Learning and Instruction 16 (2006) 170–181

Learning and Instruction

www.elsevier.com/locate/learninstruc

Commentary

When worked examples don't work: Is cognitive load theory at an Impasse?☆

Roxana Moreno

University of New Mexico, Educational Psychology Program, Simpson Hall 123, Albuquerque, NM 87131, USA

1. Worked examples and the worked-example effect

Worked examples are instructional tools to teach problem-solving skills. They usually consist of modelling the process of problem solving in a well-structured domain such as physics or mathematics by presenting an example problem and demonstrating the solution steps and final answer to the problem (Renkl, Stark, Gruber, & Mandl, 1998). When carefully designed, a sequence of worked examples is sufficient to induce problem-solving mathematics skills, even when instruction is absent (Zhu & Simon, 1987). Using worked examples in problem-solving instruction is consistent with a four-stage model of expertise that is based on the well-known *ACT-R* framework (Anderson, Fincham, & Douglass, 1997). In this model, learners who are in the first stage of skill acquisition solve problems by analogy; they use known examples of problems, and try to relate those problems to the new problem to be solved. At the second stage, learners have developed abstract declarative rules or schemas, which guide them in future problem solving. At the third stage, with sufficient practice, the schemas become proceduralised, leading to the fourth stage of expertise where automatic schemas and analogical reasoning on a large pool of examples are combined to successfully solve a variety of problem types. Empirical evidence has shown that learning with worked examples is most important during initial skill acquisition stages for well-structured domains such as physics, programming, and mathematics (VanLehn, 1996).

As can be seen from the four-stage model described, an important condition for successful problem solving is the availability of problem-type schemas, representations of problem categories with their corresponding category-specific solution procedures (Gick & Holyoak, 1983; Reed, 1993). Once a problem is identified as belonging to a known problem type, the appropriate schema is retrieved from long-term memory, and the solution procedure that is associated with that problem type is activated in working memory and used to produce a solution to the new problem. Schemas have been found to account for performance differences between experts and novices (VanLehn, 1996), a finding which led Sweller and colleagues to begin investigating practice-oriented instruction using worked examples to promote the schema acquisition of novice students (Sweller & Cooper, 1985; Sweller & Levine, 1982).

In a set of classic studies documenting the effectiveness of learning from worked examples, Sweller and colleagues found that, compared to learning by solving problems, where students are asked to engage in means–ends analyses, learning with example–problem pairs, where one example is followed by isomorphic problems to be solved, increased near transfer (Mwangi & Sweller, 1998; Sweller & Cooper, 1985; Tarmizi & Sweller, 1988; Ward & Sweller, 1990). The finding that example-based learning is more effective for problem-solving skill acquisition than the

☆ This material is based upon work supported by the CAREER 0238385 and ROLE 0231774 grants from the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the funding agency.

E-mail address: moreno@unm.edu

0959-4752/$ - see front matter
doi:10.1016/j.learninstruc.2006.02.006

standard procedure of solving problems, has been called in the cognitive load literature the *worked-example effect* (Sweller, Van Merrienboer, & Paas, 1998). The goal of more recent research, such as the studies included in this volume, is to extend research on worked-example effects by focusing on the optimal conditions for learning from examples (Renkl, Atkinson, Maier, & Staley, 2002). Rather than comparing learning outcomes between problem-solving and example–problem conditions, current research compares the learning outcomes of presenting students with different worked-example characteristics and example-based methods as derived from a cognitive load theory (CLT) framework.

2. Cognitive load theory assumptions and worked-example instruction

CLT explains learning outcomes by considering the strengths and limitations of the human cognitive architecture and deriving instructional design guidelines from our knowledge about how the human mind works (Paas, Renkl, & Sweller, 2003). The basic assumptions underlying CLT are that the human information processing system is characterized by the following: a limited working-memory capacity—the idea that only a few pieces of information can be actively processed at any one time (Baddeley, 1992); virtually unlimited capacity long-term memory—the idea that long-term memory consists of a vast number of hierarchically organized schemas (Paas et al., 2003); and automatic processing—the idea that, after being sufficiently practiced, schemas can operate under automatic processing and therefore require minimal working memory resources (Kalyuga, Ayres, Chandler, & Sweller, 2003). By taking into consideration the demands on the limited cognitive resources that are needed for schema acquisition and proceduralisation, CLT allows for predictions and explanations as to how learning can be effectively supported by teaching and instruction. Accordingly, cognitive load researchers have drawn from CLT to explain and predict how students learn from different instructional designs. For instance, the worked-example effect can be explained as being the result of a practice method that makes a more efficient use of students' limited cognitive resources than the one resulting from problem-solving practice. Rather than engaging in extensive search processes to produce the correct solution steps to solve the problem, students who study examples can use their limited cognitive resources for the induction of abstracted and generalisable problem-solving schemas that can be used to solve future problems with the same underlying structure (Sweller, 1988, 1994; Sweller et al., 1998).

Recent theoretical developments have conceptualised the idea of cognitive load further by distinguishing between extraneous, intrinsic, and germane load (Paas et al., 2003). The first source of load, namely, extraneous cognitive load (ECL), is the thrust of CLT and also the basis for the original worked example research. ECL has been defined as the cognitive load that is imposed by instructional designs that require students to engage in activities "that are not directed at schema acquisition or automation (Sweller, 1994, p. 299)." In fact, CLT was devised primarily to provide principles for the reduction of ECL, with worked examples being just one of the many instructional methods developed to reduce the extraneous load that resulted from presenting students with cognitively demanding traditional problem-solving techniques (Sweller & Cooper, 1985; Tarmizi & Sweller, 1988).

