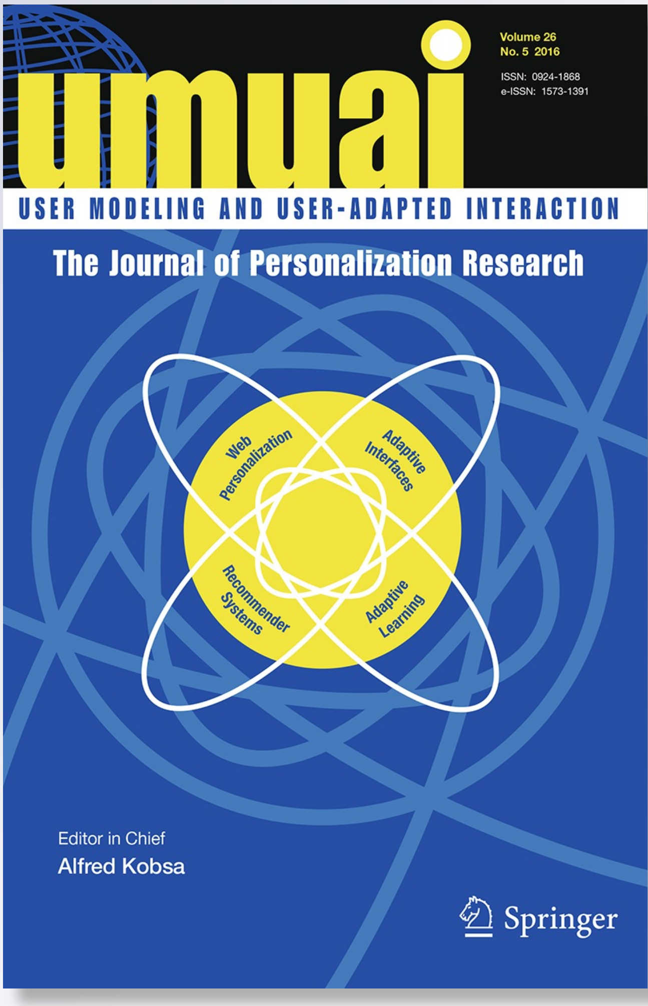
Learning with intelligent tutors and worked examples: selecting learning activities adaptively leads to better learning outcomes than a fixed curriculum

Amir Shareghi Najar, Antonija Mitrovic & Bruce M. McLaren

User Modeling and User-Adapted Interaction
The Journal of Personalization Research

ISSN 0924-1868
Volume 26
Number 5

User Model User-Adap Inter (2016)
26:459-491
DOI 10.1007/s11257-016-9181-y



User Model User-Adap Inter (2016) 26:459–491
DOI 10.1007/s11257-016-9181-y

Learning with intelligent tutors and worked examples: selecting learning activities adaptively leads to better learning outcomes than a fixed curriculum

Amir Shareghi Najar[1] · **Antonija Mitrovic**[1] · **Bruce M. McLaren**[2]

Received: 30 April 2016 / Accepted in revised form: 29 September 2016 /
Published online: 4 November 2016

Abstract The main learning activity provided by intelligent tutoring systems is problem solving, although several recent projects investigated the effectiveness of combining problem solving with worked examples. Previous research has shown that learning from examples is an effective learning strategy, especially for novice learners. A worked example provides step-by-step explanations of how a problem is solved. Many studies have compared learning from examples to unsupported problem solving, and suggested presenting worked examples to students in the initial stages of learning, followed by problem solving once students have acquired enough knowledge. This paper presents a study in which we compare a fixed sequence of alternating worked examples and tutored problem solving with a strategy that adapts learning tasks to students' needs. The adaptive strategy determines the type of the task (a worked example, a faded example or a problem to be solved) based on how much assistance the student received on the previous problem. The results show that students in the adaptive condition learnt significantly more than their peers who were presented with a fixed sequence of worked examples and problem solving. Novices from the adaptive condition learnt faster than novices from the control group, while the advanced students from the adaptive condition learnt more than their peers from the control group.

Keywords Intelligent tutoring system · Adaptive selection of learning tasks · Assistance · Self-explanation

✉ Antonija Mitrovic
tanja.mitrovic@canterbury.ac.nz

1 Department of Computer Science and Software Engineering, Intelligent Computer Tutoring Group, University of Canterbury, Christchurch, New Zealand

2 Department of Computer Science and Software Engineering, Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, USA

1 Introduction

Intelligent tutoring systems (ITS) mostly provide problem-solving opportunities, but recently there have been studies investigating the effect of combining problem solving with learning from worked examples (WEs). Early research on WEs identified the worked example effect, where students learning from examples learnt more and faster in comparison to unsupported problem solving (Sweller and Cooper 1985; Sweller 2006). Novices often have incomplete knowledge which makes problem solving difficult due to the high cognitive load, but worked examples provide solutions with associated knowledge. One of the fundamental principles of the cognitive load theory (CLT) (Sweller et al. 2011) is the *Borrowing principle*, which states that a novice learner borrows the needed information from worked examples and connects the new information with the prior knowledge. Sweller (2006) views worked examples as an instantiation of the borrowing principle. Other research also shows that WEs are an effective learning strategy if the student engages in self-explanation (Chi et al. 1994). On the other hand, in the *Randomness as genesis principle*, new information is created based on the prior knowledge in the learner's long-term memory; thereby, it loads a substantial amount of information in working memory to figure out the new information. Problem solving is the ultimate instantiation of this principle. There are three different loads for the working memory: intrinsic load, extraneous load and germane load. Intrinsic load is caused by the nature of concepts being learnt, like the difficulty of tasks; when a problem is more complex, its intrinsic load is higher. Extraneous load is more often thought of as aspects of the problems and problem solving that are not integral to solving the problem, but still part of the problem-solving environment. In contrast to extraneous load, germane load is the load needed to process information and construct schemas. Clark et al. (2006) outline different strategies and instructions to decrease extraneous and intrinsic loads, and increase the germane load.

ITSs support problem solving by providing adaptive scaffolding in terms of feedback, guidance, problem selection and other types of help. Only in recent years have studies compared learning from examples to learning with ITSs in various combinations and sequences (Schwonke et al. 2009; McLaren and Isotani 2011; Kim et al. 2007; Anthony et al. 2008; McLaren et al. 2008; Salden et al. 2010; Najar and Mitrovic 2013a, b). However, little attention has been devoted thus far to the difference between novices and advanced students in learning from examples and supported problem solving. In the early stages of learning, students should probably be given all examples (McLaren et al. 2016) or examples alternated with problem solving (Sweller and Cooper 1985). When students gain enough knowledge, they should probably be provided with problems predominantly or completely (Kalyuga et al. 2001). Research shows that students need different levels of guidance, and that the amount of assistance needs to be adapted to students' needs (Koedinger and Aleven 2007). Although there have been a few attempts to adapt examples to the student's level of expertise (Salden et al. 2009; Kalyuga and Sweller 2005), it is still an open question how much assistance should be provided to students.

In a prior study, we compared learning from examples only (EO), alternating examples and tutored problems (AEP), and tutored problems only (PO) in the domain of learning SQL database queries (Najar and Mitrovic 2013b). We scaffolded examples and problems with Self-Explanation (SE) prompts, requiring students to explain WEs or how they solved problems. The results showed that students benefitted the most from alternating examples and problems. In that study, we used a fixed sequence of examples and problems; therefore, it is possible, even likely, that some students received less or more information than they needed. This encouraged us to create an adaptive strategy that decides what type of learning activity to next present to the learner. The learning activities are problem solving, 2-step faded examples, 1-step faded examples, and worked examples, with faded steps chosen based on the student's performance. In 1-step or 2-step faded examples, students need to complete one or two clauses of an SQL query.

Our project is placed within research on adaptive selection of learning activities, which has a long history. Some of this research focused on adaptive selection of topics to be given to the student, based on his/her student model; such research is known as curriculum or instructional planning e.g. (Vassileva 1995; Peachey and McCalla 1986; Karampiperis and Sampson 2004), recommendation of learning content e.g. (Weber 1994; Santos 2011; Weber and Brusilovsky 2016), and adaptive navigation support (Masthoff 2002; Brusilovsky 2003; Hsiao et al. 2013; Sosnovsky and Brusilovsky 2015). Some of this research used worked examples as a type of scaffolding, to help the student solve a problem by providing related worked examples. For example, Burow and Weber (1996) provided worked examples from the student's own history to help when the student is having problems writing a LISP function for a related problem. Davidovic et al. (2003) also provide worked examples as a type of scaffolding, and additionally a type of learning activities in which students are required to find structural similarity between provided worked examples. In this paper, we focus on a narrower set of research projects, which focus on adaptive selection of learning activities when the sequence of problems/exercises is fixed—e.g. (Trafton and Reiser 1993). We present an overview of such research in Sect. 2.

This paper presents the study conducted to evaluate the adaptive strategy by comparing it to the best condition from the previous study, i.e. the alternating examples and tutored problems condition (AEP). We present four different modes of SQL-Tutor, the ITS that is the context of the research, in Sect. 3. The adaptive strategy is presented in Sect. 4, while Sect. 5 presents the design of our study. Please note that we have presented some results from this study in (Najar et al. 2014, 2015). Section 6 summarizes the results and provides additional findings about the effect of the two strategies on novices and advanced students, with the two types of students determined by a post-hoc split based on their scores on the pre-test. We discuss the results in Sect. 7, and conclusions and future directions are presented in Sect. 8.

2 Related work

This section starts with a short overview of the effects of self-explanation in learning from examples. Next we discuss studies that compared examples with unsupported problem solving and studies that compared examples with supported problem solving.

2.1 Self-explanation effects

Self-Explanation is a metacognitive activity in which a student explains the provided example to him/herself. Research has shown that students who explain learning material to themselves learn more than students who receive explanations (Chi et al. 1994; Brown and Kane 1988; Webb 1989; Hattie 2009). Very few students self-explain spontaneously, but can be encouraged to self-explain with carefully constructed prompts (Chi et al. 1989, 1994). Prior studies, either with a human teacher prompting self-explanations (Chi et al. 1994) or with an ITS prompting self-explanations (Aleven and Koedinger 2002; Weerasinghe et al. 2009, 2011), have convincingly shown that self-explanation is an effective metacognitive strategy resulting in deeper knowledge.

Cognitive load theory explains that WEs reduce extraneous load on working memory (Sweller et al. 2011). Extraneous load is caused by the way in which learning material is presented, and it does not directly contribute to learning (Clark et al. 2006). By reducing extraneous load, a part of working memory becomes available. If this freed working memory is then switched to germane load, learning has been shown to improve. One way of producing germane load is to prompt students to self-explain. In a study to teach concept mapping, Hilbert and Renkl (2009) showed that students who self-explained after they studied examples learnt more than students who did not engage in self-explanation. In another study, Schworm and Renkl (2006) found that self-explanation is effective for studying worked examples and solved-example problems. Solved-example problems only provide the problem formulation and the solution, while worked-out examples consist of a problem formulation, solution steps, and the final answer.

Previous research has shown that students who studied examples acquired more conceptual knowledge than procedural knowledge, and students who solved problems learnt more procedural knowledge than conceptual knowledge (Kim et al. 2007; Schwonke et al. 2009). This suggests that different types of SE are needed to scaffold problem solving and examples.

