Case based learning with worked examples in medicine: effects of errors and feedback

Robin Stark, Saarland University, r.stark@mx.uni-saarland.de
Veronika Kopp, Martin Fischer, University of Munich, Munich

Abstract: To facilitate medical students' diagnostic knowledge a case-based worked example approach was realized in a complex computer-based learning environment. To enhance the effectiveness of the approach the measures erroneous examples and elaborated feedback were additionally implemented. 153 medical students were randomly assigned to four experimental conditions of a 2x2-factorial design (errors vs. no errors, elaborated feedback vs. knowledge of correct result (KOR)). In order to assess the sustainability of the learning environment a subgroup of subjects (n = 52) was compared with a comparable group of students who did not participate in the experiment (n = 145) with respect to their performance in a regular multiple choice test. Results show that the acquisition of diagnostic knowledge is mainly fostered by providing erroneous examples in combination with elaborated feedback. These effects were independent from differences in time-on-task and prior knowledge. Furthermore, the effects of the learning environment proved sustainable.

Introduction

Diagnosing is a fundamental competence for every physician. However, students and attending physicians have enormous difficulties ascertaining the right diagnosis. Gräsel and Mandl (1993) for example showed that medical students proceed in an additive and mechanic way even at the end of their clinical studies; they tend to collect high amounts of data without building concrete hypotheses. We postulate that diagnostic knowledge consists of declarative knowledge components (conceptual knowledge; "what-information"), action-related knowledge components like strategic knowledge ("how-information") and teleological knowledge ("why-information"; van Gog, Paas, & van Merriënboer, 2004). The students in the study of Gräsel and Mandl (1993) primarily showed deficits in the latter two knowledge components. Furthermore, the three components are not sufficiently integrated into more abstract structures (schemata) allowing effective and efficient diagnostic reasoning.

In order to overcome this problem and to foster diagnostic knowledge which allows competent diagnostic reasoning, a complex computer-based learning environment was developed. This environment based on the CASUS platform which already proved effective in various studies on case-based learning in different fields of medicine (Fischer, 2000; Simonsohn, & Fischer, 2004).

Following Mandin (2001) diagnostic reasoning can be understood as schema-based problem-solving. Therefore, in order to enhance diagnostic knowledge, effective learning environments have to be provided which systematically support the students constructing and organizing schemata. Learning environments taking advantage of the *worked example effect* (Gerjets, Scheiter, & Catrambone, 2004; Renkl, 2005) are especially appropriate to achieve this challenging instructional goal. Worked examples are superior to directly teaching abstract principles as well as to actively solving training problems – at least with regard to *initial* skill acquisition (Gerjets et al., 2004). This worked example effect is attributed to the fact that studying worked examples imposes lower levels of cognitive load on the learner than solving training problems especially because no extensive search processes with regard to the correct solution steps are involved (Gerjets et al., 2004). Therefore, more cognitive resources are free for demanding processes of schema construction. Moreover, studying worked examples (in contrast to attempting to solve training problems) focuses the learner's attention on information that is relevant to schema construction. In various domains, schema-based problem solving is considered to be very effective and efficient. It proved to be a central characteristic of experts' problem solving (Gerjets et al., 2004, p. 34).

However, studies on the "self-explanation effect" (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, DeLeeuw, Chiu, & LaVancher, 1994; Renkl, 1997; Stark, 1999) clearly show that studying worked examples does not *automatically* result in effective problem-solving schemas. In order to achieve this goal, processes of cognitive and meta-cognitive self-explanation have to be fostered systematically by implementing additional instructional measures (Atkinson, Renkl, & Merrill, 2003; Große, & Renkl, 2006; Stark, 1999). In addition, the worked example method has to be adapted to the instructional context, the instructional goals and the domain in which it is implemented.

