The Journal of Experimental Education, 2002, 70(4), 293–315

From Example Study to Problem Solving: Smooth Transitions Help Learning

ALEXANDER RENKL
University of Freiburg (Germany)

ROBERT K. ATKINSON
Mississippi State University

UWE H. MAIER
Educational University of Ludwigsburg (Germany)

RICHARD STALEY
State University of New York–College at Oneonta

ABSTRACT. Research has shown that it is effective to combine example study and problem solving in the initial acquisition of cognitive skills. Present methods for combining these learning modes are static, however, and do not support a transition from example study in early stages of skill acquisition to later problem solving. Against this background, the authors proposed a successive integration of problem-solving elements into example study until the learners solved problems on their own (i.e., complete example → increasingly more incomplete examples → problem to-be-solved). The authors tested the effectiveness of such a fading procedure against the traditional method of using example–problem pairs. In a field experiment and in 2 more controlled laboratory experiments, the authors found that (a) the fading procedure fosters learning, at least when near transfer performance is considered; (b) the number of problem-solving errors during learning plays a role in mediating this effect; and (c) it is more favorable to fade out worked-out solution steps in a backward manner (omitting the last solution steps first) as compared with a forward manner (omitting the first solution steps first).

Key words: fading, learning, problem solving, transfer, worked-out examples

WORKED-OUT EXAMPLES consist of a problem formulation, solution steps, and the final solution itself. Researchers have shown that learning from such examples is of major importance for the initial acquisition of cognitive skills in well-structured domains such as mathematics, physics, and programming (for an overview, see Reimann, 1997; VanLehn, 1996). What we mean by *initial skill acquisition* can be more precisely defined by referring to Anderson, Fincham, and Douglass (1997). These authors proposed a four-stage model within Anderson's well-known ACT-R framework. They argued that skill acquisition involves four overlapping stages. In the first stage, learners solve problems by analogy; that is, they refer to known examples and try to relate them to the problem to be solved. In the second stage, learners develop abstract declarative rules, such as verbal knowledge, that guides their problem solving. After practice, they move to the third stage, in which performance becomes smooth and rapid without the use of many attention resources; that is, proceduralized rules are formed. In the fourth stage, learners who have practiced many different types of problems have many examples in mind. Hence, they can often retrieve a solution quickly and directly from memory. Anderson and his colleagues emphasized that these stages overlap in the sense that a specific learner's flexibility in using different methods (e.g., analogy or abstract rule) depends on familiarity with the specific problem at hand. From the viewpoint of skill acquisition, then, the importance of studying examples relative to problem-solving practice is very high when a student is in the first stage (analogy) or is beginning to enter the second stage (abstract rules of learning). Studying worked-out examples is no longer the preferred method when the instructional goal is to facilitate the attainment of the third stage (automatic performance), where problem-solving practice is of critical importance.

Learning from worked-out examples in initial skill acquisition is also a learning mode preferred by novices, and rightly so because it is an effective way of learning. Sweller and his colleagues (e.g., Sweller & Cooper, 1985; for an overview see Sweller, van Merriënboer, & Paas, 1998) showed that learning from worked-out examples can be more effective than learning by problem solving. Sweller and colleagues explained the often-found superiority of example learning by the argument that problem solving requires so much working memory capacity that it interferes with learning in the sense of schema acquisition. More specifically, they argued that in order to solve problems, novices (i.e., learners) use means–ends analysis, which implies that many aspects of the problem have to be focused on essentially simultaneously (e.g., actual problem state, desired problem state, difference between actual and desired problem states, relevant operators, and subgoals). Given this load, too few resources are left for the induction of abstract and generalizable problem-solving schemata (e.g., Sweller et al., 1998).

Although worked-out examples have significant advantages, their use as a learning methodology does not, of course, guarantee effective learning. According to Atkinson, Derry, Renkl, and Wortham (2000), several factors moderate the effec-

Address correspondence to Alexander Renkl, Psychological Institute, Department of Educational Psychology, University of Freiburg, Engelbergerstr. 41, D-79085 Freiburg, Germany. E-mail: renkl@psychologie.uni-freiburg.de

tiveness of learning from worked-out examples: (a) self-explanations, (b) situational factors, and (c) example design. First, the *self-explanation effect* refers to the finding that the extent to which learners profit from the study of examples depends on how well they explain the solutions of the examples to themselves (Chi, 2000; Chi, Bassok, Lewis, Reimann, & Glaser, 1989). Second, *situational factors* that can influence learning outcomes are, for example, the provision of instructional explanations during learning or goals set to the learners (e.g., learning from examples for later teaching). Third, it is important how the learning materials (examples and problems) are structured. In this article, we focused on this third aspect. More specifically, we investigated one possible approach to integrating elements of problem solving into example study. We proposed that one can combine these learning modes by successively introducing more and more elements of problem solving in example study until learners are solving the problems on their own. This rationale can also be used as a way to structure the transition from studying examples in initial skill acquisition to later problem solving in order to proceed from the second to the third stage postulated by Anderson et al. (1997).

In the next section, we discuss the literature with respect to the issue of combining example study and problem solving. Then we outline open questions and give preliminary answers that were tested in three studies, first in a field experiment and then in two more controlled laboratory experiments.

How to Combine Example Study and Problem Solving?—State of the Art

Empirical evidence has shown that learning only from (completely) worked-out examples is not as effective as learning from examples in which elements of problem solving are integrated. There are two traditional ways in which this integration can be accomplished: (a) making the solutions of examples incomplete and (b) using example–problem pairs.

Incomplete examples. Some researchers argue that incomplete examples, which the learners have to complete, effectively support the acquisition of cognitive skills (Paas, 1992; Stark, 1999; van Merriënboer, 1990; van Merriënboer & de Crook, 1992). For example, Stark (1999) conducted a controlled experiment designed to examine the extent to which the insertion of "blanks" into the solution of examples—which, in a certain sense, forced the learners to determine the next solution step on their own—fostered learning. In that study, half of the participants studied incomplete examples (experimental group), and the other half learned from complete examples (control group). In the experimental group, portions of the example solutions presented to the participants were replaced by "question marks." The learners were then asked to identity which solution step was missing. After the learners did that, or at least made the attempt, the complete solution step was presented so that learners received feedback on the correctness of their anticipation. The learning outcomes were assessed by a posttest that included near, medium, and far transfer problems. Stark's *near transfer* problems were identical in structure (same solution rationale) to the examples presented for learning but contained different surface features (cover story, objects, numbers). *Medium transfer* referred to problems with a different structure (a modified solution procedure had to be found), but similar surface features. *Far transfer* problems differed with respect to both structure and surface features. Finally, Stark required the learners to construct a problem containing certain structural and surface features; that task was also one of substantial transfer distance. Stark found that, compared with studying complete examples, studying incomplete examples fostered the quality of self-explanations and, as a consequence, the near and medium transfer of learned solution methods (for far transfer, significance level of merely 10% was reached). In addition, the learners studying incomplete examples showed better performance in the construction task.