SE prompts can be of a different nature, according to the knowledge they focus on. For instance, Hausmann et al. (2009) compare justification-based prompts (e.g. "what principle is being applied in this step?") and meta-cognitive prompts (e.g. "what new information does each step provide for you?") with a new type of prompt called step-focused prompts (e.g. "what does this step mean to you?"). They found that students in the step-focused and justification conditions learnt more from studying examples than students in the meta-cognitive prompts condition. In another study, Chi and VanLehn (1991) categorised SE as either procedural explanation (e.g. answer to "why was this step done?"), or derivation SE (e.g. answer to "where did this step come from?").

2.2 Learning from examples versus unsupported problem solving

It is still very much an open question how much assistance should be provided to students. Sweller et al. (2011) showed that maximum assistance (e.g. worked examples) is more efficient than minimal assistance (e.g. unsupported problem solving) when first learning in a new domain, which has been corroborated by prior studies such as

(Atkinson et al. 2000). Apart from the advantages of learning from examples versus unsupported problem solving, recently researchers have focused on different example-based learning strategies. Van Gog (2011) investigated the difference between worked examples only (WE), worked-examples / problem-solving pairs (WE-PS), problem-solving / worked-examples pairs (PS-WE) and problem solving only (PS) on novices (problem solving was unsupported and tasks were from domain of Ohm's law in electrical circuits). The results showed that the participants in WE and WE-PS had higher performances in the post-test than PS and PS-WE. Furthermore, mental effort and test rates in WE-PS and WE were lower than PS and PS-WE.

In a later study, Van Gog (2011) stated that the previous results on WE-PS and PS-WE might not be sufficient. Examples provided after problems had a different structure to the next problem; therefore, she opined that using identical pairs might lead to a different result. She conducted a study using modelling examples (ME) and problem solving (PS) in two conditions (PS-ME-PS-ME and ME-PS-ME-PS) in the Leap Frog game. In this game, there are two sets of frogs on two sides of the screen. Students are asked to switch frogs' sides considering the rules of the game. In modelling examples, the problem solution is demonstrated to learners by an animated video (Van Gog and Rummel 2010). After two sequences of training, students worked on two tasks, of which the second was not similar to the training tasks. There was no difference in learning performance since the students learnt the most after studying the second worked example.

In another study, student's prior knowledge had an important interaction with instructional formats. Formats which are efficient for some students might not be efficient for a student with a different knowledge level (Kalyuga 2007). In other words, if the additional information was not needed by the student, the expertise reversal effect was observed (Kalyuga et al. 2001). The expertise reversal effect is caused by redundant information, information not needed by learners as they advance in their understanding and knowledge. Information that is essential for novices often becomes redundant as the level of expertise increases. Eliminating redundant information reduces the cognitive load (Chandler and Sweller 1991). Therefore, studying a worked example under such conditions might cause unnecessary cognitive load that could interfere with learning (Kalyuga et al. 2001).

Most of the prior studies showed the worked-example effect in well-defined problem domains (e.g., mathematics, science). Rourke and Sweller (2009) presented two studies in which they investigated the worked-example effect using ill-defined problems. They hypothesised that students who learnt to recognise a designer's work from WEs could recognise other work of the same designer more easily than students who learnt from solving equivalent problems. The difference between the studies was in the participants' abilities. Students in the second study had a greater level of visual literacy skill than the students in the first study, although both studies' participants had the same knowledge level on design history. The results of both studies show that the worked-example effect can be obtained in ill-defined domains.

Kyun et al. (2013) investigated another ill-defined domain, writing essays in English for non-native speakers. In Experiment 1, the learning phase for the problem-solving condition involved writing essays for two similar questions, but participants in the WE condition saw possible answers to the first question, then wrote an essay for the second question. Researchers found a significant difference in cognitive efficiency between the two groups, with the WE condition superior to the problem-solving condition. Since the participants in Experiment 1 had high levels of literature knowledge, the researchers argue that the expertise reversal might have influenced the result. Therefore, they conducted two similar experiments using participants with less literature knowledge. Results from the second experiment indicated significant differences between the two learning strategies in cognitive efficiency and retention tests. Experiment 3 was conducted with novices, and there was a significant differences in favour of the worked-examples condition. Overall, these experiments showed that the worked-example effect increases when the level of expertise decreases in an ill-defined domain.

In the previous section, we discussed SE as a way to increase germane load and improve learning. Another approach to scaffold worked examples is to test students after examples. Roediger and Karpicke (2006) show that after studying material, testing is more effective for long-term retention than restudying; this is called the 'testing effect'. While most of the work on the testing effect has been done for memory tasks (i.e., non-problem solving tasks), Van Gog and Rummel (2010) investigated the testing effect when acquiring problem-solving skills. They created two conditions: SSSS and STST (S: study example, T: testing task). Each condition had two pairs of tasks (SS or ST) and the tasks in each pair were isomorphic. There was no significant difference in the immediate post-test, but in the delayed post-test, students who only studied examples outperformed their peers. This surprising result might be caused by WEs inducing self-explaining and self-explanation correlates with a longer retention; thus, SSSS performed better than STST in the delayed post-test.

Although research shows that students learn more effectively from WEs than from unsupported problem solving, WEs do have demonstrated drawbacks. First of all, WEs must be studied to be effective; thus, when an example is ignored, it does not promote learning. Some students either skip WEs completely, or pay little attention while studying them, perhaps because WEs are not interactive, or there is less interaction compared to problem solving. On the other hand, problems require deep processing for solution. Therefore, replacing worked examples with partly worked examples is one way to minimise ignoring worked examples (Clark et al. 2006). Partly worked examples are also known as *faded examples* (i.e. a worked example in which one or more steps are left for the student to complete). Faded worked examples may be superior to conventional problems, as partly worked examples require less effort during training and impose less cognitive load than a conventional problem to solve. Faded worked examples are kind of a compromise between worked examples and problem solving.

Paas and Van Merrienboer (1993) reported the results of a study comparing WEs, faded examples and unsupported problem solving. The study had three conditions: unsupported problems, worked examples and problems pairs, and faded examples and problems pairs. The results showed that faded examples and worked examples were superior to unsupported problem solving for attaining transfer. Learning from worked examples and learning faded examples resulted in the same learning outcomes.

2.3 Learning from examples versus tutored problem solving

Many prior studies have compared WEs to unsupported problem solving. Koedinger and Aleven (2007) criticised these prior studies because of the very different amounts of information provided in the two conditions (the unsupported problem-solving condition received no feedback upon submitting a solution). As a response to this criticism, Schwonke et al. (2009) compared a cognitive tutor (Geometry Tutor) with a new version that was enriched with faded worked examples. The result revealed an improvement in learning time from using examples. In the second experiment, they used the think-aloud protocol in order to study relevant cognitive processes. According to the result, the efficiency advantage of worked examples was replicated. Salden et al. (2010) reviewed a number of prior studies on worked examples (e.g. McLaren et al. 2008; Anthony et al. 2008) and concluded that using WEs in ITSs decreases learning time.

Trafton and Reiser (1993) conducted a study focusing on the impact of problem order and learning activity type (problem solving or WEs) on learning to program in LISP. The study was conducted in the context of BATBook, an electronic book and problem-solving environment. In addition to reading expository text on LISP, the participants could read WEs, write and evaluate LISP functions in the environment, as well as receive the full solution after three unsuccessful attempts on a problem. Twelve problems were defined for the study, so that there were six pairs of isomorphic problems. In each pair, the first problem is referred to as a source (as it introduces specific domain concepts), while the second one is referred to as the target. For each problem, the participants were required to either study a WE, or to solve the problem. The study involved four conditions: in two conditions, sources were introduced immediately before target problems, while in the other two conditions sources were presented as a block before the target problems. The other difference was whether the participants studied WEs or were required to solve the source problems. The *Alternating example* condition received pairs of problems, in which the first component was a WE, and the second was a problem to solve. The *Blocked example* condition received a block of six WEs first, and after that a block of six target problems to solve. The *Alternating Solve* condition solved source problems in a block, followed by another block of target problems. Finally, the *Blocked solve* condition received six pairs of problems which they needed to solve. The results showed that best learning outcomes where obtained when WEs were given immediately before target problems (in the Alternating example condition).

McLaren et al. (2008) discussed three studies they conducted with a stoichiometry tutor in which they compared tutored problem solving to learning from WEs combined with tutored problems. The students in the problems condition worked on solving tutored problems, while students in the examples condition observed worked examples, were prompted to self-explain, and then solved isomorphic tutored problems, in alternating fashion. In all three studies, students in the examples condition learnt faster, but there were no significant differences in learning outcome (specifically, near transfer learning). The authors suggested that one possible reason for no difference in learning is that students in the problems condition made initial problems into WEs by requesting hints, and then using the constructed solution as a template for solving the following isomorphic problems.

In a more recent study, again with the stoichiometry tutor, McLaren and Isotani (2011) compared WEs only, alternating WEs/tutored problems and all tutored problems. Surprisingly, the result showed that students benefitted most from learning with WEs only, at least with respect to learning time. However, WEs were followed by self-explanation prompts while the problems were not.

Corbett et al. (2010) investigated interleaving examples with an ITS in the domain of Genetics in which students learned about process modelling. They had four conditions: process modelling and problems (MOD), examples then problems and process modelling (ALL), examples and problems (IWE), and problem solving only (PS). They found no difference between the conditions on the robust learning and problem-solving tests. Process modelling led to greater accuracy and faster reasoning. Faster reasoning in problem solving can be achieved by using both worked examples and process modelling together. Students in all of the conditions learnt the same problem-solving knowledge which was measured by the post-test accuracy.

Although the aforementioned studies show the advantage of using WEs in conjunction with ITSs, more recent research yields opposite results. Corbett et al. (2013) reported on three empirical studies using the Genetics cognitive tutor. The studies evaluate the impact of interleaving WEs and genetic-process reasoning scaffolds in an ITS. For the genetic-process reasoning, they developed two types of tasks: process modelling tasks and solution construction tasks. The tasks were designed to precede standard genetics problem-solving tasks and ground problem-solving knowledge of students in the underlying genetics before problem solving. Each study included three conditions: a standard problem-solving condition, a scaffolded reasoning condition and an interleaved WE condition. Students in the standard problem-solving condition only completed standard problems from the Genetics Cognitive tutor. In the scaffolded reasoning condition, students had a block of scaffolded reasoning problems; the block being designed to prepare students for solving a problem. They found that interleaved WEs resulted in less basic-skill learning than problem solving. Moreover, the scaffolded reasoning condition yielded more robust understanding than problem solving. Corbett et al. (2013) findings corroborate Salden et al. (2009) study, which showed that incorporating WEs into an ITS was not superior to the ITS alone.

Overall, most of the studies show that learning from WEs resulted in reduced learning time. Although there are some studies showing higher transfer performance for faded examples, most studies have found no differences in the amount learnt. What has not been studied is how adapting when worked examples are provided in conjunction with tutored problems makes a difference. The literature seems to hint that using adaptive worked examples might be more helpful if it is reinforced with a problem-solving approach.

