# Delaying Instruction Alone Doesn't Work: Comparing and Contrasting Student Solutions is Necessary for Learning from Problem-Solving prior to Instruction

Katharina Loibl, Nikol Rummel, Institute of Educational Research, Ruhr-Universität Bochum
Email: katharina.loibl@rub.de, nikol.rummel@rub.de

**Abstract:** Recent studies have shown benefits of problem-solving prior to instruction. However, it is unclear whether these benefits are based on the cognitive processes related to the problem-solving activity prior to instruction or originate from comparing and contrasting students' solutions to the canonical solution during subsequent instruction. To separate these effects, we conducted a quasi-experimental study with 240 students varying the two factors *timing of instruction* (problem-solving prior to instruction versus instruction prior to problem-solving) and *form of instruction* (standard instruction versus instruction that compares and contrasts typical student solutions). Our results indicate that comparing and contrasting typical student solutions is a prerequisite for the effectiveness of problem-solving prior to instruction. Problem-solving prior to instruction combined with instruction where student solutions were compared and contrasted to the canonical solution outperformed all other conditions. Problem-solving prior to standard instruction was no more effective than standard instruction prior to problem-solving.

## Introduction

Can learning be best promoted by providing or by withholding instructional support? This so-called assistance dilemma (Kapur & Rummel, 2009; Koedinger & Aleven, 2007) targets the question of how to best balance the amount and timing of the instructional support given to learners. The potential benefits of delaying instruction have been shown in recent studies (Kapur, 2010, 2012; Kapur & Bielaczyc, 2012; Roll, Aleven, & Koedinger, 2009, 2011; Schwartz & Martin, 2004). In these studies students who first solved problems to a yet unknown concept before receiving instruction outperformed students who received direct instruction (i.e. instruction without previous problem-solving). It seems that solving problems which require the application of a yet unknown concept prepares students for understanding the concept in the subsequent instruction (Schwartz & Martin, 2004). It has been argued that problem-solving prior to instruction allows students to activate their prior knowledge about the domain (e.g. Kapur & Bielaczyc, 2012; Schoenfeld, 1992).

In most of the studies cited above, students worked in small groups. When solving problems prior to instruction collaboratively, students can co-construct a shared understanding that goes beyond the understanding of each individual (Moschkovich, 1996). Indeed, Sears (2006) could show that students who collaboratively solved problems prior to instruction engaged in knowledge-sharing behavior. Furthermore, they outperformed students who solved problems individually on transfer problems. The potential benefits of collaboration during problem-solving relate to the general finding that collaborative learning can promote deeper elaboration (Teasley, 1995). However, during collaborative problem-solving prior to instruction students usually invent non-canonical and incomplete solutions (Kapur & Bielaczyc, 2012). Therefore carefully designed subsequent instruction is needed to lead students towards the canonical solution.

Most research on problem-solving prior to instruction has focused on designing the problem-solving phase (e.g. with or without collaboration, Sears, 2006; with or without support, Roll, Holmes, Day, & Bonn, 2012; Westermann & Rummel, 2012a, 2012b), while the instruction phase has received less attention (Collins, 2012). As students usually fail to invent the canonical solution themselves during collaborative problem-solving, instruction is needed to ensure that students learn the correct solution method in the end. How the *form of instruction* contributes to the effectiveness of collaborative problem-solving prior to instruction has not been investigated so far. Upon closer inspection of the instruction provided in the studies by Kapur (e.g. 2010, 2012), it becomes apparent that the form of instruction might indeed be a relevant aspect: In the instruction prior to problem-solving control condition (called Direct Instruction, DI) the teacher directly presented the canonical solution (with or without explaining the structural relevant features of the formula, see Kapur & Bielaczyc, 2011). In the problem-solving prior to instruction condition (called Productive Failure, PF) the teacher compared typical student-generated solutions and contrasted them to the canonical solution during instruction in a classroom discussion. Thus, when comparing instruction prior to collaborative problem-solving and collaborative problem-solving prior to instruction, the two variables *timing of instruction* and *form of instruction* were confounded. There is reason to believe that the confounded variable (i.e. the form of instruction) is relevant for the results of problem-solving prior to instruction: When problem-solving prior to instruction is compared to an augmented instruction prior to problem-solving condition where the teacher explains the structural relevant features of the canonical solution (called Strong-DI condition), the learning differences between problem-solving prior to instruction and instruction prior to problem-solving conditions are reduced (Kapur & Bielaczyc, 2011). One might infer from this result that the form of instruction may play a crucial role to explain the beneficial effect of problem-solving prior to instruction. Although this study presents an attempt to align the instruction in both conditions, the instruction in the augmented instruction prior to problem-solving condition did not build on student-generated (i.e. erroneous) solutions. Thus, the two variables timing of instruction and form of instruction were still confounded.

