Teaching Students Self-Assessment and Task-Selection Skills with Video-Based Modeling Examples

Tamara van Gog (vangog@fsw.eur.nl)
Institute of Psychology, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Danny Kostons (danny.kostons@ou.nl)
Centre for Learning Sciences and Technologies, Open University of The Netherlands
P.O. Box 2960, 6401 DL Heerlen, The Netherlands

Fred Paas (paas@fsw.eur.nl)
Institute of Psychology, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Abstract

For self-regulated learning to be effective, students need to be able to accurately assess their own performance on a learning task, and to select an appropriate new learning task in response to that self-assessment. This study investigated the use of video-based modeling examples to teach self-assessment and task-selection skills. Students in both the experimental and control condition observed the model performing a problem solving task; students in the experimental condition additionally observed the model engaging in self-assessment and task selection. Results show that students in both conditions acquired problem-solving skills from the examples, as indicated by a substantial pretest to posttest knowledge gain. Moreover, students in the experimental condition also acquired self-assessment and task-selection skills from the examples: they demonstrated higher self-assessment and task-selection accuracy on the posttest than students in the control condition.

Keywords: Example-based learning; self-assessment; task selection; self-regulated learning.

The Role of Self-Assessment and Task-Selection Skills in Self-Regulated Learning

A major aim of many contemporary educational programs is to foster students' self-regulation skills. It is often assumed that this aim can be achieved by providing learners with the opportunity to self-regulate their learning processes. In the Netherlands, for example, a nationwide innovation was implemented in secondary education in 1999 that relies heavily on self-regulated learning (i.e., the 'study house'; http://www.minocw.nl/english/education/293/Secondary-education.html). Self-regulated learning is also assumed to result in personalized learning trajectories, in which instruction is adaptive to the individual student's needs. Such personalized instruction is expected to enhance students' motivation and learning outcomes compared to non-adaptive, fixed instruction that is the same for all students.

Unfortunately, there is little evidence for both assumptions. First of all, research has shown that students do not acquire self-regulation skills merely by engaging in self-regulated learning, rather, they need additional training or instructional support (e.g., Azevedo & Cromley, 2004; Van den Boom, Paas, Van Merriënboer, & Van Gog, 2004). Secondly, although the assumption is correct that adaptive, personalized instruction can foster learning compared to non-adaptive instruction (e.g., Camp, Paas, Rikers, & Van Merriënboer, 2001; Salden, Paas, Broers, & Van Merriënboer, 2004), it is questionable whether self-regulated learning actually results in adaptivity to students' needs.

In adaptive instructional systems, learning tasks are chosen for each individual student based on an assessment of their current level of knowledge and skill (based on several aspects of students' performance, e.g., Anderson, Corbett, Koedinger, & Pelletier, 1995; Koedinger, Anderson, Hadley, & Mark, 1997; or on a combination of their performance and invested mental effort, e.g., Camp et al., 2001; Corbalan, Kester, & Van Merriënboer, 2008; Kalyuga, 2006; Salden et al., 2004). The assessment of performance and the selection of an appropriate new learning task (i.e., based on that assessment) is conducted by the system. For self-regulated learning to be equally adaptive and effective, students themselves should be able to accurately assess their own performance and to recognize what an appropriate new task would be. Unfortunately, there is quite some evidence that students, and especially novices who lack prior knowledge of the learning tasks, are not very accurate self-assessors. Humans seem prone to several biases that affect accuracy of self-assessments (for a review, see Bjork, 1999), such as hindsight bias (i.e., once an answer or solution procedure is known, e.g., after feedback, students are more likely to think that they could have produced it themselves), or availability bias (i.e., answers that come to mind easily are not only more likely to be provided but are also more likely to be assumed to be correct). Moreover, accurate self-assessment also seems to require some domain expertise (Dunning, Johnson, Erlinger, & Kruger, 2003). Individuals with higher levels of prior knowledge are more accurate self-assessors, presumably because their experience not only provides them with more task knowledge, but also with more knowledge of the criteria and standards that good performance should meet (Dunning et al., 2003). In addition, their experience also lowers the cognitive load imposed by the task, allowing them to devote more cognitive resources to monitoring their task performance, which likely provides them with a more accurate memory representation on which to base their assessment (Van Gog & Paas, 2009).

