

Journal of Educational Psychology

Can Collaborative Learning Improve the Effectiveness of Worked Examples in Learning Mathematics?

Endah Retnowati, Paul Ayres, and John Sweller

Online First Publication, December 19, 2016. <http://dx.doi.org/10.1037/edu0000167>

CITATION

Retnowati, E., Ayres, P., & Sweller, J. (2016, December 19). Can Collaborative Learning Improve the Effectiveness of Worked Examples in Learning Mathematics?. *Journal of Educational Psychology*. Advance online publication. <http://dx.doi.org/10.1037/edu0000167>

Can Collaborative Learning Improve the Effectiveness of Worked Examples in Learning Mathematics?

Endah Retnowati, Paul Ayres, and John Sweller
University of New South Wales

Worked examples and collaborative learning have both been shown to facilitate learning. However, the testing of both strategies almost exclusively has been conducted independently of each other. The main aim of the current study was to examine interactions between these 2 strategies. Two experiments ($N = 182$ and $N = 122$) were conducted with Grade-7 Indonesian students, comparing learning to solve algebra problems, with higher and lower levels of complexity, collaboratively or individually. Results from both experiments indicated that individual learning was superior to collaborative learning when using worked examples. In contrast, in Experiment 2, when learning from problem solving using problem-solving search, collaboration was more effective than individual learning. However, again in Experiment 2, studying worked examples was overall superior to learning from solving problems, particularly for more complex problems. It can be concluded that while collaboration could be beneficial when learning under problem solving conditions, it may be counterproductive when studying worked examples.

Keywords: worked examples, cognitive load theory, collaboration, problem complexity

Across multiple domains ranging from mathematics to visual arts, researchers have demonstrated that when learning novel material, guided instruction through worked examples is more effective for novice learners than conventional problem solving strategies (see Atkinson, Derry, Renkl, & Wortham, 2000; P. A. Kirschner, Sweller, & Clark, 2006; Renkl, 2014a, 2014b; Sweller, Ayres, & Kalyuga, 2011). However, most of the research into worked examples has focused exclusively on individual learning settings. Few attempts have investigated worked examples in collaborative settings (e.g., F. Kirschner, Paas, Kirschner, & Janssen, 2011; Retnowati, Ayres, & Sweller, 2010).

A main aim of the current study (Experiment 1) was to investigate the effectiveness of collaborative learning compared with individual learning within a worked examples environment. Another aim (Experiment 2) was to compare possible interactions between studying individually or collaboratively on the one hand and studying worked examples or solving problems on the other hand. We will begin by outlining the worked example effect.

The Worked Example Effect

A worked example provides a step-by-step solution to a problem or task and is a form of explicit instruction (see P. A. Kirschner et al., 2006). Rather than trying to acquire new information through

problem-solving search or other types of discovery methods, learners are shown worked examples to study. Worked examples provide an expert's problem-solving model, from which students can study and learn (Atkinson et al., 2000). With worked examples, learners are able to focus on understanding a solution rather than focus on solving the problem (Renkl, 2014a). The worked example effect occurs when students who learn from studying worked examples subsequently obtain superior test scores to students who learn from solving problems. From this perspective, we refer to problem solving as solving problems with minimal teacher/instructor guidance on how to solve the problem.

Using algebra problems, Cooper and Sweller (1987) and Sweller and Cooper (1985) provided the first demonstrations of the worked example effect (Sweller et al., 2011). They found that students who were asked to study worked examples performed better on subsequent problem solving tests than students required to practice solving the equivalent problems. The effect was explained by the suggestion that worked examples reduced extraneous working memory load compared to solving the equivalent problems. A reduction in cognitive load facilitated the transfer of knowledge to long-term memory. These findings led to further research in mathematics and scientific domains. For example, the worked example effect was replicated in algebra (Carroll, 1994), geometry (Paas & van Merriënboer, 1994; Tarmizi & Sweller, 1988), statistics (Paas, 1992; Quilici & Mayer, 1996), and physics (Ward & Sweller, 1990) using a range of age groups and subject areas with the advantage appearing for both similar and transfer problems. Building on these initial findings, more contemporary research (for summaries, see Ayres & Sweller, 2013; Renkl, 2014a) has found the effect in nonscience domains such as visual arts (Rourke & Sweller, 2009) and English literature (Kyun, Kalyuga, & Sweller, 2013; Oksa, Kalyuga, & Chandler, 2010), as well as ongoing investigations in the science domain, such as problem solving in electrical circuits (van Gog & Kester, 2012), and geometry (Chen, Kalyuga, & Sweller, 2015, 2016a).

Endah Retnowati, Paul Ayres, and John Sweller, School of Education, University of New South Wales.

Endah Retnowati is now at the Department of Mathematics, Education Faculty of Mathematics and Sciences, Yogyakarta State University, Indonesia.

Correspondence concerning this article should be addressed to Paul Ayres, School of Education, University of New South Wales, Sydney, NSW 2052, Australia. E-mail: p.ayres@unsw.edu.au

Worked examples illustrate one of the main principles of cognitive load theory, the borrowing and reorganizing principle (see Sweller & Sweller, 2006). This principle suggests that the most effective way to obtain new information is by directly receiving it from another person who already has this information. The major mechanisms are listening to other people, reading what they write, and imitating what they do. In that sense, information is borrowed from another person's long-term memory. In the case of worked examples, information is borrowed from the long-term memory of the constructor of the worked examples. However, this information is reorganized by the learner by integrating the new information with old information stored in the learner's long-term memory (see Sweller et al., 2011). Integrating new information with old information that is already understood may assist in making sense of the new information. This conceptualization is consistent with other theories of learning that emphasize reorganization and elaboration (see Mayer, 2014). Because worked examples provide a low cognitive load environment compared with problem solving search, learning is enhanced through the construction of new schematic knowledge.

Virtually all published research into worked examples has been conducted using individual rather than collaborative learning. A notable exception was a study by Retnowati et al. (2010), who found in a single experiment using Grade 7 Indonesian students that worked examples in geometry were superior to problem solving for both individual and collaborative learners on both retention and transfer tasks that required calculations as well as providing related explanations. Qualitative data also revealed that participants believed that they understood the material more easily when using worked examples rather than problem solving. A goal of the present study was to investigate whether worked examples could be enhanced by using collaborative settings.

Collaborative Learning

Collaborative learning occurs when students learn by collaborating rather than by studying individually. It is widely used (Gillies, 2003) and considered highly desirable in the community and workplace (Barron, 2000). Considerable evidence suggests that collaborative learning has significant academic, social, and psychological benefits (Johnson, Johnson, & Smith, 1998). Multiple studies and meta-analyses have found that the various forms of collaborative or cooperative learning strategies where students work together have significant benefits over students who work individually (see Johnson, Maruyama, Johnson, Nelson, & Skon, 1981). Many of these studies have focused on learning mathematics, showing that small-group learning has led to greater mathematical outcomes than traditional methods of teaching individuals (see Davidson & Kroll, 1991). Explanations for this advantage are usually grounded in social constructivist theory or social independence theory, which emphasize that learning should be facilitated through social and collaborative activities where students construct knowledge by interactions with others and through collective goals (Johnson & Johnson, 1994; Schreiber & Valle, 2013).

Studies have been conducted to identify the factors that improve collaborative learning (for reviews, see Cohen, 1994; Kreijns, Kirschner, & Jochems, 2003; Schreiber & Valle, 2013; Van den Bossche, Gijsselaers, Segers, & Kirschner, 2006; Webb, 2009; Weinberger, Stegmann, & Fischer, 2007). It is generally agreed

that collaborative learning requires active social interactions, group goals, and individual accountability (see Slavin, 1995).

The use of problem solving activities within collaborative learning classrooms has been strongly advocated, especially by mathematics educators (see, e.g., the National Council of Teachers of Mathematics, 2000). According to De Corte (2004), one view of mathematics learning is that it is a social construction of knowledge through collaboration. An emphasis should be placed on problem solving, reasoning, and communication, forming communities of mathematical inquiry (Goos, 2004; Staples, 2007). Shared meanings of the main concepts emerge through the interactions associated with group problem solving (Plass et al., 2013), as well as learners constructing their own ideas and individual insights (Yackel, Cobb, & Wood, 1991).