We argue that comparing non-canonical student solutions to the canonical solution during instruction may help students to detect differences between their own prior ideas and the canonical solution. This process of detecting differences by comparisons is analogous to learning with contrasting cases that fosters students to distinguish between cases (e.g. Rittle-Johnson & Star, 2011; Schwartz & Martin, 2004). Detecting differences between cases or solution approaches can guide students' attention to the structural relevant features of the new content (on the effectiveness of comparing erroneous and correct examples see Durkin & Rittle-Johnson, 2012; Große & Renkl, 2007). Against this background, a classroom discussion about typical erroneous solutions may also be fruitful in instruction prior to problem-solving conditions: In such a classroom discussion the teacher can meet students at their level of knowledge and understanding (for the importance of meeting students at their level of understanding see Wittwer & Renkl, 2008) and make discrepancies between the canonical solution and possible erroneous ideas explicit (Smith, diSessa, & Roschelle, 1994). Research demonstrated that students process the canonical solution more deeply when they realize impasses and errors (van Lehn, Silver, Murray, Yamauchi, & Baggett, 2003) and that the realization of an impasse can be triggered by warning about possible errors before presenting the instructional explanation (Acuña, García-Rodicio, & Sánchez, 2010; Sánchez, García-Rodicio & Acuña, 2009).

Taking these findings together, it seems an important next step to investigate the role of taking up typical student-generated (i.e. non-canonical) solutions during instruction in problem-solving prior to instruction settings. The studies cited above indicate that students activate their prior knowledge during problem-solving which prepares them for subsequent instruction (e.g. Kapur & Bielaczyc, 2012; Schwartz & Martin, 2004). We argue that in addition to the cognitive processes related to the problem-solving activity, the form of instruction merits attention: Comparing student solutions and contrasting them to the canonical solution during instruction might be a necessary component for the effectiveness of problem-solving prior to instruction. Activating prior knowledge during problem-solving can only be effective, if students connect their prior knowledge to the new content and realize differences. Contrasting student solutions to the canonical solution helps students to connect their prior ideas to the new content and to focus on the distinguishing features, which in turn may foster the acquisition of conceptual knowledge. By contrast, problem-solving prior to instruction might be less productive for fostering procedural skills (Sleeman, Kelly, Martinak, Ward, & Moore, 1989) as it might reduce the time needed for acquiring procedural skills through practice (Klahr & Nigam, 2004; Rittle-Johnson, Siegler, & Alibali, 2001).

Against this background we hypothesized that collaborative problem-solving prior to instruction combined with subsequent instruction where student solutions are contrasted to the canonical solution is most effective to acquire conceptual knowledge, but instruction prior to collaborative problem-solving may lead to better procedural skills.

## Methods

### Study Design

Table 1: Experiment design with final sample size

| | | Form of instruction | |
|---|---|---|---|
| | | Standard instruction | Instruction that compares and contrasts typical student solutions |
| Timing of instruction | Problem-solving prior to instruction | PS-I ($N$ = 51, 3 classes) | PS-$I_{contrast}$ ($N$ = 56, 3 classes) |
| | Instruction prior to problem-solving | I-PS ($N$ = 62, 3 classes) | $I_{contrast}$-PS ($N$ = 71, 4 classes) |

To separate the effects of the sequence of problem-solving and instruction, and of comparing and contrasting student solutions to the canonical solution during instruction, we conducted a quasi-experimental study with two factors: we varied *timing of instruction* (problem-solving prior to instruction versus instruction prior to problem-solving) and *form of instruction* (standard instruction focusing on the canonical solution versus instruction that compares and contrasts typical student solutions to the canonical solution). Table 1 gives an overview of the conditions. Participants were 240 $10^{th}$ graders (13 classes) recruited from four secondary schools in Germany. For practical reasons, classes were randomly assigned to conditions as a whole. The resulting conditions did not differ significantly concerning prior knowledge as measured by a pretest ($F[3,234] = 0.47$, $p = .71$).

