The Journal of Experimental Education
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/vjxe20

The Effect of Worked Examples When Learning to Write Essays in English Literature

Suna Kyun [a] , Slava Kalyuga [a] & John Sweller [a]
[a] University of New South Wales , Australia
Published online: 13 May 2013.

To cite this article: Suna Kyun , Slava Kalyuga & John Sweller (2013): The Effect of Worked Examples When Learning to Write Essays in English Literature, The Journal of Experimental Education, 81:3, 385-408

To link to this article: http://dx.doi.org/10.1080/00220973.2012.727884

THE JOURNAL OF EXPERIMENTAL EDUCATION, *81*(3), 385–408, 2013

ISSN: 0022-0973 print/1940-0683 online
DOI: 10.1080/00220973.2012.727884

Routledge
Taylor & Francis Group

LEARNING, INSTRUCTION, AND COGNITION

The Effect of Worked Examples When Learning to Write Essays in English Literature

Suna Kyun, Slava Kalyuga, and John Sweller
University of New South Wales, Australia

Worked examples, commonly used in technical domains, are rarely used in language areas such as English literature. In 3 experiments, Korean university students for whom English was a foreign language received worked examples intended to facilitate problem solving in the ill-structured domain of English literature. During the learning phase, half of the students were presented conventional essay questions that they were asked to answer. The other half of the students were presented the same questions along with model answers that they were asked to study, followed by similar questions that they had to answer themselves. All students then were asked to answer retention, near and far transfer tests. Relatively more knowledgeable students were assigned to Experiment 1 than to Experiment 2, who, in turn, were more knowledgeable than were the students in Experiment 3. Results indicated that the effectiveness of worked examples increased with decreasing student knowledge.

Keywords *cognitive load theory, ill-structured problems, literature instruction, problem solving, worked examples*

INTRODUCTION

A worked example provides explicit guidance indicating how to solve a particular problem, consisting of the problem statement along with a possible solution (Hilbert & Renkl, 2009; Renkl, Atkinson, & Große, 2004). Instruction using worked examples is considered an effective method of initial skill acquisition and application or transfer to new problems (Schwonke, Renkl, Salden, & Aleven, 2011; Stark, Mendl, Gruber, & Renkl, 2002). The worked example effect, in turn, occurs when a randomized, controlled experiment indicates that students learn more from studying worked examples than solving the equivalent problems (e.g., Schwonke et al., 2009). It is one of the instructional effects generated by cognitive load theory (Beckmann, 2010; Sweller,

Address correspondence to John Sweller, School of Education, University of New South Wales, Sydney, NSW 2052, Australia. E-mail: j.sweller@unsw.edu.au

2010, 2011, 2012; Sweller, Ayres, & Kalyuga, 2011), an instructional theory that is based on our knowledge of human cognitive architecture.

Worked Examples in Well- and Ill-defined Domains

Whereas some problems are considered well-defined because the given state, goal-state, and problem-solving operators are clearly specified, others are considered ill-defined because the given state is incompletely specified, the goal state is specified to an even lesser extent, and the problem solving operators are unspecified (Goel, 1992a). Hence, ill-defined problems require ill-structured representations, whereas well-defined problems require well-structured representations that consist of the specified states and problem-solving operators (Goel, 1992b, Greeno, 1976). In terms of problem representation, mathematics, science, and related technical domains are classified as well-defined, whereas literature, history, social studies, and related nontechnical domains are classified as ill-defined.

There have been many studies in technical areas such as mathematics or science demonstrating the effectiveness of worked examples using well-structured, well-defined problems (see Sweller et al., 2011). Recently, the effect of worked examples has been tested in ill-defined domains such as music (e.g., Owen & Sweller, 2008), design history (e.g., Rourke & Sweller, 2009), social psychology (e.g., Hübner, Nückles, & Renkl, 2010), medical domains (e.g., Stark, Kopp, & Fischer, 2011) and learning argumentation (e.g., Schworm & Renkl, 2007). In research on worked examples in ill-defined problem areas, a written format providing solutions for problems was regarded as worked examples. Although well-defined problems have well-structured representations consisting of the specified given state, goal-stage, and problem-solving operators, ill-defined problems have ill-structured representations which not only have partially specified states or problem solving operators but also sometimes omit explicit descriptions of those components (Best, 1989, p. 455). For this reason, the nature of ill-structured representations of ill-defined problems needs to be considered when they are organized as worked examples. A worked example in an ill-defined problem area can be expected to look different to a worked example in a well-defined problem area (Schworm & Renkl, 2007). Nevertheless, the one essential commonality must be preserved: All worked examples from well- or ill-defined problem areas must provide a solution to a problem that is presented to learners.