So far, worked examples were investigated primarily in studies on initial learning in rather simple and well-structured domains. Therefore we combined the example method with elements of case-based learning which proved effective in various problem-based learning scenarios in medicine (Hmelo, Duncan, & Chinn, 2007; Schmidt, Loyens, van Gog, & Paas, 2007). In addition, two instructional measures aiming at fostering diagnostic knowledge by stimulating processes of schema induction were integrated into the learning environment: erroneous examples and elaborated feedback.

A major objective of our study was to assess the effectiveness of these instructional measures in the context of our case-based worked example approach. In addition, the influence of the two measures on time-on-task, cognitive load, and the sustainability of the learning environment in general were investigated.

Learning from errors

Learning from errors is a promising method particularly in the field of medicine where errors in the process of diagnosing can have fatal consequences (Al-Assaf, Bumpus, Carter, & Dixon, 2003). Following the classification of diagnostic errors by Graber, Gordon, & Franklin (2002), we focused on relevant and severe cognitive errors caused by inadequate knowledge or faulty data gathering, inaccurate clinical reasoning, or faulty verification (Graber et al., 2002, p. 983). According to Oser and Spychiger (2005), every error includes the chance to learn from it when learners use them purposeful as learning occasion. Therefore, they have to identify the error and understand the correct solution by comparing the incorrect solution systematically with the correct one. A study of Große and Renkl (2004) can be interpreted in terms of the necessity of implementing effective feedback methods when confronting learners with erroneous examples.

Feedback

Following van Gog et al. (2004), understanding of a procedure must involve both knowledge of its domain and of its teleology. Knowledge of a domain consists of principled knowledge about objects and events in that domain. Knowledge of the teleology of a procedure is knowledge of the rationale behind or purpose of the steps in a procedure (van Gog et al., 2004). Therefore, effective feedback should not only contain information concerning conceptual knowledge, but also information which informs the learner of the rationale behind the selection and application of operators (so called “why information”) and information about strategic knowledge used by experts selecting the operators (so called “how information”). These knowledge components also represent our model of diagnostic knowledge (see above).

Especially when learners are confronted with complex tasks, elaborated feedback which combines these kinds of information (in comparison to less informative feedback like knowledge of correct result (KOR) feedback) has positive effects on feedback reception (Jacoby, Troutman, Mazursky, & Kuss, 1984) and knowledge acquisition (e. g. Collins, Carnine, & Gersten, 1987; Krause, 2007; Narciss, 2001; Pridemore, & Klein, 1991). However, when only conceptual knowledge has to be acquired, KOR-feedback can be sufficient (e. g. Kulhavy, White, Topp, Chan, & Adams, 1985).

Time-on-task

The implementation of erroneous examples and feedback in the present study aimed at enhancing the effectiveness of the learning environment by fostering the *quality* of learning processes. However, coping with erroneous examples and processing elaborated feedback information (or compensating marginal feedback by self-explanation) and the resulting processes can result in prolongation of time-on-task, that is, quality and quantity of learning processes can be confounded. From a practical perspective, such effects are acceptable when the prolongations are moderate and at the same time associated with substantial learning effects. However, from a theoretical perspective, they are only acceptable when they do not undermine *internal validity* of the study. In spite of these potential problems, time-on-task was not restricted because of considerations about *ecological validity*: following an *integrative research approach* (Stark & Mandl, 2007), it was planned to investigate the effects of the two instructional measures under conditions which are not too far from realistic learning scenarios in practice.

Cognitive Load

From the perspective of cognitive load theory (CLT, Sweller, 1988), learning with errors makes greater demands on working memory than processing correct information, especially when only KOR-feedback is given and the learners have to explain themselves why the provided information is incorrect and draw consequences for further diagnostic steps. In a complex learning environment, these processes can lead to cognitive *overload* (Renkl, Gruber, Weber, Lerche, & Schweizer, 2003) and interfere with further knowledge acquisition. This problem can be compensated by providing elaborated feedback, at least when the design of the feedback procedure is in line with CLT (Sweller, van Merriënboer, & Paas, 1998).

Research questions

1. To what extent time-on-task is increased by erroneous examples and elaborated feedback?

We expected that both measures increase time-on-task. However, it was postulated that the potential effects on diagnostic knowledge cannot be explained by differences in time-on-task.

