The Impact of Example Comparisons on Schema Acquisition: Do Learners Really Need Multiple Examples?

Katharina Scheiter
Department of Applied Cognitive Psychology and Media Psychology, University of Tuebingen
Konrad-Adenauer-Strasse 40, 72072 Tuebingen, Germany
Tel: ++49-7071-979330, Fax: ++49-7071-979100
k.scheiter@iwm-kmrc.de

Peter Gerjets
Multimedia and Hypermedia Research Unit, Knowledge Media Research Center
Konrad-Adenauer-Strasse 40, 72072 Tuebingen, Germany
Tel: ++49-7071-979219, Fax: ++49-7071-979100
p.gerjets@iwm-kmrc.de

Julia Schuh
Virtual PhD Program: Knowledge Acquisition and Knowledge Exchange with New Media,
Konrad-Adenauer-Strasse 40, 72072 Tuebingen, Germany
Tel: ++49-7071-979322, Fax: ++49-7071-979100
j.schuh@iwm-kmrc.de

Abstract: Comparing multiple examples within problem categories is usually considered a necessary prerequisite for schema acquisition. However, there is an evident 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) – yielding that multiple examples are only helpful for schema acquisition if the examples are processed thoroughly and if additional instructional support is provided for learners. Moreover, our study gives first evidence that multiple examples may even impede performance under less optimal learning conditions.

Schema Acquisition and the Comparison of Multiple Examples per Category

It has often been argued that probably the most important prerequisite for successful problem solving in knowledge-rich domains consists in the availability of problem schemas. "Schemas are defined as mental constructs that allow patterns or configurations to be recognized as belonging to a previously learned category and which specify what moves are appropriate for that category" (Sweller & Cooper, 1985, p. 60). 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 (VanLehn, 1996). 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 content or cover story (Ross, 1989). Thus, by comparing multiple instances within a problem 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.

In this paper we will suggest that - although comparison processes are of central importance for schema acquisition - it may not be necessary to compare multiple examples within problem categories (*within-category comparison*). Rather, there may be alternative example-processing strategies that rely on single examples per problem category and that may be similarly effective, namely comparing examples across problem categories (*across-category comparison*). 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 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). However, if learners do not invest these mental resources in learning there may even be harmful effects of providing multiple examples compared to presenting only one example per problem category: If being only cursorily processed, multiple examples may result in confusion of the learner. 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. Therefore, instructional designs based on single examples per problem category may be more recommendable to account for those students who are not willing to devote a lot of effort into learning, i.e. for “lazy” as compared to hard-working students. As a consequence, 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.

In order to test these assumptions concerning (1) the effectiveness of multiple examples compared to single examples per problem category, (2) the impact of appropriately processing multiple or single examples, and (3) the differential necessity for instructional guidance in example-based learning we conducted an experimental study in which we made use of the materials and experimental setting introduced by Quilici and Mayer (1996). Their research on learning from multiple examples provided the starting point for our own attempts and is therefore described in detail in the next section.

The Study of Quilici and Mayer (1996)

Quilici and Mayer (1996) aimed at investigating the role of examples in schema construction. Their research was guided by the assumption that exposing students with examples might help them to induce a problem category that in turn would enable them to sort problems according to their structural features. In particular, Quilici and Mayer started off with the expectation that studying multiple examples per problem category would result in better schema acquisition and hence in an improved sorting performance compared to studying only one example per problem category.

In Experiment 1 Quilici and Mayer (1996) presented their participants with zero, one, or three example word problems for each of three problem categories from the domain of statistics for studying (t-test, correlation, χ^2-test) whereby each of the examples was couched into a different cover story. Subsequently, participants had to sort 12 test problems from these three problem categories according to their structural features. 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. Subjects in the zero-example condition were told to sort the problems that best went together into groups, whereas subjects in both example conditions had to assign each test problem to the example category it belonged to in their opinion. The examples were available during the sorting task.

