When Instructional Guidance is Needed

Ouhao Chen,[1] Slava Kalyuga[2] and John Sweller[2]

[1] *School of Education, Southern Cross University, Australia*
[2] *School of Education, University of New South Wales, Australia*

Studying worked examples providing problem solutions to learners usually leads to better test performance than solving the equivalent problems without guidance, demonstrating the worked-example effect. The generation effect occurs when learners who generate answers without guidance learn better than those who read answers that provide guidance. The contradiction between these results can be hypothesised to be due to differences in the element interactivity of the learning tasks. Primary school students in Year 6 participated in the experiment, which investigated the hypothesis by using geometry materials. A disordinal interaction was obtained between levels of guidance and levels of element interactivity. Higher levels of guidance facilitated learning using high element interactivity information, while lower levels of guidance facilitated learning for low element interactivity information. Cognitive load theory was used to explain these contrasting results. From an educational perspective, it was suggested that when determining levels of guidance, a consideration of element interactivity is essential.

■ **Keywords:** cognitive load theory, guidance, worked-example effect, generation effect, element interactivity

Over the past 50 years, there has been considerable debate concerning the influence of instructional guidance during teaching (Ausubel, 1964; Kirschner, Sweller, & Clark, 2006; Mayer, 2004). Contradictory evidence has been provided supporting a heavy emphasis on instructional guidance as opposed to requiring students to generate activity with much more limited guidance. In this article, we suggest a possible resolution of these contradictory findings.

A considerable number of studies have indicated that worked examples providing full problem-solution guidance to novice learners result in superior performance on subsequent tests when compared to learners solving problems with no guidance, demonstrating the worked-example effect (Renkl, 2014a, 2014b; Sweller, Ayres, & Kalyuga, 2011). The generation effect, in contrast, occurs when learners who are asked to self-generate answers rather than study externally provided answers indicate better test performance (Slamecka & Graf, 1978). The experiment reported in this article was designed to test the hypothesis that these contradictory results may be explained by different levels of element interactivity, a central concept of cognitive

load theory. We will begin by discussing the cognitive architecture that underlies cognitive load theory.

Human Cognitive Architecture

Cognitive load theory (Sweller et al., 2011) is an instructional theory based on our knowledge of human cognitive architecture. Five principles provide a structure for human cognitive architecture.

The *information store principle* indicates that human cognitive architecture requires a very large knowledge base. Long-term memory, with its apparently unlimited capacity for storing information, provides that knowledge base.

Most of the knowledge stored in long-term memory is obtained via the *borrowing and reorganising principle*. Learners obtain knowledge by observing other people's actions, listening to what they say, and reading what they write. The borrowed knowledge is reorganised by combining it with previous knowledge stored in long-term memory.

Based on the *randomness as genesis principle*, if required knowledge cannot be borrowed from others, learners can randomly generate knowledge and test it for effectiveness during problem solving. Effective knowledge can be retained in long-term memory while ineffective knowledge is jettisoned.

The amount of randomly generated knowledge needs to be limited to preserve the functionality of the information stored in long-term memory. The *narrow limits of change principle* ensures that changes to long-term memory are small and incremental, with working memory having a very limited capacity and duration. The *environmental organising and linking principle* justifies the preceding principles. It ensures that action taken is appropriate for the extant environment. While working memory is limited when dealing with novel information, it has much wider, unknown limits when dealing with familiar information taken from long-term memory (Ericsson & Kintsch, 1995). Large amounts of stored knowledge can be retrieved from long-term memory by working memory when activated by stimuli from the external environment.

Element Interactivity and Types of Cognitive Load

Element interactivity is a construct intended to reflect the degree of complexity of learning tasks for individuals with a given level of knowledge. Interactive elements are defined as elements that must be processed in working memory simultaneously (Sweller et al., 2011). An element that should be processed in working memory can be a symbol or a concept, and it is characteristically a schema — a generic knowledge structure that allows us to categorise information. The number of interacting elements in learning tasks influences the levels of element interactivity (Sweller & Chandler, 1994; Tindall-Ford, Chandler, & Sweller, 1997). In recent versions of cognitive load theory (Kalyuga, 2011), there are two types of cognitive load: intrinsic and extraneous. The concept of element interactivity can be used to define the two types of load.

