

**Relations Between the Worked Example and Generation Effects on Immediate and
Delayed Tests**

Submitted date: 29/01/2016

Abstract

The contradiction between the worked example effect that occurs when learners presented with more instructional guidance learn more than learners presented with less guidance and the generation effect that occurs when the reverse result is obtained can be resolved by the suggestion that the worked example effect is obtained using materials high in element interactivity, whereas simpler, low element interactivity materials result in the generation effect. A 2 (guidance: low vs. high) x 2 (element interactivity: low vs. high) x 2 (expertise: low vs. high) experiment investigated this hypothesis with high school trigonometry learners. On an immediate test, high guidance reflecting a worked example effect was found for novices, but a generation effect was obtained for more knowledgeable learners. In contrast, on a delayed test, a three-way interaction between guidance, element interactivity and expertise was found. This interaction was caused by a worked example effect for material high in element interactivity and a generation effect for material low in element interactivity for novices while for more knowledgeable learners, a generation effect was obtained for both low and high element interactivity materials. These results suggest firstly, that both the worked example and generation effects may be more likely on delayed than immediate tests and secondly, that the worked example effect relies on high element interactivity material while the generation effect relies on low element interactivity material.

Keywords: cognitive load theory, worked example effect, generation effect, element interactivity, immediate and delayed tests.

How much guidance should be provided to facilitate students' learning? On the one hand, within the framework of cognitive load theory, there is evidence that worked examples which provide full problem solving guidance lead to superior performance on subsequent tests of knowledge of solution procedures than actual problem solving with no guidance, demonstrating the worked example effect. On the other hand, there is evidence that requiring students to generate items using, for example, a paired associate paradigm, leads to better memory of the items on subsequent tests than externally presented answers, demonstrating the generation effect.

A possible solution to this contradiction is that these differential results are caused by different levels of element interactivity or complexity of the materials (Chen, Kalyuga, & Sweller, 2015). Problem solving tasks characteristically use complex materials while memorization tasks use much simpler materials. Cognitive load theory provides a definition and measure of complexity of instructional materials via the concept of element interactivity.

1. Cognitive Load Theory and Element Interactivity

Cognitive load theory is an instructional theory based on our knowledge of human cognitive architecture (Sweller, Ayres, & Kalyuga, 2011). This architecture constitutes a natural information processing system similar to evolution by natural selection (Sweller & Sweller, 2006) that can be described by five principles.

(a) Information Store Principle. In order to function in a complex environment, natural information processing systems must incorporate a large store of information. Long-term memory provides that store in human cognition.

(b) Borrowing and Reorganizing Principle. That store of information is largely created by borrowing information from other people by imitation, reading and listening.

(c) *Randomness as Genesis Principle*. When required information is not available from others, it must be created. A random generate-and-test procedure during problem solving performs this role in human cognition.

(d) *Narrow Limits of Change Principle*. In order to avoid the need to generate huge numbers of possible moves and to prevent significant, rapid, uncontrollable and therefore damaging changes to long-term memory, the system needs to ensure that only small changes occur at a time. Working memory provides that assurance by its very narrow capacity and duration limits when dealing with novel information from the external environment.

(e) *Environmental Organizing and Linking Principle*. In contrast to its limitations when dealing with novel information from the environment, working memory has no known limits when dealing with organized information held in long-term memory. Activated by external signals, large amounts of information can be retrieved rapidly from long-term memory to working memory allowing appropriate responses to those signals.

Based on this architecture, if the material that learners must process is complex and imposes a heavy working memory load, then it is important that instructional procedures do not unnecessarily add to that load. If the material is simpler and does not impose a heavy working memory load, then factors other than working memory load are likely to determine the effectiveness of instruction and so instructional procedures may not need to take cognitive load into account.

Element interactivity determines the extent to which information imposes a heavy cognitive load due to either its intrinsic characteristics or due to extraneous factors such as the instructional design used (Sweller, 2010; Sweller & Chandler, 1994). It is intended to provide a measure of complexity that takes into account the nature of the learning materials, known as intrinsic cognitive load, the manner in which learners interact with those learning materials, known as extraneous cognitive load, and the learners' knowledge base.

Interactive elements are defined as elements that must be processed simultaneously in working memory in order to complete a task. If those elements are intrinsic to the task at hand, the resultant working memory load is referred to as intrinsic cognitive load. That load can vary from low to high depending on the intrinsic nature of the material being dealt with. For example, when students learn the symbols of the chemical periodic table, each symbol stands for two elements that must be processed in working memory, the symbol and its name. Students can study each symbol individually with no reference to other symbols. When students try to learn the symbol for iron, *Fe*, they can do so independently of learning the symbol for copper, *Cu* and students do not need to pay attention to the relations between them. This kind of material has a low degree of element interactivity and therefore, it has a low intrinsic cognitive load. However, asking students to balance a chemical equation or solve a mathematics problem, such as $ax = b$, *solve for x*, can be high in element interactivity. For the algebra problem, novices need to consider each symbol resulting in 6 elements (including the implied multiplication and the goal symbol) and the relation of each symbol to at least one other symbol, resulting in a minimum of 12 elements that need be processed in working memory simultaneously. This task is relatively high in element interactivity material for novices compared to learning a chemical symbol and so, based on the narrow limits of change principle, imposes a heavy working memory load.

The levels of element interactivity and intrinsic cognitive load are also influenced by the expertise of learners. For an expert faced with the above algebra problem, the environmental organizing and linking principle rather than the narrow limits of change principle comes into play. Because of knowledge held in long-term memory, rather than being faced with 12 or more interacting elements, an expert may be faced with no more than a single element. An expert may immediately recognize the category to which the problem belongs and know the solution. Since the problem and solution consists of a single element,

element interactivity is very low with a very low intrinsic cognitive load. Thus, expertise can have a substantial effect on element interactivity and cognitive load.

Extraneous cognitive load also is determined by levels of element interactivity. This load is influenced by the way instructional materials are presented. If element interactivity is altered without altering what needs to be learned, we are dealing with extraneous cognitive load. This load relies entirely on the cognitive activities in which learners must be involved due to the instructional procedures used. The worked example effect provides an example.

2. The Worked Example Effect

This effect occurs if instruction requires learners to solve a problem rather than study a worked example that provides a detailed solution of a problem for a learner to study (Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 2014; Sweller & Cooper, 1985; Ward & Sweller, 1990). When solving a problem using the randomness as genesis principle, extraneous cognitive load is increased in comparison with studying a worked example using the borrowing and reorganizing principle because more elements must be considered. An unguided problem does not indicate which elements should be attended to while a worked example does, reducing the number of elements that must be processed in working memory. As a consequence, studying worked examples providing instructional guidance is usually superior to problem solving providing no guidance, demonstrating the worked example effect.

Knowledge borrowed from instructors who construct worked examples can be reorganized and transferred to long-term memory for storing, according to the information store principle. Finally, if such knowledge is successfully stored in long-term memory, it can be retrieved when needed to guide activities required for successful functioning in an external environment, according to the environmental organizing and linking principle. Many experiments have repeatedly demonstrated the worked example effect (see Renkl, 2014; Sweller et al., 2011, for recent overviews).

3. The Generation Effect

The generation effect occurs when items that are generated based on a given stimulus and an encoding rule are better remembered than the same items which are simply read (Slamecka & Graf, 1978). Paired associate lists frequently are used in generation effect studies. For example, one group may be asked to read a paired associate list consisting of words and their opposites (e.g., hot-cold) while the other group may be presented the first word of each pair and told to generate the second word. Tests of memory of the response words characteristically show superiority of the generation group.