Example–problem pairs. Sweller and his colleagues (e.g., Mwangi & Sweller, 1998; Sweller & Cooper, 1985) have conducted several classic studies documenting the effectiveness of learning from worked-out examples. However, in these studies the authors did not compare learning from examples only with learning by problem solving. Instead, the example condition usually consisted of examples followed by isomorphic problems-to-be-solved (example–problem pairs). Sweller and colleagues mainly showed that combined learning from examples and problems is more effective than learning by solving problems. Although this finding was reliably obtained when the posttest problems had identical or very similar structures (in this article, near transfer), for problems with dissimilar structures (in this article, far transfer) no effects were usually found.

Studies on learning from worked-out examples performed by other researchers (e.g., Renkl, 1997) have focused on learning from examples only. Explicit comparisons between learning from examples only and learning from example–problem pairs are, however, rare. One such study was performed by Trafton and Reiser (1993), who designed two treatments, alternating and blocked. The participants in the alternating condition were exposed to six example–problem pairs, in which each example was followed directly by an isomorphic problem, whereas participants in the blocked condition were exposed to the entire set of six examples, followed by the entire set of six practice problems. The learning outcomes were assessed by three near transfer problems (same problem structure as in the learning phase). The authors found that, as predicted, the participants in the alternating-example condition took less time and produced more accurate solutions on the transfer posttest than their counterparts in the blocked-example condition. On the basis of these findings, the authors asserted that "the most efficient way to present material to acquire a skill is to present an example, then a similar problem to solve immediately following" (Trafton & Reiser, 1993, p. 1022).

In a recent study, Stark, Gruber, Renkl, and Mandl (2000) examined whether there might be another effective variation of the traditional method of pairing examples with practice problems. Stark and his colleagues' study was motivated in part by Gräsel and Mandl's (1993) study that focused on learning diagnostic strategies in medicine. Gräsel and Mandl found that it is more effective to learn from a *cognitive model*—which can also be regarded as a kind of worked-out example—after an initial problem-solving experience. Against this background, Stark et al. (2000) argued that initial problem-solving difficulties should motivate learners to process examples that follow more deeply. Thus, Stark et al. elected to present practice problems first followed by isomorphic examples (problem–example pairs). The learning outcomes were assessed by near, medium, and far transfer problems (defined earlier as in Stark, 1999). In a comparison between learning from examples only and learning from problem–example pairs (domain: calculation of compound and real interest), the combined learning method (i.e., problem–example pairs) fostered substantially more active example processing and, as a result, improved near, medium, and far transfer performance. Taken together, combining practice problems and examples is obviously more effective than exposing learners to either sets of practice problems only or sets of examples only.

Open Questions and Answers to Be Validated

Although there can be little doubt about the effectiveness of a combined learning method, two questions still remain open: (a) Are there more effective ways of combining example study and problem solving than presenting incomplete examples or pairs of examples and problems? and (b) What is a sensible rationale for designing the transition from learning from examples in initial cognitive skill acquisition to later problem solving?

Instructional models from different paradigms propose a smooth transition from complete models (worked-out examples) to independent problem solving—that is, a fading procedure. Within the traditional cognitive perspective, which is the one adopted in this article, Anderson, Corbett, Koedinger, and Pelletier (1995), for example, delineated principles of effective tutoring from their ACT-R framework. One of these principles is to "[f]acilitate successive approximations to the target skill" (p. 181). This suggests that initially a coaching tutor removes parts of the problem-solving burden so that the learner does not have to perform all the steps. With time, the learner provides more and more of the work and the support is faded out. Such an instructional procedure is also compatible with the most prominent current situated learning model. The cognitive apprenticeship approach (Collins, Brown, & Newman, 1989) proposes a smooth transition from modeling, to scaffolded problem solving, to independent problem solving, in which instructional support fades during the transition. Taken together, irrespective of the theoretical framework, a smooth transition (fading) from complete worked-out examples to problems-to-be-solved would be preferred.

The use of incomplete examples, at least as realized in previous studies, has not incorporated such a dynamic fading component. To date, studies incorporating the "pairs arrangement" have also not used a fading component. In fact, these studies typically contain abrupt transitions from examples, as a type of model, to independent problem solving. Against this background, it is sensible to combine problem solving and example study in the following way. First, a complete example is presented (model). Second, an example is given in which one single solution step is omitted (coached problem solving). Then, the number of blanks is increased step-by-step until just the problem formulation is left, that is, a problem-to-be-solved (independent problem solving). In this way, a smooth transition from modeling (complete example) over coached problem solving (incomplete example) to independent problem solving is implemented. This rationale provides a possible answer to both questions outlined earlier (first, effective combination of example study and problem solving and, second, transition from example study to problem solving).

An important factor that should contribute to the effectiveness of a smooth transition (fading) as compared with the usual method of using example–problem pairs is that fading should reduce problem-solving errors during learning. Using example–problem pairs implies quite abrupt changes with respect to the demands placed on the learners. After a first example, the learners have to solve a whole problem totally on their own. Under a fading condition, the first problem-solving demand is to generate just a single step, and the demands are only gradually increased. Against this background, we expected that the learners would make fewer errors during learning in the fading condition. If the goal is to form rules for problem solving, instructional procedures that avoid errors (or immediately correct them if they occur) are most appropriate (e.g., Anderson et al., 1995). In other words, when the goal is to learn to solve certain types of problems that can be solved by the application of specific to-be-learned rules (near transfer), avoiding errors should be an advantage.