Element Interactivity and Intrinsic Cognitive Load

Intrinsic load is determined by an interaction between the nature of information that learners must deal with and the expertise of the learner (Sweller, 1994; Van Merriënboer, Kester, & Paas, 2006). High levels of element interactivity indicate high levels of intrinsic load, while information that is low in element interactivity has a low

intrinsic load. For instance, if foreign language students learn the translation of an English vocabulary item such as *apple,* they do not need to simultaneously consider the translation of the word *cat* or any other word in order to remember the word *apple.* Learning the vocabulary of a foreign language is a very difficult task but is low in element interactivity, as each vocabulary element can be considered without reference to other vocabulary elements. However, if Year 4 students are asked to simplify (a+b)(a-b), they may need to keep each symbol of this expression in working memory, including 11 elements, and then distribute the a and b in the left bracket into the right bracket in order to open the two brackets, implicating four elements; next, they need to do final calculations, leading to the final answer a^2-b^2, including another element. Therefore, the students need to simultaneously process about 16 interactive elements in working memory that has limited capacity (Cowan, 2001; Miller, 1956), rendering this material relatively high in element interactivity. The task is difficult, not because there are many elements that need to be processed but rather because there are many elements that need to be processed simultaneously due to their interactivity.

As mentioned above, the level of element interactivity is also determined by the level of learner expertise. For novices, the algebra question used above to explain a task that is high in element interactivity includes many more interactive elements than the same task presented to an expert. An expert who has acquired a relevant schema to solve such problems can retrieve the schema as a single entity from long-term memory, reducing the interacting element count to one. With increases in expertise, multiple elements can be incorporated into a single element using the environmental organising and linking principle, thus reducing the level of intrinsic load. What a learner treats as a single element is determined by both the nature of the information and levels of learner expertise.

Element Interactivity and Extraneous Cognitive Load

Extraneous cognitive load is unnecessary for learning and can be altered by instructional interventions. It is influenced by the way instructional tasks are presented (Sweller, Van Merrienboer, & Paas, 1998). Extraneous load can be imposed by sub-optimal teaching methods, such as those requiring mental integration of separately presented sources of information, or by using an instructionally ineffective general problem-solving method.

Extraneous load can also be described in terms of element interactivity. Sweller (2010) suggested that element interactivity was the major source of working memory load underlying extraneous as well as intrinsic cognitive load. If the number of interacting elements alters due to instructional procedures, then the change in element interactivity results in a change in extraneous load. For example, making use of the borrowing and organising principle by presenting students with well-organised solutions as worked examples will lead to a lower extraneous load compared to requiring students to self-generate solutions via the randomness as genesis principle, as randomly generating knowledge will occupy more working memory resources, causing a higher extraneous load. Students searching for appropriate moves to generate a solution to the above algebra problem will need to consider many more elements than students shown a solution to the problem.

The Worked-Example Effect

Worked examples provide a solution to a problem for a learner to study (Atkinson, Derry, Renkl, & Wortham, 2000). The typical components of a worked example include a problem statement and procedures for solving this problem. Within cognitive load theory, the traditional paradigm used to test the worked-example effect is to use worked example problem-solving pairs compared to problem-solving only pairs. A body of research has indicated positive effects of studying worked examples on solving subsequent test problems, compared to solving the equivalent problems instead. Worked examples were successfully used in many academic domains, including algebra (Cooper & Sweller, 1987; Sweller & Cooper, 1985), statistics (Paas, 1992), geometry (Paas & Van Merriënboer, 1994; Schwonke et al., 2009), physics (Reisslein, Atkinson, Seeling, & Reisslein, 2006; Van Gog, Kester, & Paas, 2011; Van Gog, Paas, & van Merriënboer, 2006), identifying artistic styles (Rourke & Sweller, 2009), and interpreting English literature (Kyun, Kalyuga, & Sweller, 2013).