In addition, the majority of prior studies using examples in ITSs were in the Geometry, Chemistry and Algebra domains. All these tutors teach well-defined tasks. For ill-defined tasks, there is no specific procedure to solve a problem; thus, WEs do not reveal essential procedures to generate solutions. As an illustration, WEs Trafton and Reiser (1993) used in teaching LISP commented how provided solutions satisfy the problem statement. Although in some cases there might be more than one solution for a well-defined task, most ill-defined tasks have multiple correct solutions (Mitrovic and Weerasinghe 2009). Rourke and Sweller (2009) showed that the worked-example effect can be obtained in ill-defined domains (compared to unsupported problem solving). Najar and Mitrovic (2013a, b, 2014) conducted a study to investigate the effect of using WEs in combination with tutored problem-solving in SQL-Tutor. Tasks in the SQL domain are ill-defined because there is no algorithm to use to generate solutions. They compared three conditions: examples only (EO), alternating examples and tutored problems (AEP), and tutored problems only (PO). After completing a problem, students received a self-explanation prompt that focused on concepts used in the problem, to make sure that students acquire conceptual knowledge. On the other hand, WEs were followed by SE prompts that focused on procedural knowledge. The study showed that the AEP and PO conditions outperformed EO in learning gain, while AEP outperformed PO in conceptual knowledge acquisition. Moreover, novices learnt most from AEP, but advanced students learnt the same from AEP and PO. Novices and advanced students learnt less from EO than AEP and PO. Therefore, interleaving examples with tutored problems is an optimal choice compared to using examples or supported problems only in SQL-Tutor. To the best of our knowledge, this is the only study that compared WEs with tutored problems in ill-defined tasks.

2.4 Adaptive worked examples

WEs improves learning gain of novices by freeing up working memory capacity, but as discussed they may become detrimental for more knowledgeable students, who seem to benefit more from problem solving (Kalyuga et al. 2001). Clark et al. (2006) suggest the process of backwards fading of worked examples, which is worked examples transitioning into problems as students gain more expertise. If the process of backward fading is adapted to the student's expertise, a higher learning gain might result. None of the aforementioned studies used adaptive examples in comparison with tutored or unsupported problem solving.

Salden et al. (2009) compared fixed faded worked-out examples with adaptive ones. Fixed faded examples were the same for all students, but the solution steps in adaptive faded examples were faded based on the student's demonstrated knowledge. They conducted two studies, one in a lab (in Germany) and the other in a classroom (in the U.S.). Their main research question was to explore whether using adaptive examples combined with problem solving compared to pure problem solving could lead to better learning. They tested three conditions: problem-solving in the cognitive tutor, the tutor enriched with fixed examples, and the tutor enriched with adaptive faded examples. The lab results indicated that adaptive examples led to better learning and higher transfer compared to the other conditions. In contrast, the classroom results led to no significant difference in the immediate post-test, but in the delayed post-test students who used adaptive examples were shown to have learnt more. They posited that the difference in the lab and classroom results might have been caused by either inherent noise in the class compared to the lab, or by not using Cognitive Tutor's mastery criterion in the class. The mastery criterion option led students to enjoy remedial problems for the concept they had not mastered yet until all learning materials were learnt. Therefore, each student completed slightly different sets of problems.

Kalyuga and Sweller (2005) proposed an adaptive model for using examples. Their model is based on cognitive efficiency (CE),[1] which is calculated from students' performance and the self-reported rating of mental effort. Kalyuga and Sweller used a different formula from what was previously proposed (Kostons et al. 2010; Van Gog and Paas 2008; Paas and Van Merrienboer 1993) as they needed to calculate CE in real time during the experiment. Performance was measured based on the number of steps the student required to solve a problem. For instance, when the problem is 2(3x – 1) = 1, some students start with 2*3x – 2*1 = 1, thus requiring more steps to solve the problem than students who start with 6x – 2 = 1 or those who provide the final answer (x = 0.5) immediately. The method was tested using the Algebra cognitive tutor enriched with WEs and faded examples. Students worked with two versions of the tutor: adaptive and non-adaptive. Students in the adaptive condition were allocated to one of the four stages of faded examples (stage 1 fully worked-out examples, stage 4 fully problem-solving tasks) based on their cognitive efficiency scores in the pre-test. All students had to proceed to the final stage of fading (stage 4) from the stage they started. In each stage, a diagnostic task decided if the student needed more information (in the forms of two worked examples or four shortened worked examples). Thirty students were randomly assigned to 15 pairs of students. One student in each pair worked with the adaptive version and the other worked with the non-adaptive version. The student in the non-adaptive condition started from the faded stage in which the other participant in the adaptive condition started. The results showed that students in the adaptive condition scored marginally significantly higher than students in the non-adaptive condition. Students in the adaptive condition showed significantly higher efficiency gains than students in the non-adaptive condition.

In the study performed by Kalyuga and Sweller (2005), the system asked students to indicate how difficult the task was for them. However, this measure is not appropriate for cognitive load (Van Gog and Paas 2008). A student may find a problem very difficult and not invest enough effort to solve it by requesting a complete solution. Instead, Van Gog and Paas (2008) suggest asking students to indicate how much effort they have invested into solving the problem.

Research has shown that students may benefit differently from tutored problems and WEs. Perhaps students with lower prior knowledge learn more from WEs than tutored problem solving, and advanced students gain more from problem solving than only studying WEs. Therefore, we expect an adaptive strategy that selects the best type of learning task (worked example or tutored problem) for students to be superior to a fixed sequence of WEs and tutored problem solving. This led us to design and conduct a study that adaptively decided on the learning task for each student based on his/her requests for assistance.

3 Modes of SQL-tutor

The study was conducted in the context of SQL-Tutor (Mitrovic 2003). The version of SQL-Tutor used in this study had four modes: problem solving, 2-step faded example,

[1] Cognitive efficiency = Performance / Mental Effort Rating.

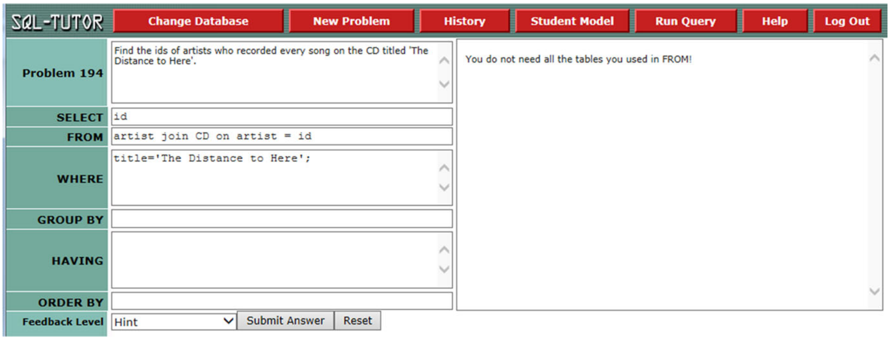


Fig. 1 Problem-solving mode of SQL-tutor

1-step faded example and worked example (Najar et al. 2014). The problem-solving mode is illustrated in Fig. 1. The 2-step faded example mode (Fig. 2) and 1-step faded example mode are similar to the problem-solving mode in which the student needs to complete two or just one clause of the SELECT statement, and the remaining clauses are already solved for the student. The fourth mode is the WE mode, which presents the complete solution and an explanation to the student (illustrated in Fig. 3). In all four modes students have access to the database schema at the bottom of the screen.

When solving problems or working with faded examples, students can choose the level of feedback they want to receive when their solution is incorrect. The level of feedback defines how much assistance is provided to the student. SQL-Tutor offers six levels of feedback: positive/negative feedback, error flag, hint, all errors, partial solution and complete solution. Positive/negative feedback is the lowest level of assistance, simply informing students whether their answer is correct or not. The message also shows how many errors students have in their solution. An error flag message identifies the clause in which the error happened. More information about the type of error is provided when a hint-type feedback is requested (illustrated in Fig. 1). The partial solution shows the correct content of the clause that the student got wrong. Feedback of type *all errors* displays hint-type messages for all errors the student has made. At the maximum level, the complete solution simply reveals the pre-specified ideal solution of the problem. When a student starts solving a new problem, the default feedback level is positive/negative. The student can submit several attempts for the same problem.

Similar to the Najar and Mitrovic (2013b) study, we presented students with a SE prompt after worked examples and problems. Schwonke et al. (2009) showed that students who worked with examples had more conceptual knowledge than procedural knowledge, and students in problem-solving condition learned more procedural knowledge than conceptual knowledge. This suggests that different types of SE are needed to scaffold problem solving and examples. Conceptual-focused self-explanation prompts (C-SE) and Procedural-focused self-explanation prompts (P-SE) are meta-cognitive prompts in which students reflect on concepts required to solve the problem and on procedural steps of learning materials respectively. Students were given C-SE prompts after problems, and P-SE prompts after examples. In the case of faded examples,

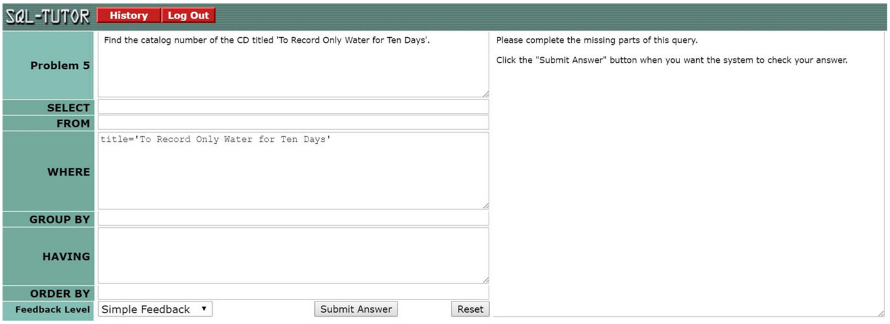


Fig. 2 The 2-step faded mode of SQL-Tutor, showing a problem with the SELECT and FROM clauses faded

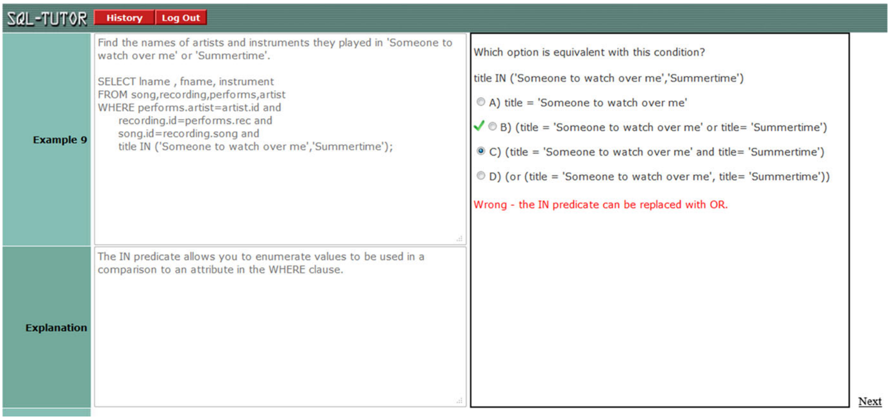


Fig. 3 Screenshot of a WE followed by P-SE

the student is required to solve one (or two) clauses of the problem. Therefore, faded examples involve limited problem-solving compared to worked examples in which the student only reads the provided solution. For that reason, we provided C-SE prompts after faded examples.

Figure 3 shows a screenshot of a situation when a student finished reading a worked example, and the system showed a P-SE prompt, located on the right side of the screen. The student gave a wrong answer to the P-SE prompt and because there is only one attempt per SE prompt, the system showed the negative feedback and revealed the correct answer. Figure 4 shows the positive feedback provided to the student who answered the C-SE prompt correctly. Once students received SE feedback, they could continue to the next task.

4 Adaptive strategy

We designed an adaptive strategy which selects a learning activity (a problem to solve, 1-step or 2-step faded examples, or a WE) for a student using cognitive efficiency.