Support for our assumption that novice students' lack of self-assessment skills leads to ineffective self-regulated learning, comes from studies that have shown that providing novice students with control over their learning process may have beneficial effects on their motivation or involvement, but often has detrimental effects on learning outcomes (see e.g., Azevedo, Moos, Greene, Winters, & Cromley, 2008; Niemic, Sikorski, & Walberg, 1996). When positive effects on learning outcomes are found, this tends to be mostly for students with higher levels of prior knowledge in the domain (e.g., Niemiec et al., 1996; Moos & Azevedo, 2008), who, as mentioned above, are also likely to be more accurate self-assessors. In addition, Kostons, Van Gog, and Paas (2010) investigated differences in self-assessment accuracy between secondary education students who differed in the amount of knowledge gained from studying in a learner-controlled instructional environment that contained heredity problems with varying levels of support at different levels of complexity. They found that the students who had gained more knowledge, had also more accurately assessed their own performance during learning.

Without accurate self-assessment, selecting an appropriate new learning task will also be very difficult. Given the central role that self-assessment and task-selection skills seem to play in self-regulated learning, an important question is whether we can teach novice students to become more accurate self-assessors and task selectors. We decided to investigate this question, using modeling examples to teach those skills.

Learning from Examples

Learning from examples is known to be a highly effective instructional strategy. Research inspired by cognitive theories such as ACT-R (Anderson, 1993) or Cognitive Load Theory (Sweller, Van Merriënboer, & Paas, 1998) has extensively investigated the effects on learning of instruction consisting of studying *worked examples*, which provide students with a written worked-out didactical solution to a problem. These studies have consistently shown that for novices, studying worked examples is more effective and/or more efficient for learning (i.e., equal or higher learning outcomes attained with lower or equal investment of time and/or effort) than (tutored) problem solving, which is known as the 'worked example effect' (Sweller et al., 1998; for further reviews, see Atkinson, Derry, Renkl, & Wortham, 2000). Studies on the worked example effect have mainly used highly structured cognitive tasks, such as algebra (e.g., Cooper & Sweller, 1987; Sweller & Cooper, 1985), statistics (e.g., Paas, 1992), geometry (e.g., Paas & Van Merriënboer, 1994), or physics (e.g., Van Gog, Paas, & Van Merriënboer, 2006), although recent studies have shown the same effect with less structured tasks such as learning to recognize designer styles in art education (Rourke & Sweller, 2009).

Research inspired by Social Learning Theory (Bandura, 1986) has mostly focused on *modeling*, that is, learning by observing another person (the model) perform a task. Models can be either adults (e.g., Schunk, 1981) or peers (e.g., Braaksma, Rijlaarsdam, & Van den Bergh, 2002; Schunk & Hanson, 1985), and they can behave didactically or naturally (i.e., possibly skipping steps, or making and/or correcting errors). Moreover, modeling examples can consist of a video in which the model is visible (e.g., Braaksma et al., 2002), a video consisting of a screen capture of the model's computer screen in which the model is not visible (e.g., McLaren, Lim, & Koedinger, 2008; Van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009), or an animation in which the model is represented by a pedagogical agent (e.g., Atkinson, 2002; Wouters, Paas, & Van Merriënboer, 2009). Like worked examples, modeling examples have also been used to teach highly structured cognitive tasks such as math (e.g., Schunk, 1981) or chemistry (e.g., McLaren et al., 2008), but they have also been widely applied with less structured tasks such as writing (e.g., Braaksma et al., 2002; Zimmerman & Kitsantas, 2002). In addition, they have been used for teaching *meta*cognitive skills such as self-regulation (e.g., Kitsantas, Zimmerman, & Cleary, 2000; Zimmerman & Kitsantas, 2002). For a more in-depth review of research on worked examples and modeling examples, see Van Gog and Rummel (in press).

This study investigated whether video-based modeling examples consisting of screen-recordings could be successfully applied for teaching secondary education students self-assessment and task-selection skills.