Most, possibly all, demonstrations of the worked-example effect used high element interactivity material that was difficult to understand by the participants. Learning tasks used for testing the worked-example effect are assumed to be high in element interactivity, at least for novice learners. A worked example provides solutions to novices under the borrowing and reorganising principle, which avoids randomly searching for solutions under the randomness as genesis principle with its increase in extraneous load. However, as discussed above, element interactivity is also dependent on the levels of a learner's expertise. With an increase in learners' levels of expertise, individual elements become integrated into more sophisticated schemas that reduce the number of interactive elements, resulting in decreased complexity of the corresponding tasks for more experienced learners. Therefore, the worked-example effect that requires complex tasks may be eliminated or even reversed, based on the effectively reduced element interactivity (Sweller et al., 2011). The reversed worked-example effect in the case of simple tasks may in fact become a form of the generation effect discussed in the next section.

The Generation Effect

The generation effect occurs when generated items are better memorised than items that are explicitly provided to learners (Slamecka & Graf, 1978). For example, instead of learners being presented with paired associates (e,g., 'tall-short') prior to a memory test, they may be provided with the word 'tall' and asked to generate its opposite. The generation effect occurs when learners who generate items in this way remember those items better than learners who are presented the same items explicitly. Many research studies have indicated the generation effect with a wide variety of tasks, such as word pairs (Slamecka & Graf, 1978), meaningful sentences (McFarland, Frey, & Rhodes, 1980), or pictures (Peynircioğlu, 1989). In most cases, the generation effect has been obtained with memorisation tasks using simple, easily understood information.

The tasks used in the demonstrations of the generation effect are usually low in element interactivity. For example, the word pair task mentioned above just requires learners to retrieve a single element, 'short', directly from long-term memory, indicating that the task has a low level of intrinsic load. When pictures were used as materials in tasks to test for the generation effect, considering that the content of the pictures usually represented familiar common objects (e.g., a table) for which the participants

already had well-established relevant knowledge, the task was also simple and low in element interactivity. A similar situation applied to studies that demonstrated the generation effect with sentences as materials. Therefore, based on its outcomes and its conditions of applicability, the generation effect is similar to the reversed worked-example effect that also occurs with low element interactivity tasks, usually due to high levels of learner expertise.

There are several explanations of the generation effect (Donaldson & Bass, 1980; Griffith, 1976; McFarland et al., 1980). Donaldson and Bass (1980) used relational theory according to which the relations between stimulus and answer were strengthened by generation, resulting in enhanced memory compared to presentation. Schweickert, McDaniel, and Riegler (1994) also indicated that the process of generation itself could improve memory and that the effect of generation resulted in a more durable memory. Alternatively, Griffith (1976) and McFarland et al. (1980) suggested that more cognitive effort was contributed to the task when generating responses resulting in stronger memory traces. Conceivably, under some formulations, this suggestion corresponds to the cognitive load theory concept of germane cognitive load.

It should be noted that the generation effect is closely related to the testing effect (Karpicke & Zaromb, 2010). The testing effect occurs when learners are given a test of previously learned information rather than being re-presented that information, very similar to the paradigm for the generation effect in which learners generate information (i.e., are tested for knowledge) rather than being presented the same information. Accordingly, there is evidence that the testing effect is heavily reliant on levels of element interactivity, with the effect reducing or disappearing when element interactivity is high (Leahy, Hanham, & Sweller, 2015; Van Gog & Sweller, 2015).

An Interaction Between Element Interactivity and Levels of Guidance

Based on the above discussion of the role of element interactivity and external guidance in occurrences of the worked-example and generation effects, it is possible to hypothesise an interaction between levels of guidance and element interactivity (Chen, Kalyuga, & Sweller, 2015, 2016a). We can hypothesise that worked examples providing comprehensive, high levels of guidance in solving high element interactivity problems will result in better performance than low levels of guidance associated with problem-solving practice that requires students to generate answers by themselves. On the other hand, using low element interactivity tasks, we also can hypothesise that generation procedures with their low levels of guidance will be superior to the use of high levels of guidance in which students directly read externally provided answers.

The current experiment was designed to test this hypothesis. A 2 (guidance: low vs. high) × 2 (element interactivity of tasks: low vs. high) mixed factorial experimental design was used in the domain of geometry. Tasks that are high in element interactivity were used to test the worked-example effect, whereas tasks that are low in element interactivity were used to test the generation effect.