Intrinsic cognitive load (ICL) is the load that depends on the difficulty of the material to be understood. Although originally proposed to be the fixed source of load (Sweller et al., 1998), there is new evidence that learning materials of high complexity is enhanced when the interacting elements are taught first in isolation and the relevant interactions are instructed later, suggesting that intrinsic load can be manipulated by instruction (Pollock, Chandler, & Sweller, 2002; Van Merriënboer, Kirschner, & Kester, 2003). Moreover, it is the opinion of this author that the fixed nature of ICL contradicts the very assumptions of CLT itself. That is, material that is complex for one individual may be very simple for another. It all depends on the schemas that have been acquired by that individual in the past and the degree to which those schemas have become proceduralised in long-term memory (Sweller, 1994).

Finally, CLT introduces the concept of germane cognitive load (GCL) as the load that results from cognitive activities that are relevant to the processes of schema acquisition and automation. Therefore, this type of load is desirable because "it contributes to, rather than interferes with learning (Sweller et al., 1998, p. 264)." For instance, within the worked-example research, some studies have examined techniques to increase example elaboration, that is, methods that prime the learner to draw inferences concerning the structure of the example, the rationale underlying solution procedures, and the goals accomplished by individual steps (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Renkl, 1997). Similarly, other researchers have examined techniques that require students to compare worked examples to increase the likelihood that they will abstract, by comparison, the structural features that examples may have in common from superficial features of the examples (Quilici & Mayer, 1996).

Two additional pivotal assumptions of CLT that will become relevant in this discussion are that (a) intrinsic, extraneous, and germane sources of cognitive load are additive and (b) designs that reduce extraneous sources of load lead to increased learning because learners are able to use the freed resources to engage in germane cognitive activities (Paas et al., 2003). Taken together, a direct implication of CLT assumptions for the design of worked examples is that example-based instruction should minimize students' use of cognitive resources in activities that are not relevant to schema acquisition and automation (sources of ICL and ECL) and maximize students' use of cognitive resources in germane activities (sources of GCL) within the limits of working memory capacity. As will be seen, the research reported in this volume takes this general CLT guideline to derive the different worked-example treatments used in the respective studies (Sweller et al., 1998). I now turn to a discussion of this work in more detail by summarizing the main findings and how each study helps advance our understanding of CLT and instructional design.

3. Overview of the empirical papers

Although CLT seems to give a satisfactory explanation for the worked-example effect, research on worked examples has shown that the mere presentation of worked examples is not enough to promote students' schema construction (Renkl, 1997). For example, students may not be able to identify relevant information in the examples or may focus on irrelevant surface features (Ross, 1989); they also often suffer from the "illusion of understanding" when studying worked examples (Renkl, 1999); and have difficulty generalizing solutions from examples to novel problems (Catrambone & Holyoak, 1989). Furthermore, despite the reduction of ECL resulting from studying rather than solving problems, learners do not to spontaneously engage in germane cognitive activities such as example elaboration or comparison (Chi et al., 1989; Gerjets, Scheiter, & Catrambone, 2004). The research presented in this volume shares the common goal of examining methods aimed at overcoming some of these limitations. This is done in three different ways. First, by decreasing the ECL of example-based learning further such as slowly increasing the amounts of solution steps that students need to solve after studying examples (Reisslein, Atkinson, Seeling, & Reisslein, 2006). Second, by decreasing the ICL from the examples themselves, such as lowering the complexity of the information contained in the example problems (Gerjets, Scheiter, & Catrambone, 2006). Third, by increasing GCL with cognitive activities that are relevant to the process of schema acquisition and automation such as asking students to give oral explanations about what steps are needed to solve a problem and why (Catrambone & Yuasa, 2006), prompting students to give written explanations about a set of elaborative questions (Große & Renkl, 2006), and presenting explanations for why the problem-solving steps are chosen or are appropriate (Catrambone & Yuasa, 2006; Van Gog, Paas, & Van Merriënboer, 2006).

3.1. Decreasing extraneous cognitive load in worked example instruction

Reisslein et al. (2006) compare three example-based methods (example–problem, problem–example, and fading) in the engineering domain with learners of high prior knowledge (HPK) and low prior knowledge (LPK). The researchers take the four-stage cognitive skill acquisition model described in the introduction (Anderson et al., 1997) to derive example-based methods for different levels of expertise. More specifically, according to this model, learners in the initial stage solve problems by referring to specific examples and attempting to relate them to the problem to be solved (similar to the example–problem method). In contrast, because learners in the final stage are able to solve problems by themselves, they can practice solving problems and later obtain feedback about their solutions (similar to the problem–example method). The third method used in this study, namely fading, was designed to transition from example study in early stages of skill acquisition to problem solving by asking students to complete an increasing number of steps. In line with the four stage model and the expertise reversal effect (Kalyuga et al., 2003), the researchers expected an interaction between method and level of prior knowledge according to which LPK learners would learn better from the example–problem and fading methods than from the problem–example pairs, whereas HPK learners would learn better from problem–example pairs than from example–problem and fading methods. They additionally predicted that learners with HPK would outperform their LPK peers on performance measures and, due to the positive effects of backward fading found in the past (Renkl, Atkinson, & Große, 2004), that learners provided with a fading method would outperform those provided with an example–problem method.

