When less is sometimes more: Optimal learning conditions are required for schema acquisition from multiple examples

Katharina Scheiter, Peter Gerjets

▶ To cite this version:

Katharina Scheiter, Peter Gerjets. When less is sometimes more: Optimal learning conditions are required for schema acquisition from multiple examples. 27th Annual Conference of the Cognitive Science Society, 2005, 2005, Stresa, Italy. pp.1943-1948, 2005. <hal-00197353>

HAL Id: hal-00197353

https://telearn.archives-ouvertes.fr/hal-00197353

Submitted on 14 Dec 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

When Less is Sometimes More: Optimal Learning Conditions are Required for Schema Acquisition from Multiple Examples

Katharina Scheiter (k.scheiter@iwm-kmrc.de)
Department of Applied Cognitive Psychology and Media Psychology, University of Tuebingen
Konrad-Adenauer-Strasse 40, 72072 Tuebingen, Germany

Peter Gerjets (p.gerjets@iwm-kmrc.de)
Multimedia and Hypermedia Research Unit, Knowledge Media Research Center
Konrad-Adenauer-Strasse 40, 72072 Tuebingen, Germany

Abstract

While it is usually claimed that multiple examples for the illustration of problem categories are a necessary prerequisite for schema acquisition, there is a lack of conclusive empirical evidence supporting this claim. Moreover, there are findings indicating that carefully designed one-example conditions may allow for profitable processes of example comparison as well. In line with this reasoning, we present an experiment – that builds up on a series of studies conducted by Quilici and Mayer (1996) and a previous experiment by our own research group (Scheiter, Gerjets, & Schuh, 2004) – that demonstrates that multiple examples may only be helpful for learning when sufficient time for their processing is provided and when learners are guided by an instruction to compare examples. On the other hand, learning from single examples proved to be less dependent on optimal learning conditions. Results are discussed in the light of current instructional design theories.

Keywords: schema acquisition; number of examples; cognitive load; comparison; problem solving

Schema Acquisition from Multiple Examples

Problem schemas are sophisticated knowledge structures that enable solving problems in knowledge-rich domains. A problem schema "allows problem solvers to group problems into categories in which the problems in each category require similar solutions" (Cooper & Sweller, 1987, p. 348). Once a problem has been identified as belonging to a known problem category the relevant schema is retrieved from memory, information that is specific to the to-be-solved problem (e.g., concrete objects, variable values) is filled into the slots of the schema (schema instantiation), and the solution procedure that is attached to the schema is executed in order to produce a solution to the problem (cf. VanLehn, 1989).

Schema-based problem solving is very efficient and therefore often seen as a marking feature of experts' problem solving. Because problem schemas are crucial for proficient problem solving, research has often focused on the question of how such schemata can be acquired. Bernardo (1994, p. 379) argues that there is "a consensus that problem-type schemata are acquired through some inductive or generalization process involving comparisons among similar or analogous problems of one type."

Therefore, a widely proposed instructional method for fostering the acquisition of problem schemata is to present multiple examples for each problem category conveyed (Cooper & Sweller, 1987; Cummins, 1992; Gick & Holyoak, 1983; Reed & Bolstad, 1991; Sweller & Cooper, 1985; Quilici & Mayer, 1996). This method enables comparisons of examples that belong to the same problem category and therefore, fosters two processes of abstraction: First, learners can determine features that appear in each of the category's examples (i.e., commonalities) and thus may be properties that a problem must possess in order to be an instance of this particular problem category. Thus, these shared properties of examples may potentially be the structural features that determine a problem's membership to a specific problem category and that cannot be altered without altering the solution procedure that applies to a problem. Second, example comparisons within a problem category may enable learners to identify features that vary between the category's examples (i.e., differences) and that are therefore obviously irrelevant with regard to the applicability of the solution principle that is attached to this particular problem category. Therefore, these varying characteristics of examples are indubitably surface features that concern only the problem's cover story (Ross, 1989). Thus, by comparing multiple instances within a category with regard to their commonalities and differences all perceived features of the examples can be hypothetically classified as either being structural or surface features.