2. To what extent diagnostic knowledge is facilitated by erroneous examples and elaborated feedback?
It was assumed that both erroneous examples and elaborated feedback have positive effects on the acquisition of diagnostic knowledge and that the combination of both measures is especially effective.

3. Which influence do both instructional measures exert on cognitive load?
It was expected that cognitive load is increased through providing erroneous examples, especially in the combination with KOR-feedback. In addition, we supposed that cognitive load and diagnostic knowledge correlate negatively.

4. To what extent effects of the learning environment are sustainable?
It was supposed that effects of the learning environment on conceptual knowledge can still be verified six weeks after the learning session.

Method

Sample and design

A total of 153 medical students (104 females and 49 males) from the universities of Munich, Germany took part in the study. All participants studied in the clinical part of the curriculum, ranging from the second to the sixth clinical semester. The average age was around 25 years (M = 25.02, SD = 3.62). Students were voluntarily recruited and received 40 Euros for participation. The subjects were randomly assigned to four conditions of a 2x2-factorial design (see table1).

Table 1: 2×2-factorial design.

Design		faktor 2: feedback	
		Elaborated	KOR
factor 1: errors	with errors	n = 36	n = 41
	without errors	n = 40	n = 36

Sustainability was assessed by comparing 145 students in the second clinical semester who did not participate in our study (control group) with the 52 subjects from the same semester (experimental group) concerning the results of a regular multiple choice test. The 52 students were selected from all experimental conditions; they were comparable with the rest of the participants concerning learning prerequisites and learning outcomes (that is diagnostic knowledge) at the end of the learning session. In addition, experimental and control group subjects did not differ with respect to relevant learning prerequisites.

Learning environment

The learning environment, integrated in the CASUS learning platform, consisted of a schema illustrating the steps of the diagnostic process and six worked examples dealing with pheochromozytoma, primary hyperaldosteronism, and renal artery stenosis. Working through these examples, the learners see themselves in the role of a (fictitious) student doing an elective working together with a general practitioner giving feedback. All examples begin with a clinical situation. In the learning environment, the protagonist starts drawing his conclusions for the diagnosis and for the further procedure. Subsequently, the general practitioner gives feedback on his diagnostic steps. Then the sequence of giving information, drawing conclusions and getting feedback goes on until at the end the final diagnosis is reached. Every example consists of three to four of such sequences containing three to four screenshots.

The additionally given schema allows to contextualize the steps of the diagnostic process by providing the problem space containing all hypertension illnesses. Additionally the schema explains the relation between the different illnesses and the underlying pathophysiological processes. Furthermore, it is shown which further information is needed for excluding and including a diagnosis, respectively.

Instructional measures

Errors. In the condition without errors, the protagonist is presented coming to the right considerations on the basis of the given information, drawing the right conclusions and finishing the case with the right diagnosis. In the condition with errors, the protagonist makes severe errors, which the students had to study. The choice of the errors was guided by the above-mentioned error taxonomy from Graber et al. (2002); in addition, it was inspired by an analysis of relevant cognitive errors in this domain. After every wrong decision, the error is corrected in form of the general practitioner's feedback.

Feedback. In the elaborated version of the feedback, the (fictitious) physician gives additional explanations to the (right or wrong) considerations and further consequences drawn by the student. Furthermore, he elucidates the diagnostic process referring to underlying pathophysiological knowledge. The conditions with KOR-feedback were very frugal in comparison to the elaborated version. Considerations, conclusions, and further procedures are only evaluated as right or wrong without further explanation. In the condition “errors and KOR-feedback” the learners had to deduce the right step in the diagnostic process from the subsequent solution step of the example. All feedback information was presented by providing written texts, no audio or video components were implemented in this learning environment.