Results showed no learning or time on task differences between the conditions, thus, contradicting prior research findings where fading methods were found to be superior to example–problem methods (Renkl et al., 2002, 2004). As expected, students who started out with HPK spent less time studying the materials and had higher transfer scores than their counterparts. Finally, based on the finding that HPK participants showed higher near transfer scores when learning from problem–example pairs than LPK participants, the authors conclude that the study demonstrated an expertise reversal effect. There was no effect for far transfer measures.

How does this study advance our understanding of CLT and instructional design? Interestingly, although the research question seems to belong to the cognitive load literature because it investigates the interaction between students' prior knowledge and methods that require completing more or less problem steps (presumably imposing more or less ECL, respectively), the only mention to CLT in this article is indirect and through the reference to the expertise reversal effect. Because there is no explicit mention about how the different treatments are hypothesized to affect the load of students while learning from the worked examples and no measure of cognitive load is used, it is difficult to make any valid conclusions about how this study supports current CLT assumptions. I will, however, make an observation that may strengthen future research in this area: The way that expertise is operationalised in this study is not aligned with CLT assumptions and may explain why the findings do not replicate other expertise reversal effects. In this study, expertise is identified with scoring above the median in an introductory engineering pretest taken a week before the treatment. Because the participants were all undergraduate students, it seems that separating experts from novices with a median split pretest measure is not a valid measure of expertise but may be a proxy measure of achievement or ability instead. Becoming an expert is a long process, which typically takes many years of deliberate practice and is not dependent on ability (Ericsson, 1996; Ericsson & Smith, 1991; Hayes, 1988). If the differences between low and HPK groups are based on ability, this alone would explain why the HPK group had overall higher scores, spent less time studying examples, and demonstrated better near transfer under more demanding conditions (problem–example) than the LPK group. In addition, because HPK students are not experts, they are not better able to transfer what they learned in isomorphic problems to more complex problems (far transfer) under more demanding conditions of learning (problem–example). Moreover, to conclude that an expertise reversal effect exists, it is necessary to have evidence that a treatment that helped a novice group learn better hindered an expert group as well as the opposite. There is no evidence in this study that the practice-example or fading conditions hurt those with HPK. In sum, to be able to infer theoretical and practical cognitive load implications from worked-example research that examines expertise reversal effects, it is most important to include valid and reliable measures of the cognitive load, which is being manipulated (in this case ECL), and students' expertise.

3.2. Increasing germane cognitive load within worked-example instruction

In an attempt to infuse more active involvement in the process of learning with examples, Catrambone and Yuasa (2006) asked students to produce self-explanations while studying worked examples to write database queries. Chi et al. (1989) had discovered in earlier research that students who attempted to establish a rationale for the solution steps presented in worked examples appeared to learn more than those who did not, a phenomenon they called the self-explanation effect. Following this self-explanation method, the researchers asked some students to study a set of worked examples and elaborate on either the conditions for the actions to be taken or the associations between conditions and actions. Another group of students learned with a passive method by studying a set of worked examples that contained explicit elaborations of the materials corresponding to the two active conditions. The authors expected to find a self-explanation effect according to which the active groups would outperform the passive group. Although no clear predictions are made about the effect of the elaboration degree (or the amount of information given to learners in the passive conditions), the authors seem to favour the hypothesis that too much elaboration/information may hurt learning by imposing too high of a cognitive load and cite prior redundancy effects and expertise reversal effects to support this idea (Kalyuga et al., 2003; Sweller, 1988). The underlying assumption is that the participants had significant enough knowledge/expertise to warrant a negative effect on learning when redundant information is presented/asked to elaborate on.

The results of this study showed that participants in the active conditions took longer to study the manual but were faster to complete the transfer tasks than those in passive conditions. In addition, the higher elaboration/more information to elaborate on groups were the most effective in improving procedural performance. There was no elaboration effect (active versus passive groups) on learning. Because the additional load imposed by the higher elaboration (more information to elaborate on) did not produce a redundancy effect, the authors conclude that, contrary to their expectation, students were not cognitively overwhelmed. In addition, the results are at odds with prior research showing that cognitive activity in the form of student elaboration helps learning from worked examples and that instructional explanations that present more rather than less information usually hurt learning (Carroll, 1998). How does this study advance our understanding of CLT and instructional design? Similar to the study of Reisslein et al. (2006), despite the overview of relevant cognitive load research in using elaboration techniques with example-based learning, no explicit relationship is made between the treatments used in the study and how those treatments are expected to impact (or not) the different cognitive load types. Despite the many references to ''cognitive load'' and the obvious relationship between GCL and the elaboration methods used in this study, this construct is not introduced or measured. Another similarity with Reisslein et al.'s study is that the authors frame the main research question around issues of expertise and redundancy. However, both expertise and prior knowledge are ill defined and not measured. More specifically, the authors are interested in determining if presenting more information or having students elaborate more on the example problems will hurt students' learning by producing too much cognitive load and interfering with schema development, particularly if the additional information is not needed by that learner. In their own terms ''sophisticated learners—which describes the learners in the present study—can actually suffer in their subsequent task performance if the training materials contain information that could be inferred reasonably easily by the learner from other material in the lesson (Catrambone & Yuasa, 2006, p. XXX).'' What is a sophisticated learner? How do we know when additional information is needed or not needed by a learner without assessing prior knowledge? How do we know how the treatments affected learning without assessing load? This research, although promising and creative, can be strengthened significantly by asking these questions up front to guide its methodology and design.