## Learning Material

To be able to compare the results of our study to those of other studies on problem-solving prior to instruction (e.g. Kapur, 2012; Roll et al., 2009; Schwarz & Martin, 2004), our learning material addressed the same concept that has been targeted in those studies: the concept of variance. Students in grade 10 of German secondary schools have not covered this topic yet.

The learning task was aligned to the task used by Kapur (2012) and was the same in all conditions: At the beginning of the first learning phase (i.e. instruction for I-PS and $I_{contrast}$-PS, problem-solving for PS-I and PS-$I_{contrast}$) students were provided with a table listing the number of goals that three fictitious soccer players had scored in the last 10 years. Students were asked to answer the question who the most consistent goal scorer was. Range and mean of the number of goals was the same for all three players to force students to think about strategies beyond their formal prior knowledge.

## Experimental Procedure

Instruction and problem-solving phases respectively took place during a lesson of 45 minutes on two consecutive days. During problem-solving, students worked in groups of three in all conditions. The same experimenter gave the instruction in all conditions. In the instruction phase of all conditions, the experimenter explained the concept and the canonical solution using the example of the goal scorers. Prior to both learning phases (i.e. prior to the first lesson) students completed a pretest on related content (e.g. mean, range, box plot, graphical representations). After both learning phases (i.e. after the second lesson), students completed a posttest.

In the problem-solving prior to instruction conditions (PS-I and PS-$I_{contrast}$), students dealt with the task to identify the most consistent goal scorer during the first lesson. During this problem-solving phase, the task asked them to invent as many solutions as possible. Students used tablet PCs to generate and exchange solution ideas. The use of tablet PCs allowed students to work individually as well as to share their ideas and focus the group's attention on selected ideas. During the second lesson, students received instruction.

In the instruction prior to problem-solving conditions (I-PS and $I_{contrast}$-PS), students first received instruction. The problem-solving phase took place during the second lesson where students solved problems isomorphic to the one discussed during instruction.

In the standard instruction conditions (PL-I and I-PL), the experimenter first presented the problem of the three soccer players and discussed the meaning of consistency with the class. This introduction was followed by a presentation of several approaches (graphical approaches, range, mean absolute deviation, and standard deviation). The class discussed the advantages and disadvantages of the different approaches (e.g. graphical approaches might be imprecise, range is sensitive to outliers). Finally the experimenter explained the structurally relevant features of the canonical solution.

In the conditions with instruction that compares and contrasts typical student solutions (PS-$I_{contrast}$ and $I_{contrast}$-PS), the experimenter presented and compared typical student-generated solutions (e.g. graphical approaches, range, number of times the soccer player scored at the mean, deviation from one year to the next with or without absolute values) and discussed whether these approaches were suitable to solve the problem by contrasting them to the canonical solution. It should be stressed, that the solutions were not the very solutions generated by students during problem-solving in this study. Rather, the solutions were *typical* student-generated solutions (taken from previous studies and pilots) that matched the solution types most often generated. Notably these solution types were similar to the ones usually generated by students in Singaporean classes in previous studies (cf. Kapur, 2012). Finally the experimenter explained the structurally relevant features of the canonical solution.

## Dependent Variables

A posttest assessed the learning outcomes after the second lesson. It included items testing for procedural skills and items testing for conceptual knowledge. Students had 30 minutes to answer the posttest items. All students finished the posttest in time.

The items testing for *procedural skills* required students to solve problems isomorphic to the one discussed during instruction. Students received 1 point for each correct calculation with a deduction of 0.5 point for computation errors. They received 1 additional point in cases where they had to compare two deviations.

Students could achieve a maximum of 4 points (i.e. 1 item required a single calculation, 1 item required the calculation of two deviations including a comparison).