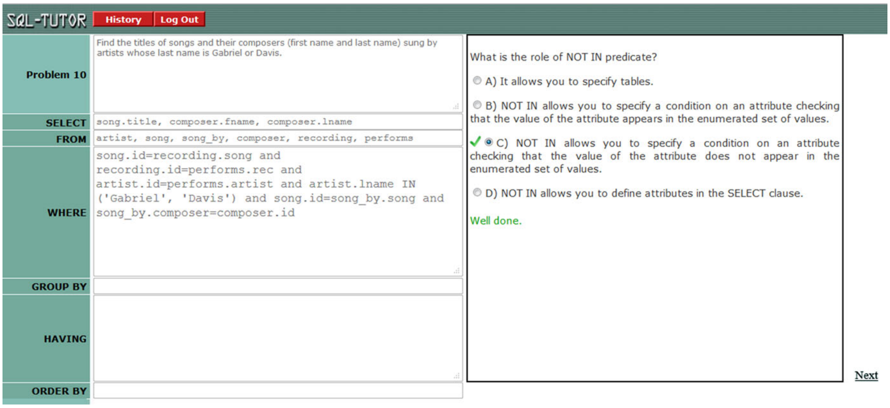


Fig. 4 Screenshot of a problem-solving page followed by C-SE

Paas and Van Merrienboer (1993) calculated cognitive efficiency from the difference between the z-scores of performance (P) and mental effort rating (R), CE $= z_P - z_R$. The problem with this approach is that CE can only be calculated after the experiment is completed. In order to determine CE during the study, Kalyuga and Sweller (2005) used mental effort (R) and performance (P) and calculated Cognitive Efficiency as CE $=$ P $\div$ R. Mental effort was self-reported by students, and performance was calculated from the number of steps students required to solve a problem.

Our adaptive strategy is also based on a measure of cognitive efficiency. The participants were asked to rate mental effort (R) after solving each problem (i.e. *How much effort did you invest to complete this task?*) on a 9-point rating scale (see Fig. 5). Like Kalyuga and Sweller (2005), we defined the critical level of cognitive efficiency as $CE_{cr} = P_{max} \div R_{max}$, where P_{max}[2] and $R_{max} = 9$. We consider CE $>=$ CE_{cr} to be high cognitive efficiency; thus, students who solved a problem with CE $>= 1$ were expected to be able to solve the next problem without needing any preparation in terms of WEs.

4.1 Estimating performance based on assistance score

In order to estimate performance P on a particular problem, we looked at how much feedback the student required from SQL-Tutor. Table 1 shows the scores (H_i) we assigned to each level (i) of feedback in SQL-Tutor. H_i represents the assistance score for feedback level i. Level 0 (H_0) presents minimum assistance (score $= 1$) and level 5 (H_5) shows the maximum assistance (score $= 6$).

A simple way to calculate the assistance score is to sum up the assistance scores of all requested feedback. In SQL-Tutor, students can ask for the same level of feedback several times. If a student received a particular feedback message and then requests it again, the message does not contain the same amount of new information. For

[2] Please note that P is scaled to the range [1,9].

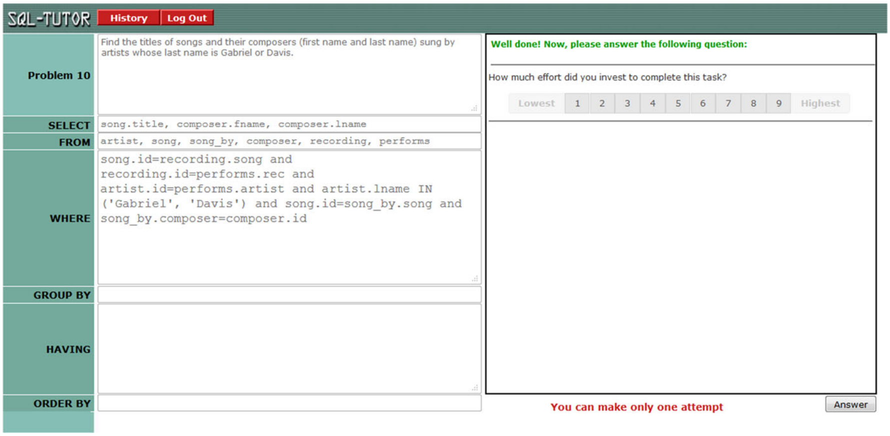


Fig. 5 Effort rating after problem solving

Table 1 Assistance scores for different levels of feedback

Name	i	H_i
Positive/negative	0	1
Error flag	1	2
Hint	2	3
Partial solution	3	4
All errors	4	5
Complete solution	5	6

instance, when a student sees a complete solution to a problem, the next time s/he asks for the complete solution, the same solution will be shown. Therefore, we multiplied the assistance score for each level of feedback by the power two series of n, with n showing the number of requests for the level of feedback (Eq. 1). Power two series converges to two.

$$\mathrm{T} = \sum_{i=0}^{5} Po\,(n_i)\,H_i, \quad \text{where}\, Po\,(n) = \sum_{j=1}^{n} \frac{1}{2^{(j-1)}} \tag{1}$$

Equation 1 does not take into account the student's behaviour after receiving feedback. For instance, the current formula shows that a student who solved a problem by receiving H_0 H_1 H_2 (without getting a partial or complete solution), received the same or more information as student B who saw a complete solution (H_5) once. Moreover, it is important to distinguish between students who complete problems with minimum assistance and students who request the complete solution in the first attempt. One way is to change the scoring system we presented in Table 1. However, changing the scoring system does not help to distinguish between students who received a complete solution in the first attempt and students who saw a complete solution after several attempts to solve the problem. For instance, students who see a complete solution after several incorrect attempts search for their mistakes when they see the complete solution. Moreover, seeing a complete solution in the first attempt encourages students to copy the solution and leads to shallow learning (Deeks 2000).

In order to take into consideration the student's behaviour, we introduced parameter B, which represents the average score of requested feedback levels, with duplicates excluded (Eq. 2).

$$\text{Student Behaviour}: \ \mathrm{B} \ = AVERAGE\,(H_m)\,, \quad m \textit{ is the set of requested hint levels} \tag{2}$$

As an example, when a student requests H_1 three times followed by H_4, the value of B is 3.5. Parameter B indicates whether the student tends to use high or low levels of assistance; for instance, if B is 2.5, the student mostly uses low feedback levels, but when B is 4.5, the student uses high levels of feedback rather than low-level feedback to solve the problem.

Equation 2 does not discriminate well between different levels of feedback. For instance, there is a small difference between $\mathrm{B} = 1$, $\mathrm{B} = 2$, $\mathrm{B} = 3$ or $\mathrm{B} = 4$. In fact, $\mathrm{B} = 4$ shows that students used a partial or a complete solution to accomplish the task, while $\mathrm{B} = 3$ shows that students definitely did not see a complete solution, but might have used a partial solution in conjunction with some other low assistance hints. Therefore, we should use different slopes for each behaviour. An appropriate function that accounts for this is *Skewness slope* (Eq. 3):

$$\text{Skewness slope}: \ \mathrm{K}\,(\mathrm{x}) = \left(\sin\left(\frac{\pi}{2}\left(\frac{x}{3}-1\right)\right)+1\right)^{2}+1 \tag{3}$$

The assistance score is now calculated as in Eq. 4:

$$\mathrm{T} = K\,(B)\sum_{i=0}^{5} H_i\,Po\,(n_i) \tag{4}$$

The assistance score T depends on the number and type of feedback messages the student obtained from SQL-Tutor while solving a problem. The maximum assistance score during problem solving is equivalent to turning the problem into a WE by asking the system for a full solution, or alternatively asking for a partial solution for each clause. When the student requests partial solution several times ($\mathrm{Po} \approx 2$), the complete solution is eventually revealed step by step. Although requesting a complete solution represents a higher assistance score, it provides the same amount of information as seeing partial solutions for each clause. Therefore, we consider multiple requests for partial solutions as maximum assistance score. Using Eq. 4, we calculate $\mathrm{T_{High}}$ to be 26 ($\mathrm{H_3} = 4$; $\mathrm{K(4)} = 3.25$, $\mathrm{Po} \approx 2$; $\mathrm{T} = 3.25 * 4 * 2$). Therefore, performance (P) can be calculated as:

$$\mathrm{P} = 26 - \mathrm{T} \tag{5}$$

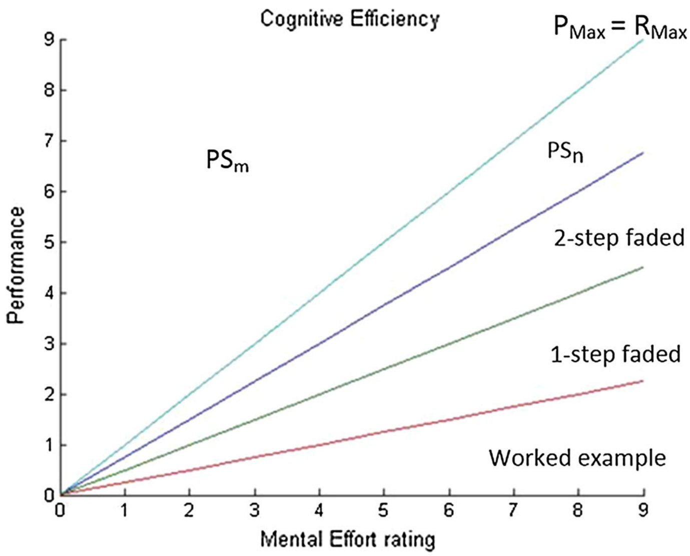


Fig. 6 Relationships between CE and preparation tasks

Note that T can have a value greater than T_{High}. Because T_{High} represents turning problems into examples, we set all assistance scores greater than T_{High} to 26. Therefore, P never becomes negative.

4.2 Adaptive selection of learning tasks

The CE score is calculated only after problems. If the student's CE is greater or equal to 1 when solving a problem, it is expected that the student will be able to solve the next (more complex) problem. However, if CE is less than one, the student might need a preparation task before solving the next problem. Figure 6 identifies the type of preparation task our adaptive strategy selects depending on the value of CE. Please note that R_{Max} is 9 (the maximum score for effort, as shown in Fig. 5). The performance (P) is also scaled to the same range (1–9), so P_{Max} is also 9.

A CE below 1 and equal or greater than 0.75 (6.75/9) shows relatively good performance on the current problem, but indicates that the student would benefit from additional practice, and the chosen preparation task is another problem to solve. Students with CE between 0.75 (6.75/9) and 0.25 (2.25/9) received 2-step or 1-step faded examples as a preparation task. As mentioned before, the steps are faded based on how much the student has learnt from the current task for each clause. Students who scored below 0.25 (2.25/9) get an isomorphic worked example before solving the next problem, as their performance on the current problem is low.

The behaviour of the adaptive strategy is illustrated in Fig. 7. When a student asked for a partial solution more than twice, or requested the complete solution, the strategy presented a worked example as a preparation task regardless of the student's CE. This is an exception from Fig. 7, because asking for such high feedback levels indicates that

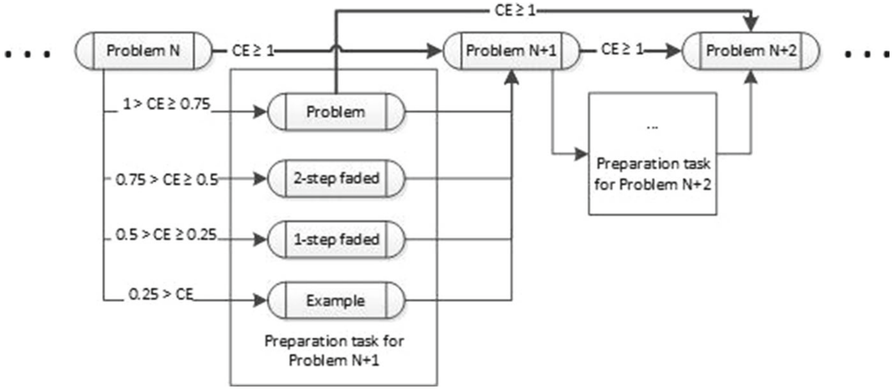


Fig. 7 Illustration of the behaviour of the adaptive strategy

the student does not have sufficient prior knowledge to solve the next problem, which is more difficult than the current problem. If a student performed well (CE >= 1) on a problem which is shown as a preparation task, the system skipped the next problem and the preparation task for the subsequent problem.

