Learning Environ Res (2009) 12:209–223
DOI 10.1007/s10984-009-9063-x

ORIGINAL PAPER

Effects of reflection prompts on learning outcomes and learning behaviour in statistics education

Robin Stark · Ulrike-Marie Krause

Received: 26 September 2007 / Accepted: 2 September 2008 / Published online: 2 September 2009

Abstract Starting from difficulties that students display when they deal with correlation analysis, an e-learning environment ('Koralle') was developed. The design was inspired by principles of situated and example-based learning. In order to facilitate reflective processes and thus enhance learning outcomes, reflection prompts were integrated into the learning environment. A total of 57 university students were randomly assigned to two experimental conditions: 28 students were prompted to give reasons for their decisions while working within the learning environment (EG 1); and 29 students dealt with Koralle without being prompted (EG 2). The control group consisted of 67 students who had already attended regular statistics lectures but had no access to the e-learning environment. EG 1 scored significantly higher in the posttest than EG 2, and the effect was practically relevant and sustainable. Reflection prompts did not influence time on task, task choices and motivational outcomes. Both experimental groups clearly outperformed the control group.

Keywords Example-based learning · Learning behaviour · Metacognition · Reflection prompts · Situated learning · Statistics education

Introduction

University training in empirical research methods and statistics in the social sciences often results in failures. Even after several statistics courses, many students display substantial misconceptions and knowledge gaps, and considerable transfer problems are common (Broers 2002; Castro Sotos et al. 2007; Tyroller 2005).

Especially in statistics, many students need individual support; often, however, high student numbers make individual tutoring difficult. To solve this problem, e-learning is

R. Stark · U.-M. Krause (✉)
Institute of Education, Saarland University, P.O. Box 151150, 66041 Saarbrücken, Germany
e-mail: u.krause@mx.uni-saarland.de

R. Stark
e-mail: r.stark@mx.uni-saarland.de

increasingly used in statistics training. In the present study, we used the e-learning approach in order to enhance students' understanding of a basic statistical concept, namely, correlation (Morris 2001). For this purpose, the situated, example-based e-learning environment 'Koralle' was developed as a supplement to regular statistics lectures.

Findings from earlier studies on example-based learning in the domain of research methods (e.g. Stark and Mandl 2002, 2005) encouraged us to integrate a prompting procedure into example-based learning. This procedure is meant to facilitate careful and mindful consideration of the learning material (for the concept of mindfulness, see Langer 1993; Salomon and Globerson 1987), especially metacognitive reflection. The main goal of this study was to examine the effectiveness of the prompting procedure.

Relevant learning prerequisites

Earlier studies showed that studying statistics in the social sciences is often hampered by students' low prior knowledge and motivation (Stark and Mandl 2000). In the present study, cognitive, metacognitive and motivational learning prerequisites were included into the analyses.

In many studies, the best predictor for effective learning was prior knowledge (Dochy 1992; Krause and Stark 2006), which influences information selection and processing and, as a consequence, learning outcomes (Alexander 1996). For example, in order to gain an in-depth understanding of a text, students must construct adequate situation models that integrate prior knowledge and new information (Kintsch 1994).

Students showing difficulties in gaining new knowledge in statistics were, to some extent, lacking metacognitive abilities, which resulted in problems with self-regulation (Stark and Mandl 2002, 2005; for an overview see Azevedo 2005). Metacognition refers to both knowledge and skills (Flavell 1976, 1979). Metacognitive knowledge can be differentiated into (1) declarative knowledge about the processes, strengths and weaknesses of one's own cognitive functions and (2) knowledge about one's own knowledge (Hasselhorn 2006). Metacognitive skills refer to procedural knowledge for regulating and controlling one's own problem-solving and learning activities (Brown and DeLoache 1978; Veenman et al. 2006). In a first pilot study with Koralle (Tyroller 2005), many students lacked metacognitive skills. Their performance resembled the behaviour of weak performers in a study by Chi et al. (1989). They rarely showed activities of monitoring, self-diagnosing and self-regulation. Choices about task structure and redundancy were made rather mindlessly. That is, the students did not show successful metacognitive processing (e.g. Schneider and Pressley 1989).

From a pedagogical perspective, this is problematic, as there is plenty of evidence showing that metacognition is important for effective self-regulation and learning (Bannert 2003; Berthold et al. 2007; Gerjets et al. 2005; Lan et al. 1993).

Concerning motivational prerequisites, we focused on topic interest. Interest is a specific relation between person and object (Krapp 2005) and results in positive evaluations of and positive emotions in the learning situation. Interest determines choices in the learning process (Krapp 2000) and is related to learning success (e.g. Hidi 1990; Schiefele et al. 1993).

Design of the learning environment

Instructional approach

Koralle's design was inspired by principles of example-based and situated learning (Stark 2004). The prompting procedure was integrated into the situated, example-based approach.