Große and Renkl (2006) tested the learning effects of presenting more than one solution method to worked-out examples in two experiments that included written and oral self-explanations, respectively. Based on CLT's assumption that worked-example instruction reduces problem-solving ECL and, therefore, frees cognitive resources to process the multiple representations, they expect multiple solution methods to foster learning from examples via increased GCL. However, they are wary that instructional support may be needed to help integrate the multiple representations with each other and to avoid a split-attention effect (Ayres & Sweller, 2005). In addition, consistent with past self-explanation effects, they expect written self-explanations prompted by the instructional program to increase GCL and learning.

The findings of the first experiment showed that multiple solutions fostered learning, but that self-explanation prompts, rather than being helpful, were detrimental to learning. The authors interpreted this contradictory finding as a consequence of lack of time, having prompted students for excessive elaborations, or having prompted for elaborations that conflicted with students' spontaneous self-explanations. The goal of the second experiment was to replicate the study on multiple solutions using a think-aloud method and varying the representational code of the solutions. Contrary to the results of the first study, no effect of multiple solutions on learning was found. Moreover, presenting multiple solutions reduced the level of anticipation and the identification of similarities between example types. The lack of replication of the second study was interpreted as a consequence of the higher intrinsic load of the materials used in the second experiment and the added extraneous load from the think-aloud method, which, taken together ''might have severely limited the cognitive resources available for activities related to germane load (Große & Renkl, 2006, p. XXX).''

How does this study advance our understanding of CLT and instructional design? Unlike the prior two studies, Große and Renkl (2006) explicitly link CLT assumptions with their research questions. This is done by clearly explaining how each method is expected to affect the different load sources and therefore the learning outcomes. However, they are similar to the prior studies in that the relevant cognitive load constructs are not measured, which limits the interpretation of their contradictory results. For example, had ICL been measured in both experiments, it would have been possible to make inferences about the impact of the ICL of the materials used in both studies and their discrepant findings on presenting students with multiple solutions. Had GCL been measured in experiment 1, it would have been possible to confirm or disconfirm the authors' hypothesis that the self-explanation activity was overloading. Finally, an interesting issue to examine in future research is to compare the cognitive load and learning consequences of studying examples that include multiple solutions with and without a spontaneous self-explanation activity. This design would allow the authors to test their hypothesis that the reversal of the self-explanation effect found in experiment 2 was in part due to the excessive additive load of presenting multiple solutions and self-explanations combined.

Van Gog et al. (2006) investigated the effectiveness of a troubleshooting training program consisting of solving conventional problems or solving worked examples either by providing *product* information, namely the solution steps, or by providing *process* information including explanations for why the problem-solving steps are chosen.

In line with the pioneering findings of Sweller and colleagues (Mwangi & Sweller, 1998; Sweller & Cooper, 1985; Tarmizi & Sweller, 1988; Ward & Sweller, 1990), the authors expected to find a worked-example effect according to which studying worked examples would result in more effective learning (better transfer and less investment of time and effort) than solving conventional problems. In addition, an interaction was expected according to which presenting process information would increase investment of effort and enhance transfer performance when studying worked examples but would increase investment of effort and reduce transfer when solving problems. The rationale for this hypothesis is that the higher effort resulting from presenting process information is assumed to be an indication of GCL, which can be only handled in worked-example conditions where ECL is lower than in problem-solving conditions. On the other hand, process information was expected to hurt transfer when solving conventional problems because this problem-solving activity already imposes high ECL on the learner.

How does this study advance our understanding of CLT and instructional design? The findings of this study support CLT in two ways. First, they replicate the worked-example effect: participants who studied worked-out solutions had higher transfer scores and reported investing less mental effort during training than those who solved conventional problems (an indication of the reduction of ECL). In addition, the findings validate the assumption that activities that help schema acquisition and proceduralisation increase cognitive load: participants in the conditions with process information felt that they had to invest more mental effort during training than participants in the conditions without this information (an indication of increase in GCL). On the other hand, contrary to what is predicted by CLT assumptions, adding process information to worked examples did not result in higher transfer performance. The authors explain this contradictory finding in two ways. First, they suggest that the format of the process information may have caused a split-attention effect. Because the text of the process-oriented worked examples was much longer than that of the product-oriented worked examples and was not integrated with the examples, it may have increased ECL. Had the additional information be given in an integrated diagram format or in audiovisual format, the results may have changed. Second, the additional process-oriented information may have increased the task complexity and therefore, increased ICL rather than the intended GCL. As can be seen, similar to Große and Renkl's (2006) study, the limitations of the cognitive load measures used in this study make it impossible to test the two alternative hypotheses that are offered. Although a strength of this research is the inclusion of an indirect measure of overall mental effort which has been extensively used in the past and found to be highly reliable (Paas, Tuovinen, Tabbers, & Van Gerven, 2003), unfortunately, this measure does not discriminate between the different cognitive load sources. Thus, we are not able to fully understand if, indeed, the lack of learning benefits of providing process information stem from compensatory effects between GCL and ECL, GCL and ICL, if the process information, rather than promote germane activity as intended, merely increased students' ECL or ICL, or if the GCL, ECL, and/or ICL added by this information exceeded overall working memory capacity.