4.3 Adaptive generation of faded examples

The faded examples are generated adaptively, based on the student's performance on the current problem. Domain knowledge is represented in SQL-Tutor as constraints (Mitrovic 2003). For each clause of the SELECT statement, there is a set of related constraints. Every time the student submits an attempt, the system analyses it and records which constraints were satisfied or violated. SQL-Tutor estimates the student's knowledge of clauses by aggregating information about violated and satisfied constraints related to a clause.

Using the student model, it is therefore possible to find out how much the student learnt about a particular clause by comparing his/her knowledge before and after the current problem. Our fading strategy determines how much the student's knowledge has changed for each of the clauses while working on the current problem, and selects the clause on which the student improved the most (or the top two clauses, if two steps are to be faded). Then the system fades one or two clauses of the next problem. If the next problem does not include the selected clause(s), the strategy fades the next clause (or two) from the sorted list. The idea is to help students rehearse what they have just learnt.

5 Experiment design

The study was conducted in a single, 100-minute long session. Figure 8 shows the design of the study. The participants were volunteers from an introductory database course at the University of Canterbury. The students had learnt about SQL in lectures

	Control	Experimental
	n = 24	n = 24
	Pre-test	
Pair 1	1^{st} task in the pair: problem	1^{st} task in the pair: problem
	2^{nd} task in the pair: example	2^{nd} task in the pair: rehearsal task (problem, 2 or 1 step faded example, worked example, or skip)
Pair 2 to 10	1^{st} task in each pair: example	1^{st} task in each pair: preparation task (problem, 2 or 1 step faded example, worked example or skip)
	2^{nd} task in each pair: problem	2^{nd} task in each pair: problem
	Each problem followed by a C-SE prompt and each example followed by a P-SE prompt	Each problem or faded example followed by a C-SE prompt and each example followed by a P-SE prompt
	Post-test	

Fig. 8 Experiment design

beforehand. The participants were informed that they would see ten pairs of tasks and that the tasks in each pair were similar. At the beginning of the session, the students took a pre-test for ten minutes. The pre-test had ten questions, eight of which were multiple-choice and two were problem-solving questions. The multiple-choice questions (worth one mark each) measured conceptual knowledge (e.g. *What clause of the SELECT statement allows the resulting table to be sorted?*). For the problem-solving questions (worth four marks each), students had to write SQL queries (e.g. *Write a query that retrieves the names of all departments located in Houston*). Therefore the maximum mark on the pre-test was 16.

Once students logged in, SQL-Tutor randomly allocated them to one of the conditions. This process yielded sample sizes of 24 in each group. The participants studied ten pairs of isomorphic tasks of gradually increasing complexity. Once students completed a problem and answered a C-SE prompt, the system asked them to self-report their cognitive load on a Likert scale from 1 (lowest effort) to 9 (highest effort). At the end of the session, students were given ten minutes to complete the post-test. However, students could start the post-test during the learning session and finish the study earlier. The post-test was isomorphic to the pre-test.

The control condition worked with a fixed set of example-problem pairs: each pair consisted of a WE followed by an isomorphic problem. The only exception is the first pair, in which the control group received a problem followed by an example; this was so that the first problem could provide the necessary information for the adaptive strategy. The control condition in this study is identical to the best condition (AEP - alternating examples/problems) from the previous study queries (Najar and Mitrovic 2013b), with the exception of the first pair.

The experimental group had pairs consisting of a preparation task followed by a problem, except for the first pair. The first pair consisted of a problem followed by a rehearsal task. Rehearsal tasks are the same as preparation tasks, but because they were provided after the isomorphic problem we called them rehearsal tasks. Our adaptive strategy (presented in Sect. 4.2) decided what type of preparation task to present to the experimental group.

In the experimental condition, the amount of information students receive was adapted to the student's performance. Therefore, novices studied isomorphic WEs when they do not know how to solve problems, while advanced students practised by solving problems only. Our previous study showed that novices benefitted the most from alternating examples and problems, while advanced students benefitted the same from alternating examples and problems and problems only (Najar and Mitrovic 2013a, a). In the experimental condition, at the maximum assistance level the adaptive strategy becomes identical to the control condition, and at the minimal assistance level it becomes the same as the problem-only condition from our previous study. Since the amount of information is adapted to the student's needs, our strategy will not cause the expertise reversal effect for the advanced students; thus our hypothesis is that the experimental group (especially advanced students) will learn more than students in the control condition. Our experimental condition can be transformed from AEP to a problems-only condition for advanced students, which would allow them to solve more problems. Moreover, the experimental group participants skip preparation tasks they have already mastered. Therefore, we hypothesised the experimental group would spend less time overall than the control group.

6 Results

Forty eight students from the University of Canterbury participated in this study. Two students from the control group were excluded from the analyses because they did not take the post-test. Thus, we had sample sizes of 22 in the control group and 24 in the experimental group. We calculated the average scores in the pre-test and the post-test and the time students spent on the system (Table 2). The students who had pre-test scores lower than 50 % (median) were considered novices and the rest were classified as advanced students.

We analysed the data to identify whether the two conditions learned differently and also to see whether the two conditions benefitted novices and advanced students differently. We start by explaining the results of the two conditions, followed by the results for novices and advanced students.

Table 2 Basic statistics for all students (standard deviation given in brackets)

Number of students	46
Avg. scores on pre-test (%)	47.69 (16.63)
Avg. scores on post-test (%)	81.94 (13.64)
Learning time (min)	65.94 (19.05)

Table 3 Basic statistics for the two groups

	Control (22)	Experimental (24)	p
Pre-test (%)	50.28 (13.7)	45.31 (18.91)	.32
Post-test (%)	77.84 (13.87)	85.68 (12.56)	.05
Improvement	$t = 9.9$, $p < .01$**	$t = 10.5$, $p < .01$**	
Pre/post-test correlation	$r = 0.55$, $p < .01$**	$r = 0.34$, $p = .1$	
Learning time (min)	73.59 (16.28)	58.92 (18.99)	$t = 2.8$, $p < .01$**
Normalised learning gain (%)	55.72 (25.19)	73.2 (19.55)	$t = 2.61$, $p = .013$*
Conceptual knowledge gain (%)	76.47 (29.75)	87.7 (17.5)	.13
Procedural knowledge gain (%)	29.51 (38.16)	61.95 (36.52)	$t = 2.94$, $p < .01$**
Number of problems solved (inc. faded)	7.0 (2.49)	8.58 (3.03)	.059
Problems solved (excl. faded examples)	7.0 (2.49)	6.96 (2.37)	.95
2-step faded		0.75 (1.15)	
1-step faded		0.88 (1.15)	
Number of WEs	7.91 (3.04)	1.79 (1.93)	$t = 8.07$, $p < .001$**
Number of attempts per problem	4.54 (1.97)	4.34 (1.7)	.72
Maximum complexity level	13.41 (5.2)	14.0 (5.3)	.71

** and * denote significance at the 0.01 and 0.05 level respectively

6.1 Analysing learning in the two conditions

The basic statistics for the two groups are presented in Table 3. There was no significant difference between the pre-test scores of the two groups. The t-test revealed a significant difference between the post-test results ($p = .05$), with the performance of the control group significantly lower than the experimental group. Thus, one of our hypothesis was confirmed.

The students in both conditions improved significantly between the pre- and post-tests, as shown by the paired t-tests reported in the *Improvement* row of Table 3. Correlations between the pre- and post-test scores are also reported in the table, but only the control condition had a significant correlation ($r = 0.55$, $p < .01$). There was also a significant difference between the mean learning times of the two groups ($p < .01$). The experimental group spent significantly less time in the intervention than the control group, as hypothesised.

The normalised learning gain[3] of the experimental group was significantly higher than that of the control group ($p = .01$). We calculated the effect size based on the normalised learning gain using Cohen's d, with the following assumption: $d \geq 0.8$ (large effect), $d \geq 0.5$ (medium effect) and $d \geq 0.2$ (small effect) (Cohen 1988). The effect size on normalised learning gain was medium ($d = 0.75$).

When we analysed normalised learning gains on the conceptual knowledge questions (questions 1–8), we found no significant difference between the groups ($p = .13$). On the other hand, the normalised learning gain on procedural knowledge (questions

[3] Normalised learning gain = (post test—pre test) / (max score—pre test).

9 and 10) of the experimental group was significantly higher compared to the control group ($p < .01$).

There was no significant difference between the two groups in the number of problems solved, either when faded examples are included or excluded. The experimental group received significantly fewer examples than the control group ($p < .01$), because the experimental group received 2-step or 1-step faded examples instead of WEs (as determined by the adaptive strategy). The average number of 2-step and 1-step faded examples solved by the experimental group was 0.75 and 0.88 respectively.

There was no significant difference in the number of attempts per problem between the two conditions. The problem complexity gradually increased from pair 1 to pair 10. There was no significant difference between the average maximum complexity levels of problems the students in the two groups solved (please note that not all students completed all problems).

As mentioned earlier, the participants received C-SE prompts after problems and P-SE after examples. We analysed the SE success rates for the two groups, which are reported in Table 4. There was no significant difference between the groups on the overall SE success rate and the P-SE success rate. The C-SE success rate of the experimental group is higher than that of the control group, but the difference is not statistically significant.

Overall, the results show that the experimental group participants, who worked with the adaptive strategy, learnt more than students who worked with a fixed sequence of examples and problems. Moreover, the experimental group spent significantly less time working with the system than students in the control condition. The results clearly show the effectiveness of our adaptive strategy in comparison with the non-adaptive sequence.

6.2 Cognitive efficiency and mental effort

Students rated their mental effort after they solved problems (not after WEs and faded examples as we could calculate CE scores after problems only), which the adaptive strategy used to calculate CE. As the mental effort rate is specified on a 9-point scale, we used non-parametric tests for this analysis. First, we used Spearman's rho test to investigate whether there is a correlation between the mental effort and the pre-test or between the cognitive efficiency and the pre-test.

Table 5 shows the results. We found a significant negative correlation between the pre-test scores and mental effort ratings ($r = -0.48$ for both groups), as well as significant correlations between the pre-test and cognitive efficiency ($r = 0.69$ for control and $r = 0.44$ for experimental group). The table also shows significant negative correlations between the mental effort and cognitive efficiency in both groups ($r = -0.67$ for control and $r = -0.73$ for the experimental group). The significant negative correlations between mental effort and CE scores could be expected because CE scores were calculated from the mental effort. Next, we used the Mann-Whitney *U* test to compare the groups on CE and the mental effort. The results are summarised in Table 5. There were no significant differences between the groups on either reported mental effort or CE scores.