Worked-out examples and problem-solving tasks

Learning by worked-out examples has proved an efficient method for learning in well-structured domains (e.g. Atkinson et al. 2000; Gerjets et al. 2004). In initial skill acquisition, examples are superior to direct teaching of abstract principles and to training by problem solving (Gerjets et al. 2004). The *worked-example effect* is attributed to the fact that studying worked examples produces low cognitive load (Sweller 1988) as no extensive search processes (*extraneous load*) are involved (Gerjets et al. 2004). Therefore, cognitive resources can be used for schema construction (*germane load*; e.g. Sweller et al. 1998). Moreover, worked examples draw the student's attention to information that is relevant for schema construction. In various domains, schema-based problem solving is considered to be very effective and efficient. It has proved to be a central characteristic of experts' problem solving (Gerjets et al. 2004, p. 34).

However, studies of the *self-explanation effect* (Chi et al. 1989, 1994) clearly show that studying examples does not *automatically* result in effective problem-solving schemas. In order to achieve this goal, processes of cognitive and metacognitive self-explanation have to be promoted systematically by additional instruction (Atkinson et al. 2003; Große and Renkl 2006; Stark 1999).

Superficial example elaboration (Renkl 1997) can be compensated by providing incomplete examples (Paas 1992; Stark 1999; van Merriënboer 1990; van Merriënboer and de Croock 1992) or by systematically combining examples with problem-solving tasks (Renkl et al. 2004; Stark et al. 2000). This combined learning approach was used in the e-learning environment Koralle.

In order to support self-regulation and autonomy, and therefore increase motivation (Deci and Ryan 2000), students were instructed to individually determine task structure and redundancy. For each problem, they could either choose to solve the problem first and afterwards study the worked-out example or study the example immediately. Besides, they could decide to solve an additional problem and thus intensify their exercise. This example-based approach was enriched by the following situated design principles.

Situated design principles

All tasks were embedded in problems that were authentic and relevant for the students (Cognition and Technology Group at Vanderbilt 1997). All data needed for problem solving were integrated into the task. Additionally, the problems were presented from multiple perspectives (Große and Renkl 2006; Spiro et al. 1991).

Furthermore, we integrated instructional aids that mainly addressed students with little prior knowledge: a detailed glossary, visual aids and guiding questions that facilitated problem solving. The guiding questions were faded out in the learning process (Collins et al. 1989; Renkl et al. 2004).

In our study, the situated, example-based approach was combined with reflection prompts.

Facilitating reflection by prompting

According to Chi (1996), thinking and learning processes can be supported by scaffolding or prompting. *Scaffolding* includes content-specific information that is necessary for a specific problem. Scaffolding could be, for example: “In order to attain goal C, goals A and B must be pursued first.” The guiding questions within the learning environment Koralle (e.g. “How strong is the correlation?”) are typical examples of scaffolding. In contrast, *prompting* is not linked to expertise in the respective field. According to Berthold et al. (2007, p. 566), prompts are “questions or hints that induce productive learning processes”. They are *strategy activators* in the sense of Reigeluth and Stein (1983), designed to overcome superficial processing. Prompts could be, for example: “What is the next step?”, “What does that mean?” and “How else could you say it?”

The effectiveness of prompting reflective processes was demonstrated in different domains, for instance in mathematical problem solving (Schoenfeld 1985). In an experimental study, prompting students to justify (and hence reflect on) their learning behaviour proved effective in encouraging knowledge application, especially with complex tasks. This kind of support was also effective in computer-based learning (Bannert 2003; Krause and Stark 2007). In a study by Berthold et al. (2007), cognitive and metacognitive learning strategies and learning outcomes were enhanced by cognitive or mixed (cognitive and metacognitive) prompts.

Instructional interventions that successfully prompt reflective activities, such as metacognitive control and other metacognitive and cognitive strategies, should be simple and neatly woven into the learning process in order not to increase extraneous load (e.g. Sweller et al. 1998) by interfering with the students’ spontaneous efforts (Bannert 2003; Hasselhorn and Hager 1998).

In our study, students were prompted to give reasons for their decisions. This was meant to encourage mindfulness and deeper processing and, therefore, enhance learning outcomes.

Evaluation perspective

The effectiveness of the prompting procedure had to be investigated by comparing the learning environment with and without prompting. Moreover, both experimental groups had to be compared with a control group in order to increase the explanatory power of the study.

Learning outcomes were conceptualised as knowledge about correlation that can be successfully applied to authentic problems of empirical research. Apart from learning outcomes, motivational effects are very important in the present context, given the lack of interest in statistics that many students display. Following the expectancy-value model (McClelland et al. 1953), motivation is often subdivided into expectancy and value aspects. Students are most likely to try hard when they feel competent and expect to be successful (Marsh and Craven 2006) and when they consider the respective activity to be worthwhile (e.g. Krapp 2005). As a consequence, the students’ subjective learning outcomes and their acceptance of the learning environment were investigated.

In order to improve the interpretability of the study, the outcome perspective was combined with a process perspective that focused on time on task, the way in which the learners deal with different options in the e-learning environment, and the reasons that the prompted students give for their decisions.

Research questions and hypotheses

The first research question was: "Does prompting students to justify their decisions during the learning process enhance learning? Are potential effects sustainable?" We hypothesised that students who are prompted to give reasons for their decisions outperform their peers who are not prompted both on the posttest and on the follow-up test. Moreover, students working with Koralle (either with or without prompting) are likely to be more successful on the posttest than a control group of students who merely attended regular lectures on the subject.