Method

Participants and Experimental Design

The participants were 50 approximately 12-year-old Year 6 students from a primary school in Chengdu, China. The school accepts students from the surrounding, working-class areas. In the first phase of the experiment, low element interactivity
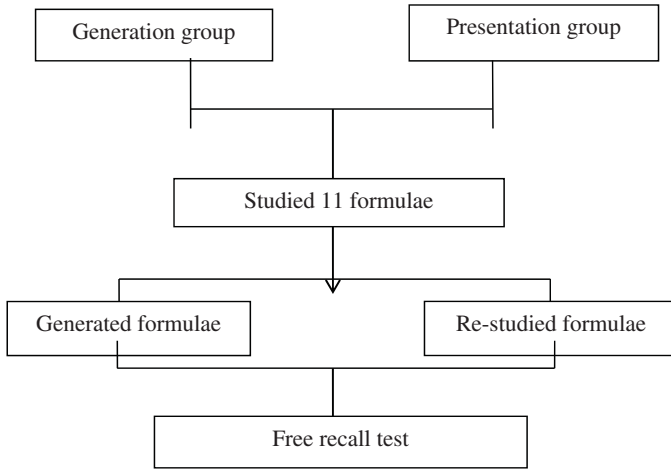
FIGURE 1

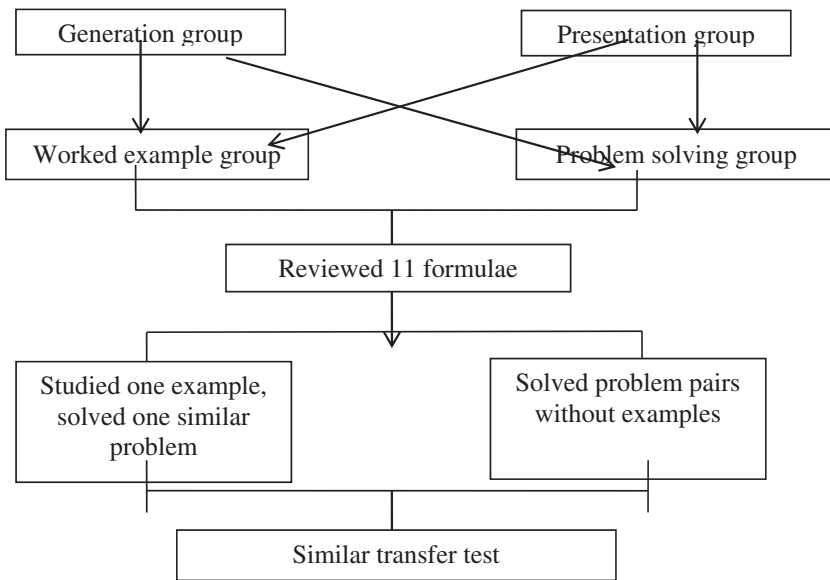Phase 1: Procedure used to test for the generation effect.

FIGURE 2

Phase 2: Procedure used to test for the worked-example effect.

tasks were used to test for the generation effect. Participants were randomly assigned to either the generation or presentation group in that first phase (see Figure 1). The second phase involved high element interactivity tasks to test for the worked-example effect. In that second phase, half of the students from the generation group of the first phase were randomly assigned to a problem-solving group and the other half to a worked-example group (see Figure 2). Similarly, half of the students from the presentation group of the first phase were allocated randomly into either the worked-example

TABLE 1

Summary of the Materials used to Test for the Generation Effect

Booklet	Content
Booklet 1	11 geometry formulae (including 2D and 3D geometric shapes)
Booklet 2	Instructions for the generation group: generate each of the formulae named below;
	For the presentation group: re-study the 11 formulae
Booklet 3	Free-recall test

TABLE 2

Summary of the Materials used to Test for the Worked-Example Effect

Booklet	Content
Booklet 1	11 geometry formulae (including 2D and 3D geometric shapes) for review
Booklet 2	For the worked-example group: two worked-example, problem-solving pairs;
	For the problem-solving group: two problem-solving, problem-solving pairs
Booklet 3	Transfer test

or problem-solving group in the second phase. Therefore, everyone participated in the low element interactivity tasks before completing the high element interactivity tasks. Three students did not complete the entire procedure of the first phase and were eliminated from the data analyses, leaving 47 students. Prior to the experiment, in their regular geometry classes, all students had studied most of the area and perimeter formulae used in this study to test for the generation effect and had received initial instruction on how to solve the problems used to test for the worked-example effect. Therefore, the information provided to these students included some revision.