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. Thus, this experiment yielded no support for the superiority of presenting multiple examples compared to a single example per problem category. 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).

This pattern of results can be explained by analyzing both example conditions with regard to the comparison processes they enable. It is assumed that in the structure-emphasizing example condition comparing examples within a problem category fosters the identification of structural features in an optimal way. If a learner compares examples within a category he or she would recognize that there are features that vary among the examples, i.e., 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 may be causally related to the solution procedure. However, the same conclusions may be derived by comparing examples across problem categories because in the structure-emphasizing condition these examples are couched in the same cover story: 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.

In contrast, in the surface-emphasizing example condition surface and structural features are confounded with each other yielding that neither comparisons within problem categories nor across problem categories allow one to distinguish between structural and surface features of examples as these features have the same distribution within and across problem categories. 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. In any case, this experiment demonstrated that the mere number of examples is far less important than the cognitive processes an instructional design allows for.

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: “Both conditions foster structural schema construction but they do so in different ways” (Quilici & Mayer, 1996, p. 156).

To conclude, the results of Quilici and Mayer (1996) failed to support the often-claimed superiority of multiple examples for schema acquisition. Rather, they provide some initial support that learning from single examples may be equally effective if these examples enable profitable processes of comparing examples across categories.

Experiment

Our own experiment that is described in the remainder of the paper provides a more fine-grained analysis to address several issues:

(1) We want to directly test whether single compared to multiple examples are equally effective for schema acquisition in case both example conditions enable profitable processes of example comparison. Even though Quilici and Mayer (1996) directly compared single- to multiple-examples conditions in their first experiment, the examples they used in this experiment did not allow for profitable example comparisons due to an unsystematic use of surface features. Therefore, the authors do not provide an answer to this question.

(2) We advocate the idea that processing strategies may often be more important determinants of learning outcomes than are instructional design decisions (Gerjets & Scheiter, 2003). In a similar vain, work on the self-explanation effect (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Renkl, 1997) has demonstrated that the way learners process examples (e.g., by elaborating the relations between an example’s content an abstract principle) is an important factor to explain performance differences between good and poor problem solvers. Thus, we want to test the aforementioned hypothesis that providing multiple examples may even be harmful for schema acquisition if learners deploy insufficient processing strategies. For this reason, we compare two groups of learners that implement either cursory or thorough example-processing strategies with regard to the effectiveness of single- as opposed to multiple-example conditions.

(3) As a consequence we also expect that multiple examples require instructional guidance that aims at fostering comparison processes. Instructional support that is directed at making learners work harder may be unnecessary for single-example conditions.

(4) We want to address some methodological drawbacks of the study of Quilici and Mayer (1996) that prevent an unambiguous interpretation of their results and that we need to rule out in order to answer our own research questions: Quilici and Mayer refrained from reporting the time participants needed for processing the examples and it remains unclear from their description whether this time was fixed for all experimental conditions or whether it was self-controlled by learners. In our own work we registered learning times and distinguished among participants processing examples thoroughly and those who only displayed a cursory example processing to answer the question whether differences in learning strategies influence the relative effectiveness of the example conditions. Furthermore, we did not ask subjects to assign the test problems to the instructional examples. Rather, the examples were removed before the sorting task and participants had to rely solely on their memory for the problem categories previously studied. We think that this procedure provides a better indicator of the quality of the schematic mental representation induced from the examples than does allowing participants to directly compare the test problems to previously studied examples.

Method

Participants

One hundred and sixty-eight students of the University of Tuebingen, Germany, participated either for course credit or payment. Average age was 23 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

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 told that they would receive examples for three problem categories from statistics. Depending on the experimental condition, they were informed that there would either be one example or three examples illustrating each problem category. Whereas some participants only received the instruction to study these 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).