As indicated earlier, the worked example effect has primarily been demonstrated using mathematics, science, and technical information in well-defined, problem-solving domains but from a theoretical perspective, it should be obtainable using any class of information, regardless of differences in the task environments and problem spaces of well-defined and ill-defined problems (Goel, 1992a; Greeno, 1976). The current work tests the hypothesis that the worked example effect can be obtained in the ill-defined area of English literature, an area in which the effect has not previously been demonstrated and in which the construction of worked examples can pose particular challenges. A contrary hypothesis is that worked examples may hamper the learning of ill-defined domains (Spiro & DeSchryver, 2009). A successful demonstration of the worked example effect using ill-defined problems would suggest that human cognitive architecture does not distinguish between learning domains and that learning and problem solving do not differ depending on the nature of the learning domain. Appendix A provides examples of literature worked examples used in the current experiments. These worked examples were used in a series of randomized, controlled experiments that tested for the worked example effect using English

literature as the subject domain. In each of three experiments, we looked at the consequences of having students study these worked examples rather than answer questions themselves. We begin by outlining the human cognitive architecture on which cognitive load theory and its recommendations concerning worked examples are based.

Human Cognitive Architecture From an Evolutionary Perspective

Cognitive load theory considers human cognitive architecture from an evolutionary perspective (Sweller, 2003; Sweller & Sweller, 2006). According to Geary (2007, 2008, 2012), knowledge can be differentiated between biologically primary and biologically secondary information. Primary knowledge is knowledge we have evolved to acquire over many generations. It can be acquired effortlessly, rapidly, and unconsciously. To acquire primary knowledge, we do not require any specific educational procedures. Learning to listen to and speak a native language provides an example. On the other hand, biologically secondary knowledge is knowledge we require for cultural reasons but which we have not specifically evolved to acquire. Secondary knowledge must be explicitly taught and learned in schools and in other educational institutions with considerable conscious effort, often requiring motivational encouragement. Without appropriate institutional and procedural support, secondary knowledge will not be acquired by most members of a society. Learning to read and write provide examples of this category of knowledge and skills as does an appreciation of literature, the topic of the present set of experiments. Unlike listening and speaking, learning to read and write or to appreciate literature will not occur without deliberate instruction because we have not evolved to automatically acquire biologically secondary skills. The human cognitive architecture on which cognitive load theory is based applies solely to biologically secondary knowledge and accordingly, the theory applies solely to biologically secondary knowledge taught in schools and other educational institutions (Sweller, 2008; Sweller & Sweller, 2006).

When dealing with biologically secondary knowledge, human cognitive architecture processes information in a manner analogous to the way in which evolution by natural selection processes information. Both systems are examples of natural information processing systems and share a common fundamental logic represented by the following five principles (Sweller & Sweller, 2006).

First, the *information store principle* states that natural information processing systems require a very large store of information to govern cognitive activity. In the case of human cognition, what we perceive, how we think, and how we solve problems are heavily determined by what has been learned and stored in long-term memory (De Groot, 1946/1965; Sweller, 2006). Second, the majority of information held in long-term memory is obtained from the long-term memory of other people through the *borrowing and reorganizing principle*. Humans listen to what other people say, read what they write, and imitate what they do (Rourke & Sweller, 2009). Third, when information cannot be borrowed and must be created, the *randomness-as-genesis principle* provides the necessary mechanism. The only mechanism to create knowledge is random generation followed by test of effectiveness. Effective, novel problem solutions are stored in long term memory while ineffective solutions are jettisoned (Sweller, 2006). Fourth, the *narrow limits of change principle* states that only a few items of novel information can be dealt with when randomly generating and testing them for effectiveness in a severely limited working memory (Torcasio & Sweller, 2010). This is why only small changes to long-term store are likely to be effective and so

changes to long-term memory must be slow and incremental. Humans use working memory when dealing with novel information from the environment, and working memory is severely limited in capacity when dealing with that information (Rourke & Sweller, 2009). Fifth, according to the *environmental organizing and linking principle*, however, when dealing with organized information from long-term memory, there are no limitations to working memory (Owen & Sweller, 2008). The ultimate purpose of human cognitive architecture and instruction is to allow us to deal with complex problems in particular environments (see Sweller & Sweller, 2006).