3.3. Decreasing intrinsic cognitive load within worked-example instruction

Gerjets et al.'s (2006) research is consistent with the recent reformulation of the ICL construct according to which, rather than fixed, the intrinsic complexity of the material to be learned can be manipulated by instruction. In particular, the researchers use modular presentation of solution procedures to reduce the ICL associated with learning from the more traditional molar examples, which are presumably high in ICL because they require students to refer to problem categories and category-specific solution procedures. Modular examples reduce ICL by allowing learners to keep only a limited number of elements active simultaneously in working memory. The researchers combined modular and molar presentation of problem solutions with either instructional explanations or self-explanation prompts. Similar to the case of Van Gog et al. (2006) and Große and Renkl (2006), this study is framed around the most recent developments of CLT. Accordingly, they expect the presentation of modular problem solutions to decrease ICL. In their words, the modular format "…should impose less intrinsic cognitive load than a molar example format and accordingly free cognitive resources that can then be used by learners to engage in example elaborations (Gerjets et al., 2006, p. XXX)." In addition, they expect that providing instructional explanations should be especially helpful for learners presented with molar examples, which are intrinsically more complex to understand and may require scaffolding, but self-explanation prompts to be more helpful for modular worked-out examples, because adding an elaboration activity to materials that are intrinsically complex may result on cognitive overload.

The results replicated prior research findings where learning with modular examples was superior to learning from molar examples (Gerjets et al., 2004). On the other hand, the first experiment showed that more elaborated instructional explanations did not foster learning for modular or molar example learning. It was concluded then that, for modular examples where ICL is low, deeper instructional explanations become superfluous to students' own elaborations. For molar examples, this is not likely to be the case due to the high ICL. However, the lower ratings on the effort scales when learning from more elaborated explanations were interpreted to indicate that deeper explanations gave these students an "illusion of understanding" (Renkl, 1999, 2002). Accordingly, the authors changed their original prediction for experiment 2 (i.e., that self-explanations would hurt students learn from molar examples). In the second experiment, it was hypothesized instead that the self-explanation prompts for molar examples would promote learning by helping students overcome the illusion of understanding found in experiment 1. The results from the second study, however, did not prove this to be true. Moreover, contrary to what was predicted, prompting for self-explanations hurt rather than helped learning with modular examples. The authors then offer a set of alternative explanations for this result including potential redundancy effects, split-attention effects, and design issues (i.e., number of prompts and feedback format and quality). They finally conclude recommending future research to investigate whether promoting anticipative reasoning is a more appropriate means to foster transfer performance than promoting principle-based explanations and taking students' prior knowledge level into consideration in this inquiry.

How does this study advance our understanding of CLT and instructional design? The authors should be commended for having tested the assumption that ICL is not fixed by comparing the learning and load outcomes of modular versus molar worked examples. The beneficial learning effects for using modular rather than molar worked examples found in this study and past studies, support the most current CLT development according to which the level of ICL can be manipulated via instruction. Similar to the rest of the research presented in this volume, an aspect where refinement is needed is the definition and measurement of the different cognitive load types. That is, although the data on learning and time on task allow us to infer the relative efficiency of each one of the treatments, no evidence exists for the validity or reliability of the subjective measures used in this study. Unlike other cognitive load research, this work demonstrates an effort to explore alternative cognitive load measures via the NASA-TLX questionnaire (Hart & Staveland, 1988). However, a limitation in using this measure is that, because it was not based on CLT, the three types of load are not explicitly assessed but other constructs are measured instead. The authors assume a mapping "between the theoretical assumptions of Cognitive Load Theory and the items of the modified version of the NASA-TLX (Gerjets et al., 2006, p. XXX)" but there is no information about the structure of this instrument and how the different factors may be related to intrinsic, extraneous, and germane load during learning. Consequently, important questions remain open. How do we know that modular examples help learning by reducing ICL if ICL is not measured? Is it possible that molar examples increase ECL instead? How do we know that more elaborated explanations are superfluous for one group but not the other? Moreover, wouldn't CLT predict superfluous information to produce a redundancy effect? How do current CLT assumptions explain the illusion of understanding effect reported here?

4. Summary of special issue findings

The empirical work presented in this volume reveals some interesting similarities and differences also found among the current cognitive load research at large, which I will point out as promising venues for further research and development. First, all studies share the common goal of improving the design of worked-example instruction by testing the learning effects of a set of innovative methods that can be embedded within example-based learning to optimise cognitive load (i.e., using fading, oral and written self-explanations, modular examples, multiple solutions, and process information). Second, all studies rely to some degree on CLT's assumptions to accomplish this goal. As can be seen, whereas some articles are entirely framed around the most recent CLT framework, other articles endorse CLT in its original form and make no assumptions regarding the different load types, yet other studies only borrow CLT concepts and effects to try to explain their findings. Third, the articles that frame research questions around current CLT developments are to some degree consistent about how certain methods are predicted to affect GCL, ICL, or ECL. For example, studying worked examples is predicted to reduce ECL as compared to problem-solving practice; student germane activities, such as producing oral or written self-explanations, are expected to increase GCL; and instructional methods aimed at helping schema acquisition, such as presenting process information, highly elaborated example explanations, or multiple example solutions, are also predicted to increase GCL.