The items testing for *conceptual knowledge* required students to decompose the canonical solution into its structurally relevant features (cf. Roll et al., 2011) and to translate between graphical and algebraic strategies: Two items presented incorrect solutions and asked students to detect the errors and to reason mathematically. For the reasoning, students had to decompose the canonical solution and refer to these structurally relevant features of the canonical solution. Students received 0.5 point for the detection of each error. They received an additional 0.5 point per detected error for correct reasoning about the structurally relevant feature. Figure 1 presents one example. Two other items required sense-making using both graphical representations and the structurally relevant features of the canonical solution. Students received 0.5 point for each structural feature correctly represented in the graphical representation. Taking all conceptual knowledge items together, students could achieve a maximum of 7 points (3 points for the first type of items, 4 points for the second type of items).

One student calculated the consistency the following way.

How did he calculate consistency? Is the method suitable to measure consistency? Explain why or why not.

$$\frac{(x_2-x_1)+(x_3-x_2)+(x_4-x_3)}{N}=\frac{(50-30)+(90-50)+(70-90)}{4}=10$$

Error: Deviation from one value to the next instead of deviation from the mean. *(0.5 point)*

Reasoning: Sensitive for sequence of data points as there is no fixed reference point. *(0.5 point)*

Error: No absolute or squared values; deviations may be negative. *(0.5 point)*

Reasoning: Positive and negative values might cancel out. *(0.5 point)*

Figure 1. Example of one item testing for conceptual knowledge with solution and coding.

In the problem-solving prior to instruction conditions (PS-I and PS-$I_{student}$) students used tablet PCs to invent their solutions during problem-solving. This enabled us to collect audio and screen recordings of students' collaborative problem-solving prior to instruction. We are currently analyzing the process of inventing and discussing solution ideas in the small groups as well as coding the quantity and quality of the invented solution ideas.

## Results

We performed a two-factorial MANOVA with the factors *timing of instruction* and *form of instruction* and the outcome variables procedural skills and conceptual knowledge. Table 2 shows the means and standard deviations.

Table 2: Means and standard deviations of the posttest results.

| Conditions | Procedural skills | Conceptual knowledge |
|---|---|---|
| PS-I | 3.24 (0.99) | 1.29 (1.02) |
| PS-$I_{contrast}$ | 2.99 (1.27) | 2.63 (1.53) |
| I-PS | 3.27 (1.02) | 1.17 (1.23) |
| $I_{contrast}$-PS | 3.41 (0.91) | 1.68 (1.35) |

For *procedural skills*, we found only a marginally significant effect for timing of instruction favoring instruction prior to problem-solving ($F[1,236] = 2.81$, $p = .095$, $\eta_p^2 = .01$). Neither the form of instruction ($F[1,236] = 0.16$, $p = .69$) nor the interaction of timing and form ($F[1,236] = 1.93$, $p = .17$) was significant.

For *conceptual knowledge* we found a small significant effect for timing of instruction favoring problem-solving prior to instruction ($F[1,236] = 10.02$, $p = .002$, $\eta_p^2 = .04$) and a large significant effect for form of instruction favoring instruction based on typical student solutions ($F[1,236] = 29.35$, $p < .01$, $\eta_p^2 = .11$). We further found a significant interaction ($F[1,236] = 5.90$, $p = .02$, $\eta_p^2 = .02$) indicating that the form of instruction has a higher effect in the problem-solving prior to instruction conditions. In order to compare the effects of the two factors and their combination, we additionally calculated posthoc comparisons (LSD) between all conditions. The pair-wise comparisons revealed significant differences between the $I_{contrast}$-PS condition to the I-PS ($p = .02$) condition and to the PS-$I_{contrast}$ condition ($p < .01$), that is, the $I_{contrast}$-PS condition significantly outperformed the condition that received standard instruction first (I-PS), but was outperformed by the PS-$I_{contrast}$ condition that combined problem-solving prior to instruction with instruction where student solutions were compared and contrasted to the canonical solution. The comparison between the I-PS condition and the PS-I condition was not significant ($p = .61$), that is, the timing of instruction had no effect when combined with standard instruction. In other words, for conceptual knowledge, problem-solving prior to instruction was only more effective than instruction prior to problem-solving if student solutions were compared and contrasted during the instruction.