Avoiding errors is not, however, necessarily productive when problems should be solved that require the modification of learned solution methods (far transfer). In this case, learned rules cannot be (directly) applied. Far transfer (e.g., Anderson et al., 1995) can be fostered by errors that trigger reflections and thereby deepen understanding of the domain (cf. VanLehn, 1996). From this perspective, fading would not foster far transfer performance. On the other hand, avoiding the demand to correct errors might reduce the cognitive load that is imposed by the problem-solving activities. According to the theory of Sweller et al. (1998), cognitive activities that contribute to a deeper understanding of the domain (e.g., self-explanations) are more likely to occur. From this perspective, fading may also foster far transfer performance.

On the basis of the aforementioned evidence, we expected that fading worked-out solution steps in contrast to using example–problem pairs fosters performance on near transfer problems (known solution methods). To what extent fading is also favorable when far transfer (new solution methods) is concerned is an open question.

EXPERIMENT 1: FIELD EXPERIMENT

As a first test of our assumptions, we conducted a small-scale field experiment in which we tested whether a smooth transition from example study to problem solving (gradual insertion of blanks into the solutions of examples) is more effective than learning by example–problem pairs as they are used in many studies on learning from examples.

Method

Sample and Design

Two ninth-grade classrooms from a German *Hauptschule* (lowest track of the German three-track system) participated in this quasi-experiment. In both classrooms, the same teacher (third author) conducted a physics lesson on electricity based on four examples/problems. In one classroom ($n = 20$) a fading procedure was used, and in the other classroom ($n = 15$) traditional example–problem pairs were used. Each example/problem involved three solution steps. Across both conditions, half of the steps were worked out, whereas the other half were to be generated. Thus, learners in both conditions were required to solve the same number of solution steps.

It should be noted that the investigation was not performed in "extra sessions" but as part of the regular physics instruction. The content domain of electricity is also part of the obligatory official curriculum for these ninth graders.

Learning Environment

In the experimental phase, the third author (a professional teacher) conducted a 45-min lesson in each classroom. Both groups worked on four examples/problems in which the cost for running a variety of electric devices for a certain time had to be determined (e.g., "An aluminum factory has a big melting furnace that is run with 1000 V. A current of 20 A has to flow through the furnace in order to melt aluminum. What does the factory have to pay per month when the furnace always runs and the kWh costs DM 0.22?"). Although the examples/problems were printed on work sheets, the problem formulation of each example/problem was read aloud by 1 of the students in the class. Following the reading of the problem formulation, the students were permitted to ask clarifying questions (of course, no questions on the solution) before working individually on the example or problem (4–6 min). At the end of each incomplete example or problem, the complete solution was presented on an overhead transparency, and, if necessary, the students corrected or supplemented their solutions. Then the teacher proceeded to the next example/problem.

In the fading classroom, the teacher presented the instruction in the following order: (a) a complete example, (b) an example with the last solution step left out, (c) an example with the last two steps omitted, and (d) a problem in which all three steps were missing ("backward rationale" of omitting solution steps). In the example–problem group, in contrast, a complete example was presented twice; each time, it was followed by a corresponding problem.

Instruments

The pretest consisted of four problems from the physics domain of electricity that were structurally equivalent to the near transfer problems in the posttest (e.g., "The electronic motor of an electronic locomotive is supplied by a voltage of 0.6 kV. On average, a current of 18 A flows through the motor. What does an eight-hour trip from Stuttgart to Hamburg cost when you assume that the *German Railway* pays DM 0.12 per kWh?"). Each pretest item included three solution steps. One point was dispensed for each correct step (partial credit). Thus, 3 points were given when a problem solution was totally correct. The maximum score to be achieved was 12.

The posttest consisted of six problems. Four near transfer problems had the same underlying structure (solution rationale, i.e., the same solution steps had to be applied in the same order) as the examples and problems used in the learning phase but different surface features (cover story, numbers). Two problems were classified as far transfer because both the underlying structure and the surface features differed (e.g., "Tanja pays for her frig DM 40 per year. One kWh costs DM 0.22. What power does the frig use if you assume that it runs all the time?").[1] Each posttest problem included three solution steps. One point was awarded for each correct step (partial credit). Hence, 3 points were given when a problem solution was totally correct. The maximum score to be achieved was 12 (near transfer) or 6 (far transfer), respectively.

Procedure

The overall procedure was identical in both classrooms. Basic knowledge of the concepts and rules of electricity was introduced in the context of regular instruction followed by a pretest that measured prior knowledge of the abstract

[1]We used only two far transfer problems because they were quite difficult and students of the German *Hauptschule* are (usually and also in this case) low achieving and are not highly motivated to engage in deep reasoning about academic issues. It was unlikely that the students would have seriously tried to solve more than two such difficult problems.

TABLE 1
Means and Standard Deviations of Pretest and Posttest Scores in Experiment 1

Variable	Fading		Example–problem pairs	
	M	*SD*	*M*	*SD*
Pretest	2.89	3.37	2.88	3.48
Posttest: near transfer	9.53[a]	3.29	7.47	2.98
Posttest: far transfer	2.18	2.24	1.27	1.66

Note. Possible ranges of the variables: pretest (0–12), near transfer (0–12), and far transfer (0–6).
[a]Differs from example–problem pairs mean on the basis of an ANCOVA ($\alpha = .05$).

rules involved in solving domain problems. Two days later, the school lessons in which the experimental variation took place were conducted. Finally, after 2 additional days, the students worked on a posttest.

Results

The means and standard deviations of the two experimental groups on the pretest and the posttest scores are presented in Table 1. The pretest performance was almost identical for both groups, $t(33) = 0.01, p > .10, \eta^2 < .01$. Hence, there was no a priori difference between groups with respect to prior knowledge.

With respect to treatment effects, we descriptively obtained higher means in the fading group for both near and far transfer. These differences were tested by an analysis of covariance (ANCOVA; controlling for prior knowledge), which presupposes that there are homogeneous slopes for the regression from the transfer measures to pretest. This precondition was not violated for near transfer or for far transfer as the corresponding statistically insignificant interaction terms indicate (near transfer: $F < 1$; far transfer: $F(1, 33) = 3.79, p > .05$). An ANCOVA test for differences in group means yielded a statistically significant difference for near transfer performance in favor of the fading condition, $F(1, 32) = 4.44, p < .05$ (adjusted means = 9.52 for fading and 7.47 for example–problem pairs). This effect was of medium practical significance, partial $\eta^2 = .12$. The group difference in far transfer performance, which was of small to medium size, partial $\eta^2 = .07$, failed to reach the accepted level of statistical significance, $F(1, 32) = 2.28, p > .10$ (adjusted means = 2.17 for fading and 1.27 for example–problem pairs). Thus, the fading procedure clearly fostered near transfer performance. We cannot, however, claim that this is also true for far transfer performance.