Despite the fact that many researchers advocate the provision of multiple examples, there is not much empirical evidence to support this claim. In particular, there is a lack in studies that directly compare single- and multiple-example conditions. An exception to that is a study conducted by Quilici and Mayer (1996), who presented their participants with zero, one, or three example word problems for each of three problem categories from the domain of statistics for studying. Each of the examples was couched into a different cover story. Subsequently, participants had to sort test problems from these three problem categories according to their structural features. The results showed that both example groups outperformed the zero-example group, whereas contrary to the initial expectations there were no performance differences between the single-example and the multiple-example groups. In order to account for this lack of difference Quilici and Mayer (1996) argued that it might not be sufficient to merely present multiple examples per problem category but that it is necessary to carefully design example combinations so that they allow for useful inferences with regard to structural and surface features of the examples presented.

Therefore, in Experiment 2 they compared two different instructional example sets that each contained three examples per problem category. In a surface-emphasizing example set the three examples that belonged to the same problem category were all couched into the same cover story and every problem category was illustrated by a different cover story. In contrast, in a structure-emphasizing example set each problem category was illustrated by examples with three different cover stories. The same three cover stories were used for every problem category. The results showed that participants in the structure-emphasizing example condition were more likely to sort the test problems on the basis of the structural similarities among the problems compared to participants in the surface-emphasizing example condition. This superiority of structure-emphasizing example sets could be demonstrated not only for sorting tasks, but also for problem solving measures where participants later had to solve isomorphic test problems by themselves (Quilici & Mayer, 1996, Experiment 4; Schorr, Gerjets, Scheiter, & Laouris, 2002). These experiments support the assumption that some sets of multiple examples may be more helpful than others, in particular, if they are designed to foster specific cognitive processes like inferring the examples' commonalities and differences. However, they do not tell us anything regarding the effectiveness of multiple examples in general.

Based on the idea that the cognitive processes which are enabled by examples are more important predictors for learning outcomes than the mere number of examples itself, we argued in a previous paper (Scheiter, Gerjets, & Schuh, 2004) that there might be alternative effective example-processing strategies that rely on single examples per problem category, namely comparing examples *across* problem categories. In the aforementioned second experiment by Quilici and Mayer (1996) across-category comparisons (based on multiple examples, however) as well as within category-comparisons have been supported in the structure-emphasizing example condition: If a learner compares examples *within* a category he or she would recognize that there are features that vary among the examples, that is, features that are related to the cover story of the problem. Because the learner knows that the examples belong to the same problem category, he or she may arrive at the conclusion that these features must be irrelevant to the problem's solution procedure, whereas the features that are common to all examples within a problem category may be causally related to the solution procedure. Because in the structure-emphasizing condition examples from different problem categories are couched in the same cover story, the same conclusions as from within-category comparisons may be derived by comparing examples *across* problem categories: Surface features are those features that the examples have in common although belonging to different categories, whereas differences between the examples may indicate structural features.

Because structure-emphasizing example sets allow for two profitable strategies of comparing instructional examples, it is not clear whether the superiority of the structure-emphasizing example condition goes back to enabling within-category comparisons, across-category comparisons, or both.

Quilici and Mayer's Experiment 3 finally provided some initial support for the suspicion that across-category comparisons may be likewise effective for schema acquisition. In this experiment participants had to study four instructional examples in order to acquire knowledge on two problem categories. These examples were presented in two training sessions. Besides the quality of the example sets (structure-emphasizing versus surface-emphasizing) the manner of presentation (mixed versus blocked) was varied. In the blocked format the first training session contained two examples belonging to one problem category and the second training session contained two examples belonging to the other category used for this experiment. However, in the mixed presentation mode participants studied one example of each category in the first training session and the remaining two examples in the second session. It was hypothesized that both manners of presentation enable different example comparisons: "In the mixed condition, students have the opportunity to notice which features differ between problem types, whereas in the blocked condition, students have the opportunity to notice which features are the same within a problem type" (Quilici & Mayer, 1996, p. 156). Thus, the blocked conditions afforded within-category comparisons, whereas the mixed conditions enabled across-category comparisons. The results first replicated the finding that structure-emphasizing example sets are superior to surface-emphasizing example sets. Second, participants in the mixed conditions performed comparably well as those in the blocked conditions, indicating that across-category comparisons may be similarly effective for schema acquisition as within-category comparisons.