The second research question was: "To what extent does the prompting procedure influence learning behaviour?" We expected that the prompting procedure would lead to increased time on task. However, as this procedure was supposed to enhance the quality of learning in terms of higher mindfulness, differences in performance are likely to persist when time on task is controlled for statistically. Concerning the students' decision behaviour, no specific effects were hypothesised.

The third research question was: "What reasons do students in the prompting group give for their decisions in the learning process?" We supposed that the majority of reasons indicate metacognitive control. However, indicators of other metacognitive strategies, as well as cognitive strategies (see Berthold et al. 2007), were expected as well. Against the background of findings by Allwood (1990), only weak to medium correlations between justification categories and learning outcomes were expected.

The fourth research question was: "Does the prompting procedure influence acceptance of the learning environment and subjective learning outcomes?" We presumed that the prompting procedure improves the students' feeling of self-regulation and autonomy and therefore promotes acceptance of the learning environment. Furthermore, as this procedure was meant to increase the learners' mindfulness, positive effects on subjective learning outcomes were expected.

Method

Subjects and design

The sample consisted of 124 students of the University of Munich, most of whom studied Education. All participants had already attended regular statistics lectures. Fifty-seven students were randomly assigned to two experimental conditions: EG 1 with prompting ($n = 28$ with 21 females, mean age = 23.89 years, $SD = 5.51$) and EG 2 without prompting ($n = 29$ with 21 females, mean age = 23.64 years, $SD = 3.63$). These students completed Koralle in the e-learning computer laboratory. The control group (67 students with 48 females, mean age = 24.22 years, $SD = 4.58$) had no access to this learning environment.

In terms of prior knowledge, the last grade received in mathematics, and topic interest, the two experimental groups and the control group were all comparable (prior knowledge: $F[2, 110] < 1$; grade in mathematics: $F[2, 110] = 1.28$, MSE = 1.62, $p > 0.20$; interest: $F[2, 110] = 1.28$, MSE = 0.13, $p > 0.20$). This was also true for the subgroup of students participating in the follow-up test that was administered in a statistics lecture ($n = 34$ with 23 students not participating; prior knowledge: $F[1, 31] = 1.52$, MSE = 2.24, $p > 0.20$; grade in mathematics: $F[1, 31] = 1.05$, MSE = 1.43, $p > 0.20$; interest: $F[1, 31] < 1$). This evidence shows that randomisation was successful. All subjects participated voluntarily; students of the experimental groups received 20 Euros.

Koralle with and without prompting

The central subject of Koralle was the influence of linearity, heterogeneous subgroups and outliers on the correlation coefficient. For each thematic unit, a pair of superficially similar but structurally different tasks was provided; in addition, an analogous pair of tasks was optionally available for each unit. Altogether, six tasks had to be worked on by all students, and six tasks were optional. Furthermore, students had to decide whether they wanted to deal with the problems by working on problem-solving tasks and afterwards receiving feedback in the form of worked-out examples, or whether they preferred to view the examples immediately.

In the prompting group, learners had to type short justifications for each decision into a text-entry field ("Please justify your decision"). To ensure authenticity and to prevent instruction from interfering too much with normal cognitive processing, students were asked for justification at the moment when they decided.

In order to increase ecological validity, time on task was not strictly controlled. Students were merely instructed to remain within a time frame of three hours, which was determined by means of a pilot study.

Instruments and data collection

All instruments used in this study were tested beforehand in a regular statistics lecture. Rating scale responses ranged from 1 (Strongly Disagree) to 6 (Strongly Agree).

Prior knowledge

Prior knowledge was assessed with a pretest that consisted of three tasks (theoretical maximum: 12; Cronbach's $\alpha = 0.71$). Task 1 required declarative knowledge about correlation (i.e. conceptual knowledge in the sense of de Jong and Ferguson-Hessler 1996). Task 2 tested whether conceptual knowledge about correlation could be adequately transformed into a scatter plot. Task 3 was more action oriented; the learners had to deal with the problem of outliers. To solve this problem, knowledge about application of concepts and procedures (procedural knowledge) and knowledge about application *conditions* (conditional knowledge in the sense of Paris et al. 1983) were needed.

Metacognitive prerequisites

Metacognitive prerequisites were recorded using 10 items that referred to different reflective activities (Cronbach's $\alpha = 0.65$; "When I am studying, I occasionally ask myself whether I understand subject matter"). The development of this scale was inspired by established instruments that include metacognitive aspects (e.g. Motivated Strategies for Learning Questionnaire, Pintrich et al. 1993; Learning and Study Strategies Inventory, Weinstein et al. 1987).

Topic interest

Interest in correlation was assessed by a rating scale that consisted of 11 items (Cronbach's $\alpha = 0.87$; "I am interested in correlation").