Discussion

Before theorizing about possible reasons for (potential) differential effects of the fading procedure on near and far transfer, we decided to see whether the respective findings could be replicated, especially because the missing statistical significance of the effect on far transfer might be due to the large intragroup variances or to more substantial reasons (see Table 1).

A replication was necessary because a field study such as Experiment 1 always has some factors that might diminish the internal validity of the findings. For example, the teacher who conducted the instruction in both classrooms was not blind with respect to the experimental expectations. Furthermore, Experiment 1 was merely a quasi-experiment (no random assignment of participants to the experimental conditions). Hence, the conditions in both classrooms might not have been totally identical except for the independent variable (fading vs. example–problem pairs). For example, the problem formulation of each example/problem was not read aloud by the same student. Finally, no data on possible processes that mediate the effects of the fading procedure on the learning outcomes were recorded. These issues were addressed in Experiment 2.

EXPERIMENT 2: LABORATORY EXPERIMENT

To conceptually replicate the results of the preceding field experiment under more controlled conditions, we ran a laboratory experiment. We also tested for one possible mediating mechanism that was discussed earlier in this article.

As outlined earlier, there are quite abrupt changes with respect to the demands placed on the learners in the example–problem conditions that may lead to a relatively high rate of problem-solving errors during learning. Hence, the number of problem-solving errors during learning should mediate a positive effect of fading on near transfer performance. This hypothesis was tested in Experiment 2. For far transfer, we had no clear expectations with respect to the effects of fading and the role of errors. Thus, Experiment 2 was exploratory regarding the issue of far transfer.

Method

Sample and Design

The participants in this study were 54 American psychology students at a large, southeastern U.S. university. They were randomly assigned to the fading or to the example–problem condition, respectively ($n = 27$ in each group). As with our field experiment, the number of unsolved solution steps (12) was held constant across both conditions.

Learning Environment

We used a computer-based learning program that had been originally developed by Renkl (1997), modified by Stark (1999), and finally adapted to the present needs by the second author. It presented worked-out examples and problems from the domain of probability calculation. An example is the following: "Jonathan has recently bought a new camera. Independently of each other, he frequently makes two errors when he takes a picture. He manages to blur the image in 40% of his photos (p = 2/5) and he forgets to activate the flash in 10% of the photos (p = 1/10) so that the pictures end up too dark. If you randomly choose one of Jonathan's developed pictures, what is the probability that it will be flawless?" The examples/problems were displayed in a step-by-step procedure (see Figures 1 and 2). On the first page of an example/problem, the problem givens were displayed. The learners could read them and then go to the next page, where a first solution step was presented or the learners were required to determine a solution step on their own (or at least to attempt it; see Figure 1). After inspecting or determining this solution step, the participants proceeded to the following page where the next solution step was added or required, and so on (see Figure 2). When the whole solution of a problem was presented or required, the next page contained the first page of a new example/problem, and this process repeated itself until the lesson was completed. In the case of omitted solution steps, the learners had to type in a solution attempt; otherwise they could not proceed. Hence, the correctness of the problem-solving attempts could be determined. Note that the correct step was always displayed when the learners went to the next page so that there was feedback on the correctness of the learners' problem-solving attempts.

On the whole, there were two sets of probability tasks. Each set consisted of four tasks with the same underlying structure (solution rationale) but different surface features (cover stories, numbers). In the fading group, the first task was a completely worked-out example. In the second task, the first solution step was omitted.

FIGURE 1. Example with a first missing solution step.

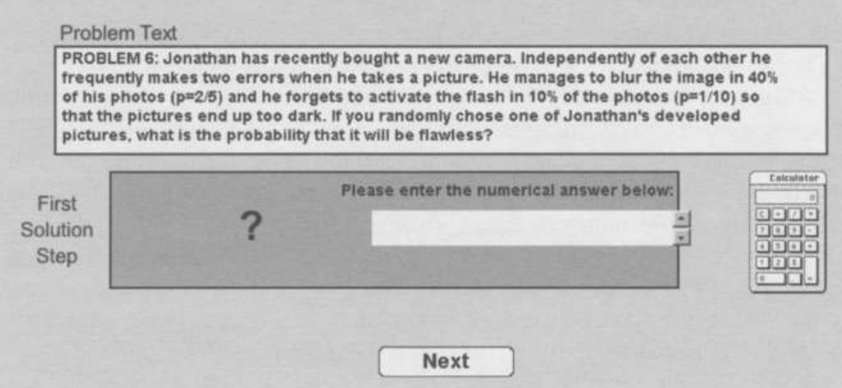
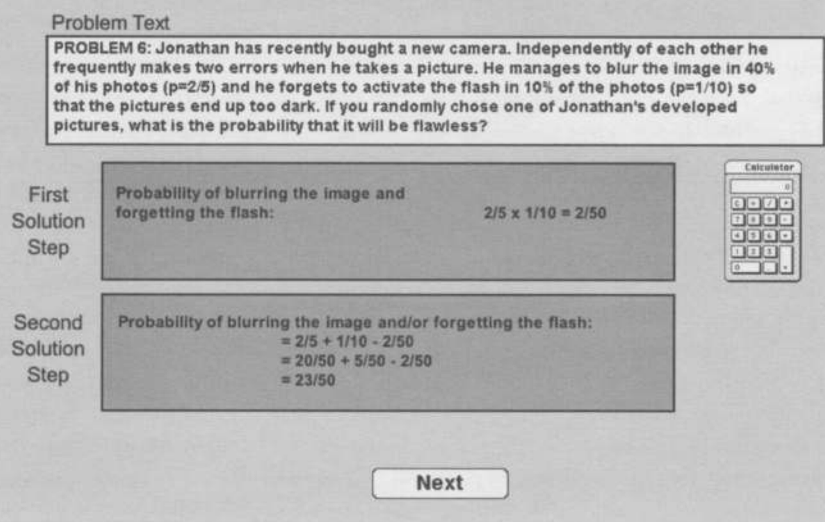


FIGURE 2. Example with a worked-out second solution step.