Table 4 Analyses of SE prompts success rates

	Control (22)	Experimental (24)	p
SE success rate	82.6 (12.2)	88.0 (12.5)	.14
Procedural SE success rate	90.3 (12.9)	90.0 (11.5)	.79
Conceptual SE success rate	73.6 (15.9)	84.0 (20.1)	.07

Table 5 Cognitive efficiency and mental effort analysis

	Control (22)	Experimental (24)	p
Correlation: pre-test and mental effort	$r = -0.48$, $p = .03$*	$r = -0.48$, $p = .02$*	
Correlation: pre-test and CE	$r = 0.69$, $p < .001$**	$r = 0.44$, $p = .03$*	
Correlation: mental effort and CE	$r = -0.67$, $p = .001$**	$r = -0.73$, $p < .001$**	
Cognitive efficiency (CE)	2.28 (2.29)	2.70 (1.85)	.09
Mental effort	4.77 (1.71)	4.38 (1.20)	.24

** and * denote significance at the 0.01 and 0.05 level respectively

6.3 Effect of preparation tasks on problem solving

We were interested to find how each preparation task affected problem solving performance, and therefore CE score on the following problem. We extracted CE scores from the previous problems (CE_1) and from the following problems (CE_2) for each preparation task. This gave us 262 pairs of (CE_1, CE_2) from four types of preparation tasks in the experimental condition, and one type of preparation task in the control condition. Because of the low number of instances, 1-step and 2-step faded examples were considered as one type, named faded examples. We also excluded data from the first pair in which students had rehearsal tasks instead of preparation tasks. Note that in the skip action students did not see any preparation task; therefore, this condition is equal to not having preparation tasks.

The results are summarised in Table 6. The table shows the average CE scores before and after each type of preparation tasks. The CE scores of the experimental group students who had a preparation task significantly improved (example: $p < .01$; faded: $p = .02$; problem: $p = .04$). However, the CE scores of students who skipped preparation tasks significantly deteriorated ($p < .01$). This could be expected as students were not prepared for the next problem. However, their average CE scores is still above 1 (mean = 2.32), which shows that students had enough knowledge to solve the next problem. This can be considered a trade-off between spending time on the preparation task that is not needed or skipping the preparation task and shortening the learning time. Figure 9 illustrates how the average CE scores changed for different types of preparation tasks.

Table 6 shows that the CE scores of the control group did not significantly change by studying examples; however, we see that CE scores of the experimental group, when examples were studied, significantly improved. This shows that the preparation tasks provided, particularly examples, were more targeted for the experimental group

Table 6 The effect of preparation tasks on CE

Condition	Action	Number of pairs	CE_1	CE_2	Difference
Experimental	Example	35	0.03 (0.05)	0.90 (1.78)	$t = -2.94$, $p < .01$**
	Faded	15	0.50 (0.14)	2.06 (2.32)	$t = -2.61$, $p = .02$*
	Problem	9	0.90 (0.03)	2.88 (2.51)	$t = -2.37$, $p = .04$*
	Skip	73	4.01 (2.77)	2.32 (2.59)	$t = 3.81$, $p < .01$*
Control	Example	130	2.04 (2.37)	1.71 (2.21)	$t = 1.50$, $p = .14$

** and * denote significance at the 0.01 and 0.05 level respectively

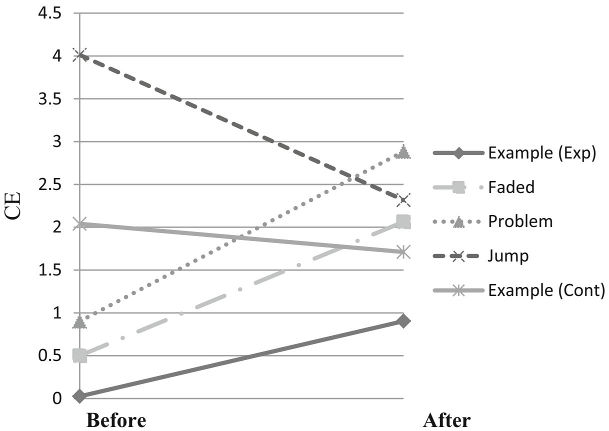


Fig. 9 CE scores on problems before and after certain types of preparation tasks

than the control group. Eventually, in the experimental group, students were given examples when their CE scores were below 0.25.

6.4 Novices and advanced students

We post-hoc divided participants into novices and advanced students. Students who scored less than the median on the pre-test (50 %) were considered as novices (control = 8, experimental = 13) and students who scored greater than the median on the pre-test were considered advanced students (control = 14, experimental = 11). We used non-parametric tests for the following analyses.

The basic statistics for novices are shown in Table 7. The Mann-Whitney U test revealed no significant difference between the two groups on the pre-test and post-test scores. The novices from the experimental group spent less time on the learning activities, but the difference is not significant. There were no significant differences between the two groups on the normalised learning gain, conceptual knowledge gain, procedural knowledge gain, number of problems solved (including faded examples) and problems solved (excluding faded examples). However, there was a significant

Table 7 Basic statistics for novices

	Control (8)	Experimental (13)	p
Pre-test (%)	36.72 (8.48)	30.29 (8.4)	.14
Post-test (%)	72.66 (15.65)	81.73 (14.54)	.18
Learning time (min)	72.75 (12.51)	59.69 (15.97)	.06
Normalised learning gain (%)	58.35 (21.65)	73.43 (20.39)	.12
Conceptual knowledge gain (%)	85.71 (13.99)	94.6 (10.57)	.19
Procedural knowledge gain (%)	32.83 (41.95)	55.39 (35.83)	.24
Number of problems solved (inc. faded)	7.62 (2.13)	8.54 (2.99)	.55
Problems solved (excluding faded)	7.62 (2.13)	6.61 (2.33)	.34
2-step faded		0.77 (1.3)	
1-step faded		1.15 (1.28)	
Number of examples	8.87 (3.56)	2.31 (2.32)	$U = 4.5$, $p < .001$**
Number of attempts per problem	4.57 (1.82)	4.57 (1.66)	.75
Maximum complexity level	14.5 (4.75)	13.54 (5.36)	.80

** denote significance at the 0.01 level

Table 8 Analysing cognitive load and cognitive efficiency for novices

	Control (8)	Experimental (13)	p
Correlation: pre-test and mental effort	$r = -0.22$, $p = .60$	$r = -0.09$, $p = .76$	
Correlation: pre-test and CE	$r = 0.5$, $p = .21$	$r = 0.33$, $p = .28$	
Cognitive efficiency (CE)	1.04 (0.33)	2.35 (1.98)	$U = 83$, $p = .02$*
Mental effort	5.59 (1.04)	4.86 (0.87)	.053

* denote significance at the 0.05 level

difference between the numbers of examples the two groups studied ($p < .01$). The novices from the experimental group studied significantly fewer examples than the control group. The table also shows no significant differences on the attempts per problem the average maximum complexity level for completed problems.

We further analysed data presented in Table 6 relating to WEs. The experimental group students who received WEs as preparation tasks were all novices.

Table 8 shows the results of cognitive load and cognitive efficiency analyses for novices. The Spearman's rho test revealed no significant correlations between the pre-test and mental effort and no significant correlations between the pre-test and cognitive efficiency for novices in the control group and experimental group. The Mann-Whitney U test was used to compare novices from the two groups on CE and mental effort. The novices in the experimental group experienced significantly higher CE than novices in the control group ($p = .02$). Moreover, novices in the experimental group indicated significantly lower mental effort than novices in the control group ($p = .05$).

Table 9 shows the result of the Mann-Whitney U test for SE success rate. We found no significant difference in SE, P-SE and C-SE success rates between novices in the control and experimental groups.

Table 9 SE prompts analyses for novices

	Control (8)	Experimental (13)	p
SE success rate	78.99 (10.8)	87.35 (13.15)	.16
Procedural SE success rate	84.39 (9.13)	91.18 (10.05)	.16
Conceptual SE success rate	73.6 (13.53)	81.07 (20.09)	.36

Table 10 Basic statistics for advanced students

	Control (14)	Experimental (11)	p
Pre-test (%)	58.04 (9.31)	63.07 (9.86)	.17
Post-test (%)	80.8 (12.37)	90.34 (8.08)	$U = 117.5$, $p = .02$*
Learning time (min)	74.07 (18.53)	58 (22.85)	.15
Normalised learning gain (%)	54.23 (27.68)	72.94 (19.49)	.05
Conceptual knowledge gain (%)	71.19 (35.23)	79.55 (20.87)	.77
Procedural knowledge gain (%)	27.62 (37.33)	69.7 (37.49)	$U = 121.5$, $p = .01$*
Number of problems solved (inc faded)	6.64 (2.68)	8.64 (3.23)	.12
Problems solved (excluding faded)	6.64 (2.68)	7.36 (2.46)	.64
2-step faded		0.73 (1)	
1-step faded		0.55 (0.93)	
Number of examples	7.36 (2.68)	1.18 (1.17)	$U = 5.5$, $p < .001$**
Number of attempts per problem	4.52 (2.11)	4.08 (1.79)	.54
Maximum complexity level	12.79 (5.58)	14.55 (5.54)	.5

** and * denote significance at the 0.01 and 0.05 level respectively

The basic statistics for advanced students are shown in Table 10. There are no significant differences between the pre-test scores and learning times. There was a significant difference in the post-test ($p = .02$), with the experimental group scoring higher than the control group. The normalised learning gain of the advanced students from the experimental group is higher compared to their peers from the control group, but the difference is not significant ($p = .051$). There was no significant difference between the two groups in the conceptual knowledge gain, but the experimental group acquired more procedural knowledge than the control group ($p = .01$). The table shows no significant differences between the two groups in the number of problems solved (including faded examples) and problems solved (excluding faded examples). The control group participants had no faded examples and therefore studied significantly more examples than the experimental group. There is no significant difference between the numbers of attempts per problem, nor between the average maximum complexity level advanced students in the two groups received.

Table 11 shows the results of cognitive load and cognitive efficiency analyses for advanced students. The Spearman's rho test revealed no significant correlations between the pre-test and mental effort, or between pre-test and CE for both groups.

Table 11 Cognitive load and cognitive efficiency analyses for advanced students

	Control (14)	Experimental (11)	p
Correlation: pre-test and mental effort	r = −0.37, p = .19	r = 0.054, p = .08	
Correlation: pre-test and CE	r = 0.49, p = .07	r = 0.50, p = .12	
Cognitive efficiency (CE)	2.99 (2.64)	3.11 (1.67)	.4
Mental effort	4.29 (1.87)	3.81 (1.31)	.43

Table 12 SE prompts analyses for advanced students

	Control (14)	Experimental (11)	p
SE success rate	84.6 (12.84)	88.76 (12.23)	.43
Procedural SE success rate	93.73 (13.73)	89.04 (13.34)	.24
Conceptual SE success rate	73.59 (17.78)	88.1 (20.89)	.07

The Mann-Whitney U test shows no significant difference between the control and experimental groups on CE and mental effort.

Table 12 shows the result of the Mann-Whitney U test for SE success rate. We found no significant difference in SE, P-SE and C-SE success rates between novices in the two groups.

Overall, the results show that novices from the experimental condition learnt the same amount of knowledge as novices from the control group, but in a shorter time. Therefore, the adaptive strategy was more efficient for novices than studying a fixed sequence of examples and problems. Advanced students in the experimental group learnt more than advanced students in the control group while spending the same amount of time. Thus, the adaptive strategy is more efficient and effective than using a fixed sequence of examples and problems.