Collaborative Learning and Evolutionary Psychology

Evolutionary psychology is used as a base for cognitive load theory, and this view of cognition can be used to provide a new perspective on some of the fundamental underpinnings of collaborative learning (Paas & Sweller, 2012; Sweller et al., 2011). A key aspect of this argument comes from the work of Geary (1995, 2008, 2012), who distinguished between two types of knowledge: biologically primary and secondary knowledge. Biologically primary knowledge is knowledge that we have evolved to acquire over many generations. It is easily and unconsciously acquired and is modular with different skills likely to have evolved during different evolutionary epochs. Examples are learning to listen, speak, recognize faces, and use general problem solving strategies. Biologically secondary knowledge is knowledge that we need to acquire for cultural reasons. We have evolved to acquire secondary knowledge as a general skill. We have not evolved to acquire particular types of secondary knowledge in the same way that we have evolved to acquire particular types of primary knowledge. Virtually every topic taught in education and training establishments provides an example of biologically secondary knowledge.

Geary argued that working in a collaborative environment may be natural and effortless, because it is a biologically primary activity that humans have evolved to engage in (Geary, 1995, 2008). However, this advantage may come at a cost (Geary, 1995, 2008), as during collaborative learning, students may tend to automatically develop their general communication and coordination skills, rather than allocating more attention to the assigned biologically secondary knowledge. While Geary (2008) acknowledges that social context and interaction with teachers and peers contribute to a student's learning, he also questions whether students can learn better in social contexts, rather than through explicit instruction.

As shown in many studies (see Johnson et al., 1998), social skills may automatically be improved through collaboration, which is consistent with Geary's argument outlined above. However, learning the content of a collaborative lesson is another matter because that content most likely requires the acquisition of biologically secondary skills (e.g., mathematics) that require conscious effort. As Geary suggested, collaboration may not necessarily produce advantages in academic outcomes if no more than an automatic improvement in collaborative skills occurs.

Collaborative Learning and Cognitive Load Theory

Similar to worked examples (as argued above), collaborative learning demonstrates another example of the borrowing and reorganizing principle (see Paas & Sweller, 2012). Knowledge can be borrowed from other members of the group, and reorganized, linking new knowledge with old knowledge stored in long-term memory. Group interactions can help individuals make sense of the information and steer the reorganization of the information accordingly (see De Corte, 2004; Plass et al., 2013). Because humans have evolved to communicate, to share, and to obtain information from each other as biologically primary skills, collaborative learning may have an advantage over individual learning in that it involves sharing information and learning from each other, as occurs in everyday life (Sweller et al., 2011).

Another advantage of collaborative learning is that it may assist in learning complex materials. Complex materials are difficult to learn because they impose a heavy working memory load (Sweller et al., 2011). However, if the learning material is shared among several group members, an individual is required to process less task-relevant information, potentially reducing working memory load (F. Kirschner, Paas, & Kirschner, 2009a). Working memory resources then can be allocated to learning about important aspects of the materials by processing relevant information communicated from other group members. Based on this view, collaboration should be effective by providing group members with information that they otherwise would need to search for themselves. This potential provision of information should reduce extraneous cognitive load. In this sense, a biologically primary activity, collaboration, may provide an advantage in acquiring biologically secondary knowledge such as mathematics. Combining the limited working memory resources of several individuals should increase the resources available to all in a manner that does not occur when students are engaged in individual learning and have to deal with all the working memory load themselves. Hence, through collaboration, individuals may be better able to learn about complex materials.

Initial experimental evidence in support of this hypothesis was found by F. Kirschner, Paas, and Kirschner (2009b) using a high-school biology topic, where an individual learning condition was compared to a collaborative learning condition consisting of three group members. During the learning phase, students were given problem-solving tasks to complete individually or collaboratively. For the collaborative learning condition, every member of a group had information about one third of the whole task only, and hence sharing was required to complete the task. In the individual condition, individual students were given the whole task to solve. Following the learning phase, all students were tested individually using retention and transfer tasks. Significant interaction effects were found. For the retention tasks individuals learned more efficiently, while for the transfer tasks collaboration led to more efficient learning. In a follow up study also with biology content, F. Kirschner et al. (2011) found that collaborative learning was more effective than individual learning on high but not low complexity tasks.

The Current Study

The evidence described so far suggests that both worked examples and collaborative learning are effective learning strategies.

This study aimed to extend the research into both strategies by combining them in an authentic learning environment. More specifically, the main research question was to investigate if the effectiveness of worked examples could be improved by using collaborative learning. If, as indicated above, the borrowing and reorganizing principle suggests that most learning is based on obtaining information from others, then the use of collaboration permits learners to not only obtain information from explicit instruction via worked examples, but also obtain information from co-learners. By studying worked examples collaboratively, learners may obtain additional information from other learners that would not be available if learning individually.

Notwithstanding the possible advantage of adding information from collaborators to the information obtained from a worked example, the expertise reversal effect suggests that for given levels of expertise and complexity (Chen, Kalyuga, & Sweller, 2016b), the provision of additional information can become redundant, resulting in an increase rather than a decrease in cognitive load (Kalyuga, Ayres, Chandler, & Sweller, 2003). Evidence for such an outcome was obtained by Nihalani, Mayrath, and Robinson (2011). They found that for novices, feedback was more effective than collaboration. For more expert learners, the addition of feedback reduced learning and reduced the advantages of expertise. For these learners, feedback was redundant and redundancy has been shown repeatedly to interfere with learning due to an increased extraneous cognitive load. Hence, there may be conditions under which the combination of collaboration and worked examples may be less advantageous.

We also investigated how problem complexity has an impact on the effectiveness of collaborative learning and worked examples. Furthermore, because of the design of the experiments it was possible under the conditions to examine if the collaborative context was superior to individual learning, and whether worked examples were superior to problem solving. Throughout the study, authentic classroom environments were used rather than laboratory conditions.

Study Hypotheses

Hypothesis 1: Worked examples will be enhanced by studying collaboratively compared to studying individually. This hypothesis was based on research that argues collaborative learning is superior to individual learning (see Johnson & Johnson, 2002). In addition considerations of evolutionary psychology suggest that humans have evolved to collaborate naturally (Geary, 1995, 2008) and that collaborative environments provide an effective way to obtain new information by directly receiving it from another person who already has this information (Paas & Sweller, 2012; Sweller & S. Sweller, 2006).

Hypothesis 2. The effectiveness of collaborative learning will be increased by task complexity. This hypothesis flows from the general research into collaborative learning and problem complexity (see Cohen, 1994). Evidence for the impact of complex tasks in collaborative learning compared to individual learning has been demonstrated by F. Kirschner, Paas, and Kirschner (2011) using biology content, and by Zhang, Ayres, and Chan (2011) using web design materials. As reported, F. Kirschner et al. (2011) also showed that effective collaborative learning requires a high intrinsic cognitive load that cannot be

tackled easily by individuals. Students who learned in groups gained a benefit by sharing the high working memory load created by the complex tasks with other group members.

Hypothesis 3: Studying worked examples would be more advantageous than conventional problem solving. This hypothesis flows from cognitive load theory and the worked example effect. It was tested in Experiment 2.

Experiment 1

This experiment tested the hypotheses by investigating the influence of problem complexity on individual and collaborative learning using a worked example strategy. Two types of algebraic problems were created with low or high levels of complexity. Problem complexity was categorized by the number of steps required to complete the solution, and the level of conceptual knowledge required. The topic, solving linear equations, was selected from the National Curriculum of Indonesia as the experiment was conducted in Indonesian schools.

With problems of differing complexity it was feasible that the order in which they were presented could influence subsequent learning. Hence, the problem sequence was counterbalanced in this experiment to avoid sequential learning effects. All participants received a set of worked examples to test Hypotheses 1 and 2. A 2 (learner grouping context: Collaborative vs. Individual) \times 2 (level of complexity: Low vs. High) \times 2 (task sequence: Low–High complexity vs. High–Low complexity) mixed experimental design was used with level of complexity the repeated measure.

Method

Participants. One hundred eighty-two students from six Year 7 mathematics classes in an Indonesian school in Magetan, East Java, participated in the study. The school followed the national curriculum, and the topics used in the experiment were mandated by the curriculum. The Indonesian national curriculum requests teachers not to use teacher-centered learning methods such as lectures but to use student-centered learning methods such as small group discussions (BNSP, 2006; Depdiknas, 2004; National Ministry of Education, 2006). The participating school indicated that the students were used to studying in small groups in all subjects with varied methods of instruction. The school also indicated that they had allocated students to the six mathematics classes randomly at the beginning of the school year. A team of three mathematics teachers taught specific topics to all six classes, indicating that all students received mathematics instruction from each teacher on set topic blocks throughout the school year.

At the beginning of the school year students were assigned to small learning groups by the mathematics teachers based on having the same gender, and of mixed ability (heterogeneous groupings). Grouping students together according to gender was part of the school's policy for students this age, as it was assumed that boys and girls interact minimally and form single-sex friendships. As friendship groupings can have positive effects on collaboration (see Hanham & McCormick, 2009), it was thus assumed that the group members had developed some level of cohesiveness and familiarity with each other, and could work collaboratively.