Method

Participants and Design

Participants were 39 Dutch secondary education students (age M = 15.08, SD = 0.48; 26 female) in the fourth year of pre-university education (the highest level of secondary education in the Netherlands, which has a duration of six years). They were novices on the content domain of the examples (heredity problems), which had yet to be taught in the formal curriculum. Participants were randomly assigned to the experimental (n = 20) or control condition (n = 19).

Materials

Pretest and Posttest The pretest and posttest consisted of 5 paper and pencil heredity problems, at five levels of complexity (see Figure 1), presented in random order. The students were informed at what level of complexity each problem was. These heredity problems could be solved by going through the following five steps: (1) translate the phenotypes (expression of genetic trait) described in the cover story into genotypes (a pair of upper and/or lower case letters representing genetic information); (2) put these genotypes into a hereditary diagram; (3) determine direction of reasoning and number of Punnett Squares; (4) fill in Punnett Square(s); (5) extract final solution from Punnett Square(s). The posttest problems were equivalent but not identical to the pretest problems; they had similar structural features and were of similar complexity, but the surface features (cover stories) differed. On both tests, participants were instructed to write down the steps they took to reach their solution.

Complexity Level	Support Level	Learning Tasks				
Complexity 1 - 2 generations - 1 unknown - 1 solution - Deductive	Completion 3 steps worked-out	Eye color	Hair structure	Shapes cat fur	Japanese Apple tree	Depression
	Completion 2 steps worked-out	Eye color	Hair structure	Sickle cell Anemia	Curve chicken beak	Guinea Pigs
	Conventional 0 steps worked-out	Eye color	Hair structure	Huntington	Milk Allergy	Cleft Lip
Complexity 2 - 2 generations - 1 unknown - Multiple solutions - Deductive	Completion 3 steps worked-out	Eye color	Hair structure	Flower color	Widow's peak	P.R.A.
	Completion 2 steps worked-out	Eye color	Hair structure	Shapes cat fur	Albinism	Pea plant
	Conventional 0 steps worked-out	Eye color	Hair structure	Tongue Curling	Japanese Apple tree	Fruit flies
Complexity 3 - 2 generations - 1 unknown - Multiple solutions - Inductive	Completion 3 steps worked-out	Eye color	Hair structure	Fruit flies	Curve Chicken beak	Wolfram syndrome
	Completion 2 steps worked-out	Eye color	Hair structure	Dog tail length	Japanese Apple tree	Milk allergy
	Conventional 0 steps worked-out	Eye color	Hair structure	Freckles	Flower Color	Earlobes
Complexity 4 - 3 generations - 1 unknown - Multiple solutions - Both ways	Completion 3 steps worked-out	Eye color	Hair structure	Albino	Shapes cat fur	Fruit flies
	Completion 2 steps worked-out	Eye color	Hair structure	Fruit flies	Tongue Curling	Flower color
	Conventional 0 steps worked-out	Eye color	Hair structure	Pea plant	Dimples	Depression
Complexity 5 - 3 generations - 2 unknowns - Multiple solutions - Both ways	Completion 3 steps worked-out	Eye color	Hair structure	Milk Allergy	Depression	Huntington disease
	Completion 2 steps worked-out	Eye color	Hair structure	Dog tail Length	Wolfram syndrome	Flower color
	Conventional 0 steps worked-out	Eye color	Hair structure	Cystic Fibrosis	Fruit flies	Rat fur

Figure 1: Overview of the task database.

Mental effort rating After each problem in the pretest and posttest, participants rated how much mental effort they invested in solving that problem on a 9-point rating scale (Paas, 1992).

(Self-)assessment After the mental effort rating, participants self-assessed their performance on a 6-point rating scale ranging from 0 (none of the five steps correct) to 5 (all steps correct). After the experiment, participants' performance was scored by the experimenter on the same scale (i.e., max. problem: 5; max. test: 25).

Task selection After self-assessment, students indicated on an overview of the task database (Figure 1) what problem they would select next. At each of five complexity levels (left column), there were three levels of support: completion problem, 3 steps worked-out (white row); completion problem, 2 steps worked-out (light gray row); conventional problem, no steps worked-out (dark gray row). At each level of support within each complexity level there were 5 tasks to choose from, which had equal structural features but different cover stories. Participants knew the complexity level of the problem they had just worked on. They did not actually get the problem they selected to work on next; test problems were the same for all students.