Subsequently to these instructions, participants received a booklet that contained three sheets of paper. On each sheet there was either one example or there were three examples which all belonged to the same problem category. Each sheet of paper represented one problem category that were not named, but only labeled in an abstract way (i.e., problem category 1, 2, and 3). The problem categories and the examples were identical to the ones used by Quilici and Mayer (1996), that is, 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 χ^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 conditions these three cover stories were used to construct structure-emphasizing example sets. That is, the three examples of each problem category were embedded in three different cover stories that were used across problem categories. Participants could study the examples as long as they liked to and we registered the time taken for example study. When they indicated having studied the examples sufficiently, these were removed from the table and the participants received the instructions for the sorting task. They were told to sort the 12 test problems according to the features that seemed to be relevant to their solution. They 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. The 12 test problems were identical to those used by Quilici and Mayer (1996) and were each printed on a single card. Always four problems came from one of the three problem categories. Similarly to the structure-emphasizing example sets, the test problems were constructed in a way that their cover stories differed within a problem category, but that the same cover stories were used across the three problem categories. 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 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. Both variables were varied between subjects resulting in four experimental conditions. Furthermore, we distinguished 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. This distinction according to the intensity of example study was used as a third factor in all statistical analyses.

As dependent variables we registered the example-processing time and the quality of the sorting performance. For the latter measure the structure and the surface score introduced by Quilici and Mayer (1996) were used. 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 and Discussion

In a first step we analyzed whether participants were comparable across conditions with regard to their domain-specific prior knowledge as indicated by their performance in the pretest (Table 1). An ANOVA (number of examples x comparison instruction x intensity of example study) revealed no significant main effects (number of examples: $F(1,159) = 1.70$; $MSe = 219.41$; $p > .15$; comparison instruction: $F < 1$; intensity of example study: $F < 1$) nor were there any interactions (comparison instruction x intensity of example study: $F(1,159) = 1.09$; $MSe = 219.41$; $p > .30$; all other Fs < 1). Therefore, any performance differences may unambiguously be attributed to the instructional design variables, to participants' strategic behavior, or to interactions among these variables.

Example-study Time

Rather naturally, an ANOVA revealed that the example-study time was prolonged, when there was more than one example per category ($F(1,156) = 175.80$; $MSe = 1.84$; $p < .001$) and when participants studied examples intensively compared to an only cursory processing ($F(1,156) = 253.49$; $MSe = 1.84$; $p < .001$). Additionally, the presence of the comparison instruction increased the time spent for learning ($F(1,156) = 4.10$; $MSe = 1.84$; $p < .05$).

Table 1: Pretest (in % correct) and time demands (in min) as a function of the number of examples, the presence of a comparison instruction, and the intensity of example study

	EXAMPLE STUDYING							
	Cursory				Intensive			
	COMPARISON INSTRUCTION				COMPARISON INSTRUCTION			
	Absent		Present		Absent		Present	
# OF EXAMPLES	1	3	1	3	1	3	1	3
Pretest	74.2	69.9	74.2	72.6	77.6	76.2	75.2	73.0
Example study time	2.7	4.1	2.6	4.9	5.3	8.2	5.0	9.5

A significant interaction between the number of examples and the comparison instruction furthermore revealed that the increase in time due to the comparison instruction was larger in the three-examples conditions than in the one-example conditions ($F(1,156) = 4.03$; $MSe = 1.84$; $p < .05$). Additionally, there was an interaction between the number of examples and the intensity of example processing indicating that multiple compared to single examples increased the learning more when learners had decided to study examples intensively ($F(1,156) = 4.71$; $MSe = 1.84$; $p < .05$). Finally, a marginally significant three-way interaction demonstrated that the increases in learning time due to presenting multiple examples were largest when examples were studied intensively and when learning was accompanied by a comparison instruction compared to all other contrasts ($F(1,156) = 3.55$; $MSe = 1.84$; $p < .10$).