Moreover, learning from multiple examples per problem category may even be harmful if these examples are not appropriately processed. In particular, studying multiple examples can be very demanding because of the vast amount of information that has to be processed simultaneously in order to identify the commonalities and differences among the examples. That is, although comparing multiple examples may be a very successful way of learning it requires a lot of time and effort to be devoted to the learning task which may in turn result in a substantial amount of cognitive load (Sweller, van Merriënboer, & Paas 1998). On the other hand, if learners process multiple examples only cursorily, they may get easily confused, which will have a negative impact on learning outcomes. Single examples are probably less vulnerable to effects of inappropriate, i.e., not sufficiently intense processing of examples because there is less information that has to be properly processed from the very beginning. This also implies that learning from multiple examples per problem category may require instructional support that aims at fostering the appropriate processing of examples, whereas probably no instructional guidance is needed when single examples are given.

We provided evidence for these claims in an experimental study in which we made use of the materials and experimental setting introduced by Quilici and Mayer (1996).

Necessary Instructional Conditions for Learning from Multiple Examples

In the study by Scheiter et al. (2004) students received either one or three examples illustrating each of three problem categories. The problem categories and the examples were identical to the ones used by Quilici and Mayer (1996). In both example conditions the cover stories of the examples were the same across problem categories thereby enabling profitable processes of comparison. In the three-examples condition the examples of each problem category were embedded in three different cover stories, which were used across problem categories (i.e., surface-emphasizing example sets). Participants could study the examples as long as they liked to.

In addition to varying the number of examples per problem category, we manipulated whether there was an additional instruction to compare the examples or not. Whereas some participants only received the instruction to study the examples carefully in order to understand them, others were additionally told to compare the examples and to particularly pay attention to the examples' commonalities and differences. This comparison instruction left open whether to compare examples across categories or within categories (in the three-examples condition).

The time students took to study the examples was registered and used to distinguish among participants who had studied examples in a cursory manner and those who had studied them intensively by means of a median split with regard to example-study time conducted within each of the four experimental conditions. The resulting variable (i.e., intensity of example study) was used as a third independent variable in the design of the study.

After participants had indicated having studied the examples sufficiently, these were removed from the table and the participants were told to sort test problems according to the features that seemed to be relevant to their solution. The problems were identical to those used by Quilici and Mayer (1996).

Analyzing participants' outcomes by means of a three-factor ANOVA (number of examples x availability of comparison instruction x intensity of example study) revealed that providing multiple examples increased the time needed for learning and was moreover accompanied with improvements regarding the identification of structural similarities only when learners studied the examples intensively and when they further received instructional support by prompting them to compare examples and to identify the commonalities and differences. Under less optimal learning conditions, i.e., when the examples were processed without any additionally instructional guidance, multiple examples even tended to worsen performance. Providing one example per problem category proved to be far less dependent on the presence of optimal instructional conditions and on adequate learning behavior. Multiple examples thus might be recommendable only under rather optimal learning conditions.

There is however a limitation to this previous study which we aim at addressing in the current study. In the previous study we distinguished between participants who either processed examples cursorily or intensively by taking into account the time they decided to devote for learning. There may however be other differences between these learners besides the time spent on examples, which may be responsible for the obtained effects. In the current study we thus assigned students to conditions with fixed learning times in order to investigate the relationship between example-processing time, number of examples, and available instructional support by means of a comparison instruction more thoroughly.

Experiment

Method

Participants. Eighty-one students of the University of Tuebingen, Germany, participated either for course credit or payment. Average age was 22.6 years. The participants had already taken or were currently enrolled in a statistics course and were thus familiar with the domain used for experimentation.

Materials and Procedure. The materials were identical to the ones in the study by Scheiter et al. (2004). Participants first had to fill in a multiple-choice questionnaire that contained 12 items dealing with basic concepts and terms of descriptive and inferential statistics (e.g., what is expressed by a correlation between two variables?, Which testing procedure can you apply to frequency data?). After having filled in the questionnaire they were either informed that there would be either one example or three examples illustrating three problem categories from statistics depending on the experimental condition. Half of the students were additionally told to compare the examples and to particularly pay attention to the examples' commonalities and differences. Subsequently, participants received a booklet that contained either one example or three examples for each problem category. In the one-example conditions there was one example problem that was typical for a t-test, one typical for a correlation, and one typical for $\chi 2$-test. The cover stories of the examples were the same across problem categories thereby enabling profitable processes of comparison. We used three different kinds of cover stories in order to control for possible effects of the surface features of the examples. In the three-examples condition, these three cover stories were used to construct surface-emphasizing example sets. That is, the three examples of each problem category were embedded in three different cover stories, which were used across problem categories.