Most experiments testing the generation effect have used simple materials that are low in element interactivity. For instance, Pyke and LeFevre (2011) used alphabetic material such as $G + 4 = K$ where “4” refers to the distance between “G” and “K” in the alphabet. The generation group generated each answer while participants in the read group were shown each answer. Both groups then were tested on their knowledge of the distance between letters. Learning this low element interactivity task was superior for the generation group.

3.1 Explanations of the Generation Effect

Single-factor theories assume that the generation effect is caused by a single factor. Two kinds of single factor have been proposed by researchers to explain the generation effect (Burns, 1990). The first assumes the generation effect enhances the processing of shared features of a stimulus and response, namely stimulus-response relational theory suggested by (Hock, Throckmorton, Webb, & Rosenthal, 1981). The second type of single-factor theory assumes the generation effect enhances the processing of features of responses only, namely, the response-oriented theory (Begg, Snider, Foley, & Goddard, 1989) which can be divided into two possible approaches: a response-specific factor, which indicates that the generation effect only improves the features of responses that are not shared with another response or

stimulus; and a response-relational factor, which suggests that the generation effect enhances the features shared among responses.

Begg et al. (1989) obtained results supporting response-oriented factor theory. If the generation effect enhances stimulus–response processing, the advantage of the generation effect should be found by testing knowledge of the stimuli as well as the responses. However, Begg et al.'s results did not support this hypothesis. The opposite view was expressed in the work of Greenwald and Johnson (1989). They found a small but significant generation effect by testing knowledge of the stimuli in unexpected free-recall, cued-recall and recognition tests. This study therefore supported the suggestion that the generation effect enhanced stimulus-response processing.

Multiple-factor theories assume that the generation effect should be explained by multiple factors rather than a single factor. According to Burns (1990), a multiple-factor theory suggests that the stimulus-response relation and response-oriented factors both should be addressed. Similarly, Hirshman and Bjork (1988) found the generation effect with cued recall (which is sensitive to stimulus-response processing as well as response-oriented processing) and free recall tests (which are sensitive to response-oriented processing), thus supporting the multiple-factor theory.

Another version of multiple-factor theory was proposed by McDaniel, Riegler, and Waddill (1990). The three factors suggested in their research were: the information provided by the letters of each word, the information paired with each word, and the information from the whole list. The second factor included two situations: an explicit cue accompanying words (e.g., *HOT-C*_ : antonym) or an implicit cue which provides only a relevant hint of the word (e.g., *TALL-S*_). The results suggested that when the targets themselves and the relations among initial words and targeted words were obvious to learners, they would

consider the information which came from the whole list, and then a generation effect could be found using a free recall test.

The generation effect is closely related to the testing effect (Karpicke & Zaromb, 2010). The paradigm for the testing effect is that learners are given a test of previously learned information rather than being re-presented that information, which is 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, some recent research studies have indicated that the testing effect is heavily reliant on levels of element interactivity with the effect more likely to be found when element interactivity is low (Leahy, Hanham, & Sweller, 2015; Van Gog & Sweller, 2015).

3.2 Immediate vs. Delayed Testing

According to Table 1 that summarizes the results of experiments testing the generation effect, most studies (35 out of 38) successfully demonstrated the effect after a delay or distraction. A smaller proportion of studies (19 out of 26) demonstrated the effect using immediate tests. While the effect seems to be obtainable on both immediate and delayed tests, it may be more likely using delayed rather than immediate tests, although as can be seen from Table 1, that conclusion is far from certain. Furthermore, even if it is more likely, it is unclear why. One explanation that has been offered is that the process of generation itself can improve memory and that the effect of generation is more durable on memory (Schweickert, McDaniel, & Riegler, 1994).

4. The Expertise Reversal Effect

Within a cognitive load theory framework, the generation effect can be related to the expertise reversal effect which provides an alternative explanation. The expertise reversal effect occurs when the effectiveness of instructional procedures reverses with increasing expertise (Kalyuga, Ayres, Chandler, & Sweller, 2003). For example, based on the worked

example effect, studying worked examples is superior to solving the equivalent problems. With increasing expertise, the advantage of worked examples first decreases, then disappears and finally reverses with problem solving or generating answers to problems becoming superior to studying worked examples (Kalyuga, Chandler, Tuovinen, & Sweller, 2001), a result leading to the fading effect (Atkinson, Renkl, & Merrill, 2003; Renkl, Atkinson, & Maier, 2000). The fading effect occurs when learning is facilitated by presenting novices with guidance and gradually fading that guidance with increasing expertise.

As indicated above, with increasing expertise, element interactivity decreases. A concept or procedure that requires novices to manipulate multiple elements simultaneously may consist of a single element for more expert learners who, in accord with the environmental organizing and linking principle, can transfer the interacting elements from long-term memory to working memory as a single element or a very small number of elements. Accordingly, the superiority of generating problem solutions over studying worked examples found using more expert learners may have the same source as the generation effect that occurs using low element interactivity information. Whether information is low in element interactivity because of the nature of the material or because of the expertise of learners may be irrelevant. The cognitive effects may be identical. Generation may be characteristically superior to presentation when dealing with low element interactivity information irrespective of the reason for the information being low in element interactivity.

This issue is central to the present study. It may, for example, be argued that the worked example and generation effects cannot be meaningfully compared because the goals of the two tasks are different. The worked example effect is concerned with learning how to solve problems while the generation effect is commonly concerned with memorization. In fact, based on the above cognitive architecture, this distinction is not viable, with all learning, whether complex or simple, requiring a transfer of information into long-term memory. The

suggestion that skill in problem solving depends on memorization predates cognitive load theory (De Groot, 1965) and has been a central concept of cognitive load theory and the worked example effect for the last 40 years. The purpose of studying worked examples is to learn problem states and associate them with their appropriate moves. It is this activity that increases problem-solving expertise.

If this conceptualization is valid, the only difference between learning to solve problems and memorizing is the complexity or element interactivity of the material. This characterization leads to novel hypotheses. For novices presented high element interactivity material, guidance may be essential. For more expert learners for whom the material is no longer high in element interactivity, guidance may be redundant, leading to the expertise reversal effect when compared with novices. Accordingly, for novices, we can hypothesize that the worked example effect will be obtained with guidance via worked examples superior to generation via problem solving but as expertise increases, the worked example effect will convert to a generation effect with generation superior to guidance via presenting answers. The two contradictory effects that usually are treated independently as unrelated effects can instead be treated as closely related effects solely dependent on levels of element interactivity and with the same goal, the transfer of knowledge to long-term memory.

Kalyuga and Singh (2015) have recently suggested differentiating instructional goals with the purpose of defining the area of applicability of cognitive load theory. The goals considered in this paper belong to the category of acquisition of domain-specific schemas in long-term memory (with differences in complexity of the schemas determined by the levels of element interactivity only) – which is the same category according to Kalyuga and Singh, and the category that is relevant to cognitive load theory (alternative categories in their framework where cognitive load theory may not apply are, for example, pre-instructional goals such as motivating and engaging learners, activating their prior knowledge etc.; or

higher-level goals such as acquisition of generalized schemas for transfer within wider domains).

5. Present Study

From the above overview, the contradiction between the worked example and generation effects can be resolved by considering element interactivity as a critical factor (Chen et al., 2015). Specifically, for high element interactivity material, we hypothesize that high levels of guidance (worked examples) should be superior to low levels of guidance (problem solving), demonstrating the worked example effect, whereas, for low element interactivity material, low levels of guidance (generation) should be superior to high levels of guidance (presentation), demonstrating the generation effect (Hypothesis 1). However, with an increase of learner expertise, the worked example effect may disappear because the high level of element interactivity should be reduced, therefore, the generation effect may be obtained for both sets of materials (Hypothesis 2). Both hypotheses were tested using a 2 (expertise: high vs. low) x 2 (guidance: high vs. low) x 2 (element interactivity: high vs. low) experimental design with the third variable as a within-subject factor.