Performance in the Sorting Task

Participants' ability to sort problems according to structural similarities as expressed in the structure score (Figure 1b) was not affected by either the number of examples per problem category or by the instruction to compare examples (both Fs < 1). However, performance improved with the time participants devoted to learning ($F(1,160) = 5.59$; $MSe = 853.70$; $p < .05$). Therefore, as already advocated by Gerjets and Scheiter (2003) example-processing strategies were better predictors for learning outcomes than were features of the instructional design. There was no interaction between the number of examples and the intensity of example processing ($F < 1$). However, the comparison instruction interacted with the intensity of example study ($F(1,160) = 3.05$; $MSe = 853.70$; $p < .10$) as well as with the number of examples $F(1,160) = 3.95$; $MSe = 853.70$; $p < .05$). In particular, there was no difference between the one- and three-example conditions when the instruction to compare examples was present ($M_{1example} = 51.98$ versus $M_{3examples} = 55.56$; $t(82) = -0.52$; $p > .60$; two-tailed), whereas performance deteriorated when the three examples were studied without this instructional support ($M_{1example} = 60.98$ versus $M_{3examples} = 48.41$; $t(82) = 2.02$; $p < .05$; two-tailed). Additionally, studying examples intensively improved performance only when it was guided by the comparison instruction ($M_{cursory} = 45.63$ versus $M_{intense} = 64.11$; $t(82) = -2.79$; $p < .01$; two-tailed) whereas there were no effects of example-processing intensity without comparison instruction ($M_{curosry} = 52.78$ versus $M_{intense} = 56.81$; $t(82) = -0.63$; $p > .50$; two-tailed). Finally, a significant three-way interaction $F(1,160) = 3.98$; $MSe = 853.70$; $p < .05$) indicated that multiple examples compared to single examples improved performance only if they were accompanied by the comparison instruction and an intensive example processing ($t(45) = -1.74$; $p < .10$, two-tailed). Without additional instructional support performance was even hindered by providing multiple examples even if they had been thoroughly studied ($t(38) = 2.10$; $p < .05$, two-tailed). If the examples were studied only cursorily, it did not matter whether they saw one or three examples (with comparison instruction: ($t(45) = 0.79$; $p > .40$; without comparison instruction: ($t(42) = 0.75$; $p > .40$; two-tailed).

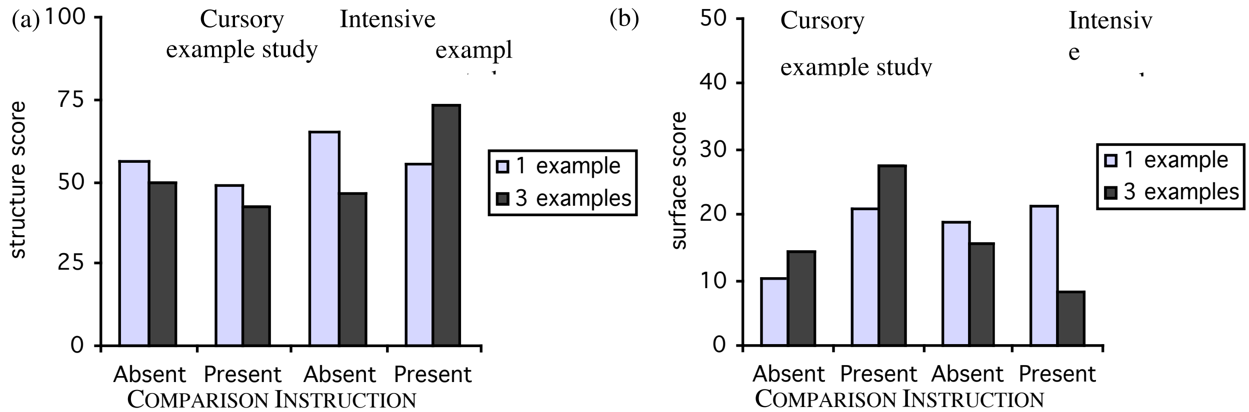

Figure 1: (a) Structure score and (b) surface score as a function of number of examples, the presence of the comparison instruction, and the intensity of example study