## Discussion

Previous studies have shown benefits of problem-solving prior to instruction for the acquisition of conceptual knowledge. These benefits may stem from the cognitive processes related to the problem-solving activity prior to instruction or they may originate from the specific form of subsequent instruction that compares and contrasts students' solutions to the canonical solution during instruction. In our study we aimed at separating the effects of *timing of instruction* (problem-solving prior to instruction versus instruction prior to problem-solving) and of *form of instruction* (standard instruction focusing on the canonical solution versus instruction that compares and contrasts typical student solutions to the canonical solution). We tested for learning effects on conceptual knowledge and procedural skills.

Our findings support the notion that problem-solving prior to instruction can prepare students for the acquisition of *conceptual knowledge* from subsequent instruction as indicated by the main effect for timing of instruction. In this regard, our study replicates the beneficial effect of problem-solving prior to instruction found by others (e.g. Kapur, 2009, 2012; Roll et al., 2011; Schwartz & Martin, 2004).

Moreover, the form of instruction appears to be of central relevance: Comparing typical non-canonical student solutions and contrasting them to the canonical solution during instruction may guide students' attention to the structurally relevant aspects of the content and thereby promotes learning. As indicated by the main effect of form of instruction and the pair-wise comparisons, comparing and contrasting typical student solutions to the canonical solution is beneficial in both settings: problem-solving prior to instruction and instruction prior to problem-solving. Similar to these results, we already showed in an earlier study (Westermann & Rummel, 2012b) that even in an instruction prior to problem-solving setting it is beneficial for learning if instruction builds on typical student solution in comparison to standard instruction.

The most interesting finding of our study is the interaction effect showing that the beneficial effect of problem-solving prior to instruction only comes to bear if the teacher (or in our study the experimenter) compares typical student solutions and contrasts them to the canonical solution during instruction. More specifically the PS-$I_{contrast}$ condition, that is problem-solving prior to instruction combined with instruction where student solutions are compared and contrasted to the canonical solution, outperformed all other conditions. This finding suggests a dual learning mechanism: In a first step, problem-solving prior to instruction prompts students to activate their prior knowledge and to generate own solution ideas (cf. Kapur & Bielaczyc, 2012). In a second step, comparing student solutions and contrasting them to the canonical solution during instruction helps students to detect differences between their own prior ideas and the canonical solution. The detection of differences guides students' attention to the structurally relevant aspects of the content (cf. Durkin & Rittle-Johnson, 2012). Focusing the attention on the most important aspects in turn helps students to process these aspects deeply and fosters the acquisition of conceptual knowledge (Renkl, 2008).

Furthermore the difference between the $I_{contrast}$-PS condition and the PS-$I_{contrast}$ condition suggests that connecting prior knowledge to the new content and detecting differences between non-canonical solutions and the canonical solution works better when students indeed activated their prior knowledge during problem-solving and generated solutions themselves. In the study cited above (Westermann & Rummel, 2012b), we only found a descriptive, but not statistically significant difference between the two conditions with different timing of instruction (before versus after problem-solving) where instruction build on typical student solutions. How do these two studies differ? First of all, the sample size of our study presented here is higher. Secondly, we conducted the previous study (Westermann & Rummel, 2012b) at two schools from the same well-educated neighborhood. Students from these schools might have been higher motivated in connecting the new content to their prior knowledge and therefore prompting them to activate their prior knowledge first might have been less important. The schools of our current study were located in four different neighborhoods resulting in a more representative sample.

In addition to the learning effects on conceptual knowledge, our findings confirm that time for practicing problem-solving after instruction is needed to foster *procedural skills* (cf. Rittle-Johnson et al., 2001): Both instruction prior to problem-solving conditions (I-PL and $I_{contrast}$-PS) outperformed both problem-solving prior to instruction conditions (PS-I and PS-$I_{contrast}$) on items testing for procedural skills. This finding is not surprising as the latter conditions had no time to practice problem-solving after learning the canonical solution. Studies that found no difference between instruction prior to problem-solving and problem-solving prior to instruction on items testing for procedural skills usually allowed practice for students in the problem-solving prior to instruction conditions after students received the canonical solution during instruction (e.g. Kapur, 2010, 2012; Roll et al., 2009). Taken together, our findings underline the importance of defining the learning goal when choosing one instructional approach over the other.