Cognitive learning outcomes

The posttest consisted of five tasks. Tasks 1 and 3 on the posttest were identical with Tasks 1 and 2 on the pretest. These two tasks (Cronbach's $\alpha = 0.68$; theoretical maximum: 7) were used as indicators of learning progress. Task 2 on the posttest demanded declarative, procedural and conditional knowledge: an unexpected correlation coefficient was presented in the context of an authentic empirical study, and the students had to identify possible reasons for this result. Task 4 was rather similar to Task 3 in the pretest, but it additionally required transfer. Task 5 was a complex transfer task in which declarative, procedural and conditional knowledge had to be applied to a statistical situation that had not directly been addressed in the e-learning environment (moderation problem, three heterogeneous subgroups); for this task, deeper understanding of the acquired concepts and procedures was necessary. The theoretical maximum score for Tasks 2, 4 and 5 was 13 points (Cronbach's $\alpha = 0.68$). In order to test the stability of potential effects of reflection prompting, the posttest was administered again about 4 weeks after the experiment (follow-up measure).

Motivational learning outcomes

Students' acceptance of the learning environment was assessed by using five items (Cronbach's $\alpha = 0.74$; "I would recommend the learning environment to my fellow students"). Subjective learning outcomes were recorded by four items (Cronbach's $\alpha = 0.75$; "By the learning environment I have learned how to interpret correlation results").

Learning behaviour

Time on task and student choices (frequency of active problem solving in obligatory tasks [0–6], number of optional tasks [0–6], and frequency of active problem solving in optional tasks [0–6]) were registered automatically by the learning environment.

The justifications typed into text-entry fields by students in the prompting group were analysed by means of content analysis. Categories were developed theoretically and adapted empirically by applying them to a test sample of four subjects. Seven disjunctive categories resulted. First, *monitoring* is the process by which learning and understanding is evaluated (e.g. "I think I have understood the first example so far"). Second, *self-regulation* involves how actions are planned and regulated (e.g. "I want to practice the correct solution procedure for this task"). Third, *epistemological aspects* refer to beliefs about learning and knowledge (e.g. "This way I can learn from my mistakes"). Fourth, *positive motivation/emotion* involves the description of positive states (e.g. "It is more exciting to work on a new example"). Fifth, *negative motivation/emotion* involves the description of negative states (e.g. "I am depressed, my solution is wrong again"). Sixth, *other reasons* cover any statement that does not fit into any of the categories. Seventh, o*mission* is used when no reason is given.

The students' statements were coded by two independent coders. Inter-rater reliability was good (Cohen's kappa = 0.75, $p < 0.01$). In the case of different codes, the first author of this study decided.

Procedure

First, the pretest was administered, and metacognitive prerequisites, interest in the subject and socio-biographic variables were recorded. The experimental sessions were conducted in the following 2 weeks. During the sessions, the students worked with Koralle, either with or without prompting. After a short break, the rating scales for acceptance and subjective learning outcomes and the posttest were administered. About 4 weeks after the last experimental session, the participants took the posttest again. This time, the control group was tested also (for practical reasons, it was not possible to test the control group twice). Afterwards, Koralle was available to all students.

Results

Learning outcomes

After the learning phase, students scored significantly and substantially higher on the two tasks that were identical in the pretest and posttest (see Table 1; main effect for time: $F[1, 55] = 52.39$, MSE = 52.79, $p < 0.01$, $\eta^2 = 0.49$).[1] The reflection prompts had no effect on this development (main effect for learning condition and interaction time × learning condition: $F[1, 55] < 1$).

The other three tasks on the posttest were compared for both experimental groups and the control group (see Table 1). Differences between the groups were statistically significant and substantial ($F[2, 121] = 106.07$, MSE = 702.19, $p < 0.01$, $\eta^2 = 0.64$). Students in the prompting group were clearly more successful than their peers who were not prompted to give reasons ($t[54.16] = 3.05$, $p < 0.01$, $\eta^2 = 0.14$). Additionally, students in the experimental conditions scored significantly higher than the control group (EG 1 vs. CG: $t[38.71] = 13.19$, $p < 0.01$, $\eta^2 = 0.72$; EG 2 vs. CG: $t[37.04] = 7.92$, $p < 0.01$, $\eta^2 = 0.56$).[2]

The prompting effect was also statistically significant concerning follow-up performance ($t[32] = 2.20$, $p < 0.05$, $\eta^2 = 0.13$). However, both experimental groups performed much better directly after the learning phase (main effect for time: $F[1, 32] = 26.50$, $p < 0.01$, $\eta^2 = 0.45$). This effect was stronger than the prompting effect ($F[1, 32] = 8.07$, $p < 0.01$, $\eta^2 = 0.20$). The interaction between learning condition and time was not significant ($F[1, 32] = 1.23$, *ns*).

When the follow-up performance of both experimental groups was compared with the performance of the control group, the experimental groups were still clearly superior (EG 1 vs. CG: $t[23.36] = 7.92$, $p < 0.01$, $\eta^2 = 0.56$; EG 2 vs. CG: $t[15.14] = 4.03$, $p < 0.01$; $\eta^2 = 0.31$).

Acceptance of the learning environment and subjective learning outcomes were rather high in both groups (see Table 1). Concerning both aspects, differences between the two experimental groups were not significant ($t[53] = -0.48$ and -0.90, respectively). Hence, reflection prompts did not influence acceptance and subjective learning outcomes.

[1] Effect sizes were classified according to the conventions specified by Cohen (1988): $\eta^2 \leq 0.05$: small effect; $0.06 \leq \eta^2 \leq 0.13$: medium effect; $\eta^2 \geq 0.14$: large effect.