As indicated above, whether a test is immediate or delayed also may be an important factor. Almost all studies demonstrating the worked example effect tested participants immediately after the study phase, whereas, most generation effect studies used delayed tests. Therefore, based on the above literature survey, we also expected that the generation effect may be more likely to be obtained using a delayed test (Hypothesis 3).

6. Method

This experiment was designed to test the above three hypotheses. The subject area used to test for the worked example effect consisted of simplifying trigonometry expressions. For novices, that task was high in element interactivity and so was predicted to yield a worked example effect. For more expert learners for whom the same information was low in element

interactivity, a generation effect was predicted. Simpler memorization tasks requiring the reproduction of trigonometry formulae that were low in element interactivity were expected to yield a generation effect for both novices and experts.

6.1 Participants

Fifty Year 10 students from a secondary school in Chengdu, China were used as relatively less knowledgeable participants. They were approximately 16 years old. Six students' data were not complete, so they were excluded from the final analysis, leaving 44 students. No students had studied trigonometry in class and so all were regarded as novices. Another 52 Year 11 students, from the same school were recruited as more knowledgeable participants. They were approximately 17 years old. Two students' data were excluded as they were incomplete, leaving 50 students. In class, all Year 11 students previously had studied the trigonometry formulae used in this study to test for the generation effect. Similarly, all students had been taught how to simplify trigonometry expressions used to test for the worked example effect in their previous semester. Therefore, they were regarded as relatively expert learners with respect to the formulae as well as the problems used in this experiment.

In the first phase of the experiment (low element interactivity materials), students were randomly assigned to form the low guidance, generation group and the high guidance presentation group (see Figure 1). In the second phase (high element interactivity materials), half of the students from the generation group were randomly assigned to the low guidance problem solving group and the other half to the high guidance worked example group (see Figure 1). The same allocation method was used to assign students from the presentation group to either the worked example or problem-solving groups.

6.2 Materials

Ten trigonometry formulae were chosen from textbooks used in secondary school to test for the generation effect. There were five sine formulae and five cosine formulae (see Table 2).

6.2.1 Levels of Element Interactivity of Both Sets of Materials for Year 10

Students

The examples provided below indicate the levels of element interactivity for materials used in this study by counting the number of interactive elements. The formula, $\sin(A+B) = \sin A \cos B + \cos A \sin B$ was one of the formulae used to test for the generation effect in the first phase of the experiment. As the Year 10 students did not have relevant knowledge of this formula, elements of the equation had to be memorized. For a simple memory test (as opposed to a test requiring students to use the formula), each of the elements is independent of all other elements. Whether a particular element is remembered or forgotten should not affect memory of any of the other elements. Accordingly, the element interactivity count is 1 because the elements do not interact.

Simplifying a trigonometry formula below was used as high element interactivity material. The following is a worked example used in the experiment:

Simplify the following formula:

$$\begin{aligned} & \sqrt{2} \sin\left(\alpha + \frac{\pi}{4}\right) \\ &= \sqrt{2} \left(\sin \alpha \cos \frac{\pi}{4} + \cos \alpha \sin \frac{\pi}{4} \right) \\ &= \sqrt{2} \left(\frac{\sqrt{2}}{2} \sin \alpha + \frac{\sqrt{2}}{2} \cos \alpha \right) \\ &= \sin \alpha + \cos \alpha \end{aligned}$$

This worked example was one of the tasks used to test for the worked example effect. Year 10 students who had not learned previously how to simplify trigonometry expressions needed to remember the single symbols of the expression of the first line (e.g., sin, “+”), totaling 11

elements, including multiplication that is implied without a specific symbol. Next, in order to obtain the expression of the second line, students had to recall the relevant formula ($\sin(A+B) = \sin A \cos B + \cos A \sin B$), totaling 16 elements, and apply it to the first line resulting in a formula including 19 elements. Then they needed to calculate the value of sine and cosine indicated in the third line, totaling 8 elements. Finally, they needed to open the final brackets to give the last line, involving 17 elements. Therefore, in total, it can be estimated that about 71 interacting elements were involved in this task, which is a very complex (high in element interactivity) task for Year 10 students, compared to the trigonometry formula considered above. Because they interact, if any of these elements were forgotten or incorrect, the other elements could not be processed. Of course, relevant prior knowledge may reduce the number of interacting elements considerably.

6.2.2 Levels of Element Interactivity of Both Sets of Materials for Year 11

Students

The test materials were identical to those used for year 10 students. When counting the number of interacting elements to evaluate the effective level of element interactivity in the instructional materials of this experiment, the relatively higher level of learner expertise was taken into account. As was the case for Year 10 students, each part of a formula could be learned independently of the other parts and so for Year 11 students when testing for the generation effect, the element interactivity count is one.

Compared to Year 10 students, a reduction in the number of interacting elements can be expected for materials used to test for the worked example effect. For example, in the case of the task used above to illustrate the procedure, Year 11 students are likely to have already acquired the relevant knowledge for simplifying trigonometry expressions. They can transfer that knowledge from their long-term memory as single units to be processed in working

memory, thus reducing the number of interacting elements to a much lower level than that of Year 10 students, rendering this material low in element interactivity for Year 11 students.

6.2.3 Booklets for the First Phase and the Second Phase

In the first phase of the experiment, three A4-sized papers were distributed to students to test the generation effect. The first paper contained the ten basic trigonometry sine and cosine formulae which were randomly arranged in order (see Table 2 above). This paper was common to the generation and presentation groups. The second paper for the students in the presentation group was identical to the first paper, whereas those in the generation group were required to generate relevant formulae by themselves based on the names of each of the ten formulae that were provided as cues. The last paper, common to both groups, was blank and used for a free recall test. Students had to write out as many of the formulae as they could remember from the previous papers.

In the second phase, three A4-sized papers were provided again to test for the worked example effect. The first paper was the same as the first paper used in the first phase. Its function was revision of the previously learned formulae that were required for the problems used in this phase. The second paper distinguished the worked example group from the problem-solving group. Specifically, two worked examples, each followed by a similar problem for students to solve, were provided for the worked example group. In contrast, for the problem-solving group, the same four problems had to be solved by students themselves with no worked examples provided. The last paper was used for the test and contained five problems, similar to the problems in the second paper, for students to solve. The example used above to illustrate the calculation of the interactive elements for high element interactivity material shows an example used in the second (acquisition stage) paper for the worked example group. The paper presented to the problem-solving group was similar except

the solutions to problems 1 and 3 were not provided. An example of a test problem in the final paper is:

Simplify the following expression:

$$\left(\cos \frac{\pi}{12} - \sin \frac{\pi}{12}\right)\left(\cos \frac{\pi}{12} + \sin \frac{\pi}{12}\right).$$

6.3 Procedures

The first phase of the experiment testing the generation effect lasted 35 minutes (See Figure 2.). One Year 10 and one Year 11 class was used. Prior to studying the first paper, students were re-seated according to the group into which they were randomly placed (7 minutes).

The study stage was 10 minutes in length. After being randomly assigned to the experimental condition and re-seated, students were permitted to study the first paper. Note taking was allowed on this paper if needed. After 10 minutes, all students handed in this paper.