These five principles constitute a human cognitive architecture that provides us with instructional implications. On the basis of this architecture, the primary function of instruction is to build or alter schemas in long-term memory. If there is no positive change in long-term memory, nothing has been learned and instruction has failed. The most efficient changes of knowledge held in long-term memory can be generated through the borrowing and reorganizing principle. The majority of human learning occurs through this principle resulting in knowledge stored in long-term memory. That knowledge can be used to efficiently direct cognitive activity and solve problems through the environmental organizing and linking principle.

Worked examples provide the ultimate instantiation of the borrowing and reorganizing principle (Sweller, 2006). In contrast, problem solving equally provides an instantiation of the randomness-as-genesis principle. Presenting learners with information through worked examples reduces the random generate and test procedure that is characteristic of problem solving. Being presented with information rather than attempting to generate it should reduce working memory load. Only the elements contained within a worked example need to be considered rather than the potentially large number of elements that could be generated during problem solving. Thus, by using the borrowing and reorganizing principle rather than the randomness-as-genesis principle, working memory load can be reduced. The worked example effect has been demonstrated on many occasions in technical domains such as mathematics, science, and computer programming (e.g., Carrol, 1994; Cooper & Sweller, 1987; Crippen & Earl, 2007; Gerjets, Scheiter, & Catrambone, 2006; Große & Renkl, 2007; Kalyuga, Chandler, Tuovinen, & Sweller, 2001; Kyun & Lee, 2009; Paas & van Merriënboer, 1994; Quilici & Mayer, 1996; Sweller & Cooper, 1985; van Gog, Paas, & van Merriënboer, 2006, 2008; Ward & Sweller, 1990; Zhu & Simon, 1987).

Worked Example Effect

The classical worked example effect is tested by comparing two conditions. In the problem-solving condition, learners are taught a new area such as new geometry theorems and then, during an acquisition phase, presented with a series of problems requiring the use of those theorems. The problems are frequently presented in pairs of structurally similar problems with minor changes in, for example, the size of angles. In the worked example condition, the initial teaching phase is identical to the teaching phase for the problem-solving group but differs during the acquisition phase. The first of each pair of acquisition problems is usually presented as a worked example. The problem is presented along with a detailed solution. In the test phase, both groups are tested using a conventional, problem-based test in which a series of problems must be solved. The worked example effect is obtained if the worked example condition that has been required to solve fewer problems than the problem-solving condition, nevertheless, obtains higher test scores.

While according to human cognitive architecture as used by cognitive load theory, the worked example effect should be equally obtainable in ill-defined as well-defined areas (Rourke & Sweller, 2009), it has been suggested the success of explicit instructional guidance approaches such as worked examples in well-structured domains cannot extend to ill-structured domains (e.g., Spiro & DeSchryver, 2009). However, there is no evidence that learning and problem solving differ substantially depending on the learning domain. According to our current understanding of human cognitive architecture, the procedures by which we learn and solve problems are identical for well- and ill-defined problems. In both cases, learning to solve problems is a domain specific activity in which students learn to recognize a problem state and learn appropriate moves for that state. We learn to solve problems by recognizing problem states and the appropriate moves associated with them. Unless this information is stored in long-term memory, problem-solving skill does not develop. The clearest way to ensure that learners have acquired this information is to present them with worked examples. Accordingly, worked examples should be presented to novice problem solvers and those worked examples should be just as effective in ill-defined as in well-defined areas.

In this context, Hübner, Nückles, and Renkl (2010) found that writing instruction using a worked example (i.e., a written example of a learning journal to use their terms) enhanced students' comprehension in the domain of social psychology. While there is frequently an assumption that writing automatically can contribute to learning (Applebee, 1984; Emig, 1977; Tynjälä, Mason, & Lonka, 2001), Hübner and colleagues (2010) attempted to use a worked example to reinforce students' understanding in a domain-specific area. In their study, after a videotaped lecture on social psychology, students were asked to either write a journal on the relevant issue treated in the lecture or to study a worked example. Students who received a worked example to study performed significantly better in a transfer comprehension test than those who practiced writing a journal themselves.