These preexisting groups that had been created 3 months earlier by the school, independent of this study, formed the basis for

creating the two grouping treatments. Each group was assigned at random to either stay as a group or become uncoupled to study individually. This process produced 79 individual learners and 27 collaborative groups (22 groups of 4, 5 groups of 3, $n = 103$). Both groups and individual learners were then randomly assigned to a specific task sequence of either low–high or high–low complexity. Due to absenteeism 168 students (88 girls, 80 boys) actually participated, with an average age of 12.6 years ($SD = 0.46$). In the low–high complexity sequence, 38 students completed the task individually and 45 students completed the task collaboratively (9 groups of 4, 3 groups of 3). In the high–low complexity sequence, there were 33 students in an individual and 52 students in a collaborative context (10 groups of 4, 4 groups of 3).

Materials. Two types of algebra problems were created based on solving linear equations with differing levels of complexity. Both task types required students to solve a linear equation. The low complexity problem was presented in algebraic notation, but the high complexity problem required an equation to be derived, as it was presented as a word problem. The requirement to translate the words into equations increased complexity.

An example of a low-complexity problem is “Solve $3n + 10 = 85$, for n .” An example of an equivalent high-complexity problem is “Three times the number of Dina’s marbles when added to 10 equals eighty-five. How many marbles does Dina have?” The high-complexity problem required more solution steps, as not only does the equation have to be constructed, a conceptually demanding task, but it also has to be solved. Consequently, this word problem was considered higher in element interactivity (Sweller, 2010; Sweller & Chandler, 1994), because several variables have to be considered simultaneously to construct the equation, although the given problem context may describe operators (symbols) in the constructed equation more meaningfully. In contrast, the low complexity problem does not have this additional task; hence, the algebra rules can be applied in a straightforward fashion. The students in this study had some previous experience with linear equation solving and word problems, but mostly with fewer variables, and not with a combination of constructing and solving equations. For each problem type, instructional and testing materials were constructed.

An instructional materials booklet was designed using a worked example approach. The worked example material used problem pairs, consisting of a worked example and a similar problem to be solved (see Sweller & Cooper, 1985; Trafton & Reiser, 1993). The worked example provided a problem statement and a step-by-step solution to the problem (i.e., algorithm, explanation, final answer) and was written on the left side of the page. The paired problem to be solved was positioned on the right side of the page and consisted of the problem statement only. Final answers for these problems, but not step-by-step solutions, were provided on the same page of the booklet to allow students to know whether they had correctly solved the problem, providing some support consistent with previous research (see Cooper & Sweller, 1987). The relevant instruction was provided directly above each problem. All instructions were in the students’ native Indonesian. Appendixes A and B show examples, translated into English, of the format of the low-complexity and the high-complexity worked examples respectively.

The learning material of low-complexity problems consisted of four worked example problem pairs. Hence, the worked example

condition required learners to study 4 worked examples and solve 4 problems overall, whereas the problem solving condition required all 8 problems to be solved. The similar and transfer tests required 4 and 3 problems to be solved, respectively. The internal consistency of the similar test using Cronbach's alpha was .84, and .75 for the transfer test. The transfer test problems consisted of modified equations requiring more solution steps than the similar test problems. The learning material of high-complexity problems consisted of 3 worked example problem pairs. Hence, the worked example condition required learners to study 3 worked examples and solve 3 problems overall, whereas the problem solving condition required all 6 problems to be solved. The similar and transfer tests consisted of 3 and 2 problems, respectively. Cronbach's alpha was .86 for the similar test, and .71 for the transfer test. The transfer test problems had the additional requirement of calculating a subgoal before the goal could be calculated.

To measure cognitive load during acquisition, a self-rating scale of difficulty was used based on the scale developed by Paas (see Paas, 1992; van Gog & Paas, 2008). Furthermore, consistent with recent research, which suggested that multiple recordings produce the most consistent results (see van Gog, Kirschner, Kester, & Paas, 2012), every page of the instructional material had a subjective rating question, written on the bottom line of the page, that asked, "How easy or difficult was it to study and solve these problems? Circle your answer on a scale from 1 = *Extremely easy* to 9 = *Extremely difficult*." The cognitive load measures collected on each page were added and then averaged to describe the overall student's cognitive load experience in this phase.

Procedure. Before the experimental stage started, all students underwent a preparation period. This initial session was conducted by one of the researchers who was a native Indonesian mathematics teacher. First, students practiced translating a word problem containing one operator into an equation, based on the statement "Bob has 3 more marbles than Wina." The purpose was to activate students' prior knowledge about translating a simple sentence containing a variable into an equation along with the basic algebra rules. The researcher used explicit instruction to explain how to solve this problem.

Second, to familiarize students with instruction using worked examples, four pairs of worked examples, using the same format as in the main experiment, were provided. Each pair consisted of an example to study followed by a similar problem to solve and dealt with translating a simple sentence into an equation. This practice lasted 15 min and then the results of the constructed formula were discussed with the teacher. Immediately afterward, three worked-example pairs for solving simple linear equations by applying one algebra rule were given (e.g., solve $a + 20 = 65$). The results of this 15-min practice were then also discussed with the teacher. This discussion was based on student questions with the researcher responding to the questions without further elaboration.

To complete the preparation period, the teacher then provided an example of a word problem (complex problem), similar to the first problem of the high complexity problem in the learning material. This problem was written on the blackboard, and students were shown how to translate the word problem into a linear equation. The teacher explained that two or more steps were required to transform the linear equation in such a way that it could be solved. However, the step-by-step solution and the final answer were not

shown. The whole preparation period was repeated for each class (6 times) by the same researcher.

In the first stage of the experiment (Stage I), students in the Low-High sequence were presented the low-complexity materials first, whereas those in the High-Low sequence were presented the high-complexity materials first. This stage consisted of three phases: acquisition, similar test, and transfer test, which were completed without pauses between them.

Students in each class were separated into two classrooms according to their grouping classifications to begin the acquisition phase, with each classroom supervised by both a teacher from the school and the researcher. First, each student received a worked example booklet specific to their learning condition. Twenty min were allocated for all groups completing low-complexity problems, and 30 min were allocated for high-complexity problems. Before learning commenced, the supervising teacher explained the rules for studying individually or collaboratively, reading from a common script for each strategy.

For individual study, students were told to put an effort into understanding the learning material individually and were not permitted to ask any questions of the other students or the teacher during learning. For collaborative study, students were told by the teacher to discuss the learning material together by reading the task together, eliciting understanding, helping each other, and making sure every member understood the learning material. They were not permitted to ask any questions of other group members or the teacher during learning. For both groups it was also explained how students should complete the cognitive load measures that would appear on each page of their booklet. No feedback was provided during or after the acquisition phase.

Directly following the acquisition phase, the similar and transfer tests were completed individually. All students were given the maximum time period and did not receive any feedback. Fifteen min and 20 min were given to complete the low-complexity similar test and transfer tests, respectively. To complete the high-complexity similar and transfer tests, 20 min were given for each test. After the transfer test, students were given a 15-min break.

Stage II was completed directly after the break, and students switched to the alternate complexity level materials. If students had initially completed the low complexity problems, they then completed those with high complexity next, and vice versa. Allocated times depended on the material and activity as described in Stage I.

During the acquisition phase, group answers were allowed for some groups, and therefore these data were not analyzed, as individual responses were not available for all participants. Scoring for the similar and transfer tests used the following guidelines: For a low-complexity problem, each successful answer had to complete two steps showing two algebraic manipulations. If the answer was entirely correct, a score of 2 was given. If only one step (one strategy) was correctly applied, a score of 1 was given. If the answer did not show any algorithmic validity, a score of 0 was given. For a high-complexity problem, each correct answer had to include three steps. The first was creating the linear equation, while the second and third steps were solving the equation. If the answer was entirely correct, a score of 3 was given. If the equation was correctly created (the first step correct) but only partially solved (1 correct step), a score of 2 was given. If the equation was correctly created (the first step correct) but incor-

rectly solved (0 correct steps), a score of 1 was given. If the equation was incorrectly created but was solved correctly, a score of 1 was given. If the answer did not demonstrate any correct steps, a score of 0 was given. To enable a comparison to be made between the two types of problems, the total scores for each measure were converted into a proportion.

Results and Discussion

A 2 (Collaborative vs. Individual) \times 2 (Low–High vs. High–Low complexity sequence) \times 2 (Low- vs. High-complexity) ANOVA with repeated measures on the last variable was used to analyze the data. The means (and standard deviations) of test performance and cognitive load ratings are summarized in Table 1.