[2] Adjusted degrees of freedom because of heterogeneous variance. For all comparisons between experimental and control group, *p* values were also calculated using a non-parametric test (Mann–Whitney *U*) because of differences in sample sizes. The *U*-test results did not differ from the results of the *t*-tests.

Table 1 Cognitive and motivational learning outcomes: means and standard deviations

Outcome	Experimental group with prompting EG 1 ($n = 28$)		Experimental group without prompting EG 2 ($n = 29$)		Control group ($n = 67$)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Pretest (tasks 1, 2) (max: 7)	2.78	1.25	2.27	1.42	–	–
Posttest (tasks 1, 3) (max: 7)	4.22	1.12	4.00	1.28	–	–
Posttest (tasks 2, 4, 5) (max: 13)	8.80	2.02	7.03	2.37	3.28	1.42
Follow-up test (tasks 2, 4, 5) (max: 13)[a]	7.64	2.34	5.86	2.30	–	–
Acceptance of the learning environment (min–max: 1–6)	4.78	0.79	4.86	0.53	–	–
Subjective learning outcomes (min–max: 1–6)	4.52	0.83	4.69	0.53	–	–

[a] EG 1: $n = 20$, EG 2: $n = 14$

Learning behaviour

Time on task varied between 30 and 105 min in the prompting group and between 30 and 81 min in the group without reflection prompts. The difference between means (about 6 min) was not statistically significant ($t[55] = 0.85$). Time on task was not associated with learning outcomes in either experimental group ($r = -0.03$ and -0.09, respectively).

In both groups, most obligatory tasks were completed in the form of problem-solving tasks (see Table 2). The small difference between both conditions was not significant ($t[55] = 1.23$). Optional tasks were only rarely requested in either condition (here, no significance testing was necessary, as the means did not differ). Concerning the optional tasks, there was a clear preference for active problem solving, especially in the prompting group. However, the difference between the two experimental groups concerning this preference was only marginal and not statistically significant ($t[55] = 1.13$).

In summary, the experimental groups displayed a similar learning behaviour in both time on task and task choice.

Table 3 provides descriptive statistics for the justification categories. Most reasons can be summarised into the categories of 'monitoring' and 'self-regulation'. Justifications categorised as 'epistemological aspects' ranked second. Motivational or emotional states were verbalised less often. Justifications that could not be coded and no entries were rare exceptions.

Justification categories and learning outcomes were nearly unrelated (self-monitoring: $r = 0.04$, self-regulation: $r = 0.02$, epistemological aspects: $r = 0.03$;

Table 2 Learning behaviour: means and standard deviations

Learning behaviour	With prompting		Without prompting	
	M	*SD*	*M*	*SD*
Time on task (in hr:min:sec)	1:14:47	0:29:02	1:08:29	0:27:18
Obligatory problem-solving tasks (min–max: 0–6)	5.10	1.10	4.60	2.00
Optional tasks (min–max: 0–6)	1.60	1.10	1.60	1.60
Optional problem-solving tasks (min–max: 0–6)	1.40	1.20	1.00	1.40

Table 3 Justification categories: frequencies, numbers of cases, means and standard deviations

Category	*f*	*N*	*M*	*SD*
Monitoring	143	27	5.30	2.91
Self-regulation	86	27	3.19	1.88
Epistemological aspects	73	23	3.17	1.97
Positive motivation/emotion	52	22	2.36	1.18
Negative motivation/emotion	22	16	1.38	0.50
Other reasons	7	4	1.75	0.50
Omission	6	5	1.20	0.45

positive motivation/emotion: $r = -0.03$; negative motivation/emotion: $r = -0.19$). All correlations were non-significant.

Discussion

Learning outcomes

Before the students worked with Koralle, many of them failed to solve simple correlation tasks on the pretest. Even basic knowledge that is addressed in regular statistics lectures was not familiar to them, or it could not be applied. This finding corresponds with many studies on statistics learning (e.g. Broers 2002; Castro Sotos et al. 2007; Krause 2007).

After the learning phase, most students of both experimental groups were able to cope successfully with simple and even with more complex tasks. In a study by Krause (2007) that involved the same learning environment and the same tests (but investigated effects of feedback and cooperative learning), comparable learning progress was achieved. Although these effects have to be interpreted carefully due to well-known problems of repeated measurement, they confirm the worked-example effect (e.g. Atkinson et al. 2000; Gerjets et al. 2004) and the effectiveness of examples in combination with problem-solving tasks (e.g. Renkl et al. 2004; Stark et al. 2000).

The effectiveness of this combined method (which, in a sense, is a special type of the effective *completion method*; Paas 1992; van Merriënboer and de Croock 1992) was enhanced by a simple prompting procedure. It is assumed that, because of the low complexity of the tasks used for assessing learning progress (e.g. Bannert 2003), prompts had no effects on these tasks. However, concerning the more complex tasks that required declarative, procedural and conditional knowledge, the prompting procedure proved effective. Still, prompting was less effective than a feedback intervention that also aimed at increasing reflective processes in the same instructional context (Krause 2007; Krause et al. 2009).