Similar to other efforts to optimise worked-example design in recent years, another commonality among the studies in this volume is that the interventions seem to produce no significant learning effect, especially when examining effects on far problem-solving transfer. Maybe this is evidence that there is no way to optimise cognitive load further in example-based learning, or, like the American saying goes "we can't squeeze blood out of a turnip." Therefore, a fruitful venue for future research may be to take this research to the next level by focusing on techniques aimed at promoting far transfer abilities. However, this change of focus does not go without challenges. It seems unlikely that far transfer effects can be found within the current research designs and paradigms, which are limited to brief, one-time interventions, and where students are presented with a minimum amount of isomorphic worked examples and only tested immediately after instruction.

Perhaps the most problematic issue that is transpiring across the research in this volume is that the reported findings for worked-example methods are mostly at odds with the predictions derived from CLT and past cognitive load effects. In particular, Reisslein et al.'s (2006) results contradict the findings of Renkl et al. (2002, 2004) where fading methods foster learning as compared to using example–problem pairs. Catrambone and Yuasa's (2006) results fail to replicate the robust self-explanation effect found in the past (Atkinson, Renkl, & Merrill, 2003; Chi et al., 1989; Renkl et al., 1998). Große and Renkl's (2006) first experiment demonstrates that prompting students for written self-explanations can even have a negative effect on learning, and the learning effects of presenting students with multiple solutions are reversed between their first and second experiment. Although Van Gog et al. (2006) are able to support CLT's prediction that presenting process information increases cognitive load as compared to only presenting product information, this increased load was not associated with better learning outcomes, therefore challenging the idea that this activity was a source of GCL. Finally, Gerjets et al.'s (2006) research contradicts two CLT hypotheses. First, it contradicts the hypothesis that higher elaboration of instructional explanations should increase GCL and learning (instead students reported lower levels of mental effort when learning with more elaborated explanations and learning was not affected). Second, contrary to what is predicted by CLT, prompting for self-explanations does not affect learning from molar examples and hurts rather than helps learning with modular examples.

What is concluded from this set of inconsistent findings? As can be seen, although all articles in this volume present reasonable arguments for which the instructional methods used in the respective studies should affect cognitive load and promote learning, once the results are shown to be at odds with these theoretically driven arguments, the speculative post hoc interpretations leave us with more questions than answers and cast doubt over the validity of CLT. For example, Große and Renkl (2006) conclude that the lack of a positive effect for presenting multiple solutions in their second study might have been in part due to the added extraneous load from the think-aloud method. So, are self-explanations a source of GCL or ECL? Is this dependent on the results of each research study? If so, what is the predictive value of CLT? Another example is that of Van Gog et al.'s (2006) interpretation for not finding evidence that adding process information to worked examples is a source of GCL. They conclude that the additional process information may have increased ICL rather than the intended GCL. So, does the presentation of process information increase GCL or ICL? Can CLT help us answer this question and predict learning outcomes? A last example of the unreliable predictability of results in this field is shown by Gerjets et al.'s (2006) abrupt change in their predictions from the first to the second experiment. Although presenting more elaborated explanations was originally expected to increase learning (especially from modular examples) by increasing GCL, because this effect was not found in the first experiment, it was concluded that the additional information was either unnecessary (for the modular group) or not processed sufficiently (for the molar group). In addition, self-explanations were originally expected to decrease learning from molar examples by exceeding overall cognitive capacity due to the addition of GCL to an already high ICL condition. However, based on the lack of overload found in the first experiment, the authors expected self-explanations to promote learning instead. Once the second experiment shows no self-explanation effect for the molar group and a negative learning effect for the modular group, it is concluded that the self-explanation condition may have produced redundancy and/or split-attention effects arising from the design of the prompts and feedback (all sources of ECL). Taken together, the unstable nature of these varied interpretations makes one wonder: Is CLT at an impasse? Or does this mean that our current methods do not allow us to appropriately test CLT?

5. Conclusion and suggestions for future research

The articles in this special issue are a sample of the extremely creative and prolific research programs that CLT has ignited all over the world. Nevertheless, the puzzling results that are reported here suggest that we may be going faster than we should. Despite their promise, there is strong evidence that worked examples don't always work and yet, the cognitive load field is unable to produce reliable explanations for why this is the case. Even when reductions of intrinsic and extraneous loads are carefully considered in worked-example designs and the learner is presumably left with enough resources to engage in germane activities, these activities do not necessarily result in better learning. Where do we go from here?

A first direction is methodological in nature. Despite the overall theoretical convergence among most CLT researchers, it should be clear that a major shortcoming in this field is the lack of uniform operationalisation and measurement of pivotal CLT constructs such as mental effort, difficulty, expertise, prior knowledge, and the different types of cognitive load. This is one of the most challenging areas within CLT and should become a research priority if we are to validate its assumptions. For example, due to these methodological limitations, there is no evidence to date, supporting one of CLT's main tenets: that the three cognitive load types are additive.