The generation or presentation stage was 10 minutes in length. The generation group and the presentation group received the second paper from the experimenter separately. The generation group had to generate all of the formulae they had studied in the first paper, whereas students in the presentation group were required to again study the formulae of the first paper. No one could hand in the paper before 10 minutes had passed. 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 the papers.

The free recall test stage was 8 minutes in length. Students were required to write out as many of the formulae that they had studied in the first and second papers as they could. Students could only hand in their test paper after 8 minutes had elapsed. Therefore, if students finished early, they were required to review their answers. When scoring the test, one mark was allocated for a correctly recalled formula. Therefore, the maximum score in the

free recall test was 10. Each student's score out of 10 was converted into a percentage for analysis providing the scores that tested knowledge of the low-element interactivity material.

In the second, worked example effect phase, the time allocated was also 35 minutes (See Figure 3.). The second phase followed immediately after the first phase. Prior to the experiment, the students in the generation and presentation groups had already been randomly assigned to form the worked example and problem solving groups for both the Year 10 and Year 11 classes. Students were re-seated according to the group to which they had been placed (7 minutes).

The study stage was 10 minutes in length. The procedure for this stage was identical to that used for the equivalent stage in the first phase of the experiment, testing the generation effect.

The worked-example or problem-solving stage was 10 minutes in length. The general procedure was identical to that used in the first phase. The students in the worked example group were required to study the worked example of Problem 1 demonstrating how to use the sine formula to simplify a trigonometry expression and then to solve a similar problem (Problem 2). All students then had to solve Problem 2 which was similar to Problem 1. A similar procedure was used for Problem 3 that was a worked example of how to use the trigonometry formulae in the first paper and Problem 4 that was a similar problem to Problem 3 but that students had to solve. The students in the problem solving group were required to solve the same four problems (Problems 1 - 4) used in the worked example group by themselves, with none of the problems presented as worked examples.

The test stage was 8 minutes in length. Again, the general procedure was identical to that used in the first phase. The test required students to solve five problems which were similar to those in the second paper. A maximum of three marks could be allocated for each of the problems (one for correctly using the formula; one for correctly calculating the sine

value; and one for correctly calculating the cosine value). Each student's total score out of 15 (five problems each with a maximum score of three) was converted to a percentage score for analysis. For Year 10, the internal reliability of the immediate test using Cronbach's α was .76 after deleting the 4th test question. This question was deleted because 75% of the students were able to fully solve this problem (deleting the question increased the internal reliability from .70 to .76). For Year 11, the internal reliability of the same immediate test for the worked example effect using Cronbach's α was .49, after eliminating the 4th test question, as around 80% of the students could fully solve the problem. These scores provided the dependent variable measuring knowledge of the high element interactivity material.

For Year 10 students, a week after the immediate test, 36 students took the delayed tests for the worked example and generation effects (the data of two students were excluded as they were not complete, leaving 34 students for the final analysis). The content of the delayed tests was exactly the same as the immediate tests. Using Cronbach's α , the internal reliability of the test for the worked example effect was .75, after deleting the 4th test question, with around 65% of students able to fully solve this problem. For Year 11 students, a week after the immediate test, 50 students from the same class were tested again with the same tests from the previous week (four students of this class were absent, leaving 46 students for the final analysis). The internal reliability of the delayed test for the worked example effect using Cronbach's α was .29, after deleting the 4th test question, as over 80% of the students could solve this problem.

The low α level for the Year 11, worked example test requires an explanation. The low consistency of the four test questions used may be due to the relative expertise of participants. Learners in Year 11 were more knowledgeable than those in Year 10. That knowledge allowed many to fully solve many of the problems. Where a full solution was not obtained, learners tended to obtain a score of 0 rather than 1 or 2, resulting in scores on each problem

having a wide range. Many learners obtained a score of 3 on one problem but a score of 0 on another, a common occurrence on problem solving tasks. The consequence is low consistency using Cronbach's α .

7. Results

7.1 The Immediate Test

Means and standard deviations of percentage test score results may be found in Table 3. These results were analyzed using a 2 (levels of expertise) x 2 (levels of guidance) x 2 (levels of element interactivity) ANOVA with repeated measures on the element interactivity factor. For the high element interactivity test, all means, standard deviations and statistical analyses were based on the four test questions remaining after eliminating Question 4.

The main effect of expertise was significant, $F(1, 90) = 190.82$, $MSe = 437.78$, $p < .001$, $\eta_p^2 = .680$. The percentage correct test scores for the more knowledgeable Year 11 students were higher than for the less knowledgeable Year 10 students. The main effect of guidance was not significant, $F(1, 90) = .21$, $MSe = 437.78$, $p = .649$, $\eta_p^2 = .002$. The main effect of element interactivity was significant, $F(1, 90) = 124.61$, $MSe = 421.94$, $p < .001$, Wilks' Lambda = .419, $\eta_p^2 = .581$. The low element interactivity material percentage correct test scores were higher than the high element interactivity test scores.

With respect to the 2-way interactions, the interaction between guidance and element interactivity was not significant, $F(1, 90) = 1.69$, $MSe = 421.94$, $p = .198$. Wilks' Lambda = .982, $\eta_p^2 = .018$. The interaction between expertise and element interactivity was significant, $F(1, 90) = 47.94$, $MSe = 421.94$, $p < .001$. Simple effects tests indicated that for high element interactivity materials, Year 11 students significantly outperformed Year 10 students, $t(92) = 16.18$, $SE_{diff} = 3.88$, $p < .001$, $d = 1.71$ (see Figures 4 and 5). Similarly, for low element interactivity materials, Year 11 also significantly outperformed Year 10 students, $t(92) = 4.56$, $SE_{diff} = 4.72$, $p < .001$, $d = .86$. The significant interaction was obtained because the effect

size for the high element interactivity information was much larger than for the low element interactivity information. As can be seen in Figures 4 and 5, the difference between Years 10 and 11 is much larger for high than low element interactivity material. This result indicates that Year 11 students were much better able to handle the high element interactivity information than the Year 10 students but this difference was considerably attenuated for the low element interactivity information.

A significant interaction between expertise and guidance was also found, $F(1, 90) = 4.38$, $MSe = 437.78$, $p = .039$, $\eta_p^2 = .046$. Simple effects tests indicated that for Year 10 students, the mean score for high guidance was significantly higher than for low guidance, $t(42) = 3.18$, $SE_{diff} = 3.46$, $p = .002$, $d = .87$ (see Figure 4), reflecting a worked example effect, while for Year 11, the mean score for low guidance was significantly higher than for high guidance, $t(48) = 2.16$, $SE_{diff} = 3.61$, $p = .04$, $d = .59$ (see Figure 5), reflecting a generation effect and explaining the interaction. More expert learners performed better with less guidance while novices did not.

The three-way interaction of element interactivity, guidance and expertise was not significant, $F(1, 90) = .489$, $MSe = 421.94$, $p = .486$, Wilks' Lambda = .995, $\eta_p^2 = .005$.

7.2 The Delayed Test

Means and standard deviations of percentage test score results may be found in Table 4. These results were analyzed using a 2 (levels of expertise) x 2 (levels of guidance) x 2 (levels of element interactivity) ANOVA with repeated measures on the element interactivity factor. All means, standard deviations and statistical analyses in Table 4 were based on the four test questions remaining after eliminating Question 4.