Cognitive load during acquisition. A significant complexity effect was found, $F(1, 164) = 41.80$, $MSE = 1.40$, $p < .001$, $\eta_p^2 = .203$. The low-complexity materials in the acquisition phase were rated significantly easier ($M = 3.05$, $SD = 1.63$) than the high-complexity materials ($M = 3.89$, $SD = 1.89$). However, no significant main effect was found for learner grouping or task sequence (for both, $F < 1$, ns.). Nor was there a significant interaction between learner grouping and task sequence, nor between learner grouping and problem complexity (for both, $F < 1$, ns.).

A significant interaction effect between problem complexity and task sequence was found, $F(1, 164) = 18.11$, $MSE = 1.40$, $p < .001$, $\eta_p^2 = .099$. Simple effect tests indicated that students reported a higher increase in cognitive load for high-complexity problems compared to low-complexity problems when the task sequence was Low–High, $F(1, 82) = 61.80$, $MSE = 1.38$, $p < .001$, $\eta_p^2 = 0.43$, compared to when the sequence was High–Low, $F(1, 84) = 1.958$, $MSE = 1.45$, $p = .165$, $\eta_p^2 = .023$. As inspection of the means indicates, the higher increase in cognitive load for the higher-complexity problems under a Low–High sequence is primarily due to the relatively low load imposed by low complexity problems when they are presented first.

Similar test results. There was a main effect for task complexity, $F(1, 164) = 50.76$, $MSE = 0.06$, $p < .001$, $\eta_p^2 = .236$. Students scored significantly higher on the low-complexity problems ($M = 0.63$, $SD = 0.34$) than the high-complexity problems ($M = 0.44$, $SD = 0.35$). There was no significant main effect for learner grouping context ($F < 1$, ns.) or task sequence ($F < 1$, ns.). There was no interaction between learner grouping and complexity, $F(1, 164) = 2.18$, $MSE = 0.06$, $p = .142$, $\eta_p^2 = .013$, and all other interaction measures were non-significant (all $F < 1$, ns.).

Transfer test results. A main effect of complexity was found, $F(1, 164) = 13.58$, $MSE = 0.07$, $p < .001$, $\eta_p^2 = .076$. The scores

for the low-complexity transfer test ($M = 0.44$, $SD = 0.37$) were significantly greater than those for the high-complexity transfer test ($M = 0.34$, $SD = 0.35$). A learner grouping effect was also found, $F(1, 164) = 7.54$, $MSE = 0.19$, $p = .007$, $\eta_p^2 = .044$. Learning individually ($M = 0.46$, $SD = 0.38$) resulted in better transfer results than learning collaboratively ($M = 0.32$, $SD = 0.34$). A significant interaction between problem complexity and learner grouping context was also found, $F(1, 164) = 19.51$, $MSE = 0.07$, $p < .001$, $\eta_p^2 = .106$ (see Figure 1). Simple effects tests indicated that learning individually resulted in better performance in high-complexity tasks than collaborative learning, $F(1, 166) = 20.59$, $MSE = 0.13$, $p < .001$, $\eta_p^2 = .11$. However, no difference was found for low-complexity tasks ($F < 1$, ns.).

A nonsignificant difference between task sequences was found, $F(1, 164) = 3.57$, $MSE = 0.19$, $p = .06$, $\eta_p^2 = .021$, although the High–Low sequence generating higher scores ($M = 0.41$, $SD = 0.31$) than the Low–High sequence ($M = 0.34$, $SD = 0.32$). A significant interaction effect between the learner grouping context and task sequence was found, $F(1, 164) = 3.89$, $MSE = 0.19$, $p = .05$, partial $\eta^2 = 0.02$. The simple effects test results indicated that individual learning resulted in a better performance than group learning, when the learning sequence was High–Low, $F(1, 83) = 12.35$, $MSE = 0.17$, $p = .001$, $\eta_p^2 = .13$. When the task sequence was Low–High, no significant differences were found, ($F < 1$, ns.). Moreover, a significant interaction between the task complexity and task sequence was found, $F(1, 164) = 4.15$, $MSE = 0.07$, $p = .043$, $\eta_p^2 = .025$. The simple effects test results indicated a significant difference of low and high complexity transfer performance when the learning sequence was Low–High, $F(1, 82) = 16.20$, $MSE = 0.08$, $p = .001$, $\eta_p^2 = .17$. When the task sequence was High–Low, no significant differences were found, $F(1, 84) = 3.48$, $MSE = 0.07$, $p = .07$, $\eta_p^2 = .04$).

An important aim of this experiment was to create two types of tasks based on simultaneous equations that had two levels of complexity. The results indicated that this aim was supported as students scored significantly higher on the low-complexity problems compared to high-complexity problems on both tests. Additionally, students also experienced a considerably lower cognitive load when learning using low-complexity problems compared to high-complexity problems. Therefore, it is likely that the higher-complexity task, with more steps for the solution, has a higher level of element interactivity.

The first hypothesis of this experiment (Hypothesis 1) predicted that students would benefit from studying worked examples collaboratively rather than individually. No evidence was found during testing to support this prediction. In contrast, a number of

Table 1
Means (and Standard Deviations) for Test Results and Cognitive Load Ratings in Experiment 1

Condition	Cognitive load (1–9)		Similar test (0–1)		Transfer test (0–1)	
	Low-complexity	High-complexity	Low-complexity	High-complexity	Low-complexity	High-complexity
Low–High sequence						
Collaborative	2.71 (1.44)	4.49 (2.26)	.64 (.35)	.42 (.32)	.47 (.41)	.17 (.33)
Individual	2.89 (1.56)	3.92 (1.87)	.57 (.36)	.44 (.35)	.38 (.38)	.34 (.38)
High–Low sequence						
Collaborative	3.31 (1.67)	3.46 (1.60)	.65 (.30)	.41 (.34)	.41 (.35)	.23 (.28)
Individual	3.30 (1.85)	3.73 (1.63)	.64 (.33)	.47 (.38)	.51 (.35)	.59 (.41)

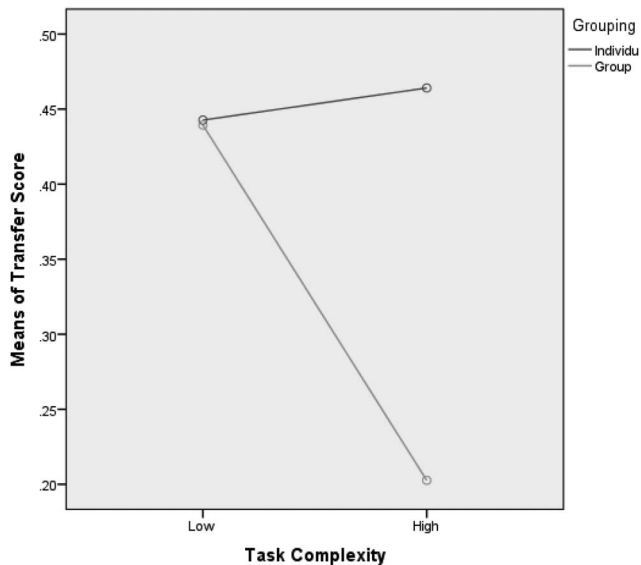


Figure 1. Interaction between task complexity and grouping context on transfer test scores in Experiment 1.

results indicated a reverse effect. On the transfer test, students who studied individually performed significantly higher than those who studied collaboratively, and more specifically on the higher-complexity tasks. Furthermore, for the High–Low order of study, individual study resulted in a significant advantage.

Hypothesis 2 predicted that the effectiveness of collaborative learning would be increased by task complexity. No support for this hypothesis was found, as collaborative learning was not found to be superior to individual learning on any specific task. The significant interaction effect on transfer problems indicated that individual study was more advantageous than collaborative study, for high-complexity problems—the reverse of what was expected. It was concluded, based on these results in the current context, that when using worked examples, collaborative study was a disadvantage.

Experiment 2

The results of Experiment 1 suggested that worked examples may not be enhanced by collaborative learning, but it was notable that worked examples were not compared with a problem solving control group as is usually the case in worked-examples research. Because we did not test for the worked example effect in Experiment 1, it is possible that the worked example approach was unsuitable for this match of topic and learner, leading to no learning advantage for using worked examples. To rule out this possibility, Experiment 2 included a problem-solving treatment. This enabled the first two hypotheses to be tested again using a 2 (instructional strategy: Worked Example vs. Problem Solving) \times 2 (grouping contexts: Collaborative vs. Individual) \times 2 (level of complexity: Low vs. High) mixed experimental design. Because in Experiment 1 few differences were found by balancing the sequencing of problem types, only the complex-simple sequence (High–Low complexity task sequence) was used in Experiment 2, as it produced the most significant interactions.