The prompting effect was sustainable; follow-up testing confirmed the superiority of the prompted learners. However, performance had decreased in both experimental groups 4 weeks after the learning session. Although not all subjects were present at the follow-up session, performance decreases cannot be attributed to (negative) selection effects because the participants did not differ from the rest of the experimental groups concerning several learning prerequisites. Apparently, students had forgotten parts of what they had learned. In regular statistics education, forgetting can be countered by repetition and practice.

However, the main goal—to increase understanding of correlation—was achieved: Both experimental groups clearly outperformed the control group that was comparable with respect to various learning prerequisites and had attended regular statistics lectures but had no access to the e-learning environment. Several factors influencing the differences between the experimental groups and the control group must be looked at carefully. Firstly, students in the control group received no additional training. Secondly, selection effects might have influenced the results. Thirdly, assessment of learning prerequisites followed the maxim of parsimony and, therefore, might not have been precise enough. Finally, making a payment to students in the experimental groups might have enhanced test motivation. Especially in instructional contexts where motivation is generally low, extrinsic motivation can play an important role in learning and performance (e.g. Stark and Mandl 2000).

Still, even in the light of these possible limitations, working with Koralle was associated with enhanced understanding of basic correlation concepts and principles, and effectiveness was increased by reflection prompts. Given the fact that the prompting intervention was very economical, effects are remarkable and need replication.

No treatment effects could be identified in motivational outcomes. Acceptance of the learning environment and subjective learning outcomes were remarkably high in both experimental groups. Similar results were found in the study by Krause (2007) that involved the same learning environment. In other intervention studies on example-based learning, for instance in the domain of medicine (e.g. Kopp et al. 2007), acceptance and subjective learning outcomes were also rather high. Of course, these results might have been influenced by response biases; however, at least there is no evidence for *negative* motivational effects stemming from the intervention.

Learning behaviour

Reflection prompts neither influenced time on task nor the students' task choices. In both experimental groups, most learners preferred to start with problem solving and studied the example solutions afterwards. Starting by problem solving can reveal illusions of competence and increase the subsequent intensity and quality of self-explanation (Stark et al. 2000). In a study on case-based learning in medicine, starting with problem solving proved more successful than beginning by studying model solutions (Mandl and Gräsel 1997). Because examples can be studied in an active and self-regulated way as well (e.g. Chi et al. 1989) and problems can be solved in a rather mechanical and superficial way, the students' preference for problem solving in our study is only a weak indicator of productive learning. However, it shows that students were ready to try hard, which confirms that motivation was relatively high.

In complex computer-based learning environments, students often ignore optional features and aids (Dillon and Gabbard 1998), even though they might profit from these options (Stark and Mandl 2002). In our study, choosing more optional tasks might have resulted in better learning outcomes (Pirolli and Anderson 1985). However, as Koralle is highly structured and, to some extent, repetitive (especially concerning the optional tasks), students possibly experienced the additional activities as boring or superfluous.

Learners' justifications in the prompting group suggest that the prompting procedure primarily induced activities of metacognitive control (monitoring, self-regulation). Besides, epistemic knowledge was activated, including both metacognitive and cognitive aspects which, as a rule, are interrelated (Berthold et al. 2007). Hence, in line with the study by Berthold et al. (2007), the prompts seem to have fostered metacognitive and—at least to some extent—cognitive strategies. In contrast to our expectations, the justification categories were not related to learning outcomes. The scatter plots did not reveal any factors which might be responsible for this. We can only put forward the tentative assertion that the prompting intervention promoted cognitive and metacognitive processes, such as higher attention and deeper reflection, which led to better information processing. It would be naïve to assume that the justifications given by the students are exact representations of the underlying (and at least partly implicit) processes. From the literature on methodical problems of thinking-aloud procedures (Ericsson and Simon 1993), we know that learners, even if they are inclined to follow the experimental instructions, neither *want* to verbalise all their thoughts nor are *capable* of doing this while working on a task. Besides, justifications do not contain any information about the validity of reflections, their contextualisation, consequences that are drawn and other factors relevant for effects on learning. Both from an intra-individual and from an inter-individual perspective, a justification category can include various processes. Therefore, from our perspective, Pearson correlation coefficients are neither necessary nor sufficient conditions for proving the effectiveness of processes induced by the prompting procedure; they only provide first hints.

Outlook

In order to examine the working principle of reflection prompts, additional measures—especially the students' mental effort—should be included during the learning process. In addition, cognitive and metacognitive strategies should be recorded retrospectively using standardised instruments. In order to cope with the discussed problem of heterogeneous justification categories, detailed, theory-guided case analyses are indicated.

Even if we cannot sufficiently explain why and how the prompts in our study worked, they definitely worked. Still, the effect needs to be replicated before practical recommendations are made.

In the light of the results of our study and those of Krause (2007), we presume that the implemented principles of situated, example-based learning could prove effective for teaching other pivotal concepts and principles of statistics and empirical research methods, such as the logic of statistical significance testing, which is a challenge for many students (Cohen 1994).

References

Alexander, P. A. (1996). The past, present, and future of knowledge research: A re-examination of the role of knowledge in learning and instruction. *Educational Psychologist, 3*, 89–92.