The main effect of expertise was significant, $F(1, 76) = 288.67$, $MSe = 342.54$, $p < .001$, $\eta_p^2 = .792$. The percentage correct test scores were higher for the more knowledgeable Year 11 students than for the less knowledgeable Year 10 students. The main effect of guidance

was significant, $F(1, 76) = 5.78$, $MSe = 342.54$, $p = .019$, $\eta_p^2 = .071$, with low guidance scores higher than high guidance scores. The main effect of element interactivity was significant, $F(1, 76) = 74.99$, $MSe = 349.45$, $p < .001$, Wilks' Lambda = .503, $\eta_p^2 = .497$. The low element interactivity material percentage correct test scores were higher than the high element interactivity test scores.

With respect to the 2-way interactions, the interaction between guidance and element interactivity was significant, $F(1, 76) = 5.53$, $MSe = 349.45$, $p = .021$, Wilks' Lambda = .932, $\eta_p^2 = .068$. Simple effect tests indicated that for low element interactivity materials, low guidance (generation) was superior to high guidance (presentation), $t(78) = 2.03$, $SE_{diff} = 6.08$, $p = .04$, $d = .44$, indicating a generation effect, while there was no significant difference for high element interactivity materials.

The interaction between expertise and element interactivity was significant, $F(1, 76) = 21.01$, $MSe = 349.45$, $p < .001$, Wilks' Lambda = .783, $\eta_p^2 = .217$. Simple effects tests indicated that for high element interactivity materials, Year 11 significantly outperformed Year 10, $t(78) = 16.25$, $SE_{diff} = 3.93$, $p < .001$, $d = 1.77$. with a similar result with a reduced effect size for low element interactivity materials, $t(78) = 7.64$, $SE_{diff} = 4.77$, $p < .001$, $d = 1.32$. The advantage Year 11 students have over Year 10 students is greater for high than low element interactivity, explaining the significant interaction. The interaction between guidance and expertise was not significant, $F(1, 76) = .023$, $MSe = 342.54$, $p = .879$, $\eta_p^2 = 0$.

The 3-way interaction of element interactivity, guidance and expertise was significant, $F(1, 76) = 6.254$, $MSe = 349.45$, $p = .015$, Wilks' Lambda = .924, $\eta_p^2 = .076$. Subsequent 2-way interaction tests acting as simple effects tests for the 3-way interaction, indicated a significant interaction between guidance and element interactivity for Year 10 students (see Figure 6.), $F(1, 32) = 12.66$, $MSe = 282.78$, $p = .001$, Wilks' Lambda = .717, $\eta_p^2 = .283$.

Following the significant, 2-way interaction, further simple effects tests were conducted. For

the low element interactivity material testing for the generation effect, the effect of guidance was significant, $t(32) = 2.63$, $SE_{diff} = 8.05$, $p = .01$, $d = 0.83$. The mean percentage correct scores indicated that low guidance (generation) was superior to high guidance (presentation) demonstrating a generation effect. For the high element interactivity material testing for the worked example effect, the effect of guidance was close to significance, $t(32) = 1.97$, $SE_{diff} = 3.98$, $p = .06$, $d = 0.65$. (It might be noted that while we have used a 2-tailed test, we had a clear directional prediction for this test theoretically permitting the use of a 1-tailed test.) The mean percentage correct scores indicated that high guidance (worked examples) was superior to low guidance (problem solving) demonstrating a worked example effect.

In contrast to the significant 2-way interaction for Year 10 students, there was no significant interaction between guidance and element interactivity for the Year 11 students, (see Figure 7.), $F(1, 44) = .012$, $MSe = 397.942$, $p = .915$. Wilks' Lambda = 1.000, $\eta_p^2 = 0$. Instead for the Year 11 students, the main effect of guidance was significant, $F(1, 44) = 4.40$, $MSe = 298.65$, $p = .042$, $\eta_p^2 = .091$ with both high and low element interactivity information resulting in higher test scores after low guidance.

The contrasting 2-way interactions between the Year 10 and Year 11 students explain the significant 3-way interaction. For Year 10 students, a generation effect was obtained on low element interactivity information while a worked example effect was obtained on high element interactivity information. For Year 11 students, a generation effect was obtained on both sets of information.

8. Discussion

It was hypothesized that high guidance would be superior to low guidance using materials high in element interactivity, resulting in the worked example effect, whereas low guidance was predicted to be superior to high guidance for materials low in element interactivity, resulting in the generation effect (Hypothesis 1). This result was obtained on the

delayed test for the Year 10, less knowledgeable students. On the immediate test, high guidance suggesting the worked example effect was obtained with no evidence of a generation effect. It was further predicted that with an increase in learner expertise, the worked example effect would disappear due to the expertise reversal effect (Hypothesis 2). This result was obtained on both immediate and delayed tests for the Year 11, more knowledgeable students. Lastly, based on previous results, it was predicted that the generation effect would be more likely on a delayed rather than an immediate test (Hypothesis 3). That hypothesis was confirmed. The results of the immediate test for Year 10 showed no statistically significant generation effect, but a significant effect was obtained on the delayed test in accord with the hypothesis, whereas, for Year 11, the generation effect was obtained on both the immediate and the delayed tests.

Considering the effect of expertise on performance with materials high in element interactivity, this experiment demonstrated an expertise reversal effect. High guidance in the form of worked examples was redundant for relatively more experienced Year 11 learners who already had acquired relevant knowledge. As a consequence, the worked example effect for Year 10 was replaced by a reversed worked example effect with problem solving superior to worked examples for Year 11.

As indicated above, the relation between the generation and worked example effects closely resembles the expertise reversal effect. It may be plausible to argue that the same theoretical explanation can be used for both. If so, an alternative explanation may be available for the generation effect. The reversed worked example effect found when using more expert learners normally is explained using the redundancy effect, an explanation that is used for all versions of the expertise reversal effect. It is assumed that some forms of instruction that are beneficial for novices lose their effectiveness and may even have deleterious effects with increases in expertise. The negative effects are theorized to be due to

redundancy. If, for example, worked examples are redundant for more expert learners, then processing those worked examples increases the number of interacting elements compared to problem solving, leading to problem solving superiority.

Since increased expertise reduces element interactivity, we might predict that if low element interactivity information is presented to learners, generation might be superior to study. This relation was obtained in this experiment. For Year 11, all of the materials were low in element interactivity due to higher levels of expertise, and so the generation effect was obtained universally. This result for the worked example materials, in conjunction with Year-10 participants, provided a classic version of the expertise reversal effect.

A limitation of the current study is the low internal reliability of the delayed test for Year 11 students. Replications clearly are required. Nevertheless, it should be noted that the inferential statistical tests used in experimental studies also constitute tests of reliability. The fact that the test items may be testing different constructs does not eliminate their validity as tests of knowledge.

The results obtained in this experiment also have clear educational implications. Firstly, the required level of instructional guidance cannot be determined without considering the level of complexity of learning materials which is dependent on the level of element interactivity. For simple content with low element interactivity, including more complex information processed by more expert learners, requiring students to self-generate may be superior to studying explicitly provided answers, especially for longer-term retention. On the other hand, for complex materials, explicit instruction is essential for novice learners. Secondly, contrary to some claims about the ineffectiveness of explicit instruction for longer-term learning, demonstrating worked examples on how to deal with complex materials to novice learners may have a durable, longer-term, positive effect. Thirdly, according to

reviews of the generation effect studies (see Table 1), few studies have addressed curriculum-based materials and this research will broaden the research base of the generation effect.