Contrary to the previous experiment, participants had a fixed amount of time available for studying the examples that depended on experimental condition. After this time had exceeded, the examples were removed from the table and the participants conducted the sorting task used in the previous experiment. They were instructed to sort 12 problems according to features relevant to their solution. There were always four test problems that belonged to one problem category and each of these problems had a different cover story. The same four cover stories were used across categories. Participants were informed that they could build as many categories as they wanted and that the problems might not divide evenly upon the categories. Participants who had been told to compare examples with regard to their commonalities and differences received this comparison instruction again for the sorting task. Participants who had been told to compare examples with regard to their commonalities and differences received this comparison instruction again for the sorting task. There were no time limits for working on the sorting task.

Design and Dependent Measures. As a first independent variable we varied the number of examples per problem category (one versus three) that were presented to participants. The second independent variable consisted in the presence or absence of the comparison instruction for learning as well as for the sorting task. Learning time was fixed at three different levels as a third independent variable. Learners in the one-example conditions were either allowed 150 or 300 seconds for studying the examples, learners in the three-examples conditions either had available 300 or 600 seconds. These time limits correspond to the mean times used by learners in the previous experiment (Scheiter et al. 2004), who had studied one example cursorily (150 sec.) or intensively (300 sec.) or three examples cursorily (300 sec.) or intensively (600 sec.).

This research design allowed for answering three questions: (1) How do learning time and a comparison instruction affect learning from one example? (2) How do learning time and a comparison instruction affect learning from three examples per category? (3) How do the number of examples and a comparison instruction affect learning when the time for studying the examples is fixed?

As dependent variables we registered the sorting time and the quality of the sorting performance by determining the structure and the surface score introduced by Quilici and Mayer (1996). The rationale for determining these scores is as follows: From the 12 test problems (12 * 11)/2 = 66 pairs of problems can be build, with 18 pairs containing problems that are members of the same problem category and 12 pairs containing problems that share the same cover story. The remaining 36 pairs contained problems that share neither structure nor surface features. In order to determine the structure score for a person one has to count the number of problem pairs that have been correctly identified as sharing the same structural features by putting them into the same category and dividing this number by 18 (i.e., the highest possible score). The structure score expresses a participant's ability to categorize problems according to their structural similarities and can therefore be seen as a measure for the successful acquisition of problem schemata. On the contrary, the surface score indicates a participant's tendency to sort problems according to surface similarities and is determined by counting the number of problem pairs that have been assigned to the same problem category although only sharing the same cover story and by dividing this number by 12. For the ease of interpretation the scores were transformed into percentages.

Results

The results section is divided into three parts each addressing one of the three research questions. First, we thus analyzed the one-example conditions for effects of learning time and the comparison instruction, and then we conducted the same analyses for the three-example conditions. We end by contrasting the outcomes of the one- and the three-example conditions with identical learning times.

How do learning time and a comparison instruction affect learning from one example per category? An ANOVA (learning time x comparison instruction) for the one-example conditions (columns one and two in Table 1) revealed that neither learning time nor the comparison instructed affected the participants' ability to sort problems according to their structural features as reflected in the structure score (learning time: $F < 1$; comparison instruction $F(1, 38) = 1.39$; $MSE = 773.53$; $p > .20$; interaction: $F < 1$). However, providing more learning time slightly decreased learners' tendency of being misled by the problems' surface features in the sorting task ($F(1, 38) = 2.88$; $MSE = 506.01$; $p < .10$). The comparison instruction had no effect on the structure score ($F < 1$) nor was there an interaction ($F(1, 38) = 1.57$; $MSE = 506.01$; $p > .20$). Finally, learning time affected the efficiency by which participants sorted problems: Participants, who were allowed to spend more time processing the examples, also took more time to complete the sorting task ($F(1, 38) = 8.91$; $MSE = 124184.77$; $p = .005$). There was no main effect of the comparison instruction or an interaction for the sorting time (both Fs < 1).