Materials

Three booklets were used to test for the generation effect (see Table 1). The first booklet contained 11 geometry formulae chosen from textbooks used in primary and secondary schools to teach geometry. There were three surface area formulae (for 3D geometric shapes), four perimeter formulae, and four area formulae (for 2D geometric shapes; see Appendix A). All of the formulae were presented in algebraic format. The second booklet was different for the generation and the presentation groups. For the generation group, students were required to generate the formula based on the name of the formula used to describe the formula, whereas the presentation group was required to restudy the 11 formulae presented in the same manner as in the first booklet. The third booklet was blank for all students to write out within a limited time the formulae that they had studied in the previous two booklets.

Another three booklets were developed to test for the worked-example effect (see Table 2). The first booklet was the same as the one used to test for the generation effect. The purpose of this booklet was to help students review the formulae needed to solve the following problems. The second booklet was used to provide differential instruction for students in the worked example and problem-solving groups. For the worked-example group, the first and the third of four questions about calculating the area of composite shapes were presented as worked examples. The second and the

fourth questions were similar to the first and the third questions respectively, but no answers were provided. Students needed to solve the second and fourth questions themselves. For the problem-solving group, all four questions were provided as problems for students to generate solutions, with no worked examples provided. The last booklet was a test paper with five questions that were similar to the questions in the second booklet. Appendices B and C indicate the information in the worked-example booklet and the test questions, respectively.

Evaluating levels of element interactivity. When counting the number of interacting elements to evaluate the level of element interactivity, learner expertise always must be taken into account. When memorising a formula, each symbol that is familiar to the participants can be learned independently of every other symbol, and so the elements do not interact, resulting in an element interactivity count of 1. When calculating the area of a given composite shape, Year 6 students are likely to have already acquired relevant knowledge allowing them to perceive and calculate the areas of a rhombus and a trapezium as single units to be processed in working memory. They are unlikely to need to consider the basic lines that form a rhombus or trapezium, as these individual elements have already been integrated into previously acquired schemas that can be regarded as a single entity when calculating the area of a rhombus or trapezium. But Year 6 students are unlikely to be ready to retrieve the formulae for the areas of a rhombus or a trapezium together as a single entity, nor are they likely to readily recognise a rhombus or trapezium within a complex shape consisting of a combination of both geometric shapes. In order to solve a question concerning a diagram consisting of a rhombus and trapezium combined, the more complex schema required is unlikely to have been acquired by these Year 6 students. When faced with a figure consisting of a rhombus and a trapezium, we can estimate that they must recognise the rhombus within the complex figure (1st element), recognise the trapezium (2nd element) and realise that the total area of the shape (3rd element) requires the addition (4th element) of both areas (5th and 6th elements). Therefore, there were approximately six interactive elements for tasks used to test the worked-example effect, whereas the number of interactive elements was for tasks used to test the generation effect.

Procedure

The test for the generation effect phase of the experiment lasted for one class period of 35 minutes (see Figure 1). Prior to studying the first booklet, students were reseated according to the group into which they were randomly placed (7 minutes).

The study stage. After being reseated, students began studying the 11 formulae in the first booklet. They could make notes in this booklet if they needed to. After 10 minutes, all students handed in the booklet.

The generation or presentation stage. The experimenter distributed the second booklet to students in the generation and presentation groups separately. For the generation group, students generated the 11 formulae according to the cues provided (the name of the formula), whereas students in the presentation group were required to again study the 11 formulae included in the first booklet. No one could hand in this booklet before 10 minutes had elapsed. Any students who completed their task in less than 10

minutes were told to review the material again. After 10 minutes, all students handed in this booklet.