Allwood, C. M. (1990). On the relation between justification of solution method and correctness of solution in statistical problem solving. *Scandinavian Journal of Psychology, 31*, 181–190.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. W. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*, 181–214.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Combining fading with prompting fosters learning. *Journal of Educational Psychology, 95*, 774–783.

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist* (Special Issue on Computers as Metacognitive Tools for Enhancing Student Learning), *40*, 199–209.

Bannert, M. (2003). Effekte metakognitiver Lernhilfen auf den Wissenserwerb in vernetzten Lernumgebungen [Effects of metacognitive learning aids on knowledge acquisition in networked learning environments]. *Zeitschrift für Pädagogische Psychologie, 17*, 13–25.

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction, 17*, 564–577.
Broers, N. J. (2002). Selection and use of propositional knowledge in statistical problem solving. *Learning and Instruction, 12*, 323–344.
Brown, A., & DeLoache, J. S. (1978). Skills, plans and self-regulation. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 3–35). Hillsdale, NJ: Lawrence Erlbaum.
Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review, 2*, 98–113.
Chi, M. T. H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology, 10*, S33–S49.
Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145–182.
Chi, M. T. H., DeLeeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439–477.
Cognition and Technology Group at Vanderbilt. (1997). *The Jasper project: Lessons in curriculum, instruction, assessment and professional development*. Mahwah, NJ: Erlbaum.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*, 997–1003.
Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction* (pp. 453–494). Hillsdale, NJ: Erlbaum.
De Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*, 105–113.
Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227–268.
Dillon, A., & Gabbard, R. (1998). Hypermedia as an educational technology: A review of the quantitative research literature on learner comprehension, control, and style. *Review of Educational Research, 68*, 322–349.
Dochy, F. J. R. C. (1992). *Assessment of prior knowledge as a determinant for future learning. The use of prior knowledge state tests and knowledge profiles*. Utrecht, The Netherlands: Uitgeverij Lemma B. V.
Ericsson, K. A., & Simon, H. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
Flavell, J. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231–235). Hillsdale, NJ: Lawrence Erlbaum.
Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist, 34*, 906–911.
Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science, 32*, 33–58.
Gerjets, P., Scheiter, K., & Schuh, J. (2005). Instruktionale Unterstützung beim Fertigkeitserwerb aus Beispielen in hypertextbasierten Lernumgebungen [Instructional support for skill acquisition from examples in hypermedia-based learning environments]. *Zeitschrift für Pädagogische Psychologie, 19*(1/2), 23–38.
Große, C. S., & Renkl, A. (2006). Effects of multiple solution methods in mathematics learning. *Learning and Instruction, 16*, 122–138.
Hasselhorn, M. (2006). Metakognition [Metacognition]. In D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* [Handbook of educational psychology] (pp. 480–485). Weinheim, Germany: Psychologie Verlags Union.
Hasselhorn, M., & Hager, W. (1998). Kognitive Trainings auf dem Prüfstand: Welche Komponenten charakterisieren erfolgreiche Fördermaßnahmen? [Cognitive trainings put to test: What components characterise successful facilitation methods?]. In M. Beck (Ed.), *Evaluation als Maßnahme der Qualitätssicherung: Pädagogisch-psychologische Interventionen auf dem Prüfstand* [Evaluation as a means of quality management: Pedagogical-psychological interventions put to test] (pp. 85–98). Tübingen, Germany: DGVT-Verlag.
Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research, 60*(4), 549–571.
Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*, 294–303.
Kopp, V., Stark, R., & Fischer, M. R. (2007). Förderung von Diagnosekompetenz in der medizinischen Ausbildung durch Implementation eines Ansatzes zum fallbasierten Lernen aus Lösungsbeispielen

[Enhancing diagnostic skills in medical training by case-based learning from worked examples]. *GMS Zeitschrift für Medizinische Ausbildung*, *24*(2): Doc 107.

Krapp, A. (2000). Individuelle Interessen als Bedingung lebenslangen Lernens [Individual interests as a condition for life-long learning]. In F. Achtenhagen & W. Lempert (Eds.), *Lebenslanges Lernen im Beruf—seine Grundlegung im Kindes-und Jugendalter, Band 3: Psychologische Theorie, Empirie und Therapie* [Professional lifelong learning—its foundation in childhood and adolescence, Volume 3: Psychological theory, empiricism and therapy] (pp. 54–75). Opladen, Germany: Leske + Budrich.

Krapp, A. (2005). Basic needs and the development of interest and intrinsic motivational orientations. *Learning and Instruction, 15*, 381–395.

Krause, U.-M. (2007). *Feedback und kooperatives Lernen* [Feedback and cooperative learning]. Münster, Germany: Waxmann.

Krause, U.-M., & Stark, R. (2006). Vorwissen aktivieren [Activating prior knowledge]. In H. Mandl & H. F. Friedrich (Eds.), *Handbuch Lernstrategien* [Handbook of learning strategies] (pp. 38–49). Göttingen, Germany: Hogrefe.

Krause, U.-M., & Stark, R. (2007, August). *Problem-oriented learning and reflection prompts in teacher education: Enhancing knowledge about cooperative learning*. Paper presented at the 12th Biennial Conference of the European Association for Research on Learning and Instruction (EARLI), Budapest, Hungary.