Instruments and data collection

Prior knowledge. The prior knowledge test measuring domain specific *conceptual knowledge* contained 21 multiple choice questions, most of which have been used in the German medical examination. Basing on analyses of reliability, two tasks were excluded from further analysis. So 19 points could be achieved in this measure. The reliability of the test was sufficient (Cronbach’s Alpha = .63).

Diagnostic knowledge. The construct of diagnostic knowledge comprised conceptual, strategic and teleological knowledge. The 19 conceptual knowledge tasks of the prior knowledge were presented again after the learning phase (Cronbach’s Alpha = .60). In order to measure strategic knowledge, ten key feature problems (maximum: 29 points) were applied (Bordage, Brailovsky, Carretier, & Page, 1995; Fischer, Kopp, Holzer, Ruderich, & Jünger, 2005). The reliability was .72 (Cronbach’s Alpha). Additionally, problem solving questions were presented which functioned as a measure for strategic and teleological knowledge. Students had to generate their first leading diagnosis upon the information of a given clinical case scenario. Furthermore, they were asked to explain their decision and to describe the underlying pathophysiological processes (Cronbach’s Alpha = .73; Maximum: 20 points). The three tests measuring diagnostic knowledge correlated with Pearson coefficients between .45 and .63. The reliability of the aggregated diagnostic knowledge measure (Maximum: 68 points) was .85 (Cronbach’s Alpha).

Cognitive load was assessed by a rating scale of Paas und Kalyuga (2005) with nine items in which the learners had to evaluate the difficulty of the task and their mental effort on seven steps (from 1 “very low” to 7 “very high”). The rating scale was presented during the learning session after the third and sixth example. Cronbach’s Alpha was .75 and .83, respectively.

Sustainability of the learning environment was assessed by a regular multiple choice exam which consisted of thirty questions measuring primarily conceptual knowledge in the domain of internal medicine (Cronbach`s Alpha = .68; Maximum: 30 points). Seven of these questions dealt with the content matter of the learning environment (arterial hypertension); the rest of the tasks had to do with other aspects of internal medicine.

Time-on-task

Time-on-task was recorded automatically by the computer-based learning environment.

Procedure

After a short introduction students completed the multiple choice test measuring prior knowledge followed by the learning session in which the students worked with the learning environment and gave their ratings on the cognitive load scale. After a pause of 15 minutes, the students worked on the diagnostic knowledge tests. Six weeks after the experiment, the regular multiple choice exam took place.

Results

Time-on-task

Students learning with elaborated feedback worked significantly longer with the learning environment than students in the two KOR-conditions (see table 2). The main effect “feedback” was significant and substantial ($F(1,149) = 36.15$; $p < .01$; $\eta^2 = .20$). Errors had only a small effect on time-on-task which did not reach statistical significance ($F(1,149 < 1)$. The interaction missed statistical significance as well ($F(1, 149) < 1$). Time-on-task was not associated with diagnostic knowledge. In the four groups, the correlations varied between .01 and .09.

Effects on diagnostic knowledge

Concerning prior knowledge, no significant group differences were found. However, descriptive differences in prior knowledge and time-on-task were statistically controlled in the subsequent analyses.

There were neither ceiling nor floor effects. With respect to conceptual knowledge, students in the condition “with errors-elaborated feedback” achieved the highest scores. Students working with erroneous examples combined with KOR-feedback did rather poorly. The scores of the other two groups differed only marginally and lay in-between the scores of the two “extreme groups”. Concerning the other aspects of diagnostic knowledge, the same descriptive pattern appeared (see table 2).

Table 2: Time-on-task, conceptual, strategic and teleological knowledge and cognitive load (both times of measurement) in the four learning conditions: means and standard deviations (in brackets).