In the third task, the first two steps were omitted ("forward rationale" of omitting solution steps). The fourth task was essentially a problem-solving task (all three steps were missing). In the example–problem group, two pairs consisting of a completely worked-out example followed by a problem-solving task were presented.

Instruments

We used a pretest to assess prior knowledge. It consisted of nine relatively simple problems involving probability calculation (e.g., "When rolling a 6-sided die what is the probability that '2' or '4' will appear?"). For each correct solution, 1 point was awarded (no partial credit). The maximum score was 9.

The learning outcomes were assessed by a posttest that included 13 problems. In addition to a very simple warm-up problem, which was ignored for further analysis, we used 6 near transfer items and 6 far transfer items. Compared with the examples/problems studied during the learning phase, the near transfer problems had the same underlying structure (solution rationale) but different surface features (cover story, numbers; e.g., "While preparing a batch of rolls at the local bakery, the baker's assistant forgot to add salt to 30% of the rolls and, independent of this event, he burned 40% of the rolls. If the head baker arrives to examine the quality of his assistant's work by randomly testing a roll, what is the probability that it is edible; that is, that it has the right amount of salt and is not burned?"). Far transfer problems differed with respect to both structure and surface features (e.g., "When driving to work, Mrs. Fast has to pass the same traffic light twice—once in the morning and once in the evening. It is green in 70% of the cases. What is the probability that she can pass through a green light in the morning but has to stop in the evening?").

Each posttest problem included three solution steps. One point was awarded for each correct step (partial credit). Thus, a totally correct problem solution was awarded 3 points. The maximum score was 18.

Procedure

The participants worked in group sessions lasting about 90 min. During that time, they worked individually in front of a computer. First, a pretest on prior knowledge in probability calculation was presented. To provide or reactivate basic knowledge that allowed the participants to understand the worked-out examples, we gave an instructional text on basic principles of probability calculation to the participants. After reading this instructional text, the participants were to study the worked-out examples and problems provided by the computer program. In this phase, the experimental variation took place (fading vs. example–problem pairs). The time spent for learning was recorded. Finally, the participants worked on a posttest.

Results

The means and the standard deviations of the two experimental groups for the pretest (prior knowledge), the time spent studying the examples and problems (learning time), the proportion of correct solution steps generated during learning, and posttest performance with regard to near transfer and to far transfer are presented in Table 2. The small difference between the pretest scores in favor of the example–problem group was neither statistically, $t(52) = -0.49$, $p > .10$, nor practically significant, $\eta^2 = .01$. Hence, the groups were a priori comparable with respect to prior knowledge. In addition, the learning time did not significantly differ between groups, $t(52) = 0.28$, $p > .10$, $\eta^2 < .01$. Thus, possible group differences with respect to learning could not be simply attributed to time on task.

With respect to treatment effects, we descriptively obtained substantially higher means in the fading group for the proportion of correct solution steps and for near transfer. We made comparisons between the experimental conditions using ANCOVAs with prior knowledge as the covariate. The precondition of homogeneous slopes for regressions from transfer performance to pretest was not violated (corresponding tests of heterogeneity of slopes for near transfer and for far transfer: both *F*s < 1). The ANCOVAs yielded a statistically significant difference for near transfer performance, $F(1, 51) = 4.58$, $p < .05$ (adjusted means = 10.00 for fading and 7.59 for example–problem pairs). This effect was of medium size,

TABLE 2
Means and Standard Deviations of Pretest Scores, Learning Time (Min), Correctness of Solution Steps During Learning, and the Posttest in Experiment 2

	Fading		Example–problem pairs	
Variable	*M*	*SD*	*M*	*SD*
Pretest	5.00	2.13	5.29	2.33
Learning time	31.15	10.83	30.37	9.41
Correctness of solution steps	7.97[a]	3.79	6.22	3.98
Posttest: near transfer	9.70[a]	5.80	7.89	6.35
Posttest: far transfer	6.96	4.54	7.81	4.35

Note. Possible ranges of the variables: pretest (0–9), correctness of solution step (0–12), near transfer (0–18), and far transfer (0–18).
[a]Differs from example–problem pairs mean on the basis of an ANCOVA (α = .05).

partial η^2 = .08. For far transfer, we obtained a negligible effect, partial η^2 = .01, that did not reach the level of statistical significance ($F < 1$; adjusted means = 7.19 for fading and 7.59 for example–problem pairs). A third ANCOVA revealed that there was also a statistically significant difference between groups with respect to the proportion of correct solution steps, $F(1, 51) = 7.62$, $p < .05$ (no heterogeneous regression slopes, $F < 1$; adjusted means = 8.15 for fading and 6.04 for example–problem pairs). This difference was of medium size, partial η^2 = .13.

To test the mediation hypothesis that fading fosters learning outcomes (at least near transfer) because fewer errors occur during learning, we performed an additional ANCOVA for near transfer performance in which the proportion of correct solution steps was included as a covariate in addition to prior knowledge (no heterogeneous regression slopes of the regressions from near transfer on correctness of solution steps, $F < 1$). The mediation hypothesis would have been supported if the group effect (more or less totally) disappeared in this case (cf. Baron & Kenny, 1986). This proved to be true. The F statistic and the partial η^2 for the group effect were negligible sizes of .23 and .01, respectively (adjusted means = 9.03 for fading, 8.57 for example–problem pairs).

Discussion

In the present laboratory experiment, we conceptually replicated the effectiveness of our fading procedure for near transfer. We obtained this converging result even though the present study and our first investigation differed with respect to the type of learners (low-track students vs. university students), the learning domain (physics/electricity vs. mathematics/probability calculation), the learning setting (school lesson vs. computer-based learning in the laboratory), and the kind of fading out worked-out solution steps (backward vs. forward). We interpreted the stability of this finding despite these very different context conditions as an indicator that our fading procedure has a reliable effect.