The free recall test stage. A blank paper was given to students to write out as many of the formulae that they had studied in the previous two booklets as they could. Students could only hand in their test booklet after 8 minutes had elapsed. Therefore, if students finished early, they were required to review their answers. When scoring the test, a correct formula was awarded one mark. Therefore, the maximum score in the free recall test was 11. Each student's score out of 11 was converted into a percentage score testing for the effects of guidance on low element interactivity material.

The worked-example effect phase of the experiment also lasted for one class period of 35 minutes that occurred 4 hours later on the same day after the generation effect phase (see Figure 2). Students were reseated according to the group to which they had been allocated (7 minutes).

The study stage. The procedure for this stage lasted for 10 minutes and was identical to that used for the equivalent stage in the generation effect phase of the experiment.

The worked-example or problem-solving stage. The general procedure was identical to that used in the generation effect phase. Students in the worked-example group studied the first and the third questions as examples of how to calculate the area of composite shapes before solving the second and the fourth questions, which were similar to the first and the third questions; namely, the first question was a worked example followed by a similar problem (question 2), with the same pattern for questions 3 and 4. For the problem-solving group, the students were required to solve the same four questions used in the worked-example group without any worked examples provided. Students could only submit this booklet after 10 minutes had passed. If they finished sooner, they were required to review the booklet.

The test stage. Again, the general procedure was identical to that used in the generation effect phase. Five questions that were similar to the questions used in the second booklet were designed as the test questions (see Appendix C). Eight minutes were allocated for the test. Students could obtain a maximum of 4 marks for each of the problems (1 mark for constructing the line to divide the composite shape into two basic geometry shapes; 1 for correctly using the area formula of one of the basic geometry shapes; 1 for correctly using the area formula of the other basic geometry shape; 1 for adding the two areas). Each student's total score out of 20 (five problems each with a maximum score of 4) was converted to a percentage score testing for the effects of guidance on high element interactivity material.

Results

Means and standard deviations of percentage test scores may be found in Table 3. These results were analysed using a 2 (levels of guidance) × 2 (levels of element interactivity) ANOVA with repeated measures on the second factor. It needs to be noted that because the two tasks differed, the main effect of levels of element interactivity had no theoretical or practical interest and is included purely in order to allow a test of a possible interaction effect between the two factors. An advantage of either high or

TABLE 3
Mean (*SD*) Percentage Correct Test Score Results

Guidance	Low element interactivity	High element interactivity
High (*N* = 22)	71.4 (20.15)	90.9 (10.08)
Low (*N* = 25)	80.9 (13.82)	85.6 (16.09)

low element interactivity material can be a function of the differing tests rather than indicating the difficulty of different levels of interactivity and so is not informative. In contrast, the main effect of differing levels of guidance eliminates the effect of the differing tests and so is potentially informative. The interaction tests whether low levels of guidance result in superior scores to high levels of guidance using the same low element interactivity test, while simultaneously testing whether high levels of guidance result in superior scores to low levels of guidance using the same high element interactivity test. All means and standard deviations are presented in Table 3 and the following analyses were based on the five test questions.

The main effect of guidance was not significant, $F(1, 45) = .466$, MSe $= 226.56$, $p = .498$, $\eta_{p^2} = .010$. The main effect of element interactivity was significant, $F(1, 45) = 13.71$, MSe $= 250.39$, $p = .001$. Wilks' Lambda $= .766$, $\eta_{p^2} = .234$. The mean percentage correct scores indicated that high element interactivity task scores were higher than low element interactivity task scores, noting that this test removes the variance due to element interactivity, but the comparison itself is not informative because the tests differ. Finally, the interaction between guidance and element interactivity, which was of primary interest in this experiment, was significant, $F(1, 45) = 5.16$, $p = .028$, Wilks' Lambda $= .766$, $\eta_{p^2} = .103$.

Following the significant interaction, simple effects tests were conducted. Note that one-tailed *t* tests were used as we had a clear directional hypothesis for both tests. For the low element interactivity tasks testing for the generation effect, the effect of guidance was significant, $t(46) = -2.02$, $SE_{diff} = -9.88$, $p = .025$ (one-tailed), $d = .57$. The mean percentage correct scores indicated that the low guidance condition was superior to the high guidance condition, demonstrating a generation effect.