	M (*SD*)			
	with errors		without errors	
	Elaborated	KOR	elaborated	KOR
Time-on-task	44.60 (*14.70*)	31.37 (*7.75*)	45.49 (*16.87*)	33.70 (*10.02*)
Conceptual knowledge (max. 19 points)	14.69 (*2.35*)	12.83 (*2.28*)	13.38 (*2.72*)	13.03 (*2.76*)
Strategic knowledge (max. 29 points)	20.68 (*3.02*)	17.81 (*2.96*)	18.32 (*4.04*)	18.85 (*4.42*)
Strategic and teleological knowledge (max. 20 points)	11.93 (*3.65*)	9.93 (*3.53*)	10.86 (*3.70*)	11.63 (*3.27*)
Cognitive load (t1)	3.36 (*0.57*)	3.57 (*0.56*)	3.36 (*0.79*)	3.11 (*0.56*)
Cognitive load (t2)	3.34 (*0.75*)	3.59 (*0.67*)	3.37 (*0.86*)	3.10 (*0.71*)

Concerning conceptual knowledge assessed by multiple choice questions, the main effect "errors" ($F(1,145) = 3.88$; $p < .05$; $\eta^2 = .03$) was significant. On average, more conceptual knowledge was acquired when erroneous examples were provided. Neither the main effect "feedback" ($F < 1$) nor the interaction ($F(1,145) = 2.10$; *n. s.*) were significant.

With respect to strategic knowledge measured by key feature problems, again the main effect „errors" was significant ($F(1,145) = 3.99$; $p < .05$; $\eta^2 = .03$), qualified by a significant "errors x feedback"-interaction ($F(1,145) = 7.58$; $p < .01$; $\eta^2 = .05$). The main effect "feedback" was not significant ($F < 1$). As expected, students profited from erroneous examples especially when they received elaborated feedback.

A different picture emerged when the results in strategic and teleological knowledge assessed by problem solving questions were analyzed. Here the feedback-factor proved significant $F(1,145) = 4.40$; $p < .05$; $\eta^2 = .03$). The interaction was significant, too ($F(1,145) = 5.96$; $p < .05$; $\eta^2 = .04$). The main effect „errors" had no influence ($F < 1$). Only when erroneous examples were provided, elaborated feedback was beneficial here.

To sum up, acquisition of conceptual knowledge was enhanced by erroneous examples. When erroneous examples were combined with elaborated feedback, they also fostered acquisition of strategic knowledge. The third aspect of diagnostic knowledge was supported by elaborated feedback, especially when erroneous examples were presented. These effects were independent from prior knowledge and differences in time-on-task.

Cognitive load

Table 2 shows that both times, learners in the condition „with errors-KOR" showed the highest load scores; in the condition „without errors-KOR" the lowest load scores occurred. The other two conditions differed only marginally. At t1, the main effect "errors" was significant ($F(1,149) = 5.05$; $p < .05$; $\eta^2 = .03$). At t2, the descriptive findings were comparable; however, the level of statistical significance was missed ($F(1,148) = 3.56$; $p < .10$). The main effect "feedback" was not significant (t1 and t2: $F(1,149) < 1$). However, both times, the interaction between errors and feedback was significant (t1: $F(1,149) = 5.18$; $p < .05$; $\eta^2 = .03$; t2: $F(1,148) = 4.53$; $p < .05$; $\eta^2 = .03$). As expected, providing erroneous examples in combination with KOR-feedback enhanced cognitive load.

Under all learning conditions, cognitive load and diagnostic knowledge correlated negatively. However, the relation was not very strong, ranging from -.11 to -.49.

Sustainability

In the multiple choice questions not dealing with arterial hypertension, students participating in the experiment reached 15.7 of 23 points ($SD = 3.12$), the control group 14.8 points ($SD = 3.11$). This difference was not significant ($F(1,195) = 2.99$; *n.s.*). However, in the arterial hypertension questions, the experimental group reached 5.59 of 7 points ($SD = 0.95$), whereas the control group reached only 5.07 points ($SD = 1.14$). This effect was significant ($F(1,195) = 8.49$; $p < .01$) but not substantial ($\eta^2 = .04$).