Modeling examples The four modeling examples consisted of a recording of the model's computer screen along with a spoken explanation by the model of what s/he was doing. The gender of the models was varied: two examples were by two different male models, and two examples were by two different female models (see Table 1). In the experimental condition, the modeling examples consisted of three "phases":

(1) Problem solving: The model performed the problem solving task. Two models worked on problems of complexity level 1, and two models worked on problems of complexity level 2 (i.e., of the five complexity levels present in the task database and in the pretest and posttest; see Table 1). The quality of the models' performance varied between the examples: one example showed a model accurately solving the problem, but in the other three examples the models made one or more errors (see Table 1). This was done to create variability in phases 2 and 3 of the examples, that is, in the model's self-assessment scores and task selections (i.e., if the model would not make any errors or would detect and correct them immediately, they would always have the highest possible self-assessment score).

Table 1: Overview of modeling example characteristics.

Example	Model	Performance	Complexity
1	Male 1	0 errors	Level 1
2	Female 1	2 errors	Level 1
3	Male 2	4 errors	Level 2
4	Female 2	1 error	Level 2

(2) Self-assessment: Following task performance, the model rated invested mental effort on the 9-point rating scale and assessed their performance on the 6-point rating scale, assigning themselves one point for each correct step. The models' self-assessment was always accurate.

(3) Task selection: Then, the model selected a new task based on a combination of the performance score and the mental effort score. The models used a table (see Figure 2) in which the relationship between performance and mental effort scores was depicted, which could be used to infer a recommended 'step size' for task selection (e.g., performance of 4 and mental effort of 3 means a step size of +2). A positive step size means a recommendation to select a more challenging task (i.e., less support or higher complexity level), a step size of 0 means repeating a comparable task (i.e., same level of support and same complexity level), and a negative step size means a recommendation to select a simpler task (i.e., higher level of support or lower level of complexity). This kind of task selection algorithm based on performance and mental effort scores has proven to lead to an effective learning path in

studies on adaptive, personalized task selection (e.g., Camp et al., 2001; Corbalan et al., 2008; Kalyuga, 2006; Salden et al., 2004). The models' task selection was always accurate.

Participants in the control condition observed only the model's problem solving (phase 1). In the time in which the participants in the experimental condition observed the model's self-assessment and task selection, participants in the control condition were instructed to indicate whether the model made any errors during task performance, and if so, what the errors were and what the correct step would have been.

Performance 4-5	**+2**	**+1**	**0**
2-3	**+1**	**0**	**-1**
0-1	**0**	**-1**	**-2**
	1, 2, 3	**4, 5, 6**	**7, 8, 9 Effort**

Figure 2: Determining task selection step size.

Procedure

The experiment was conducted in a computer room at the participants' school. First, all participants completed the pretest on paper. Participants were given four minutes to complete each problem, followed by one minute for assessing their performance (a previous study had shown this to be sufficient time for solving the problem; Kostons et al., 2010). Participants were not allowed to proceed to the next problem before the time was up; time was kept by the experimenter using a stopwatch. After completing the pretest, participants studied the modeling examples on the computer; each participant had a head set for listening to the model's explanations. In the experimental condition, the modeling examples showed participants the task performance, self-assessment, and task selection by the model. In the control condition, participants only observed the task performance by the model and then indicated whether errors were made and if so, what the correct step was. This part was computer-paced, participants had to view the examples in the order in which they were offered and could not pause, stop, or replay the examples. Finally, all participants completed the posttest on paper, according to a similar procedure as the pretest.

Data Analysis

Self-assessment accuracy on each posttest problem was determined by computing the absolute difference between participants' objective performance score and their self-assessment of their performance. The lower this difference, the more accurate participants' self-assessment was (i.e., 0 = 100% accurate). We did not compute or analyze self-assessment accuracy on the pretest, because participants managed to solve very few problems on that test. When one is not able to perform a task at all, it is not very difficult to assess one's own performance as 0. This would be highly accurate, but would have led to a substantial overestimation of participants' self-assessment accuracy, as it is not very indicative of self-assessment accuracy on tasks that they were –at least partly- able to solve.