How do learning time and a comparison instruction affect learning from three examples per category? The same analyses were conducted for the three-examples conditions (columns three and four in Table 1). Learning time positively influenced participants' ability to sort problems according to their structural features ($F(1, 35) =$

4.19; MSE = 790.51; $p < .05$). While it seemed that the comparison instruction also increased the structure score, the respective contrast failed to reach statistical significance ($F(1, 35) = 2.06$; MSE = 790.51; $p > .15$). There was no interaction between the two factors ($F < 1$). Participants' tendency to sort problems according to relevant surface features was decreased by both, the comparison instruction ($F(1, 35) = 6.53$; MSE = 658.68; $p < .02$) and, though less pronounced, by learning time ($F(1, 35) = 3.42$; MSE = 658.68; $p < .10$). There was no interaction between the two factors ($F < 1$). Finally, there were no significant effects for sorting time (comparison instruction: $F(1, 35) = 1.12$; MSE = 89329.37; $p > .25$; learning time: $F < 1$; interaction: $F(1, 35) = 1.07$; MSE = 89329.37; $p > .30$).

How do the number of examples and a comparison instruction affect learning when the learning time is fixed? This final analysis assessed the effects of the number of examples whose processing took place within a fixed amount of time (i.e., 300 sec) and either was or was not guided by the comparison instruction (columns two and three in Table 1). An ANOVA (number of examples x comparison instruction) revealed no significant effects for the structure score (number of examples: $F < 1$; comparison instruction: $F(1, 34) = 1.85$; MSE = 713.74; $p > .15$; interaction: $F(1, 34) = 1.26$; MSE = 713.74; $p > .25$). Additionally, there were no significant main effects for the surface score (number of examples: $F(1, 34) = 1.95$; MSE = 537.06; $p > .15$; comparison instruction: $F(1, 34) = 2.30$; MSE = 537.06; $p > .10$). However, a significant interaction between the two factors ($F(1, 34) = 4.82$; MSE = 537.06; $p < .05$) indicated that without a comparison instruction three examples compared to one example per problem category slightly increased the participants' tendency to sort problems according to their surface features ($t(15) = -1.87$; $p = .08$). There were no differences between the two example conditions when additional instructional support was given ($t(15) = 0.87$; $p = .40$). Finally, learners needed less sorting time when multiple examples were provided $F(1, 34) = 4.09$; MSE = 137900.66; $p = .05$). Sorting time was unaffected by the comparison instruction and there was no interaction (both Fs < 1).

Summary and Discussion

Many researchers have claimed that learning from multiple examples is superior to having only one example for the illustration of each problem category. However, direct tests of this assumption have yet failed to provide empirical support for this claim (Gerjets, Scheiter, & Tack, 2000; Quilici & Mayer, 1996; Scheiter et al., 2004). Moreover, it can be assumed that learning from multiple examples more heavily relies on optimal learning conditions compared to learning from single examples.

In the current study learning from one example per problem category was improved the more learning time was available in that learners tended to be less easily misguided by the problems' surface features. However, these improvements in sorting performance were accompanied by the fact that learners took more time to complete the sorting task. Accordingly, providing more learning time for studying one example per problem category did not prove to be a very efficient means of fostering schema acquisition. On the other hand, performance was left unaffected by the availability of a comparison instruction. These results partially resemble findings from the Scheiter et al. (2004) study, where sorting performance proved to be rather stable across different one-example conditions. This robustness of learning from single examples may be especially helpful for situations in which full control of possible moderating factors (e.g., the time available for learning, the time learners are willing to spend with the instructional materials, the absence or presence additional instructional guidance) cannot be guaranteed.

In the three-example conditions favorable learning conditions like extended learning times or additional instructional support improved learning outcomes to a larger extent than in the one-example conditions. In particular, they also increased the effectiveness of the instruction by inducing a problem schema based on the problems' structural features rather than on their irrelevant surface features. The latter tendency of assigning meaning to surface features was reduced to zero percent, once learners were given sufficient time to process multiple examples and an instruction guided them to compare the examples. On the other hand, when not given sufficient time and without

Table 1: Performance as a function of the number of examples, learning time, and the presence of a comparison instruction

		Number of examples			
		1		3	
		Learning time		Learning time	
		150 sec	300 sec	300 sec	600 sec
Without comparison instruction	Structure score (%)	38.4	46.9	36.1	63.3
	Surface score (%)	29.9	9.3	36.5	14.2
	Sorting time (sec)	565.0	977.1	742.4	847.0
With comparison instruction	Structure score (%)	56.7	49.0	57.8	67.7
	Surface score (%)	17.5	14.4	8.3	0.0
	Sorting time (sec)	755.3	996.3	740.2	645.4

further guidance, learners' problem schemas were based on surface features (36.1 %) to the same extent as they were based on structural features of the problems (36.5%).