Discussion

Our results show that the acquisition of diagnostic knowledge is mainly fostered by providing erroneous examples in combination with elaborated feedback. When only KOR-feedback is given, erroneous examples are detrimental. These results were independent from prior knowledge and time-on-task. Against the background of the rather short time-on-task invested by learners in all conditions, even the rather *small* effects of the two instructional measures are remarkable. They confirm pedagogical considerations about the learning potential of errors (Oser & Spychiger, 2005). The unfavourable effect of erroneous examples in combination with KOR-feedback on knowledge acquisition replicates findings of studies in which erroneous examples were provided *without* informative feedback measures (Große & Renkl, 2004). In addition, they are in line with current feedback research in which the effectiveness of informative feedback measures is stressed (Krause, 2007), at least in complex learning. For most students, the information that a specific diagnostic conclusion or procedure is *not* adequate seems not to be sufficient to induce deep conceptual understanding. The information deficit in the conditions with KOR-feedback, especially when the learners are confronted with erroneous examples, has to be compensated by effective self-explanation (Chi et al., 1989; Renkl, 2005; Stark, 1999). However, as Renkl (1997) showed, a lot of learners process example information rather superficially. As a result, the full potential of the learning method is not tapped. In fact, at least in the context of *complex* learning environments, the given feedback has to inform the students *why* the respective step in the problem solving process is wrong and *which* specific conclusion or procedure is adequate.

The unfavourable effects of erroneous examples in combination with KOR-feedback also correspond with CLT (Sweller, 1988): this learning condition was associated with the highest load scores. As these scores in our study were negatively related with knowledge acquisition, the conclusion is suggestive that this learning condition reduced the worked example effect by imposing too much extraneous load on the learners (Sweller et al., 1998). It has to be noted that our study does not allow for insights concerning the effects of diagnostic errors made by the learners themselves but only for insights of erroneous example information studied by the learners. According to Oser and Spychiger (2005), errors made by the learners themselves might be comparably productive, at least when immediate and elaborated feedback is secured and a learning context is given in which a positive "Fehlerkultur" is realized. However, analyzing the effects of learner errors makes necessary completely different study designs.

Six weeks after the experiment was over, effects of the learning environment on conceptual knowledge were still significant. The fact that this effect only occurred when the learners had to answer multiple choice questions focussing on content aspects represented in the learning environment can be interpreted as a kind of validation of the sustainability effect. From a pedagogical perspective, this finding is astonishing because the learning sessions were short and the learning environment is not specialised on fostering conceptual knowledge but on diagnostic knowledge which also includes strategic and teleological knowledge components. In further studies, sustainability analyses should include *all* aspects of diagnostic knowledge. In addition, in order to gain more distinguished insights into the sustainability of the instructional measures, further analyses should include larger experimental groups which enable differentiated comparisons.

If the positive findings of our study can be replicated, the learning environment should be integrated into the regular clinical curriculum at university. Concerning further studies, thinking aloud procedures (e. g. Ericsson, & Simon, 1993) should be employed in order to investigate effects of erroneous examples and feedback on the quality of the learning *process*. Only by recording the learners' self-explanations, more *direct indicators* for pedagogically desired processes of schema induction can be won.

References

Al-Assaf, A. F., Bumpus, L. J., Carter, D., & Dixon, S. B. (2003). Preventing Errors in Healthcare: A Call for Action. *Hospital Topics, 81*, 5-12.

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

Bordage, G., Brailovsky, C., Carretier, H., & Page, G. (1995). Content validation of key features on a national examination of clinical decision-making skills. *Academic Medicine, 70* (4), 276-281.

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.

Chi, M. T. H., DeLeeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439-477.

Collins, M., Carnine, D., & Gersten, R. (1987). Elaborated corrective feedback and the acquisition of reasoning skills: A study of computer-assisted instruction. *Exceptional Children, 54,* 254-262.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis. Verbal reports as data. (Rev. ed.*). Cambridge, MA: MIT Press.

Fischer, M. R. (2000). CASUS - An Authoring and Learning Tool Supporting Diagnostic Reasoning. *Zeitschrift für Hochschuldidaktik, 1,* 87-98.