It has to be noted, however, that a conceptual replication is not the same as a direct empirical replication. Thus, after Experiment 2, there remained at least some uncertainty as to whether a direct replication of the findings would also succeed. In addition, an open question arose from the fact that we used two ways of fading out worked-out solution steps—a backward and forward procedure—across Experiments 1 and 2. Because the context conditions in our two studies varied substantially, we could not compare the relative effectiveness of these two procedures. This comparison is necessary in order to answer several important questions, principal among them whether the specific type of fading procedure significantly influences learning outcomes or whether it is of minor importance.

EXPERIMENT 3: LABORATORY EXPERIMENT

To address the two open questions that were mentioned in the preceding discussion, relating to the issues of direct replication and type of fading procedure, we conducted a third experiment. To directly replicate the findings of Experiment 2, we implemented identical conditions (example–problem pairs and "forward" fading). In addition, we used the condition of "backward" fading in an effort to examine for potential differences between the two types of fading.

Specifically, we addressed the following two main questions in Experiment 3: (a) Can the results with respect to the (missing) effects of fading on errors during learning, near transfer, and far transfer be replicated? and (b) Do forward fading and backward fading have different learning effects?

Method

Sample and Design

The participants in this study were 45 American students enrolled in several educational psychology courses at a small, northeastern U.S. liberal arts college. They were randomly assigned in equal numbers to the forward fading, backward fading, or the example–problem condition, respectively (n = 15 in each group). As with our two previous experiments, the number of unsolved solution steps (12) was held constant across the three conditions.

Learning Environment

For the forward fading and example–problem groups, we used exactly the same computer-based learning program as in Experiment 2. For the backward fading condition, the program for the forward fading group was modified. Whereas the first solution step in the second task was omitted for the forward fading group, the last solution step was omitted for the backward fading group. In the third task, the first two steps were omitted for the forward fading group, whereas the last two solution steps were omitted for the backward fading condition. The fourth task was essentially a problem-solving task (all three steps were missing) for both fading conditions.

Instruments

The instruments used in this experiment were the same as those used in Experiment 2.

Procedure

The procedure for this experiment was identical to that for Experiment 2.

Results

As previously mentioned, this experiment addressed two basic questions. First, for the replication of the effects of fading, we compared the two fading conditions with the example–problem group [contrast: (forward fading + backward fading) vs. example–problem]. Second, in comparisons of the two fading conditions, the example–problem group was omitted in the corresponding analyses (contrast: forward fading vs. backward fading). In other words, we performed a priori orthogonal contrasts in order to address the research questions.

The means and the standard deviations of the three conditions for the pretest (prior knowledge), the time spent studying the examples and problems (learning time), the proportion of correct solution steps generated during learning, and posttest performance with regard to near transfer and to far transfer are presented in Table 3.

We first report the analyses on the replication question. The small difference between the pretest scores in favor of the two fading conditions was not statistically significant, $t(43) = 1.38$, $p > .10$, and it was of minor effect size, $\eta^2 = .04$. Consequently, the two fading conditions and the example–problem condition can be considered a priori comparable with respect to prior knowledge. In addition, we found no statistically significant difference in learning time between the average of the fading conditions and the example–problem condition, $t(43) = 1.38$, $p > .10$. This effect was also only of minor effect size, $\eta^2 = .04$. Accordingly, possible group differences with respect to learning could not be simply attributed to time on task.

With respect to examining the data for treatment effects, we descriptively obtained substantially higher means in the two fading groups for near transfer, for far transfer, and for the proportion of correct solution steps. As in our previous experiments, we used ANCOVAs (controlling for prior knowledge) to com-

TABLE 3
Means and Standard Deviations of Pretest Scores, Learning Time (Min), Correctness of Solution Steps During Learning, and the Posttest in Experiment 3

	Forward fading		Backward fading		Example–problem pairs	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pretest	6.67	1.72	6.53	1.77	5.80	2.04
Learning time	34.73[c]	6.55	28.07	6.06	28.27	7.47
Correctness of solution steps	10.47	1.92	10.93	1.98	7.60[b]	4.15
Posttest: near transfer	12.40	4.98	11.60	5.10	6.53[a]	4.97
Posttest: far transfer	6.00	3.93	7.87	3.07	4.07[a]	2.91

Note. Possible ranges of the variables: pretest (0–9), correctness of solution step (0–12), near transfer (0–18), and far transfer (0–18).
[a]Differs from fading conditions mean on the basis of an ANCOVA ($\alpha = .05$). [b]Differs from fading conditions mean on the basis of a Mann–Whitney U test ($\alpha = .05$). [c]Differs from backward fading mean on the basis of a *t* test ($\alpha = .05$).

pare the experimental conditions. The precondition of homogeneous slopes for regressions from the posttest measures to pretest was not violated for near transfer (heterogeneity of slopes, $F < 1$) or for far transfer, $F(1, 41) = 1.05$, $p > .10$. Using ANCOVAs, we obtained statistically significant differences between the fading conditions and the example–problem group for near transfer, $F(1, 42) = 9.88$, $p < .05$ (adjusted means = 11.41 for fading and 7.13 for example–problem pairs), as well as for far transfer, $F(1, 42) = 5.45$, $p < .05$ (adjusted means = 6.75 for fading and 4.25 for example–problem pairs). The practical significance of this difference was strong for near transfer, partial $\eta^2 = .19$, and medium for far transfer, partial $\eta^2 = .12$.

Thus, the positive effect of fading on near transfer was replicated. In contrast to our previous studies, we also found a positive effect on far transfer. To further investigate the reasons for this finding, we performed the following post hoc analysis. We compared only forward fading with example–problem pairs—the conditions that were identical with those in Experiment 2—so that the statistical comparisons were fully comparable across both experiments. In this case, we obtained results identical to those in Experiment 2—that is, a statistically insignificant effect on far transfer, $F(1, 27) = 1.22$, $p > .10$, with minor practical significance, partial $\eta^2 = .04$ (adjusted means = 5.72 for fading and 4.35 for example–problem pairs). Testing the backward condition against the control group yielded a statistically significant difference, $F(1, 27) = 10.20$, $p < .05$ (adjusted means = 7.70 for fading and 4.24 for example–problem pairs). The practical significance of this effect was high, partial $\eta^2 = .27$. Hence, the statistically significant effect on far transfer can be attributed primarily to the backward fading condition (see also the descriptive statistics in Table 3).