7 Discussion

We analysed the data from two perspectives: comparing the two conditions and analysing novices and advanced students separately. Students who worked with the adaptive strategy learnt more and faster than students in the control condition. The participants in the experimental condition skipped the next problem if their cognitive efficiency was greater than 1, but students in the control group could not skip any tasks. Therefore, it is not surprising that the experimental condition spent less time. However, at the same time there was no significant difference between the maximum complexities, which suggests that students in both groups attempted the same problems. Therefore, we conclude that the experimental group learnt faster and acquired more knowledge than the control group.

We could not predict which group would learn faster before the study. The experimental group studied fewer examples than the control group; therefore, it was possible that the control group would spend less time working with the system than the experimental group.

Although post-test scores and normalised learning gains show that the experimental group learnt more than the control group, both groups acquired the same amount of conceptual knowledge. As students in both groups saw the same number of pairs, we could expect that the control group would learn more conceptual knowledge than the experimental group. Schwonke et al. (2009) showed that students learn more conceptual knowledge from studying examples than solving problems. In our study, students in the control group studied more examples than their peers from the experimental group. Therefore, no significant difference in conceptual knowledge reveals that our strategy has provided the right amount of conceptual knowledge for the experimental group. Moreover, the experimental group was more successful in answering C-SE prompts than the control group. C-SE prompts were only provided after problems and faded examples; as a result, the control group only saw C-SE prompts after problems. Therefore, a possible reason for a better C-SE success score by the experimental group is that the experimental group received faded examples before some C-SE prompts.

The control group acquired less procedural knowledge than the experimental group, while there was no significant difference between the two groups in the number of problems they solved (excluding faded examples). The experimental group had an adapted mixture of problems, faded examples and examples instead of examples only. Moreover, we faded solution steps adaptively. In faded examples, students worked on those parts of solutions that corresponded to recently learnt concepts (i.e. clauses of the SELECT statement). Therefore, not only did the adaptive learning strategy perform better than the fixed condition, but also applying the adaptive fading strategy improved students' procedural knowledge more than studying worked examples. The advantage of our adaptive strategy is also corroborated by the CE analysis. The experimental group experienced higher cognitive efficiency than the control group. Although the difference is not significant.

We conducted paired t-tests to investigate whether or not CE scores had significantly changed after the preparation task. The results of the experimental group show that average CE scores significantly improved when students had examples, faded examples or problems. However, when students did not have a preparation task, the CE score significantly deteriorated. This could be expected because students were not prepared for the next problem. Nevertheless, the average CE score of such students was still above 1, which shows that their performance was above their cognitive load score. This was a trade-off in favour of learning time: students skipped the preparation tasks unless their CE scores dropped below 1. We believe this is the beauty of our adaptive model: it does not compel advanced students to do redundant tasks; consequently, the model avoids the expertise reversal effect (Kalyuga 2007). All students in the control group had examples as preparation tasks. The result shows CE scores of the control group did not significantly change; therefore, having examples in the control group was not as effective as having examples in the experimental group. The difference was whether or not the students had prior knowledge, not in the examples themselves which were the same. In the experimental group, students who had low prior knowledge were given examples to study, but in the control group, all students had examples.

Next we investigated how novices and advanced students benefitted from the proposed adaptive strategy. Novices in the experimental group acquired the same amount of knowledge, but did so faster than novices in the control group. Our previous study (Najar and Mitrovic 2013a) suggested AEP for novices and AEP and PO for advanced students. The results of this study show that novices who worked with the adaptive strategy learnt the same as students in the control group, but in a shorter time. Moreover, novices in the experimental group had significantly higher CE scores than the control group and our strategy induced significantly less cognitive load than the control condition. The results clearly show that our strategy is a better option for novices than using a fixed sequence of examples/problems.

We did not see a significant difference in learning times for the advanced students of the two groups. We think that the time the experimental group saved due to skipping was almost covered by the time that the control group saved by skipping studying examples or studying examples very quickly (the average time per WE was only a minute and a half). Advanced students in the experimental group learnt more than advanced students in the control group. Perhaps advanced students in the control group suffered from an illusion of understanding or expertise reversal. C-SE success rate analysis reveals that advanced students in the experimental group were more successful than those in the control group. As the experimental group saw C-SE prompts after faded examples, not only after problems like the control group, we conclude that students learnt more conceptual knowledge from solving faded examples than solving problems. Therefore, our strategy worked for the advanced students.

When we look at the results for novices and advanced students, we can see that advanced students in the experimental group solved slightly more problems than the control group, while novices solved slightly fewer problems than the control group. Our strategy clearly shows that advanced students should be provided more problems than novices while, in a fixed sequence of examples and problems, both advanced students and novices had to solve the same number of problems and examples.

8 Conclusions

Prior research shows that students, particularly novices, learn more from examples than unsupported problem solving. While most of the studies that compared examples to ITSs indicate that students learn the same from worked examples and ITSs in domains with well-defined tasks, a few studies show that examples are not as effective as alternating examples and problems or problems only, e.g. (Kalyuga et al. 2001). However, research shows that in the early stages of learning, using examples only or alternating examples and problems is superior than using problems only (Najar and Mitrovic 2013b). This suggests that students benefit differently from worked examples and problem solving in different stages of learning.

In the current study, we compared the best condition from the previous study (AEP), a fixed sequence of examples and problems, with a novel adaptive strategy. We discussed a new approach to measuring problem-solving performance on the basis of assistance scores. Using performance and mental effort rates, we could calculate the cognitive efficiency. The adaptive strategy used cognitive efficiency scores to choose appropriate learning tasks for students. The strategy provided worked examples, faded examples and problems to students when they needed them. The fading strategy was also adaptive: the system faded the solution steps related to the most recently learnt knowledge, thus giving students additional opportunities to strengthen such knowledge.

Prior research also used cognitive efficiency to provide appropriate learning tasks (Kalyuga and Sweller 2005), but they used students' performance which was based on how many steps students required to solve testing tasks (equations). In our study, we measured cognitive efficiency based on how much assistance students received when solving problems. Therefore, our strategy uses different features to make a decision.

In our study, the control group worked with a fixed sequence of examples and problems, while the experimental group worked with the adaptive strategy. The results show that the experimental group learnt more and faster than the control group. For novices, the adaptive strategy led to faster learning, while for advanced students the adaptive strategy produced a higher learning gain than the fixed sequence of WEs and problem solving. Overall, the results prove that the adaptive strategy outperforms the non-adaptive one. Further work is required to investigate the effect of adaptive selection of learning activities when the sequence of problems/WEs is not fixed.

Another limitation of our study is related to the control condition, which received pairs consisting of WEs and problems. We chose this control condition as it was the best condition from our previous study (Najar and Mitrovic 2013b); however, it would be interesting to see how the experimental condition compares to a different control, where learning activities (WEs and problems) are selected randomly.

We explained a new formula to measure assistance. Using assistance scores, we can have a better understanding of how much assistance students should receive. We used assistance scores to calculate cognitive efficiency. Perhaps one of the other benefits of assistance scores is in identifying novices and advanced students while they solve problems. Note that students may have different knowledge levels for solving different problems. Knowing students' knowledge levels about a problem can help us to provide proactive feedback messages.

Prior research has shown that adaptive faded examples are superior to non-adaptive faded examples (Salden et al. 2009), but their fading strategy was based on students' performance in answering self-explanation prompts. In our study, we used the student model to see how much students learnt about each clause, then faded the clause(s) students learnt most recently. Although we used self-explanation prompts in our study, in contrast with (Salden et al. 2009), the adaptive strategy does not use self-explanation scores. The adaptive strategy can be used for any ITSs that provide multi-level feedback, while the model proposed by (Kalyuga and Sweller 2005) is designed for a specific domain (algebra).

A limitation of the presented study is its singular domain focus (SQL), and it is not possible to determine whether or not the reported results can be generalised. Moreover, our adaptive strategy fades only one or two clauses of a solution. Although it allows WEs to gradually transition to problems, a fading strategy that fades more clauses (i.e. steps) may be even more effective.

The presented adaptive strategy determines the type of task based on the student's cognitive efficiency score on the previous problem. The performance element of cognitive efficiency is calculated from the assistance score. In future research, performance scores can be calculated more precisely by using assistance scores and self-explanation scores; we expect that would further improve the adaptive strategy.

References

Aleven, V., Koedinger, K.R.: An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. Cognit. Sci. **26**(2), 147–179 (2002)

Anthony, L., Mankoff, J., Mitchell, T., Gross, M.: Developing Handwriting-based Intelligent Tutors to Enhance Mathematics Learning. Doctoral Dissertation, Carnegie Mellon University (2008)

Atkinson, R.K., Derry, S.J., Renkl, A., Wortham, D.: Learning from examples: instructional principles from the worked examples research. Rev. Educ. Res. **70**(2), 181–214 (2000)

Brown, A.L., Kane, M.J.: Preschool children can learn to transfer: learning to learn and learning from example. Cognit. Psychol. **20**(4), 493–523 (1988)

Brusilovsky, P.: Adaptive navigation support in educational hypermedia: the role of student knowledge level and the case for meta-adaptation. Br. J. Educ. Technol. **34**(4), 487–497 (2003)

Burow, R., Weber, G.: Example explanation in learning environments. In: Frasson, C., Gauthier, G., Lesgoed, A. (eds.) Proceedings of the 3rd International Conference Intelligent Tutoring Systems, Montreal, Canada, pp. 457–465. Springer, Berlin (1996)

Chandler, P., Sweller, J.: Cognitive load theory and the format of instruction. Cognit. Instr. **8**(4), 293–332 (1991)

Chi, M., De Leeuw, N., Chiu, M.H., LaVancher, C.: Eliciting self-explanations improves understanding. Cognit. Sci. **18**(3), 439–477 (1994)

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., Glaser, R.: Self-explanations: how students study and use examples in learning to solve problems. Cognit. Sci. **13**(2), 145–182 (1989)

Chi, M.T.H., VanLehn, K.A.: The content of physics self-explanations. J. Learn. Sci. **1**(1), 69–105 (1991)

Clark, R.C., Nguyen, F., Sweller, J.: Efficiency in learning: evidence-based guidelines to manage cognitive load. Jossey-Bass, San Francisco (2006)

Corbett, A., Kauffman, L., MacLaren, B., Wagner, A., Jones, E.: A cognitive tutor for genetics problem solving: learning gains and student modeling. J Educ Comput Res. **42**, 219–239 (2010)

Corbett, A., MacLaren, B., Wagner, A., Kauffman, L., Mitchell, A., d Baker, R.S.: Differential Impact of Learning Activities Designed to Support Robust Learning in the Genetics Cognitive Tutor. In: Lane, H.C., Yacef, K., Mostow, J. (Eds.) Proceedings of 16th Artificial Intelligence in Education, Memphis, pp. 319–328 (2013)

Davidovic, A., Warren, J., Trichina, E.: Learning benefits of structural example based adaptive tutoring systems. IEEE Trans Educ. **46**(2), 241–251 (2003)

Deeks, A.: Web Based Assignments in Structural Analysis. Nanyang Technological University, Deeks, Centre for Educational Development (2000)

Hattie, J.: Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement. Routledge, New York (2009)

Hausmann, R., Nokes, T., VanLehn, K., Gershman, S.: The design of self-explanation prompts: the fit hypothesis. In: Proceedings of 31st Annual Conference of the Cognitive Science Society, pp. 2626–2631 (2009)

Hilbert, T.S., Renkl, A.: Learning how to use a computer-based concept-mapping tool: self-explaining examples helps. Comput. Hum. Behav. **25**(2), 267–274 (2009)