Fischer, M. R., Kopp, V., Holzer, M., Ruderich, F., & Jünger, J. (2005). A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. *Medical Teacher, 27,* 450-455.

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional Examples to Reduce Intrinsic Cognitive Load: Molar versus Modular Presentation of Procedures. *Instructional Science, 32,* 33-58.

Graber, M., Gordon, R., & Franklin, N (2002). Reducing Diagnostic Errors in Medicine: What's the Goal? *Academic Medicine, 77,* 981-92.

Gräsel, C., & Mandl, H. (1993). Förderung des Erwerbs diagnostischer Strategien in fallbasierten Lernumgebungen. [*Fostering the acquisition of diagnostic strategies in case-based learning environments.*] *Unterrichtswissenschaft, 21,* 355-370.

Große, C. S., & Renkl, A. (2004). Learning from worked examples: What happens if errors are included? In P. Gerjets, J. Elen, R. Joiner & P. Kirschner (Eds.), *Instructional design for effective and enjoyable computer-supported learning* (pp. 356-364). Tübingen: Knowledge Media Research Center.

Große, C. S., & Renkl, A. (2006). Effects of multiple solution methods in mathematics learning. *Learning and Instruction, 16,* 122-138.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99-107.

Jacoby, J., Troutman, T., Mazursky, D., & Kuss, A. (1984). When feedback is ignored: Disutility of outcome feedback. *Journal of Applied Psychology, 69* (3), 531-545.

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

Kulhavy, R. W., White, M. T., Topp, D. W., Chan, A. L., & Adams, J. (1985). Feedback complexity and corrective efficiency. *Contemporary Educational Psychology, 10*, 285-291.

Mandin, H. C. (2001). Schemes for Problem Solving - Hypertension. *The Meducator, 1(1), Handbook for Practitioners - Medical Education Series 1.*

Narciss, S. (2001). *Informative feedback as a bridge from instruction to learning in computer-based trainings.* Paper presented at the 9th European Conference for Research on Learning and Instruction, Fribourg, Switzerland.

Oser, F., & Spychiger, M. (2005). *Lernen ist schmerzhaft: Zur Theorie des Negativen Wissens und zur Praxis der Fehlerkultur. [Learning is painful. On the theory of negative knowledge and the practice of error management.*] Weinheim: Beltz.

Paas, F., & Kalyuga, S. (2005). *Cognitive Measurements to Design Effective Learning Environments.* Paper presented at the International Workshop and Mini-conference on Extending Cognitive Load Theory and Instructional Design to the Development of Expert Performance, Heerlen, Netherlands.

Pridemore, D. R., & Klein, J. D. (1991). Control of feedback in computer-assisted instruction. *Educational Technology Research & Development, 39,* 27-32.

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

Renkl, A. (2005). The worked-out-example principle in multimedia learning. In R. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (pp. 229-246). Cambridge, UK: Cambridge University Press.

Renkl, A., Gruber, H., Weber, S., Lerche, T., & Schweizer, K. (2003). Cognitive Load beim Lernen aus Lösungsbeispielen. [*Cognitive Load when Learning by Worked Examples.*] *Zeitschrift für Pädagogische Psychologie, 17,* 93-101.

Schmidt, H. G., Loyens, S. M. M., van Gog, T., & Paas, F. (2007). Problem-Based Learning is Compatible with Human Cognitive Architecture: Commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 91-97.

Simonsohn, A., & Fischer, M. R. (2004). Evaluation eines fallbasierten computergestützten Lernsystems (CASUS) im klinischen Studienabschnitt. [Evaluation of a case-based computer-based learning system (CASUS) during the course of clinical studies.] *Deutsche Medizinische Wochenschrift, 129*, 552-556.

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

Stark, R., & Mandl, H. (2007). Bridging the gap between basic and applied research by an integrative research approach. *Educational research and evaluation, Special issue, 13*(3), 249-261.

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

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

Van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2004). Process-oriented worked examples: Improving transfer performance through enhanced understanding. *Instructional Science, 32,* 83-98.