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.
- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning From Studying Examples to Solving Problems: Effects of Self-Explanation Prompts and Fading Worked-Out Steps. *Journal of Educational Psychology, 95*, 774-783.
- Begg, I., Snider, A., Foley, F., & Goddard, R. (1989). The generation effect is no artifact: Generating makes words distinctive. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 977-989.
- Begg, I., Vinski, E., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory & Cognition, 19*, 487-497.
- Burns, D. J. (1990). The generation effect: A test between single-and multifactor theories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1060-1067.
- Burns, D. J. (1992). The consequences of generation. *Journal of Memory and Language, 31*, 615-633.
- Burns, D. J. (1996). The item-order distinction and the generation effect: The importance of order information in long-term memory. *The American Journal of Psychology, 109*, 567-580.
- Burns, D. J., Curti, E. T., & Lavin, J. C. (1993). The effects of generation on item and order retention in immediate and delayed recall. *Memory & Cognition, 21*, 846-852.

- Buyer, L. S., & Dominowski, R. L. (1989). Retention of solutions: It is better to give than to receive. *The American Journal of Psychology*, 353-363.
- Carroll, M., & Nelson, T. O. (1993). Effect of overlearning on the feeling of knowing is more detectable in within-subject than in between-subject designs. *The American Journal of Psychology*, 227-235.
- Chechile, R. A., & Soraci, S. A. (1999). Evidence for a multiple-process account of the generation effect. *Memory*, 7, 483-508.
- Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology*, 107, 689-704.
- Crutcher, R. J., & Healy, A. F. (1989). Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 669-675.
- De Groot, A. D. (1965). Thought and choice in chess. *The Hague: Mouton*.
- De Winstanley, P., & Bjork, E. (1997). Processing instructions and the generation effect: A test of the multifactor transfer-appropriate processing theory. *Memory*, 5, 401-422.
- DeWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition*, 32, 945-955.
- Dick, M. B., Kean, M. L., & Sands, D. (1989). Memory for internally generated words in Alzheimer-type dementia: Breakdown in encoding and semantic memory. *Brain and Cognition*, 9, 88-108.
- Donaldson, W., & Bass, M. (1980). Relational information and memory for problem solutions. *Journal of Verbal Learning and Verbal Behavior*, 19, 26-35.
- Erdfelder, M. C. S. E. (1998). Determinants of Positive and Negative Generation Effects in Free Recall. *The Quarterly Journal of Experimental Psychology: Section A*, 51, 705-733.

- Fiedler, K., Lachnit, H., Fay, D., & Krug, C. (1992). Mobilization of cognitive resources and the generation effect. *The Quarterly Journal of Experimental Psychology*, *45*, 149-171.
- Flory, P., & Pring, L. (1995). The effects of data-driven and conceptually driven generation of study items on direct and indirect measures of memory. *The Quarterly Journal of Experimental Psychology*, *48*, 153-165.
- Gardiner, J. M. (1988). Generation and priming effects in word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 495-501.
- Gardiner, J. M. (1989). A generation effect in memory without awareness. *British Journal of Psychology*, *80*, 163-168.
- Gardiner, J. M., & Arthurs, F. S. (1982). Encoding context and the generating effect in multitrial free-recall learning. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *36*, 527-531.
- Gardiner, J. M., Dawson, A. J., & Sutton, E. A. (1989). Specificity and generality of enhanced priming effects for self-generated study items. *The American Journal of Psychology*, 295-305.
- Gardiner, J. M., Gregg, V. H., & Hampton, J. A. (1988). Word frequency and generation effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 687-693.
- Gardiner, J. M., & Hampton, J. A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 732-741.
- Gardiner, J. M., & Rowley, J. M. (1984). A generation effect with numbers rather than words. *Memory & Cognition*, *12*, 443-445.

- Ghatala, E. S. (1983). When does internal generation facilitate memory for sentences? *The American Journal of Psychology*, 75-83.
- Glisky, E. L., & Rabinowitz, J. C. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 193-205.
- Graf, P. (1982). The memorial consequences of generation and transformation. *Journal of Verbal Learning and Verbal Behavior*, 21, 539-548.
- Greenwald, A. G., & Johnson, M. M. (1989). The generation effect extended: Memory enhancement for generation cues. *Memory & Cognition*, 17, 673-681.
- Grososky, A., Payne, D. G., & Campbell, K. D. (1994). Does the generation effect depend upon selective displaced rehearsal? *The American Journal of Psychology*, 53-68.
- Hertel, P. T. (1989). The generation effect: A reflection of cognitive effort? *Bulletin of the Psychonomic Society*, 27, 541-544.
- Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 484-494.
- Hock, H. S., Throckmorton, B., Webb, E., & Rosenthal, A. (1981). The effect of phonemic processing on the retention of graphemic representations for words and nonwords. *Memory & Cognition*, 9, 461-471.
- Johnson, M. K., Raye, C. L., Foley, H. J., & Foley, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *The American Journal of Psychology*, 94, 37-64.
- Johnson, M. M., Schmitt, F. A., & Pietrukowicz, M. (1989). The memory advantages of the generation effect: Age and process differences. *Journal of Gerontology*, 44, 91-94.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38, 23-31.

- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579-588.
- Kalyuga, S., & Singh, A. M. (2015). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 1-22*.
- 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.
- Kinoshita, S. (1989). Generation enhances semantic processing? The role of distinctiveness in the generation effect. *Memory & Cognition, 17*, 563-571.
- 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.
- Liu, I. M., & Lee, Y. S. (1990). Memorial consequences of generating words and non-words. *The Quarterly Journal of Experimental Psychology, 42*, 255-278.
- Lutz, J., Briggs, A., & Cain, K. (2003). An examination of the value of the generation effect for learning new material. *The Journal of General Psychology, 130*, 171-188.
- MacLeod, C. M., & Daniels, K. A. (2000). Direct versus indirect tests of memory: Directed forgetting meets the generation effect. *Psychonomic Bulletin & Review, 7*, 354-359.
- McDaniel, M. A., Riegler, G. L., & Waddill, P. J. (1990). Generation effects in free recall: further support for a three-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 789-798.
- McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language, 27*, 521-536.

- McElroy, L. A. (1987). The generation effect with homographs: Evidence for postgeneration processing. *Memory & Cognition*, *15*, 148-153.
- McElroy, L. A., & Slamecka, N. J. (1982). Memorial consequences of generating nonwords: Implications for semantic-memory interpretations of the generation effect. *Journal of Verbal Learning and Verbal Behavior*, *21*, 249-259.
- 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.
- McFarland, C. E., Warren, L. R., & Crockard, J. (1985). Memory for self-generated stimuli in young and old adults. *Journal of Gerontology*, *40*, 205-207.
- McNamara, D. S., & Healy, A. F. (1995). A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems. *Journal of Memory and Language*, *34*, 399-416.
- McNamara, D. S., & Healy, A. F. (2000). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory and Language*, *43*, 652-679.
- Mulligan, N. W., & Duke, M. D. (2002). Positive and negative generation effects, hypermnesia, and total recall time. *Memory & Cognition*, *30*, 1044-1053.
- Nairne, J. S., Puse, C., & Widner, R. L. (1985). Representation in the mental lexicon: Implications for theories of the generation effect. *Memory & Cognition*, *13*, 183-191.
- Nicolas, S. (1996). The generation effect in a word-stem completion task: The influence of conceptual processes. *European Journal of Cognitive Psychology*, *8*, 405-424.
- Olofsson, U., & Nilsson, L. G. (1992). The generation effect in primed word-fragment completion reexamined. *Psychological Research*, *54*, 103-109.