With regard to participants' tendency to sort problems according to their surface similarities the following pattern of results could be observed (Figure 1b): There were no main effects of any of the three factors (number of examples: $F < 1$; intensity of example study: $F < 1$; comparison instruction: $F(1,160) = 1.33$; $MSe = 700.60$; $p > .25$). Furthermore, there was no interaction between the number of examples and the comparison instruction ($F < 1$) or between the number of examples and the intensity of example study ($F(1,160) = 2.62$; $MSe = 700.60$; $p > .10$). However, there was a marginally significant interaction between the comparison instruction and the intensity of example study ($F(1,160) = 3.04$; $MSe = 700.60$; $p < .10$). Participants who had refrained from processing examples thoroughly more often sorted problems according to surface features when they had additionally been given an instruction to compare examples ($M_{absent} = 12.50$ versus $M_{present} = 24.29$; $t(89) = -2.07$; $p < .05$; two-tailed). However, when examples were studied intensively the comparison instruction had no further impact on performance. ($M_{absent} = 17.50$ versus $M_{present} = 15.09$; $t(75) = 0.42$; $p > .60$; two-tailed). The three-way interaction was not significant ($F < 1$).

Summary and Conclusions

The study presented was intended to shed light on the role of multiple examples in schema acquisition. It is usually claimed that the availability of multiple examples per problem category with different surface features is necessary in order to identify the structural commonalities of examples and to acquire problem schemata.

However, our study yielded evidence 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. Given these results, multiple examples may be recommendable only under rather optimal learning conditions, whereas providing one example per problem category proves to be far less vulnerable to these learning conditions. Under less optimal learning conditions, i.e., when the examples are processed without any additionally instructional guidance, multiple examples may even tend to worsen performance. With regard to participants' tendency to be misled by surface features the comparison instruction helped to reduce this tendency only if learners were willing to spend sufficient time on processing the examples. To conclude, multiple examples do not seem to be necessary for schema acquisition. Moreover, single examples may even be less dependent on the presence of optimal instructional conditions and on adequate learning behavior.

There are however certain limitations to the study and open questions that need to be addressed in future studies. First, the performance in the sorting task is an indicator only of a learner's ability to categorize problems according to their structural features, while it does not tell us anything with regard to whether learners will actually be able to solve problems on their own after having studied single or multiple examples. Prior own studies (Gerjets, Scheiter, & Tack, 2000) have also revealed that – when using multiple *worked-out* examples and when measuring *problem-solving* performance rather than the ability to categorize tasks – single examples might be as helpful as multiple examples. However, it would be good to replicate these findings with different materials. The second issue is related to the first in that it addresses the question whether multiple examples may show their advantage only when more demanding tasks are used. That is, it might well be that the effort to process multiple rather than single examples pays off only when we compare the two instructional conditions with regard to far transfer performance. Similarly, multiple examples might be superior in cases where the to-be-taught domain is more complex and therefore harder to understand. Here, multiple as opposed to single examples may be necessary to receive a thorough illustration of the principles underlying the domain. Third, in the current study we distinguished between participants who either processed examples cursorily or intensively by taking into account the time they decided to devote for learning. From an instructional perspective it is important to know how we can have *all* learners process examples thoroughly, e.g. by presenting learners with incomplete rather than completely worked-out examples and having them fill in the gaps by themselves. If learners can be scaffolded to study examples more intensively, they might also profit from multiple (worked-out) examples. These issues we would like to pursue in future studies that compare multiple over single examples under different learning conditions.

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.

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

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. (2003). Goals and strategies as moderators between instructional design and cognitive load: Evidence from hypertext-based instruction. *Educational Psychologist, 38*, 33-41.

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* (pp. 166-171). 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.

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

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.

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* (pp. 810-815). Mahwah, 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 Merrienboer, 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.

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

Acknowledgements

We would like to thank the participants of the practical course on experimental psychology during the Summer 2002 and Flora Brehm for conduction the experiments.