Hsiao, I.H., Bakalov, F., Brusilovsky, P., König-Ries, B.: Progressor: social navigation support through open social student modeling. New Rev. Hypermed. Multimed. **19**(2), 112–131 (2013)

Kalyuga, S.: Expertise reversal effect and its implications for learner-tailored instruction. Educ. Psychol. Rev. **19**, 509–539 (2007)

Kalyuga, S., Chandler, P., Tuovinen, J., Sweller, J.: When problem solving is superior to studying worked examples. Educ. Psychol. **93**(3), 579–588 (2001)

Kalyuga, S., Sweller, J.: Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. Educ. Technol. Res. Dev. **53**(3), 83–93 (2005)

Karampiperis, P., Sampson, D.G.: Adaptive Instructional Planning Using Ontologies. In: Proceedings of 4th IEEE International Conference Advanced Learning Technologies Joensuu, Vol. **4**, pp. 126-1 (2004)
Kim, R.S., Weitz, R., Heffernan, N.T., Krach, N.: Tutored Problem Solving vs . " Pure " Worked Examples. In: Proceedings of 31st Annual Conference of the Cognitive Science Society, pp. 3121–3126. Cognitive Science Society, Austin, TX (2007)
Koedinger, K., Aleven, V.: Exploring the assistance dilemma in experiments with cognitive tutors. Educ. Psychol. Rev. **19**(3), 239–264 (2007)
Kostons, D., van Gog, T., Paas, F.: Self-assessment and task selection in learner-controlled instruction: differences between effective and ineffective learners. Comput. Educ. **54**(4), 932–940 (2010)
Kyun, S., Kalyuga, S., Sweller, J.: The effect of worked examples when learning to write essays in English literature. J. Exp. Educ **81**(3), 385–408
Martin, B., Mitrovic, A.: Using learning curves to mine student models. In: Ardissono, L., Brna, P., Mitrovic, A. (eds.) Proceedings of 10th International Conference User Modeling, pp. 79–88. Springer, Edinburgh (2005)
Masthoff, J.: Design and evaluation of a navigation agent with a mixed locus of control. In: Cerri, S.A., Gouarderes, G., Paraguacu, F. (eds.) Proceedings of International Conference Intelligent Tutoring Systems, pp. 982–991. Springer, Berlin (2002)
McLaren, B.M., van Gog, T., Ganoe, C., Karabinos, M., Yaron, D.: The efficiency of worked examples compared to erroneous examples, tutored problem solving, and problem solving in classroom experiments. Comput. Hum. Behav. **55**, 87–99 (2016)
McLaren, B., Isotani, S.: When Is It Best to Learn with All Worked Examples? In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) Proceedings of 15th International Conference Artificial Intelligence in Education, pp. 222–229. Springer, Berlin (2011)
McLaren, B.M., Lim, S.-J., Koedinger, K.R.: When and how often should worked examples be given to students? New results and a summary of the current state of research. In: Proceedings of Conference Cognitive Science Society, pp. 2176–2181. Cognitive Science Society, Austin, TX (2008)
Mitrovic, A.: An intelligent SQL tutor on the web. Int. J. Artif. Intell. Educ. **13**(2–4), 173–197 (2003)
Mitrovic, A., Martin, B.: Evaluating the effectiveness of feedback in SQL-Tutor. In: Kinshuk, Jesshope, C., Okamoto, T. (eds.) Proceedings International Workshop on Advanced Learning Technologies, pp. 143–144. Palmerston North (2000)
Mitrovic, A., Martin, B.: Scaffolding and fading problem selection in SQL-Tutor. In: Hoppe, U., Verdejo, F., Kay, J. (eds.) Proceedings of 11th International Conference Artificial Intelligence in Education, pp. 479–481. Australia, Sydney (2003)
Mitrovic, A., Weerasinghe, A.: Revisiting Ill-Definedness and the Consequences for ITSs. In: Dimitrova, V., Mizoguchi, R. (eds.) Proceedings of 14th International Conference on Artificial Intelligence in Education, pp. 375–382. IOS Press, Brighton (2009)
Najar, A.S., Mitrovic, A.: Do novices and advanced students benefit differently from worked examples and ITS? In: Wong, L.H., Liu, C.-C., Hirashima, T., Sumedi, P., Lukman, M. (eds.) Proceedings of International Conference ICCE, pp. 20–29. Bali (2013a)
Najar, A.S., Mitrovic, A.: In: Yacef, K., et al. (eds.) Examples and Tutored Problems: How can Self-Explanation Make a Difference to Learning? Proceedings of 16th International Conference AIED, pp. 339–348. Springer, Heidelberg (2013b)
Najar, A.S., Mitrovic, A.: Examples and tutored problems: is alternating examples and problems the best instructional strategy? Res. Pract. Technol. Enhanc. Learn. **9**(3), 439–459 (2014)
Najar, A.S., Mitrovic, A., McLaren, B.M.: Adaptive support versus alternating worked examples and tutored problems: which leads to better learning? In: Dimitrova, V., et al. (eds.) User Modeling, Adaptation, and Personalization, pp. 171–182. Springer International Publishing, Aalborg (2014)
Najar, A.S., Mitrovic, A., McLaren, B.M.: Examples and tutored problems: adaptive support using assistance scores. In: Proceedings of 24th International Joint Conference on Artificial Intelligence Buenos Aires, pp. 4317-4323. AAAI Press (2015)
Paas, F.G.W.C., Van Merrienboer, J.J.G.: The efficiency of instructional conditions: an approach to combine mental effort and performance measures. Hum. Factors **35**, 737–743 (1993)
Peachey, D., McCalla, G.: Using planning techniques in intelligent tutoring systems. Int. J. Man-Mach. Stud. **24**, 77–98 (1986)
Roediger, H.L., Karpicke, J.D.: Test-enhanced learning taking memory tests improves long-term retention. Psychol. Sci. **17**(3), 249–255 (2006)

Rourke, A., Sweller, J.: The worked-example effect using ill-defined problems: learning to recognise designers' styles. Learn. Instr. **19**(2), 185–199 (2009)
Salden, R., Aleven, V., Renkl, A., Schwonke, R.: Worked examples and tutored problem solving: redundant or synergistic forms of support? Topics Cognit. Sci. **1**(1), 203–213 (2009)
Salden, R., Koedinger, K., Renkl, A., Aleven, V., McLaren, B.: Accounting for beneficial effects of worked examples in tutored problem solving. Educ. Psychol. Rev. **22**(4), 379–392 (2010)
Santos, O.C.: (Ed.) Educational Recommender Systems and Technologies: Practices and Challenges: Practices and Challenges. IGI Global, Hershey (2011)
Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., Salden, R.: The worked-example effect: not an artefact of lousy control conditions. Comput. Hum. Behav. **25**, 258–266 (2009)
Schworm, S., Renkl, A.: Computer-supported example-based learning: when instructional explanations reduce self-explanations. Comput. Educ. **46**, 426–445 (2006)
Sosnovsky, S., Brusilovsky, P.: Evaluation of topic-based adaptation and student modeling in QuizGuide. User Model. User-Adapt. Interact. **25**(4), 371–424 (2015)
Sweller, J.: The worked example effect and human cognition. Learn. Instr. **16**(2), 165–169 (2006)
Sweller, J., Ayres, P., Kalyuga, S.: Cognitive Load Theory. Springer, New York (2011)
Sweller, J., Cooper, G.A.: The use of worked examples as a substitute for problem solving in learning algebra. Cognit. Instr. **2**, 59–89 (1985)
Trafton, J. G., Reiser, B. J.: The Contributions of Studying Examples and Solving Problems to Skill Acquisition. In: Proceedings of 15th Annual Conference of the Cognitive Science Society, ACM Press, pp. 1017–1022 (1993)
Van Gog, T.: Effects of identical example-problem and problem-example pairs on learning. Comput. Educ. **57**(2), 1775–1779 (2011)
Van Gog, T., Kester, L., Paas, F.: Effects of worked examples, example-problem, and problem-example pairs on novices' learning. Contemp. Educ. Psychol. **36**(3), 212–218 (2011)
Van Gog, T., Paas, F.: Instructional efficiency: revisiting the original construct in educational research. Educ. Psychol. **43**(1), 16–26 (2008)
Van Gog, T., Rummel, N.: Example-based learning: integrating cognitive and social-cognitive research perspectives. Educ. Psychol. Rev. **22**(2), 155–174 (2010)
Vassileva, J.: Reactive instructional planning to support interacting teaching strategies. In: Proceedings of 7th Conference Artificial Intelligence in Education, pp. 334–342 (1995)
Webb, N.M.: Peer interaction and learning in small groups. Educ. Res. **13**(1), 21–39 (1989)
Weber, G.: Examples and remindings in a case-based help system. In: European Workshop on Advances in Case-Based Reasoning, pp. 165–177. Springer, Berlin (1994)
Weber, G., Brusilovsky, P.: ELM-ART-an interactive and intelligent web-based electronic textbook. Int. J. Artif. Intell. Educ. **26**(1), 72–81 (2016)
Weerasinghe, A., Mitrovic, A.: Facilitating deep learning through self-explanation in an open-ended domain. Knowl.-Based Intell. Eng. Syst. **10**(1), 3–19 (2006)
Weerasinghe, A., Mitrovic, A., Martin, B.: Towards individualized dialogue support for ill-defined domains. Int. J. Artifici. Intell. Educ. **19**(4), 357–379 (2009)
Weerasinghe, A., Mitrovic, A.: Facilitating adaptive tutorial dialogues in EER-tutor. In: International Conference on Artificial Intelligence in Education, pp. 630–631. Springer, Berlin (2011)

Amir Shareghi Najar is an expert data scientist at Hivetec Australia. He is also a co-founder and the technical team lead of Elimuplus where he pursues his vision of helping students to learn faster and more efficient by using Intelligent Educational Systems. Dr Shareghi Najar holds PhD degree in Computer Science from the University of Canterbury, New Zealand. His research interests include worked-out examples, self-explanation prompts, eye-tracking applications in education, and adaptive contents strategies.

Antonija Mitrovic University of Canterbury, Department of Computer Science and Software Engineering, Private Bag 4800, Christchurch 8140, New Zealand Dr A. Mitrovic is Professor at the University of Canterbury, founder and director of the Intelligent Computer Tutoring Group. Dr Mitrovic received her PhD in Computer Science from the University of Nis, Yugoslavia, in 1994. She is an associate editor of the following journals: International Journal on Artificial Intelligence in Education, IEEE Transactions on Teaching and Learning Technologies, and Research and Practice in Technology Enhanced Learning. She was awarded the Distinguished Researcher Award in 2011 by the Asia-Pacific Society for Computers in

Education, and served as the President of the International Society on Artificial intelligence in Education 2013-2015. Dr Mitrovic's primary research interests are in student modelling. She has authored over 230 papers and edited several books.

Bruce M. McLaren Carnegie Mellon University, Human-Computer Interaction Institute, 5000 Forbes Avenue, Pittsburgh, PA 15213 Dr. B. M. McLaren, an Associate Research Professor at Carnegie Mellon University, is particularly interested in intelligent tutoring systems, e-learning principles, collaborative learning, and educational games. Dr. McLaren holds a Ph.D. and M.S. in Intelligent Systems from the University of Pittsburgh, an M.S. in Computer Science from the University of Pittsburgh, and a B.S. in Computer Science (*cum laude*) from Millersville University. He has over 130 publications, including 25 journal articles, and serves as the President-Elect (2017–2019) of the International Society for Artificial Intelligence in Education.