With the introduction of a problem solving treatment it was possible to test for a worked example effect (see Atkinson et al., 2000; P. Kirschner et al., 2006; Renkl, 2014a, 2014b). In other words, it was predicted that worked examples would be advantageous compared to conventional problem solving (Hypothesis 3).

Method

Participants. One hundred twenty-two students from four Year 7 classrooms in an Indonesian school, in Kudus, Central Java, participated in the study. Consistent with the sample in the previous experiment, the school had a similar organization and followed the same national curriculum. Consistent with Experiment 1, a team of three mathematics teachers taught all classes according to set topics. Also similar to the participants in Experiment 1, collaborative learning was reported as a common learning strategy, not only in mathematics classes but also in other subjects. As was the case in Experiment 1, the small groups were created at the beginning of the school year independently of this experiment, and composed of mixed ability students with the same gender. Students were assumed to be familiar with each other since they had been in the same groupings for more than five months.

First, students were randomly allocated into individual ($n = 63$) or collaborative ($n = 60$, 9 groups of 4, and 8 groups of 3) learning conditions, and then randomly assigned into worked example or problem solving groups. Five students were excluded from the analysis because they did not complete all experimental stages, leaving 118 students (46 girls, 72 boys) with an average age of 12.50 years ($SD = 0.55$). Thirty students studied worked examples individually, 29 solved problems individually, 31 (4 groups of 4, and 5 groups of 3) studied worked examples collaboratively, and 28 (5 groups of 4, 2 groups of 3, and 1 group of 2) solved problems collaboratively.

Learning materials and procedure. The materials used in this experiment were identical to Experiment 1, except that a new conventional problem-solving group was introduced. Problem-solving acquisition booklets for both levels of complexity were designed, based on the worked examples booklets. Where in Experiment 1 for each problem pair, the first problem had a fully worked example given, this solution was no longer provided. Instead, this problem now had to be solved by participants during the acquisition phase. Hence, for the worked example condition, students studied a problem, and solved a similar problem; for the problem solving condition, both problems had to be solved without solutions being shown. Each problem pair was placed on a single page, positioned identically to the worked example material except that no solutions were shown. Students were instructed to solve each problem. Equivalent to the worked example booklet, the final answer of every problem was provided on each page.

The similar test and the transfer test materials were identical to the low- and high-complexity problems used in Experiment 1, and the allocated times remained the same. The internal consistency of the tests was measured again using Cronbach's alpha for this sample. For the low-complexity problems, the values were .88 for the similar test and .80 for the transfer test. For the high-complexity problems, the values were .82 for the similar test and .67 for the transfer test.

The procedures used in this experiment were identical to Experiment 1. The only difference, apart from introducing additional

problem-solving groups, was that only the complex-simple sequence (High–Low complexity task sequence) was used for the two types of problems, as this sequence previously produced significant interactions.

Results and Discussion

A 2 (Worked Example vs. Problem Solving) \times 2 (Collaborative vs. Individual) \times 2 (Low- vs. High-complexity) ANOVA with repeated measures on the last variable was used to analyze the data. The means (and standard deviations) of test performance and cognitive load ratings are summarized in Table 2.

Cognitive load during acquisition results. A main effect of instructional strategy was obtained, $F(1, 114) = 90.96$, $MSE = 3.82$, $p < .001$, $\eta_p^2 = .44$, where the worked example conditions ($M = 3.80$, $SD = 1.56$) generated significantly lower cognitive load scores (difficulty scale) than the problem solving conditions ($M = 6.23$, $SD = 2.08$). No significant effect for learner grouping contexts was found, $F(1, 114) = 1.97$, $MSE = 3.82$, $p = .164$, $\eta_p^2 = .02$. The low-complexity problems ($M = 4.47$, $SD = 1.93$) generated significantly less cognitive load than high-complexity problems ($M = 5.56$, $SD = 1.71$), $F(1, 114) = 23.5$, $MSE = 2.99$, $p < .001$, $\eta_p^2 = .17$.

A significant 3-way interaction was found, $F(1, 114) = 3.99$, $MSE = 2.99$, $p = .048$, $\eta_p^2 = .034$. Simple effects tests showed that individual learners experienced a significantly higher cognitive load than learners in the collaborative context when learning the high-complexity problems using worked examples, $F(1, 59) = 4.34$, $MSE = 2.45$, $p = .041$, $\eta_p^2 = .069$ (see Figure 2), but no other significant effects were found.

Similar test results. A significant worked example effect was found, $F(1, 114) = 24.93$, $MSE = 0.111$, $p < .001$, $\eta_p^2 = .18$, as studying worked examples ($M = 0.53$, $SD = 0.27$) was found to be superior to problem solving ($M = 0.32$, $SD = 0.28$). No significant effect for the learner grouping context was found ($F < 1$, ns.). However, there was an interaction effect between instructional strategy and learner grouping context, $F(1, 114) = 5.92$, $MSE = 0.111$, $p = .017$, $\eta_p^2 = .049$ (see Figure 3). Simple effects tests revealed that there were no significant differences between the learner grouping contexts when students studied worked examples, $F(1, 59) = 2.56$, $MSE = 0.10$, $p = .115$, $\eta_p^2 = .042$. There was nonsignificant difference between means, with a small to medium effect size, $F(1, 55) = 3.32$, $MSE = 0.12$, $p = .07$, $\eta_p^2 = .057$, in favor of collaborative learning when students studied through the problem-solving format. That difference can be assumed to have been the primary cause of the significant interaction.

A significant effect of complexity was found, $F(1, 114) = 23.72$, $MSE = 0.04$, $p < .001$, $\eta_p^2 = .172$. Students performed significantly higher in low-complexity problems ($M = 0.49$, $SD = 0.29$) than in high-complexity problems ($M = 0.36$, $SD = 0.25$).

A 3-way interaction effect was also found, $F(1, 114) = 8.83$, $MSE = 0.04$, $p = .004$, $\eta_p^2 = .072$, caused by the significant differences found in low-complexity tests (see Figure 4). For worked examples, individual study was superior to collaborative study, $F(1, 59) = 7.01$, $MSE = 0.08$, $p = .01$, $\eta_p^2 = .106$, replicating the results of Experiment 1, but when problem solving, collaborative study was superior to individual study, $F(1, 55) = 4.67$, $MSE = 0.09$, $p = .035$, $\eta_p^2 = .078$. No significant differences were found for high-complexity problems ($F < 1$, ns., for both).

Transfer test results. There was no worked example effect, $F(1, 114) = 2.06$, $MSE = 0.104$, $p = .154$, $\eta_p^2 = .018$, nor a learner grouping context effect ($F < 1$, ns.). However, there was a significant interaction between the instructional strategy and the learner grouping context, $F(1, 114) = 8.60$, $MSE = 0.104$, $p = .004$, $\eta_p^2 = .070$ (see Figure 5). The simple effects test indicated a significant difference for worked examples, $F(1, 59) = 4.81$, $MSE = 0.08$, $p = .032$, $\eta_p^2 = .075$, where individual study again was superior to collaborative study. For problem solving, a significant effect again was found in favor of collaborative learning, $F(1, 55) = 3.96$, $MSE = 0.13$, $p = .05$, $\eta_p^2 = .067$.

A main effect of complexity was found, $F(1, 114) = 33.64$, $MSE = 0.03$, $p < .001$, $\eta_p^2 = .23$. Students performed significantly higher in the low-complexity transfer problems ($M = 0.31$, $SD = 0.27$) than in the high-complexity problems ($M = 0.18$, $SD = 0.23$). A significant interaction effect between the instructional strategy and problem complexity was also found, $F(1, 114) = 9.75$, $MSE = 0.03$, $p = .002$, $\eta_p^2 = .08$ (see Figure 6). The simple effects tests indicated that for the high-complexity transfer problems, worked examples led to a significantly higher performance than problem solving, $F(1, 116) = 7.96$, $MSE = 0.06$, $p = .006$, $\eta_p^2 = .064$, but for low-complexity transfer problems, there were no significant differences between the learning strategies ($F < 1$, ns.).