- Payne, D. G., Neely, J. H., & Burns, D. J. (1986). The generation effect: Further tests of the lexical activation hypothesis. *Memory & Cognition*, *14*, 246-252.
- Pyke, A. A., & LeFevre, J. A. (2011). Calculator use need not undermine direct-access ability: The roles of retrieval, calculation, and calculator use in the acquisition of arithmetic facts. *Journal of Educational Psychology*, *103*, 607-616.
- Rabinowitz, J. C., & Craik, F. I. (1986). Specific enhancement effects associated with word generation. *Journal of Memory and Language*, *25*, 226-237.
- Reardon, R., Durso, F. T., Foley, M. A., & McGahan, J. R. (1987). Expertise and the generation effect. *Social Cognition*, *5*, 336-348.
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*, 1-37.
- Renkl, A., Atkinson, R. K., & Maier, U. H. (2000). From studying examples to solving problems: Fading worked-out solution steps helps learning. In L. Gleitman, & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 393-398). Mahwah, NJ: Lawrence Erlbaum Associations.
- Richardson-Klavehn, A., Gardiner, J. M., & Java, R. I. (1994). Involuntary conscious memory and the method of opposition. *Memory*, *2*, 1-29.
- Schmidt, S. R. (1992). Evaluating the role of distinctiveness in the generation effect. *The Quarterly Journal of Experimental Psychology*, *44*, 237-260.
- Schmidt, S. R., & Cherry, K. (1989). The negative generation effect: Delineation of a phenomenon. *Memory & cognition*, *17*, 359-369.
- 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.

- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 592-604.
- Slamucka, N. J., & Fevreiski, J. (1983). The generation effect when generation fails. *Journal of Verbal Learning and Verbal Behavior*, 22, 153-163.
- Soloway, R. M. (1986). No generation effect without semantic activation. *Bulletin of the Psychonomic Society*, 24, 261-262.
- 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., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology*, 4, 434-458.
- 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.
- Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, 7, 1-39.

Tables

Table 1 Research Studies of the Generation Effect

Experiment	Immediate	Delayed
Begg, Vinski, Frankovich & Holgate (1991)	√	√
Burns (1992)	×	
Burns (1996)		√
Burns, Curti, & Lavin (1993)	×	√
Buyer & Dominowski (1989)		√
Carroll & Nelson (1993)		×
Chechile & Soraci (1999)	√	
Crutcher & Healy (1989)		√
De Winstanley & Bjork (1997)		√
De Winstanley & Bjork (2004)	√	√
Dick, Kean & Sands (1989)	√	
Donaldson & Bass (1980)	√	
Fiedler, Lachnit, Fay, & Krug (1992)	√	
Flory & Pring (1995)		√
Gardiner (1988)		√
Gardiner (1989)	√	
Gardiner & Arthurs (1982)		√
Gardiner, Dawson & Sutton (1989)		√
Gardiner, Gregg, & Hampton (1988)		√
Gardiner & Hampton (1985)		√
Gardiner & Hampton (1988)		√
Gardiner & Rowley (1984)		√
Ghatala (1983)	√	
Glisky & Rabinowitz (1985)		√
Graf (1981)		√
Greenwald & Johnson (1989)	√	√
Grosofsky, Payne & Campbell (1994)	√	√
Hertel (1989)		√
Java (1994)	√	
Johnson, Raye, Foley & Foley (1981)	√	√
Johnson, Schmitt & Pietrukowicz (1989)		√
Kinoshita (1989)		√
Liu & Lee (1990)	√	
Lutz, Briggs & Cain (2003)		×
MacLeod & Daniels (2000)		√
McDaniel, Waddill & Einstein (1988)	×	
McElroy (1987)	√	
McElroy & Slamecka (1982)	×	
McFarland, Jr., Frey & Rhodes (1980)	√	√
McFarland, Jr., Warren & Crockard (1985)		√

McNamara & Healy (1995)		√
McNamara & Healy (2000)		√
Mulligan & Duke (2002)	×	
Nairne, Pusey & Widner, Jr. (1985)		√
Nicolas (1996)		√
Olofsson & Nilsson (1992)		√
Payne, Neely & Burns (1986)	×	
Rabinowitz & Craik (1986)	√	√
Reardon, Durso, Foley & McGahan (1987)	√	
Schmidt (1992)		√
Schmidt & Cherry (1989)		×
Schweickert, McDaniel & Riegler (1994)		√
Slamecka & Fevreski (1983)	√	√
Soloway (1986)	√	
Steffens & Erdfelder (1998)	×	

Note. “√” refers to an experiment that obtained the generation effect; “×” refers to an experiment that did not obtain the generation effect

Table 2 Formulae Used to Test the Generation Effect

sine formulae	cosine formulae
$\sin(A+B)=\sin A\cos B+\cos A\sin B$	$\cos(A+B)=\cos A\cos B-\sin A\sin B$
$\sin(A-B)=\sin A\cos B-\cos A\sin B$	$\cos(A-B)=\cos A\cos B+\sin A\sin B$
$\sin(-A)=-\sin A$	$\cos(-A)=\cos A$
$\sin(\pi +A)=-\sin A$	$\cos(\pi +A)=-\cos A$
$\sin(\pi -A)=\sin A$	$\cos(\pi -A)=-\cos A$

Table 3 Mean (SD) Percentage Correct Test Score Results of the Immediate Test

Expertise	Guidance	Low Element Interactivity	High Element Interactivity
Year 10	High (N=23)	70.4 (33.77)	22.1 (10.54)
	Low (N=21)	71.4 (25.93)	11.1 (12.43)
Year 11	High (N=25)	87.6 (16.15)	76.7 (28.05)
	Low (N=25)	97.2 (8.91)	82.7 (16.12)

Table 4 Mean (SD) Percentage Correct Test Score Results of the Delayed Test

Expertise	Guidance	Low Element Interactivity	High Element Interactivity
Year 10	High (N=17)	44.7 (26.72)	19.6 (11.00)
	Low (N=17)	65.9 (19.70)	11.8 (12.17)
Year 11	High (N=24)	88.3 (22.20)	75.7 (22.78)
	Low (N=22)	95.5 (8.00)	83.7 (16.96)