## Limitations and Outlook

Although our study yields interesting results, we would like discuss some limitations and give an outlook to future research. Inspired by the in vivo research paradigm advocated of the Pittsburgh Science of Learning Center (Koedinger, Corbett, & Perfetti, 2012), we conducted our study in the field with real learners and real learning content, which promotes the external validity of the study. However, this also yielded some problems: The implementation in schools forced us to conduct a quasi-experimental study for organizational reasons. Thus, prior differences between conditions cannot be completely excluded due to randomizing at the class level.

Another aspect that has to be considered is the fact that the experimenter who taught the instruction in our study was very familiar with the material used during instruction, the structurally relevant features of the canonical solution, and the typical student solutions. This knowledge might relate to student achievement (Tchoshanov, 2011). Tchoshanov showed that teacher content knowledge is associated with lesson quality and student achievement in mathematics. Especially when building on student-generated solutions it seems crucial to be familiar with these solutions. In order to ensure a smooth implementation in the field, teachers might need to be provided with new resources and strategies (Meder, Schüpbach, & Krause, 2011) as building on student-generated solutions imposes high demands on the teacher.

When focusing on the effect of connecting the new content to the prior knowledge it should be noted that the solutions used in the instruction phase of the $I_{contrast}$-PS condition and the PS-$I_{contrast}$ condition were *typical* student-generated solutions (taken from previous studies and pilots) that matched the solutions most often generated in the problem-solving prior to instruction conditions and not the very own solution of the students. Yet, until this date, it has not been systematically investigated whether using the very own solutions of students in comparison to typical student-generated solutions during instruction would further help students to connect their prior knowledge to the new content and to detect differences between their intuitive solutions and the canonical solution.

Solution approaches invented prior to instruction are generally incomplete or erroneous (e.g. Kapur & Bielaczyc, 2012). Nevertheless, the diversity, that is the number of different solution ideas, seems to have a positive effect on posttest performance (Kapur, 2012; Kapur & Bielaczyc, 2012). While Kapur and colleagues claim that the positive effect of diversity is independent of the quality of the solution ideas, others did find that the quality of the invented solutions matters (Wiedmann, Leach, Rummel, & Wiley, 2012). In accordance with the finding of Wiedmann and colleagues, we hypothesize that the more knowledge components are shared between the invented solutions and the canonical solution, the easier it should be to connect the prior ideas to the new content during instruction. As indicated by our findings, the connection between prior ideas and the canonical solution may lead to deeper processing and in turn promote learning. We recorded process data of the problem solving prior to instruction conditions (PS-I and PS-$I_{contrast}$) that allow us to code the quantity and quality of the invented solution ideas. For future analysis, we aim at testing for possible relations between these codings and learning outcomes.

## References:

Acuña, S. R., García-Rodicio, H., & Sánchez, E. (2010). Fostering active processing of instructional explanations of learners with high and low prior knowledge. *European Journal of Psychology of education, 26*(4), 435-452.

Collins, A. (2012). What is the most effective way to teach problem solving? A commentary on productive failure as a method of teaching. *Instructional Science, 40*(4), 731-735.

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*(3), 206-214.

Große, C. S. & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction, 17*(6), 612-634.

Kapur, M. (2010). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science*, *39*(4), 561-579.

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*(4), 651-672.

Kapur, M., & Bielaczyc, K. (2011). Classroom-based Experiments in Productive Failure. In L. Carlson, C. Hoelscher, & T.F. Shipley (Eds.) *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2812-2817). Austin, TX: Cognitive Science Society.

Kapur, M., & Bielaczyc, K. (2012). Designing for Productive Failure. *The Journal of the Learning Sciences, 21*(1), 45-83.

Kapur, M., & Rummel, N. (2009). The assistance dilemma in CSCL. In: A. Dimitracopoulou, C. O'Malley, D. Suthers, & P. Reimann (Eds.), *Computer supported collaborative learning practices - CSCL2009 community events proceedings*, Vol 2 (pp. 37–42). Berlin: International Society of the Learning Sciences.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*(10), 661–667

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*(3), 239–264.

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*(5), 757–798.