Task selection accuracy on the posttest was determined by computing the absolute difference between the complexity level that would be recommended based on the objective performance assessment and the complexity level participants chose.

Results

For all analyses, a significance level of .05 was used, and Cohen's *d* is reported as a measure of effect size, with 0.2, 0.5, and 0.8 corresponding to small, medium, and large effect sizes, respectively (Cohen, 1988).

Acquisition of Problem-Solving Skills

Participants' mean performance score on the pretest was 2.08 (SD = 3.58), and on the posttest it was 14.31 (SD = 6.43), so all students acquired procedural skills for solving heredity problems from the modeling examples. A *t*-test showed no significant difference between the control condition (M = 12.05, SD = 7.12) and the experimental condition (M = 12.40, SD = 6.40) in the knowledge gain from pretest to posttest, $t(37) = 0.16$, *ns*.

Acquisition of Self-Assessment Skills

A *t*-test on the mean self-assessment accuracy scores on the posttest, showed that participants in the experimental condition were more accurate (i.e., lower score; M = 0.70, SD = 0.53) than participants in the control condition (M = 1.26, SD = .85), $t(37) = 2.51$, $p = .016$ (two-tailed), $d = 0.79$.

Acquisition of Task-Selection Skills

Data from 1 participant in the experimental condition were excluded from this analysis because of too many missing values. A *t*-test on the mean task-selection accuracy scores on the posttest, showed that participants in the experimental condition were more accurate (i.e., lower score; M = 0.81, SD = 0.60) than participants in the control condition (M = 1.21, SD = 0.54), $t(36) = 2.15$, $p = .038$ (two-tailed), d = 0.70.

Discussion

This study showed that students can not only acquire problem solving skills from studying modeling examples, but also self-assessment and task selection skills, which are considered to play an important role in the effectiveness of self-regulated learning.

We chose modeling examples as a means to teach self-assessment and task-selection skills, because research has shown that example-based learning is a powerful instructional strategy. Thus far, in educational settings, examples have mostly been used for teaching cognitive skills, and this study adds further evidence that they are useful for teaching metacognitive skills as well (see also Kitsantas et al., 2000; Zimmerman & Kitsantas, 2002). We did not, however, compare whether teaching self-assessment and task-selection skills via modeling examples was more effective than teaching those skills in some other way (e.g., via practice after having been explained the assessment and selection 'rules', i.e., how to come to a performance assessment score and how to combine performance and mental effort scores to select a new task), so the effectiveness of examples compared to other means of teaching self-assessment and task-selection skills might be explored in future research.

Our control condition received no self-assessment and task-selection training at all, but engaged in a filler task (finding and fixing errors) which may have been relevant for the acquisition of problem solving skills (see Große & Renkl,, 2007) and which we expected to direct students' attention towards assessment of performance (of the model) to some extent. Further analysis of data from the control condition was beyond the scope of this paper but could be interesting in its own right. For example, one might expect that students with better ability to find and correct errors would have better self-assessment skills and/or would show more knowledge gain. In addition, it might be interesting to establish whether the errors made by the models had any effects on students' test performance (especially for those students who were not able to find and fix errors).

A question we cannot address based on our data that would be interesting to address in future research concerns the relationship between students' levels of task knowledge and the accuracy self-assessment and task-selection skills. Even though there was some variability in pretest scores, these were in general very low. Problem-solving skills did increase from pretest to posttest. We cannot rule out that the increase in problem-solving skills might have increased students' self-assessment and task-selection accuracy in the control condition, we only know that the training in the experimental condition led to significantly higher accuracy than attained in the control condition. A problem that occurs in trying to establish gains in assessment and task selection accuracy is that it is hard to establish the level of these skills at pretest, because –as mentioned above- it is easy to rate performance as 0 when one is not able to perform a task at all. Although this is a highly accurate self-assessment, it probably does not reflect a high level of self-assessment skill. Therefore, a design in which students have lower and higher levels of prior knowledge at the start of the experiment would be required to address this question.