No overall support was found for Hypothesis 1 that students would benefit from studying collaboratively rather than individually when using worked examples. Instead, the reverse result was obtained, with individual study superior to collaborative study on both similar and transfer tests for high complexity problems. Interestingly, this superiority was associated with a higher cognitive load for individual study. Normally, a lower cognitive load is associated with improved performance. Future work will be re-

Table 2
Means (and Standard Deviations) for Test Results and Cognitive Load Ratings in Experiment 2

Condition	Cognitive load (1–9)		Similar test (0–1)		Transfer test (0–1)	
	Low-complexity	High-complexity	Low-complexity	High-complexity	Low-complexity	High-complexity
Worked examples						
Collaborative	3.16 (1.32)	4.06 (1.34)	.49 (.27)	.48 (.24)	.24 (.27)	.19 (.20)
Individual	3.07 (1.80)	4.90 (1.77)	.68 (.30)	.47 (.27)	.36 (.22)	.30 (.23)
Problem solving						
Collaborative	5.43(1.99)	6.68 (1.85)	.55 (.31)	.31 (.30)	.40 (.30)	.16 (.35)
Individual	6.21(2.60)	6.59 (1.86)	.31 (.32)	.16 (.23)	.22 (.29)	.07 (.13)

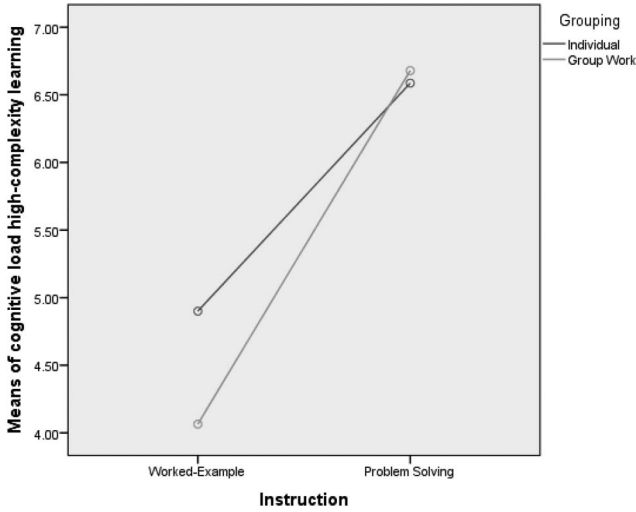


Figure 2. Interaction between instructional strategy and grouping context on cognitive load during learning high complexity material in the acquisition phase in Experiment 2.

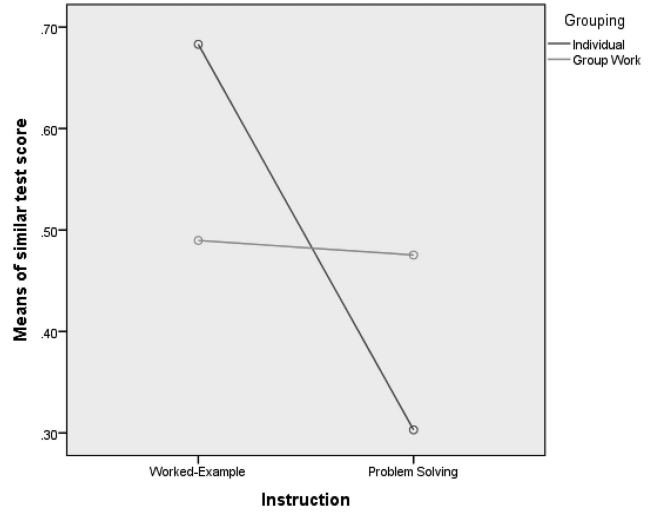


Figure 4. Interaction between instructional strategy and grouping context for similar test scores on low-complexity task in Experiment 2.

quired to establish whether this result is replicable. Nevertheless, based on interaction effects, there was some advantage for collaboration. Collaborative learning was superior to individual learning on similar and transfer tests when students initially used the problem solving strategy.

There was no support for Hypothesis 2 that the effectiveness of collaborative learning would be increased by task complexity. It was found that for the problem solving strategy, collaborative learners performed better than individual learners on similar tests for the low-complexity problems. For the worked example strategy, however, individual learners performed better than collaborative learners for low-complexity similar tests. No effects were obtained using high-complexity problems.

Hypothesis 3 predicted that students would benefit from studying worked examples rather than problem solving. An overall main effect was found in support of this hypothesis for the similar test problems. Although no main effect for worked examples was found on the transfer test phase, an interaction effect indicated that on high-complexity transfer problems, those who studied worked examples scored higher than those who initially solved problems. Furthermore, during the similar test phase, the cognitive load was found to be lower when studying worked examples rather than solving problems.

In summary, this experiment confirmed that the worked example strategy was superior to the problem solving strategy. When the material was higher in complexity, learning by worked examples in an individual setting was advantageous compared to col-

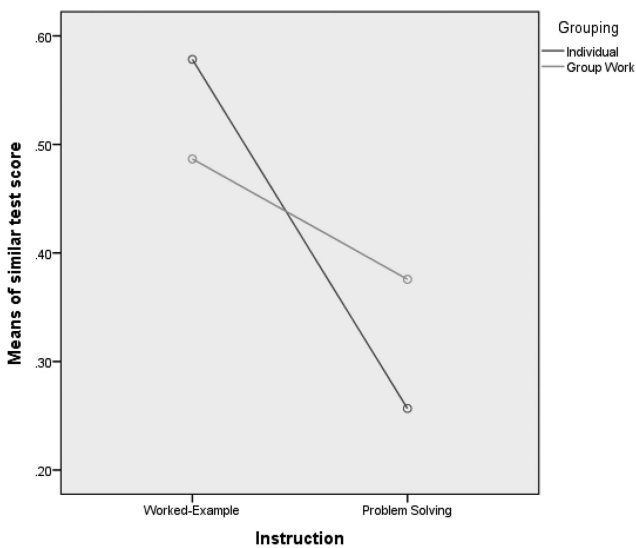


Figure 3. Interaction between instructional strategy and grouping context on similar test scores in Experiment 2.

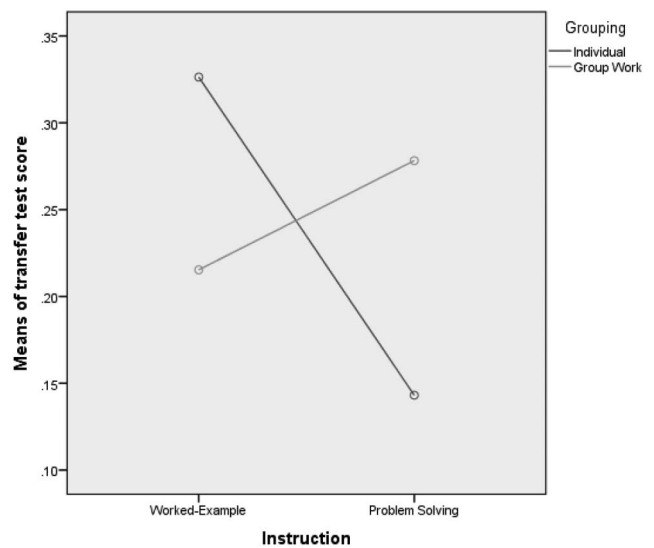


Figure 5. Interaction between instructional strategy and grouping context on transfer test scores in Experiment 2.

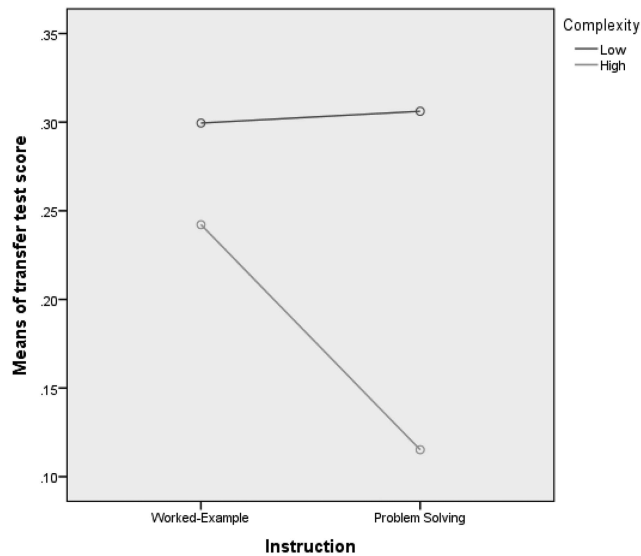


Figure 6. Interaction between instructional strategy and task complexity on transfer test scores in Experiment 2.

laborative learning. When the material was lower in complexity, some benefits were found for problem solving in groups.

General Discussion

Summary of Evidence in Support of the Hypotheses

Three hypotheses were tested over the two experiments, and the overall evidence is summarized below.

The first hypothesis predicted that *worked examples will be enhanced by studying collaboratively compared to studying individually*. This hypothesis was examined in both experiments and was not supported. When all students were tested individually on the similar and transfer tests, no superiority was found. In contrast, the results of Experiment 1 (transfer test performance in a worked example environment) showed that individual learners had a significant advantage over collaborative learners, while Experiment 2 indicated an advantage for individual study on both similar and transfer tasks for high complexity problems. Consequently, Hypothesis 1 was rejected.