For the correctness of solution steps, we found heterogeneous regression slopes, $F(1, 41) = 9.59$, $p < .05$. Whereas in the example–problem group, there was a substantial relationship between prior knowledge and correctness of solution steps (unstandardized $b = 1.05$; $r = .69$; $p < .05$), this was not the case for the fading condition (unstandardized $b = .11$; $r = .13$; $p > .10$). The lack of a relationship in the latter group can be explained by a ceiling effect: 14 of the 30 learners in the fading condition solved all steps correctly (minimum = 5 correct solution steps). This ceiling effect is positive news from an educational view; however, it is negative news with respect to statistical issues (violation of preconditions for parametric tests). Hence, we used the nonparametric Mann–Whitney U test for group comparison. It revealed that significantly fewer errors were made in the fading condition, $U = 139$, $z = 2.16$, $p < .05$.

Finally, the hypothesis that the effect of fading is mediated by the correctness of solution steps was tested. We again used ANCOVAs in which the proportion of correct solution steps was included as covariate in addition to prior knowledge. This was possible because there were no heterogeneous slopes of the regressions from near transfer or far transfer on correctness of solution steps (both Fs < 1; homogeneity with respect to pretest; see above). For far transfer, the group differences remained more or less unaffected by the inclusion of correctness of solution steps as covariate, $F(1, 41) = 5.87$, $p < .05$ (former $F = 5.45$), partial $\eta^2 = .13$ (former partial $\eta^2 = .12$; adjusted means = 6.93 for fading and 4.07 for example–problem pairs). Thus, there is no evidence that the effect of fading on far transfer is mediated by the correctness of solution steps. In the case of near transfer, the F statistic dropped from 9.88 to 4.65, but it was still statistically significant, $F(1, 41) = 4.65$, $p < .05$ (adjusted means = 10.85 for fading and 7.69 for example–problem pairs). The effect size dropped from .19 to .10 (partial η^2). Because the effect of fading did not (more or less totally) disappear after controlling for the mediating variable, one should test whether it was at least reduced to a statistically significant amount. For this purpose, we used the test for statistical significance of mediation effects proposed by MacKinnon (1999; MacKinnon & Dwyer, 1993).[2] We obtained a z score of 2.32 ($p < .05$). This pattern of results indicates that the correctness of solution steps contributed to the mediation of the effect on near transfer, but additional mechanisms had to be assumed.

In the following section, we present the analyses pertaining to the second research question that related to the difference between forward and backward fading (for the descriptive statistics, see Table 3). Prior knowledge was comparable in both groups, and no significant pretest difference was found, $t(28) =$

[2]This test procedure included the computation of two regression equations: Mediator = a·Independent + error_1 and Dependent = c·Independent + b·Mediator + error_2. Then the statistically significance of the product ($a \cdot b$) that corresponds to the mediation effect is determined ($z = a \cdot b / se_{ab}$).

0.21, $p > .10$, $\eta^2 < .01$. However, learning time differed statistically significantly, $t(28) = 2.89$, $p < .05$ (see Table 3). The practical significance of this time difference was high ($\eta^2 = .23$). The learners in the forward condition needed over 6 min more for studying the examples/problems. Because we did not find statistically significantly heterogeneous slopes between pretest and correctness of solution steps, $F(1, 26) = 1.81$, $p > .10$, near transfer, $F(1, 26) = 3.72$, $p > .05$, and far transfer ($F < 1$), we performed ANCOVAs for group comparisons with respect to the performance measures. We did not find significant differences between the fading conditions with respect to correctness of solution steps ($F < 1$, partial $\eta^2 = .02$; adjusted means = 10.46 for forward fading and 10.94 for backward fading) and to near transfer ($F < 1$, partial $\eta^2 = .01$; adjusted means = 12.31 for forward fading and 11.69 for backward fading). The group difference in far transfer performance, which was of small to medium size, partial $\eta^2 = .07$, failed to reach the accepted level of statistical significance, $F(1, 27) = 2.12$, $p > .10$ (adjusted means = 5.98 for forward fading and 7.88 for backward fading). Thus, the two fading conditions did not differ with respect to their effectiveness, but the backward fading condition was more efficient (i.e., required less learning time).

Discussion

Although there is little doubt after the third experiment that near transfer is fostered by the fading procedure, the far transfer issue remains more open. Interestingly, when a forward procedure was used (Experiment 2 and the corresponding group in Experiment 3), there was very little indication that far transfer was fostered. The use of a backward procedure resulted in a statistically significant effect in Experiment 3 and—although we obtained no statistically significant effect in Experiment 1—we found at least descriptively a higher solution rate of far transfer problems under the backward procedure (see Table 1). A sensible preliminary conclusion from this pattern of results is that forward fading does not substantially foster far transfer, whereas backward fading has the potential to accomplish this objective. Whether this is actually true has to be tested in further studies.

Beyond the question of far transfer effects by backward fading, this type of fading procedure is more favorable than forward fading because it is more efficient. The learners in the backward condition spend less time on the examples without having disadvantages in transfer performance.

Because we obtained differences between the two fading procedures, we took a closer look at these two conditions. We identified three major differences between the two conditions that are, of course, interrelated:

1. *Delay of problem-solving.* With backward fading, problem-solving demands are delayed longer because the last step of the second problem is faded out, whereas in forward fading, the first step is faded out. From a cognitive load

perspective, the delayed problem solving may be an advantage because this approach may prevent cognitive overload.

2. *Delay of feedback.* However, by delaying the demand for problem solving, learners have to wait longer for feedback about their competence in solving the problems at hand. Thus, the delay for feedback is shorter in the forward fading condition.

3. *Problem-specific information as support.* When one is determining the first solution step of problems in the forward fading condition, no problem-specific information designed to support problem solving is available to the learner. In contrast, in the backward condition, "contextual information" from preceding solution steps is available, which may provide some scaffolding for problem solving. With respect to the experimental comparisons, the outlined "instructional" differences also imply that the forward condition was more similar to the example–problem condition than to the backward condition because the two former conditions diverged only after the initial to-be-determined solution step was presented.

It is our contention that the aforementioned differences between the two fading conditions cause a divergence with respect to efficiency and possibly to far transfer. It may even be assumed that the error reduction in both conditions is affected by different causes (e.g., earlier feedback in the forward condition, longer delay of problem solving, and less cognitive load in backward fading). To clarify this question, further experimentation—including some fine-grained analysis of learning processes—is necessary.