Figures

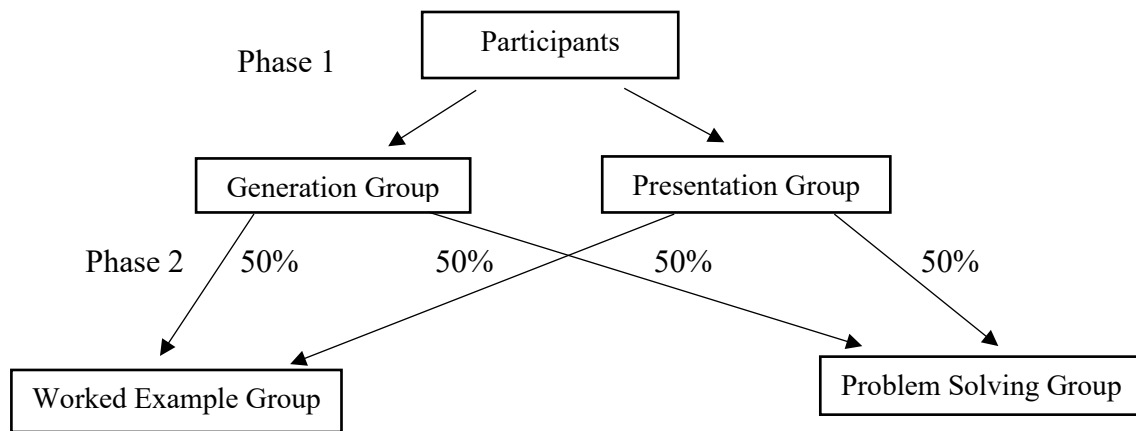


Figure 1. Procedure for Randomly Assigning Participants

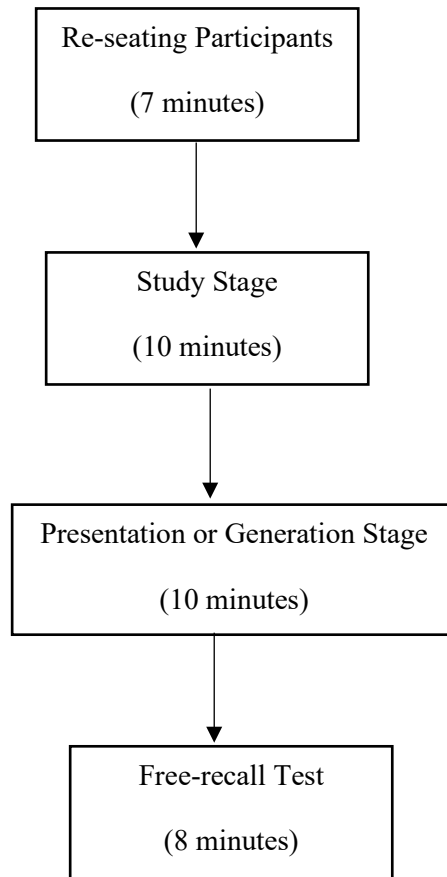


Figure 2. Procedure for Testing the Generation Effect (Phase 1)

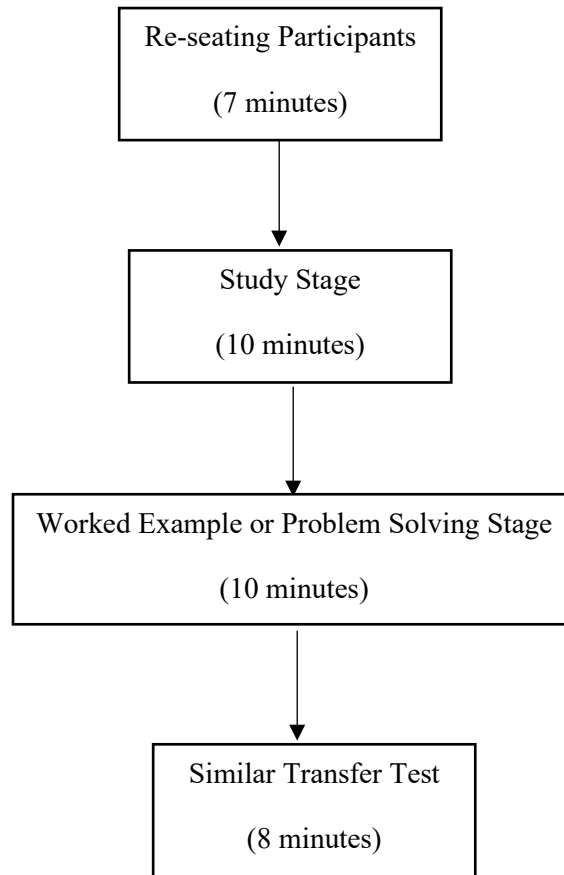


Figure 3. Procedure for Testing the Worked Example Effect (Phase 2)

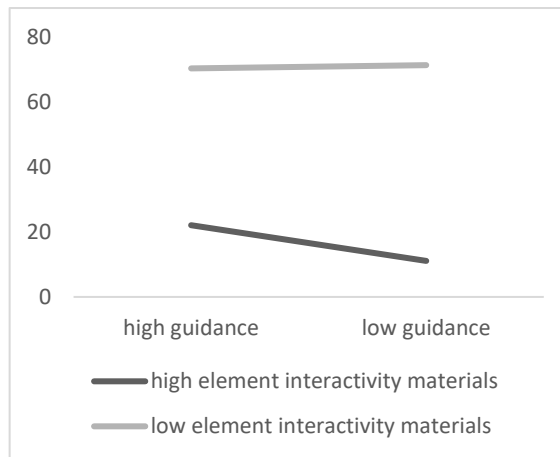


Figure 4. Relation Between Guidance and Element Interactivity on The Immediate Test for Year 10 Students

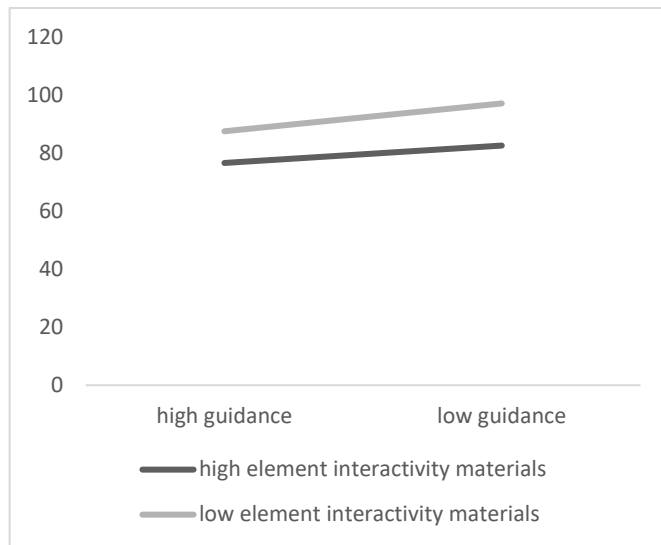


Figure 5. Relation Between Guidance and Element Interactivity on The Immediate Test for Year 11 Students

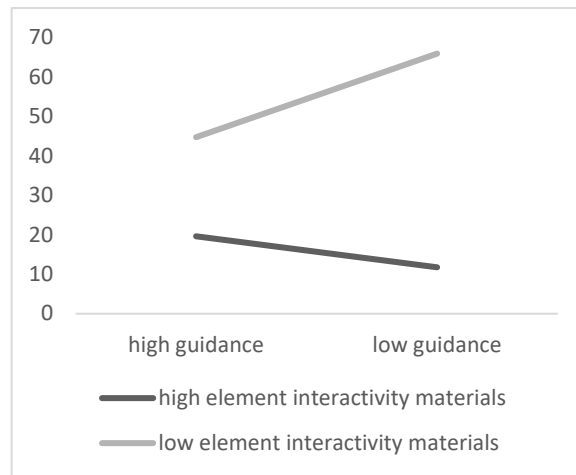


Figure 6. Relation Between Guidance and Element Interactivity on The Delayed Test for Year 10 Students

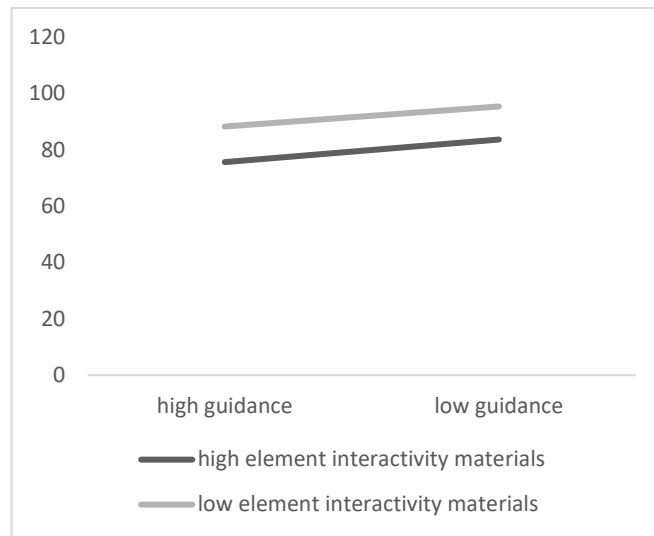


Figure 7. Relation Between Guidance and Element Interactivity on The Delayed Test for Year 11 Students