The second hypothesis predicted that *the effectiveness of collaborative learning is increased by task complexity*. This hypothesis was examined in Experiments 1 and 2. As reported, these experiments used two tasks with different levels of complexity. It was predicted that there would be interactions between task complexity and the effectiveness of collaborative learning. Although interactions were found, follow-up tests indicated that collaborative learning was not superior to individual learning on the more complex tasks. In fact, in Experiment 1 (transfer test performance), evidence emerged that for the more complex tasks individual learning was superior to collaborative learning. Thus, Hypothesis 2 was rejected.

The final hypothesis predicted that *studying worked examples would be more advantageous than conventional problem solving*. This hypothesis was examined in Experiment 2. The results indi-

cated evidence in support of the effectiveness of a worked example strategy compared to a problem solving strategy (the worked example effect), as predicted. Worked examples were found to be more effective during the similar test phase. Furthermore, students using the worked example strategy experienced a lower cognitive load for the similar test phase. A significant interaction was also found. Students who originally studied worked examples had higher scores on the high complexity transfer problems than those who originally were asked to solve problems.

As summarized above, no evidence was found that when using worked examples, collaborative learning was significantly superior to individual learning. In contrast, some evidence emerged that individual contexts were superior. A perspective from evolutionary educational psychology (Geary, 1995, 2002) can be used to explain why collaborative learning was rarely superior to individual learning. Geary argues that in social interactions, students develop their biologically primary knowledge rather than the assigned biologically secondary knowledge. In other words, they become more adept at their social interaction, which is an evolutionary primary skill, rather than the assigned mathematics task, which is an evolutionary secondary skill, and requires considerable conscious effort to learn. It was expected that group interactions would have generated superior sense making and reorganization of the information provided. However, there was no evidence for this suggestion. It is possible that worked examples provide sufficient information, rendering collaboration unnecessary.

The study also examined if the effectiveness of collaborative learning was important when dealing with complex tasks. High-complexity problems were argued to increase active social interaction during collaborative learning. As Hypothesis 2 was rejected, it can be concluded that for the complexity levels used, neither low- nor high-complexity tasks improved collaborative learning compared with individual learning. In fact, it was found on several occasions that for high-complexity tasks, individual learning led to higher performance than collaborative learning.

Moreover, it was found that collaborative learning only had an advantage over individual learning during problem solving. Collaborative learners scored higher than individual learners after having acquired their initial knowledge through problem solving. It is also notable that the collaborative advantage occurred only on low-complexity materials. It is possible that the low-complexity problem solving imposed a lower cognitive load and thus could be managed in a collaborative learning setting.

The study also tested for a worked example effect. The evidence obtained in this study is consistent with cognitive load theory research, demonstrating that overall the worked example strategy was superior to a problem solving strategy. Worked examples in general can be used in individual or collaborative learning contexts, replicating a previous finding (Retnowati et al., 2010). It is important to note, however, that the various interactions identified in this study indicated that the worked example strategy was best used in individual rather than collaborative settings, particularly for high-complexity problems.

Collaborative learning creates conditions where students in a group are expected to discuss the learning material, which can be done by giving/receiving elaborated explanations (Cohen, 1994; Webb, 1991, 2009). However, worked examples contain step-by-step explanations to reach a problem solution, so discussing worked examples may have a redundant element (Chandler &

Sweller, 1991). Worked examples are unnecessary if members of the group can “borrow” (using the cognitive load theory borrowing and reorganizing principle) the information required to learn or solve the given problem from the other group members. Similarly, group interactions may well help enhance the reorganization of new information.

While the current results were theoretically coherent and largely consistent, they will require replication in different contexts using different populations and materials. We have established that at least under some circumstances, collaboration when studying worked examples has negative rather than positive effects, while collaboration using problem solving can have positive effects. As far as we are aware, this finding is novel. We have interpreted these findings in terms of redundancy (Nihalani et al., 2011). Learners studying worked examples do not need additional information from collaborators to assist them when studying. Such additional, redundant information may have negative rather than positive effects, leading to an expertise reversal effect (Kalyuga et al., 2003). In contrast, when problem solving in the absence of worked examples, information from collaborators may be beneficial. Whether collaboration when studying worked examples is advantageous under different circumstances requires additional data. For example, exceptionally complex worked examples may benefit from a collaborative approach.

One potential limitation of the study occurred because we wanted to examine an authentic learning environment, and therefore some recommended steps, such as group processing training to prepare for effective collaboration, were not followed (see Johnson & Johnson, 1994). It is feasible that the group processes conducted by this sample were not sufficient to optimize the impact of collaboration. Nevertheless, the groups had been working together in mathematics classes for 3 (Experiment 1) and 5 (Experiment 2) months and so had experience learning in their groups. Furthermore, the finding that collaboration was superior to individual study for the problem-solving strategy suggests that there were benefits, and therefore a certain amount of effective collaborative behavior can be assumed to have been present. To collect additional data on this issue was outside the scope of the present study, but is a topic for further investigation. Furthermore, replicating this study with collaborative groups further prepared according to the steps often recommended for effective collaboration should be also be informative.

Many effective collaboration tasks consist of realistic ill-structured problems (Hmelo-Silver, 2004), have high complexity (F. Kirschner et al., 2009a), or cannot be completed by individuals (Cohen, 1994). In contrast, the tasks chosen for this study (middle school equation solving) did not contain all these characteristics. Nevertheless, we did test for the effect of complexity, and significant differences were found between the two complexity levels. Future studies could include richer problem solving tasks with more ill-defined goals as recommended. Also, delayed tests could be included in future studies to judge the permanence of learning, although it should be noted that worked examples have been found to provide robust learning longevity (see Chen et al., 2016a).

With respect to educational implications, our main research question was Can collaborative learning improve the effectiveness of worked examples? Under the given conditions, the answer is no. Worked examples seem to be most effective in individual settings. Asking learners to discuss worked examples may be redundant be-

cause they have already obtained the necessary information from an instructor via the worked example. Regarding problem complexity, individual study seemed to be most appropriate for the complex problems, although collaboration was helpful when problem solving (the inferior strategy), presumably because collaboration permitted learners to obtain missing information from other learners. Hence, collaboration may be advantageous when problem solving because to some extent it is able to provide learners with missing guidance.

In conclusion, there appear to be limits to the conditions under which collaborative learning is effective. Those limits should be considered when encouraging learners to study collaboratively.

References

- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*, 181–214. <http://dx.doi.org/10.3102/00346543070002181>
- Ayres, P., & Sweller, J. (2013). The worked example effect. In J. A. C. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 408–410). Oxford, England: Routledge.
- Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences*, *9*, 403–436. http://dx.doi.org/10.1207/S15327809JLS0904_2
- BNSP. (2006). *Panduan penyusunan kurikulum tingkat satuan pendidikan jenjang pendidikan dasar dan menengah* [Guideline for developing curriculum for elementary and middle education]. Jakarta, Indonesia: Badan Nasional Standar Pendidikan.
- Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology*, *86*, 360–367. <http://dx.doi.org/10.1037/0022-0663.86.3.360>
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*, 293–332. http://dx.doi.org/10.1207/s1532690xci0804_2
- Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology*, *107*, 689–704. <http://dx.doi.org/10.1037/edu0000018>
- Chen, O., Kalyuga, S., & Sweller, J. (2016a). Relations between the worked example and generation effects on immediate and delayed tests. *Learning and Instruction*, *45*, 20–30. <http://dx.doi.org/10.1016/j.learninstruc.2016.06.007>
- Chen, O., Kalyuga, S., & Sweller, J. (2016b). The expertise reversal effect is a variant of the more general element interactivity effect. *Educational Psychology Review*. Advance online publication. <http://dx.doi.org/10.1007/s10648-016-9359-1>
- Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, *64*, 1–35. <http://dx.doi.org/10.3102/00346543064001001>
- Cooper, G., & Sweller, J. (1987). The effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, *79*, 347–362. <http://dx.doi.org/10.1037/0022-0663.79.4.347>
- Davidson, N., & Kroll, D. L. (1991). An overview of research on cooperative learning related to mathematics. *Journal for Research in Mathematics Education*, *22*, 362–365. <http://dx.doi.org/10.2307/749185>
- De Corte, E. (2004). Mainstreams and perspectives in research on learning (mathematics) from instruction. *Applied Psychology*, *53*, 279–310. <http://dx.doi.org/10.1111/j.1464-0597.2004.00172.x>
- Depdiknas. (2004). *Kurikulum 2004 untuk sekolah menengah pertama dan madrasah tsanawiyah* [2004 Curriculum for junior high school and Islamic junior high school]. Jakarta, Indonesia: Departemen Pendidikan Nasional.
- Geary, D. C. (1995). Reflections of evolution and culture in children’s cognition. Implications for mathematical development and instruction.