The finding that fading has reliable effects on near transfer but not on far transfer may have something to do with the mediating mechanism that was identified in this study. The analyses showed that the effect on near transfer is (at least in part) mediated by the number of errors committed during learning. "Error-reducing" instructional procedures such as direct instruction (e.g., Rosenshine & Stevens, 1986) or drill-and-practice tutorials are known to effectively foster "low-level" learning or near transfer (cf. also Greeno, Collins, & Resnick, 1996). Because our fading procedure is a method of reducing errors during learning, it is understandable why it fosters primarily near transfer performance. This interpretation is compatible with the finding from Experiment 3 that the effect of fading on far transfer that was found was obviously not mediated by the correctness of the solution steps. For far transfer to be fostered, error reduction is not sufficient. Instead, some deliberate reflection that deepens understanding may be necessary (cf. VanLehn, 1996).

GENERAL DISCUSSION

The following significant contributions to the field of research on example-based learning were made:

1. A new feature for the design of materials for example-based learning—namely fading—was introduced. Fading (a) integrates example study and problem solving and (b) builds a bridge between example study in early phases of cognitive skill acquisition to problem solving in later stages.
2. In particular, fading as a feature of example-based learning appears to be effective, at least with respect to near transfer. The finding was replicated and shown to be stable across context variables such as field versus laboratory.
3. The number of problem-solving errors plays a role in mediating the effects of fading on near transfer.
4. It is more favorable to fade out worked-out solution steps in a backward manner than in a forward manner.

An important question that emerged from our research involves the generalizability of the present results. Research on learning from worked-out examples has yet to establish if there are systematic differences between content areas in which worked-out examples can be sensibly constructed (see Atkinson et al., 2000). To date, besides programming, example-based research has focused almost exclusively on mathematics and "mathematized" contents from domains such as physics and economics. Similarly, we also obtained converging results about the two mathematized content (sub)domains that we used—that is, physics/electricity and mathematics/probability calculation. Hence, we think it is appropriate to assume that the present results are generalizable across learning contents that can be taught by mathematized worked-out examples. The extent to which our fading procedure can be modified so that it can be applied to other, nonmathematized content domains and the extent to which such a modified instructional procedure would yield positive effects, however, cannot be evaluated on the basis of the present findings.

The present work has also clear practical implications. There are not only two laboratory experiments demonstrating the effectiveness of fading worked-out solution steps but also an experiment that shows how to implement the fading procedure in the field. This field study has high ecological validity because the lesson including fading was performed as regular school instruction and the findings were obtained with learning content from the obligatory curriculum. In addition, the description in the section "Learning Environment" (Experiment 1) shows how a backward fading procedure can be integrated into classroom instruction with very simple means such as work sheets and overhead transparencies. Thus, a major advantage is that learning from worked-out examples using a fading procedure can be easily implemented and is compatible with ordinary framework conditions in schools. This latter point is not trivial at all because the successful implementation of other currently intensively discussed instructional procedures such as situated learning or problem-based learning is often impeded by the organizational context conditions in schools. In sum, one can claim that fading is a procedure that is effective and easy to implement and that can be recommended to teachers.

Although the present work brought about some significant insights, a number of new research questions also surfaced. The issue of the extent to which fading can foster far transfer needs further investigation. Researchers should test whether backward fading actually has more potential to foster far transfer than forward fading. In addition, it may well be that other fading procedures are even more effective, especially with respect to far transfer. For example, one could first omit the solution step that is (on the average) the easiest one for the learners to determine, then the second easiest one, and so on. According to Sweller's cognitive load theory (e.g., Sweller et al., 1998), this procedure would reduce problem-solving burden so that cognitive capacity for self-explanations is left which, in turn, could foster far transfer. This leads to another important issue for further study. The mediating mechanisms of the fading effect—both forward and backward—need further investigation. In particular, researchers should explore the extent to which self-explanations play a role in mediating fading effects beyond the error-reduction mechanism. Taken together, follow-up studies should address the following aspects: fostering of far transfer, different fading procedures, and mediating mechanisms of different fading procedures.

In summary, with this study we have provided strong evidence for the effectiveness of our new rationale for the integration of example study and problem solving. However, for us to deeply understand the way this works and to optimize the use of this rationale, further experiments are necessary.

REFERENCES

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4,* 167–207.

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

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.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Chi, M. T. H. (2000). Self-explaining: The dual process of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Erlbaum.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182.

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.

Gräsel, C., & Mandl, H. (1993). Förderung des erwerbs diagnostischer strategien in fallbasierten lernumgebungen [Promoting the acquisition of diagnostic strategies in case-based learning environments]. *Unterrichtswissenschaft, 21,* 355–369.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 15–46). New York: Macmillan.

MacKinnon, D. P. (1999). *Statistical mediation.* Retrieved from: http://www.public.asu.edu/~davidpm/ripl/mediate.htm

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17,* 144–158.

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

Merriënboer, J. J. G. van, & de Crook, M. B. M. (1992). Strategies for computer-based programming instruction: Program completion vs. program generation. *Journal of Educational Computing Research, 8,* 212–234.

Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction, 16,* 173–199.

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

Reimann, P. (1997). *Lernprozesse beim wissenserwerb aus beispielen: Analyse, modellierung, förderung* [Learning process in knowledge acquisition from examples: Analysis, modeling, enhancement]. Bern, Switzerland: Huber.

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

Rosenshine, B. V., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376–391). New York: Macmillan.

Stark, R. (1999). *Lernen mit lösungsbeispielen. Einfluß unvollständiger lösungsbeispiele auf beispielelaboration, motivation und lernerfolg* [Learning by worked-out examples: The impact of incomplete examples on example elaboration, motivation, and learning outcomes]. Bern, Switzerland: Huber.

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 learning method: How effective is the combination of worked-out examples and problems to-be-solved?]. *Zeitschrift für Pädagogische Psychologie, 14,* 206–218.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2,* 59–89.

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

Trafton, J. G., & Reiser, B. J. (1993). The contributions of studying examples and solving problems to skill acquisition. In M. Polson (Ed.), *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47,* 513–539.