Krause, U.-M., Stark, R., & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction, 19*, 158–170.

Lan, W. Y., Bradley, L., & Parr, G. (1993). The effects of a self-monitoring process on college students' learning in an introductory statistics course. *The Journal of Experimental Education, 62*, 26–40.

Langer, E. J. (1993). A mindful education. *Educational Psychologist, 28*, 43–50.

Mandl, H., & Gräsel, C. (1997). Gestaltung konstruktivistischer Lernumgebungen in der Medizin [Designing constructivist learning environments in medical science]. In S. Höfling & H. Mandl (Eds.), *Lernen für die Zukunft—Lernen in der Zukunft: Wissensmanagement in der Bildung* [Learning for the future—learning in the future: Knowledge management in education] (pp. 155–165). München, Germany: Hanns-Seidel-Stiftung e.V.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective. *Perspectives on Psychological Science, 1*, 133–163.

McClelland, D. J., Atkinson, J. W., Clark, R. H., & Lowell, E. L. (1953). *The achievement motive*. New York: Appleton-Century-Crofts.

Morris, E. (2001). The design and evaluation of link: A computer-based learning system for correlation. *British Journal of Educational Technology, 32*, 39–52.

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*, 429–434.

Paris, S. G., Lipson, M. Y., & Wixon, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology, 8*, 293–316.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement, 53*, 801–813.

Pirolli, P. L., & Anderson, J. R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 11*, 136–153.

Reigeluth, C. M., & Stein, F. S. (1983). The elaboration theory of instruction. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status* (pp. 335–382). Hillsdale, NJ: Erlbaum.

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

Renkl, A., Atkinson, R. K., & Große, C. S. (2004). How fading worked solution steps works—a cognitive load perspective. *Instructional Science, 32*, 59–82.

Salomon, G., & Globerson, T. (1987). Skill may not be enough: The role of mindfulness in learning transfer. *International Journal of Educational Research, 11*, 623–637.

Schiefele, U., Krapp, A., & Schreyer, I. (1993). Metaanalyse des Zusammenhangs von Interesse und schulischer Leistung [Meta-analysis of the correlation between interest and academic performance]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 25*, 120–148.

Schneider, W., & Pressley, M. (1989). *Memory development between 2 and 20*. New York: Springer.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. New York: Academic Press.

Spiro, R., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology, 31*(5), 24–33.

Stark, R. (1999). *Lernen mit Lösungsbeispielen: Einfluss unvollständiger Lösungsbeispiele auf Beispielelaboration, Lernerfolg und Motivation* [Learning with worked-out examples: The impact of incomplete examples on example elaboration, learning outcomes and motivation]. Göttingen, Germany: Hogrefe.

Stark, R. (2004). Implementing example-based learning and teaching in the context of vocational school education in business administration. *Learning Environments Research, 7*, 143–163.

Stark, R., Gruber, H., Renkl, A., & Mandl, H. (2000). Instruktionale Effekte einer kombinierten Lernmethode: Zahlt sich die Kombination von Lösungsbeispielen und Problemlöseaufgaben aus? [Instructional effects of a combined teaching approach: Does the combination of worked-out examples and problem-solving tasks pay off?]. *Zeitschrift für Pädagogische Psychologie, 14*, 205–217.

Stark, R., & Mandl, H. (2000). Training in empirical research methods: Analysis of problems and intervention from a motivational perspective. In J. Heckhausen (Ed.), *Motivational psychology of human development* (pp. 165–183). Amsterdam, The Netherlands: Elsevier.

Stark, R., & Mandl, H. (2002). Konzeptualisierung und Evaluation einer komplexen netzbasierten Lernumgebung im Kontext der universitären Ausbildung in empirischen Forschungsmethoden [Conceptualisation and evaluation of a complex net-based learning environment in university training of empirical research methods]. *Unterrichtswissenschaft, 30*, 315–330.

Stark, R., & Mandl, H. (2005). Lernen mit einer netzbasierten Lernumgebung im Bereich empirischer Forschungsmethoden: Effekte zusätzlich implementierter Maßnahmen und Bedeutung von Lernvoraussetzungen [Learning within a net-based learning environment in the field of empirical research methods: Effects of additional interventions and impact of learning prerequisites]. *Unterrichtswissenschaft, 33*(1), 3–29.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.

Tyroller, M. (2005). *Effekte metakognitiver Prompts beim computerbasierten Statistiklernen* [Effects of metacognitive prompts in computer-based statistics learning]. Retrieved July 21, 2009, from http://edoc.ub.uni-muenchen.de/archive/00005514/01/Michael_Tyroller.pdf.

Van Merriënboer, J. J. G. (1990). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of Computing Research, 6*, 265–285.

Van Merriënboer, J. J. G., & de Croock, M. B. M. (1992). Strategies for computer-based programming instruction: Program completion vs. program generation. *Journal of Educational Computing Research, 8*, 365–394.

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*, 3–14.

Weinstein, C. E., Palmer, D. R., & Schulte, A. C. (1987). *Learning and study strategies inventory*. Clearwater, FL: H & H Publishing Company.