Other important questions for future research in this area concern whether training either self-assessment or task-selection skill would automatically lead to improvements in the other skill or whether both need training as in our experimental condition, as well as whether acquired self-assessment and task selection skills can transfer to other tasks in the same domain or even to other domains. We assume that spontaneous transfer is not very likely or would not be very effective, as assessment criteria and standards will differ for different types of task. However, we do expect that experience with self-assessment and task selection through training in one task or domain may facilitate acquisition of those skills for other tasks or domains (i.e., transfer in the sense of preparation for or accelerated future learning; Bransford & Schwartz, 1999).

Last but certainly not least, the most important question for future research is whether students can apply the self-assessment and task selection skills they acquired from modeling examples in a self-regulated learning environment in which they are allowed to select which problems to work on. If so, one would expect training self-assessment and task-selection skills to improve learning outcomes attained as a result of self-regulated learning.

Acknowledgments

The first author was supported by a Veni grant from the Netherlands Organization for Scientific Research (NWO; 451-08-003).

References

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

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

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology, 94*, 416-427.

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

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*, 523-535.

Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development, 56*, 45-72.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher and A. Koriat (Eds.), *Attention and performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). Cambridge, MA: MITPress.

Braaksma, M. A. H., Rijlaarsdam, G., & Van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology, 94*, 405-415.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*, 61-101.

Camp, G., Paas, F., Rikers, R. M. J. P., & Van Merriënboer, J. J. G. (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort, and mental efficiency. *Computers in Human Behavior, 17*, 575-595.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

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

Corbalan, G., Kester, L., & Van Merriënboer, J .J. G. (2008). Selecting learning tasks: Effects of adaptation and shared control on efficiency and task involvement. *Contemporary Educational Psychology, 33*, 733-756.

Dunning, D., Johnson, K., Erlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*, 83–87.

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction, 17*, 612-634.

Kalyuga, S. (2006). Assessment of learners' organised knowledge structures in adaptive learning environments. *Applied Cognitive Psychology, 20*, 333-342.

Kitsantas, A., Zimmerman, B. J., & Cleary, T. (2000). The role of observation and emulation in the development of athletic self-regulation. *Journal of Educational Psychology, 92*, 811–817.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8*, 30-43.

Kostons, D., Van Gog, T., & Paas, F. (2010). Self-assessment and task selection in learner-controlled instruction: Differences between effective and ineffective learners. *Computers & Education, 54,* 932-940.

McLaren, B. M., Lim, S., & Koedinger, K. R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2176-2181). Austin, TX: Cognitive Science Society.

Moos, D. C., & Azevedo, R. (2008). Self-regulated learning with hypermedia: The role of prior domain knowledge. *Contemporary Educational Psychology, 33*, 270–298.

Niemiec, R. P., Sikorski, C., & Walberg, H. J. (1996). Learner-control effects: A review of reviews and a meta-analysis. *Journal of Educational Computing Research, 15*, 157–174.

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

Paas, F., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*, 122-133.

Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognize designers' styles. *Learning and Instruction, 19*, 185-199.

Salden, R. J. C. M., Paas, F., Broers, N. J., & Van Merriënboer, J. J. G. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in Air Traffic Control training. *Instructional Science, 32*, 153-172.

Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: A self-efficacy analysis. *Journal of Educational Psychology, 73*, 93-105.

Schunk, D. H., & Hanson, A. R. (1985). Peer models: Influence on children's self-efficacy and achievement. *Journal of Educational Psychology, 77*, 313-322.

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

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–295.

Van den Boom, G., Paas, F., Van Merriënboer, J. J. G., & Van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: Effects on students' self-regulated learning competence. *Computers in Human Behavior, 20*, 551-567.

Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior, 25*, 785-791.

Van Gog, T., & Paas, F. (2009). Effects of concurrent performance monitoring on cognitive load as a function of task complexity. In N. Taatgen, & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1605-1608). Austin, TX: Cognitive Science Society.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction, 16*, 154-164.

Van Gog, T., & Rummel, N. (in press). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*.

Wouters, P., Paas, F., & Van Merriënboer, J. J. G. (2009). Observational learning from animated models: Effects of modality and reflection on transfer. *Contemporary Educational Psychology, 34*, 1-8.

Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*, 660-668.