For the high element interactivity material testing for the worked-example effect, the effect of guidance was not significant, $t(46) = 1.60$, $SE_{diff} = 6.68$, $p = .06$ (one-tailed), $d = .46$. The mean percentage correct scores indicated that the worked-example condition had higher means than the problem-solving condition, although the difference between the means was not significant. While not significant, that difference nevertheless did contribute to the disordinal interaction of the ANOVA.

Discussion

The results of this study supported the hypothesis of a disordinal interaction of guidance and element interactivity. Specifically, for low element interactivity material testing for the generation effect, low guidance was superior to high guidance, demonstrating a generation effect, whereas for high element interactivity tasks testing for the worked-example effect, that difference disappeared, to be replaced by a non-significant difference in the opposite direction. While a significant worked-example effect was not obtained, the difference between the means was in the predicted

direction, thus contributing to the statistically significant interaction. (It is possible to obtain a significant interaction with neither simple effect test being significant.)

The failure to obtain the worked-example effect needs some further discussion. Generally, for novice learners who do not have relevant schemas for a given task, full external guidance is essential. With increases in learner levels of knowledge, the need for external guidance such as that provided by worked examples is reduced. For learners used in this experiment, the relevant knowledge required for the tasks had been partially developed prior to the experiment. They still required some external guidance, but not as much as complete novices, which could explain why the worked-example effect in this study was not obtained. Of course, with an increased number of participants increasing the power of the test, a significant worked-example effect might be obtained.

Considering that relative to the levels of element interactivity of the corresponding tasks, the participants in this study could be regarded as both novices (in relation to the high element interactivity tasks used in the second phase) and experts (in relation to the low element interactivity tasks used in the first phase), the results could also be interpreted as an example of the expertise reversal effect (Kalyuga, 2007). According to this effect, there is an interaction between different instructional procedures (e.g., instructional methods with different levels of instructional guidance provided to learners) and levels of learner expertise. An instructional method that is suitable for novices may not be effective for more knowledgeable learners and vice versa. For tasks that are high in element interactivity, high guidance for novices is usually superior to low guidance, but the effectiveness of external guidance reverses for learners with higher levels of expertise for whom low guidance tends to be superior to high guidance (a reversed worked-example effect). In the current experiment, we did not vary expertise, but rather varied element interactivity. Since increases in expertise result in decreases in element interactivity, our results may be due to exactly the same factors that lead to the expertise reversal effect (Chen, Kalyuga, & Sweller, 2016b).

While we have attributed the results of the current experiment to differences in element interactivity between learning the geometry formulae and learning to solve problems using the formulae, there are many other differences between the two sets of tasks. Any of those differences could also be responsible for the obtained interaction and, for that reason, the current experiment does not constitute an unambiguous test of the hypothesis. At best, the results are in accord with the literature, which also seems to suggest that the generation effect is more likely to be obtained using low element interactivity material, and the worked-example effect is more likely to be obtained using high element interactivity material. Additional research will be required to confirm that the relevant variable actually is element interactivity.

In addition, it should be noted that the traditional worked-example, problem-solving paradigm used does not consist of a pure presentation of information because it includes generative activities during problem solving. Nevertheless, it does require more guidance than the pure problem solving used in the control group. However, future studies might be conducted using worked examples only compared to problem solving only.

From an educational perspective, this study may inform teachers that first, for simple tasks, learners may need to generate rather than study externally presented answers, whereas for complex tasks, additional external guidance is important for students who have just begun to acquire the relevant schemas. The degree of guidance

that should be provided to students depends on the nature of the tasks, namely their levels of element interactivity. This factor appears to be important when determining levels of instructional guidance.

Acknowledgment

Financial Support

Conflicts of Interests

None.

Ethical Standards

Ethics approval (No. HC15240) for the procedures used in this study was provided by the University of New South Wales, Sydney, Australia as an integral part of the first author's PhD program.

References

Atkinson, 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.

Ausubel, D.P. (1964). *Education and the structure of knowledge.* Chicago: Rand McNally.

Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology, 107*, 689–704.

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.

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. doi:10.1007/s10648-016-9359-1

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

Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences, 24*, 154–176.

Donaldson, W., & Bass, M. (1980). Relational information and memory for problem solutions. *Journal of Verbal Learning and Verbal Behavior, 19*, 26–35.

Ericsson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211–245.

Griffith, D. (1976). The attentional demands of mnemonic control processes. *Memory & Cognition, 4*, 103–108.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*, 1–19.

Karpicke, J.D., & Zaromb, F.M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language, 62*, 227–239.

Kirschner, P., Sweller, J., & Clark, R. (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.

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. doi:10.1080/00220973.2012.727884

Leahy, W., Hanham, J., & Sweller, J. (2015). High element interactivity information during problem solving may lead to failure to obtain the testing effect. *Educational Psychology Review, 27*, 291–304.

Mayer, R.E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*, 14–19.

McFarland, C.E., Frey, T.J., & Rhodes, D.D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning and Verbal Behavior, 19*, 210–225.

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

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

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

Peynircioğlu, Z.F. (1989). The generation effect with pictures and nonsense figures. *Acta Psychologica, 70*, 153–160.

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

Renkl, A. (2014a). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*, 1–37.

Renkl, A. (2014b). The worked-out examples principle in multimedia learning. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.). New York: Cambridge University Press.

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

Schweickert, R., McDaniel, M.A., & Riegler, G. (1994). Effects of generation on immediate memory span and delayed unexpected free recall. *The Quarterly Journal of Experimental Psychology, 47*, 781–804.

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior, 25*, 258–266.

Slamecka, N.J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 592–604.

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

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*, 123–138.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory.* New York: Springer.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185–233.

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

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

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*, 257–287.

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology, 36*, 212–218.

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

Van Gog, T., & Sweller, J. (2015). Not new but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review, 27*, 247–264.

Van Merriënboer, J.J., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: Balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology, 20*, 343–352.
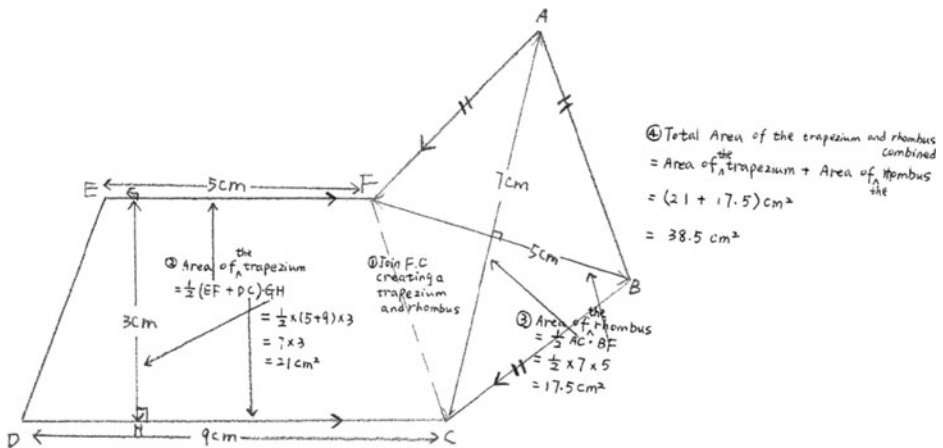
Appendix A

An Example of a Formula Used for the Generation Group

$$\text{Area} = a \times b$$
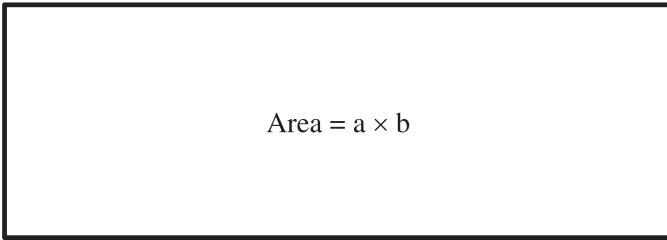
Appendix B

An Example from Booklet 2 Used for the Worked-Example Group

Appendix C

An Example of Test Questions Used to Test the Worked-Example Effect