- American Psychologist*, 50, 24–37. <http://dx.doi.org/10.1037/0003-066X.50.1.24>
- Geary, D. C. (2002). Principles of evolutionary educational psychology. *Learning and Individual Differences*, 12, 317–345. [http://dx.doi.org/10.1016/S1041-6080\(02\)00046-8](http://dx.doi.org/10.1016/S1041-6080(02)00046-8)
- Geary, D. C. (2008). An evolutionarily informed education science. *Educational Psychologist*, 43, 179–195. <http://dx.doi.org/10.1080/00461520802392133>
- Geary, D. (2012). Evolutionary educational psychology. In K. Harris, S. Graham, & T. Urdan (Eds.), *APA Educational Psychology Handbook* (Vol. 1, pp. 597–621). Washington, DC: American Psychological Association.
- Gillies, R. M. (2003). Structuring cooperative group work in classrooms. *International Journal of Educational Research*, 39, 35–49. [http://dx.doi.org/10.1016/S0883-0355\(03\)00072-7](http://dx.doi.org/10.1016/S0883-0355(03)00072-7)
- Goos, M. (2004). Learning mathematics in a classroom community of inquiry. *Journal for Research in Mathematics Education*, 35, 258–291. <http://dx.doi.org/10.2307/30034810>
- Hanham, J., & McCormick, J. (2009). Group work in schools with close friends and acquaintances: Linking self-processes with group processes. *Learning and Instruction*, 19, 214–227. <http://dx.doi.org/10.1016/j.learninstruc.2008.04.002>
- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16, 235–266.
- Johnson, D. W., & Johnson, R. T. (1994). *Learning together and alone: Cooperative, competitive and individualistic learning*. Boston, MA: Allyn & Bacon.
- Johnson, D. W., & Johnson, R. T. (2002). Learning together and alone: Overview and meta-analysis. *Asia Pacific Journal of Education*, 22, 95–105. <http://dx.doi.org/10.1080/02188790220220110>
- Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college: What evidence is there that it works? *Change: The Magazine of Higher Learning*, 30, 26–35. <http://dx.doi.org/10.1080/00091389809602629>
- Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). The effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, 89, 47–62. <http://dx.doi.org/10.1037/0033-2909.89.1.47>
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38, 23–31. http://dx.doi.org/10.1207/S15326985EP3801_4
- Kirschner, F., Paas, F., & Kirschner, P. A. (2009a). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review*, 21, 31–42. <http://dx.doi.org/10.1007/s10648-008-9095-2>
- Kirschner, F., Paas, F., & Kirschner, P. A. (2009b). Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. *Computers in Human Behavior*, 25, 306–314. <http://dx.doi.org/10.1016/j.chb.2008.12.008>
- Kirschner, F., Paas, F., & Kirschner, P. A. (2011). Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Applied Cognitive Psychology*, 25, 615–624. <http://dx.doi.org/10.1002/acp.1730>
- Kirschner, F., Paas, F., Kirschner, P. A., & Janssen, J. (2011). Differential effects of problem-solving demands on individual and collaborative learning outcomes. *Learning and Instruction*, 21, 587–599.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41, 75–86. http://dx.doi.org/10.1207/s15326985ep4102_1
- Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, 19, 335–353. [http://dx.doi.org/10.1016/S0747-5632\(02\)00057-2](http://dx.doi.org/10.1016/S0747-5632(02)00057-2)
- Kyun, S., Kalyuga, S., & Sweller, J. (2013). The effect of worked examples when learning to write essays in English literature. *Journal of Experimental Education*, 81, 385–408. <http://dx.doi.org/10.1080/00220973.2012.727884>
- Mayer, R. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). New York, NY: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139547369.005>
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- National Ministry of Education. (2006). *Peraturan Menteri Pendidikan Nasional Republik Indonesia No. 22 Tahun 2006* (Regulation of the Minister of National Education of the Republic of Indonesia No. 22 of 2006).
- Nihalani, P., Mayrath, M., & Robinson, D. (2011). When feedback harms and collaboration helps in computer simulation environments: An expertise reversal effect. *Journal of Educational Psychology*, 103, 776–785. <http://dx.doi.org/10.1037/a0025276>
- Oksa, A., Kalyuga, S., & Chandler, P. (2010). Expertise reversal effect in using explanatory notes for readers of Shakespearean text. *Instructional Science*, 38, 217–236. <http://dx.doi.org/10.1007/s11251-009-9109-6>
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology*, 84, 429–434. <http://dx.doi.org/10.1037/0022-0663.84.4.429>
- Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review*, 24, 27–45. <http://dx.doi.org/10.1007/s10648-011-9179-2>
- Paas, F., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, 86, 122–133. <http://dx.doi.org/10.1037/0022-0663.86.1.122>
- Plass, J. L., O’Keefe, P. A., Homer, B. D., Case, J., Hayward, E. O., Stein, M., & Perlin, K. (2013). The impact of individual, competitive, and collaborative mathematics game play on learning, performance, and motivation. *Journal of Educational Psychology*, 105, 1050–1066. <http://dx.doi.org/10.1037/a0032688>
- Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, 88, 144–161. <http://dx.doi.org/10.1037/0022-0663.88.1.144>
- Renkl, A. (2014a). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38, 1–37. <http://dx.doi.org/10.1111/cogs.12086>
- Renkl, A. (2014b). The worked examples principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 391–412). New York, NY: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139547369.020>
- Retnowati, E., Ayres, P., & Sweller, J. (2010). Worked example effects in individual and group work settings. *Educational Psychology*, 30, 349–367. <http://dx.doi.org/10.1080/01443411003659960>
- Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognise designers’ styles. *Learning and Instruction*, 19, 185–199. <http://dx.doi.org/10.1016/j.learninstruc.2008.03.006>
- Schreiber, L. M., & Valle, B. E. (2013). Social constructivist teaching strategies in the small group classroom. *Small Group Research*, 44, 395–411. <http://dx.doi.org/10.1177/1046496413488422>
- Slavin, R. E. (1995). *Cooperative learning: Theory, research, and practice* (2nd ed.). Boston, MA: Allyn & Bacon.
- Staples, M. (2007). Supporting whole-class collaborative inquiry in a secondary mathematics classroom. *Cognition and Instruction*, 25, 161–217. <http://dx.doi.org/10.1080/07370000701301125>

- Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review*, 22, 123–138. <http://dx.doi.org/10.1007/s10648-010-9128-5>
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer. <http://dx.doi.org/10.1007/978-1-4419-8126-4>
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12, 185–233. http://dx.doi.org/10.1207/s1532690xci1203_1
- Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, 2, 59–89. http://dx.doi.org/10.1207/s1532690xci0201_3
- Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology*, 4, 434–458. <http://dx.doi.org/10.1177/147470490600400135>
- Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology*, 80, 424–436. <http://dx.doi.org/10.1037/0022-0663.80.4.424>
- Trafton, J. G., & Reiser, B. J. (1993). *The contributions of studying examples and solving problems to skill acquisition*. Paper presented at the Fifteenth Annual Conference of the Cognitive Science Society, Hillsdale, NJ.
- Van den Bossche, P., Gijsselaers, W. H., Segers, M., & Kirschner, P. A. (2006). Social and cognitive factors driving teamwork in collaborative learning environments: Team learning beliefs and behaviors. *Small Group Research*, 37, 490–521. <http://dx.doi.org/10.1177/1046496406292938>
- van Gog, T., & Kester, L. (2012). A test of the testing effect: Acquiring problem-solving skills from worked examples. *Cognitive Science*, 36, 1532–1541. <http://dx.doi.org/10.1111/cogs.12002>
- van Gog, T., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and frequency of mental effort measurement: Evidence in favour of repeated measures. *Applied Cognitive Psychology*, 26, 833–839. <http://dx.doi.org/10.1002/acp.2883>
- van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, 43, 16–26. <http://dx.doi.org/10.1080/00461520701756248>
- Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, 7, 1–39. http://dx.doi.org/10.1207/s1532690xci0701_1
- Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education*, 22, 366–389. <http://dx.doi.org/10.2307/749186>
- Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology*, 79, 1–28. <http://dx.doi.org/10.1348/000709908X380772>
- Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concept and assessment. *Learning and Instruction*, 17, 416–426. <http://dx.doi.org/10.1016/j.learninstruc.2007.03.007>
- Yackel, E., Cobb, P., & Wood, T. (1991). Small-group interactions as a source of learning opportunities in second-grade mathematics. *Journal for Research in Mathematics Education*, 22, 390–408. <http://dx.doi.org/10.2307/749187>
- Zhang, L., Ayres, P., & Chan, K. (2011). Examining different types of collaborative learning in a complex computer-based environment: A cognitive load approach. *Computers in Human Behavior*, 27, 94–98. <http://dx.doi.org/10.1016/j.chb.2010.03.038>

(Appendices follow)