Meder, L., Schüpbach, H., & Krause, A. (2011). Sind innovative Lehr- und Lernformen für Schüler wie auch für Lehrkräfte vorteilhaft? Internationale Vergleichsstudie zum Zusammenhang von Lehr-/Lernformen, Unterrichtsqualität und psychischen Belastungen der Lehrkräfte [Are innovative teaching and learning methods for teachers as beneficial as for students? International comparative study about the relations between teaching/ learning methods, quality of instruction and psychological stress in classroom teachers]. In Landesstiftung Baden-Württemberg (Ed.), *Programm Bildungsforschung* (pp. 197-223). Stuttgart: Landesstiftung Baden-Württenberg.

Moschkovich, J. N. (1996). Moving up and getting steeper: Negotiating shared descriptions of linear graphs. *Journal of the Learning Sciences, 5*(3), 239-277.

Renkl, A. (2008). Lehren und Lernen im Kontext der Schule [Teaching and learning in school contexts]. In Renkl (Ed.), *Lehrbuch Pädagogische Psychologie* (pp. 109–153). Bern: Huber.

Rittle-Johnson, B., Siegler, R., & Alibali, M. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*(2), 346-362.

Rittle-Johnson, B., & Star, J. R. (2011). The power of comparison in learning and instruction: Learning outcomes supported by different types of comparisons. In B. Ross & J. Mestre (Eds.), *Psychology of Learning and Motivation: Cognition in Education* (Vol. 55, pp. 199-226). San Diego: Elsevier.

Roll, I., Aleven, V., & Koedinger, K. R. (2009). Helping students know 'further' - increasing the flexibility of students' knowledge using symbolic invention tasks. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 1169-1174). Austin, TX: Cognitive Science Society.

Roll, I., Aleven, V., & Koedinger, K. R. (2011). Outcomes and mechanisms of transfer in invention activities. In L. Carlson, C. Hoelscher, & T.F. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 2824-2829). Boston, Massachusetts: Cognitive Science Society.

Roll, I., Holmes, N., Day, J. & Bonn, D. (2012). Evaluating metacognitive scaffolding in Guided Invention Activities. *Instructional Science, 40*(4), 691-710.

Sánchez, E., García Rodicio, H., & Acuña, S. R. (2009). Are instructional explanations more effective in the context of an impasse? *Instructional Science, 37*(6), 537–563.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for Research on Mathematics Teaching and Learning* (pp. 334-370). New York: MacMillan

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

Sears, D. A. (2006). *Effects of innovation versus efficiency tasks on collaboration and learning* (Doctoral dissertation, Stanford University, California). Retrieved November 07, 2012, from http://www.stat.auckland.ac.nz/~iase/publications/dissertations/06.Sears.Dissertation.pdf

Sleeman, D., Kelly, A. E., Martinak, R., Ward, R. D., & Moore, J. L. (1989). Studies of diagnosis and remediation with high school algebra students. *Cognitive Science, 13*, 551-568.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences, 3*(2), 115-163.

Tchoshanov, M. A. (2011). Relationship between teacher knowledge of concepts and connections, teaching practice, and student achievement in middle grades mathematics. *Educational Studies in Mathematics, 76*(2), 141-164.

Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology, 31*(2), 207-220..

van Lehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction, 21*(3), 209–249.

Westermann, K. & Rummel, N. (2012a). Delaying instruction: evidence from a study in a university relearning setting. *Instructional Science, 40*(4), 673-689.

Westermann, K. & Rummel, N. (2012b). New evidence on productive failure - Building on students' prior knowledge is key! In J. van Aalst, K. Thompson, M. J. Jacobson & P. Reimann (Eds.), *The future of learning: Proceedings of the 10th international conference of the learning sciences (ICLS 2012) – Volume 2, Short Papers* (pp. 266-270). Sydney, Australia: ISLS.

Wiedmann, M., Leach, R. C., Rummel, N. & Wiley, J. (2012). Does group composition affect learning by invention? *Instructional Science, 40*(4), 711-730.

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

## Acknowledgements

This research was funded by the Center of Educational Studies of the Professional School of Education at the Ruhr-Universität Bochum, Germany. We would like to especially thank Manu Kapur for providing us with his study materials and many background information concerning his own studies. We are thankful for the participating schools for the organizational efforts. We thank our student research assistants, Katja Goepel, Christian Hartmann, and Andreas Vogel for their help with collecting and coding the data.