Comparing the conditions with single or multiple examples to each other furthermore revealed that there were no other benefits achieved by presenting multiple examples besides reducing the time needed to sort the problems. That is, the recognition of a problem's assumed category membership was faster when learning from multiple examples, which may indicate a more stable or higher automated schema. However, this schema was by no means more accurate than when learning from single examples only. Moreover, when not accompanied by an instruction to compare examples multiple examples even tended to increase the participants' tendency to be misled by surface features. This again confirms findings from the Scheiter et al. (2004) study that only under optimal instructional conditions learning from multiple examples may be superior to learning from single examples. In fact, as simple contrasts for the structure score revealed the one-example condition in which only 150 seconds were provided as learning time and in which there was no instructional guidance was outperformed only by the two three-example conditions in which the learning time was four times as long. Note that this is a time span only half of the learners had been willing to invest when being given the choice to determine their own learning time in the Scheiter et al. study (2004).

We found this lack of a general advantage of multiple examples in both our studies, although we used structure-emphasizing example sets that were carefully designed to allow for useful inferences (Quilici & Mayer, 1996). Accordingly, we recommend presenting less instructional materials to learners, which should however be carefully adjusted to the cognitive processes assumed to foster schema acquisition. Presenting multiple examples may result in a so-called redundancy effect (Sweller et al. 1998), because information on structural features is repeatedly presented across all examples of a problem category. Nevertheless, a learner has to process and compare all the information first, before he or she will recognize that – after the first example has been processed – subsequent examples do not contain any new relevant information. Processing redundant information thus demands cognitive resources (e.g., time) no longer available for other, more helpful cognitive processes and thus may even hinder learning. This reasoning sheds some doubts on instructional design recommendations given in constructivist approaches according to which experience with multiple instances illustrating a common principle from different perspectives is required to develop higher-level knowledge structures (e.g., Cognitive Flexibility Theory, Spiro & Jehng, 1991). According to our findings instructional designers have to carefully decide whether the new aspects delivered through multiple examples are worth the danger that can result from presenting partially redundant information.

Acknowledgments

We thank Antonia Baumeister, Eva Schmetz, and Brigitte Vollmann for conducting the experiment.

References

Bernardo, A. B. I. (1994). Problem-specific information and the development of problem-type schemata. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 379-395.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation of mathematical problem-solving transfer. *Journal of Educational Psychology, 79*, 347-362.

Cummins, D. D. (1992). Role of analogical reasoning in the induction of problem categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1103-1124.

Gerjets, P., Scheiter, K., & Tack, W. H. (2000). Resource-adaptive selection of strategies in learning from worked-out examples. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1-38.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88*, 144-161.

Reed, S. K., & Bolstad, C. A. (1991). Use of examples and procedures in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 753-766.

Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 456-468.

Scheiter, K, Gerjets, P., & Schuh, J. (2004). The impact of example comparisons on schema acquisition: Do learners really need multiple examples? In Y. B. Kafai, W. A. Sandoval, N. Enyedy, A. S. Nixon, & F. Herrera (Eds.), *Proceedings of the 6th International Conference of the Learning Sciences.* Mahwah, NJ: Erlbaum.

Schorr, T., Gerjets, P., Scheiter, K., & Laouris, Y. (2002). Designing sets of instruction examples to accomplish different goals of instruction. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Spiro, R. J., & Jehng, J.-C. (1990). Cognitive flexibility and hypertext: Theory and technology for the nonlinear and multidimensional traversal of complex subject matter. In D. Nix & R. J. Spiro (Eds.), *Cognition, education, and multimedia.* Hillsdale, NJ: Erlbaum.

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

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

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of cognitive science*. Cambridge, MA: MIT-Press.