frac{3}{2}$ m and y is three times x.



Rated likelihood for solving additional problems in %
Number of arguments expressing independence
100
90
80
70
60
50
40
30
20
10
0
Conc+Conc
Conc+Proc
Conc-Conc
Conc-Proc
Proc+Proc
Proc+Conc
Proc-Proc
Proc-Conc
Likelihood ratings
arguments expressing independence



Rated likelihood for solving additional problems in %
Number of arguments expressing independence
100
90
80
70
60
50
40
30
20
10
0
Conc+Conc
Conc+Proc
Conc-Conc
Conc-Proc
Proc+Proc
Proc+Conc
Proc-Proc
Proc-Conc
ratings low prior knowledge
ratings high prior knowledge
arguments low prior knowledge
arguments high prior knowledge