A second promising direction is to carefully reconsider some of the conceptualisations and assumptions laid down by CLT under the light of what we can learn from the worked-example research. As pointed out in the past (Gerjets & Scheiter, 2003), a limitation of CLT is that it assumes a one-to-one mapping between the instructional design and its resulting pattern of loads without taking into consideration moderating variables that may be interfering with this direct mapping. Atkinson, Renkl, and colleagues (Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 2005) offered some guidance to help identify factors that moderate worked-example effects. More specifically, they argue that example-based learning is dependent on the design of examples (inter and intra-example features) and individual differences in example processing, especially those related to students' self-explanations. Based on the findings reported in this volume and other worked-example research conducted in the past, I suggest extending the list of mediating factors to the following three sources: worked-example design, amount of essential processing during learning, and individual differences (Moreno, 2005).

First, worked-example learning will depend on the design of the worked examples themselves. This mediating source of worked-example effects has been divided into intra-example and inter-example design effects in the past (Atkinson et al., 2000). For instance, different intra-example designs may present more or less information in the solution steps (Catrambone & Yuasa, 2006; Gerjets et al., 2006), one or multiple solutions (Große & Renkl, 2006), subgoal highlighting (Catrambone, 1998), more or less integrated verbal and visual representations of the problem solutions (Moreno & Durán, 2004; Mwangi & Sweller, 1998), and mixed modality representations of problem solutions (Moreno & Mayer, 2002; Moreno, Mayer, Spires, & Lester, 2001; Mousavi, Low, & Sweller, 1995). Additionally, different inter-example designs may present more or less number of worked examples (Cooper & Sweller, 1987; Gick & Holyoak, 1983), more or less variability of surface and structure features (Paas & Van Merriënboer, 1994; Quilici & Mayer, 1996; Ross, 1989), and different sequencing of worked example and practice problems (Trafton & Reiser, 1993).

Second, the effectiveness of worked examples depends on the type of cognitive activity promoted by the worked-example method. If this activity is essential to the instructional objective (in this case to promote schema acquisition and proceduralisation), it should lead to increased learning. Examples are using self-explanation prompts (Chi et al., 1989), using mixed-initiative problem solving (Moreno et al., 2001), presenting incomplete examples (Paas, 1992; Renkl et al., 2002), and providing learners with subgoals (Atkinson, Catrambone, & Merrill, 2003; Catrambone, 1996, 1998; Gerjets et al., 2004).

The cognitive load and learning effects resulting from different example designs and cognitive activities have been extensively researched. However, what is becoming clearer is that these effects are highly dependent on students' characteristics. Therefore, individual differences that are relevant to CLT should be taken into consideration when trying to derive theoretical and practical implications for example-based learning. For instance, according to CLT, learners' level of prior knowledge is extremely important to be able to categorize instructional information and activities as intrinsic, extraneous, or germane, and to predict learning outcomes (Kalyuga et al., 2003; Sweller, 1988; Sweller et al., 1998). In other words "a cognitive load that is germane to a novice may be extraneous for an expert" (Paas, Renkl, & Sweller, 2004, p. 2).

However, the learning benefits of example-based methods that prompt students to engage in essential cognitive processing are not only dependent on students' prior knowledge. For example, drawing on CLT and the popular theory of multimedia learning (Mayer & Moreno, 2003), I have proposed a cognitive-affective theory of learning with media that includes self-regulation and motivation factors as learning mediators (CATLM; Moreno, 2005). According to a CATLM, some media may be perceived as more interesting than others, therefore producing positive learning benefits by influencing students to spend more effort on the task (Lester, Towns, & FitzGerald, 1999; Tang & Isaacs, 1993). Similarly, some methods embedded in the media may be perceived to be more supportive than others, therefore producing positive learning effects by reducing students' fear of failure or increasing their self-efficacy (Cennamo, 1993). Importantly, learning is affected by students' interest in the subject domain and motivation to engage in the learning task. Those who have domain-specific interest and a mastery orientation to learning show deeper learning strategies and higher levels of monitoring, self-regulation, and learning than those who are less motivated to learn or who are not interested in the domain to be learned (Pintrich & Schunk, 2001).

In addition, learners may use self-regulation to control both, their motivation and cognitive processing. When students are aware of the strengths and limitations of their knowledge, strategies, and motivation, they are better able to regulate their own learning by planning and monitoring the cognitive processes needed for understanding (Bruning, Schraw, & Ronning, 1999). In line with this thesis, Gerjets and Scheiter (2003) propose to integrate learner goals and processing strategies in the CLT framework, especially when worked-example instruction requires learner control. For instance, when studying worked examples, learners vary in the degree of elaboration as measured by cluster analyses of the type of elaborations during a think-aloud session (Renkl et al., 2002). That is, learning from worked examples is highly dependent on the quality of students' elaborations, such as if they are indicative of surface versus deep processing. It is also found that students who engage in higher metacognitive elaborations during example learning report higher mental effort values but not higher transfer scores than those who process the examples with higher cognitive elaborations and those who process examples superficially (Renkl et al., 2002). In sum, to advance our understanding about who learns from different worked-example designs and methods and how, new CLT developments should specify the mediating effects of students' individual differences, especially those related to their prior knowledge, metacognition, and motivation.

In closing, we should concede as cognitive scientists, that valid criticisms can be raised against any existing theory of cognition and that such criticism is essential to progress. Theories and constructs are useful only as long as they evolve in their heuristic, explanatory, and predictive functions, and the evolution of theories is best carried out in an environment of constructive scrutiny, where no theory is accepted uncritically. How do current CLT assumptions and methods serve our need to better understand the science and practice of instructional design? How does the taxonomy of loads help predict and explain learning with different methods for different learners? How do the principles described by Sweller (2006) advance our understanding about how students learn from examples? These questions, like self-explanation prompts, are aimed at helping us gain awareness of the common problem structure underlying the studies reported in this special issue. Hopefully, the answers to these questions will guide cognitive load researchers into new methods, viable theoretical alternatives for interpreting their findings, and serve to promote continued progress in the development of knowledge about a wide spectrum of learning phenomena.

References

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

Atkinson, R. K., Catrambone, R., & Merrill, M. M. (2003). Aiding transfer in statistics: Examining the use of conceptual equations and elaborations during subgoal learning. *Journal of Educational Psychology*, *95*, 762–773.

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.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology*, *95*, 774–783.

Ayres, P., & Sweller, J. (2005). The split-attention principle. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 135–146). Cambridge, UK: Cambridge University Press.

Baddeley, A. (1992). Working memory. *Science*, *255*, 556–559.

Bruning, R., Schraw, G. J., & Ronning, R. R. (1999). *Cognitive psychology and instruction*. Upper Saddle River, NJ: Prentice-Hall.

Carroll, J. M. (Ed.). (1998). *Minimalism beyond the Nurnberg Funnel*. Cambridge, MA: MIT Press.

Catrambone, R. (1996). Generalizing solution procedures learned from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1020–1031.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General*, *127*, 355–376.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1147–1156.

Catrambone, R., & Yuasa, M. (2006). Acquisition of procedures: The effects of example elaborations and active learning exercises. *Learning and Instruction*, *16*, 139–153.

Cennamo, K. S. (1993). Learning from video. Factors influencing learners' preconceptions and invested mental effort. *Educational Technology, Research, and Development*, *41*, 33–45.

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.

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

Ericsson, K. A. (1996). The acquisition of expert performance. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts, sciences, sports, and games* (pp. 1–50). Mahwah, NJ: Erlbaum.

Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits*. New York: Cambridge University Press.

Gerjets, P., & Scheiter, K. (2003). Goal configurations and processing strategies as moderators between instructional design and cognitive load: Evidence from hypertext-based instruction. *Educational Psychologist*, *38*, 33–41.

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science*, *32*, 33–58.

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction*, *16*, 104–121.

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

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

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of experimental and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North Holland.

Hayes, J. R. (1988). *The complete problem solver*, 2nd ed. Mahwah, NJ: Erlbaum.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*, 23–31.

Lester, J., Towns, S. G., & FitzGerald, P. J. (1999). Achieving affective impact: Visual emotive communication in lifelike pedagogical agents. *International Journal of Artificial Intelligence in Education*, *10*(3–4), 278–291.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*, 43–52.

Moreno, R. (2005). Instructional technology: Promise and pitfalls. In L. PytlikZillig, M. Bodvarsson, & R. Bruning (Eds.), *Technology-based education: Bringing researchers and practitioners together* (pp. 1–19). Greenwich, CT: Information Age Publishing.

Moreno, R., & Durán, R. (2004). Do multiple representations need explanations? The role of verbal guidance and individual differences in multimedia mathematics learning. *Journal of Educational Psychology*, *96*, 492–503.

Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology*, *94*, 598–610.

Moreno, R., Mayer, R. E., Spires, H., & Lester, J. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, *19*, 177–213.

Mousavi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, *87*, 319–334.

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

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

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*, 1–4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, *32*, 1–8.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*, 63–71.

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

Pintrich, P. R., & Schunk, D. H. (2001). *Motivation in education: Theory, research and practice*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, *12*, 61–86.

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. (1993). A schema-based theory of transfer. In D. K. Detterman, & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 39–67). Norwood, NJ: Ablex.

Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and Instruction*, *16*, 92–103.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *12*, 1–29.

Renkl, A. (1999). Learning mathematics from worked-out examples: Analyzing and fostering self-explanations. *European Journal of Psychology of Education*, *14*, 477–488.

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*, 529–556.

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

Renkl, A., Atkinson, R. K., & Große, C. S. (2004). How fading worked solution steps works: A cognitive load perspective. *Instructional Science*, *32*, 59–82.

Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *The Journal of Experimental Education*, *70*, 293–315.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology*, *23*, 90–108.

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.

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

Sweller, J. (1994). Cognitive load, learning difficulty, and instructional design. *Learning and Instruction*, *4*, 295–312.

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction*, *16*, 165–169.

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., & Levine, M. (1982). Effects of goal specificity on means–ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 463–474.

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

Tang, J. C., & Isaacs, E. (1993). Why do users like video? Studies of multimedia-supported collaboration. *Computer-Supported Cooperative Work: An International Journal*, *1*, 163–196.

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology*, *80*, 424–436.

Trafton, J. G., & Reiser, B. J. (1993). The contributions of studying examples and solving problems to skill acquisition. In M. Polson (Ed.), *Proceedings of the fifteenth annual conference of the Cognitive Science Society* (pp. 1017–1022). Hillsdale, NJ: Erlbaum.

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.

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

Van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist*, *38*, 5–13.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*, 1–39.

Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction*, *4*, 137–166.