In the present context, an essay question is considered to be an ill-defined problem requiring an ill-defined solution and model answers to an essay question can be classed as worked examples. We hypothesized that although a model answer is one of a large number of possible answers to an essay question, students will learn more from studying that answer than by writing an essay themselves in ill-defined areas such as history, literature, and social studies. We conducted the present experiments to test this hypothesis using children's English literature as the curriculum domain.

EXPERIMENT 1

Experiment 1 tested the hypothesis that students learning about children's English literature would obtain higher scores on test essays if they had access to model essays that answer questions during the learning phase compared with students who were presented the same essay questions that they had to answer themselves. Cognitive load theory places its emphasis on domain-specific learning. The information store principle assumes that skill derives from the acquisition of a large number of domain-specific schemas held in long-term memory. In the present context, we assume that students will learn how to answer particular essay questions associated with children's English literature. We hypothesise that initially, studying worked examples will facilitate learning more than generating answers because the reduction in working memory load associated with studying

an answer using the borrowing and reorganizing principle imposes a lower load than generating an answer using the randomness-as-genesis principle.

In the present experiment, three different perspectives of a model essay to each question were given to students before they tried to solve problems on their own. Through the process of studying these model essays (defined as worked examples), students were expected to learn what constitutes a good answer to the particular question in this domain.

METHOD

Participants

The participants were 63 (59 Korean, 2 French, and 2 Chinese) female university students (ages ranged from 19 to 25 years) enrolled in the course "Children's English Literature" at the Sookmyung Women's University, Seoul, Korea. Almost all students had formal English ability scores and the few students who did not have these scores were eliminated from the analysis. Their English test average scores were 258.8 for the Computer-Based Test of the Test of English as a Foreign Language (maximum score: 300), 100.13 for the Internet-Based Test of the Test of English as a Foreign Language (maximum score: 120), and 895 for the Test of English for International Communication (maximum score: 990).

The students' language proficiency was sufficiently high to discuss and write their thoughts in English. On the contrary, their experience of formal education in English literature varied. Almost all students had taken English literature subjects including children's English literature as college courses. Among them, some had taken more than four literature subjects as college courses. Students were randomly assigned to one of the two conditions except that their knowledge of English and English literature was taken into account to ensure equivalent levels of knowledge in the two groups (see the "Experimental Procedure" section). There were 32 participants in the worked-example group and 31 in the problem-solving group.

Materials

The paper-and-pencil based materials consisted of a pretest questionnaire, four pages of learning materials including a mental effort rating questionnaire used in the learning phase, and two pages of retention and transfer test materials used in the posttest phase.

Pretest questionnaire

The pretest questionnaire was used to collect information about each participant's age, grade, affiliated schools and colleges, the number of previously taken college courses in English literature including children's English literature, and English test scores on the Computer-Based Test of the Test of English as a Foreign Language, the Internet-Based Test of the Test of English as a Foreign Language, and the Test of English for International Communication.

Learning materials

The learning materials were produced by the professor who taught this subject. The materials were actual teaching materials used in a realistic teaching and learning environment. A separate set of learning materials was developed for the worked example and problem-solving formats. Four essay questions were included in each format based on two pairs of similar problems. In the worked example format, a worked out, model answer to the essay question was presented first (Appendix A) followed by a similar problem-solving practice exercise. In the problem-solving format, the same problem-solving questions were included but model answers were not provided. Before each paired set of worked examples and problems, the worked example format included the following instruction: "Question 1 and 2 are closely related to each other. After studying Question 1 and its possible answers, answer Question 2. When you answer Question 2, you can, of course, make use of answers to Question 1. Please note the time limit for studying and answering Questions 1 and 2." The problem-solving condition included the following instructions: "Please answer the two following questions. Note that the two questions are very similar and so you should feel free to answer the two questions in a similar manner. Please note the time limit for each question."

Rating of mental effort

We used Paas and van Merriënboer's (1994) 9-point subjective rating scale to measure levels of cognitive load following the learning acquisition phase. Students were required to rate the question, "How easy or difficult was it to study and solve these tasks?" from 1 (*extremely easy*) to 9 (*extremely difficult*). It was hypothesized that difficulty ratings and hence cognitive load would be higher in the essay writing than the essay studying condition because of the increased difficulty in generating an answer through the random generate and test principle than analyzing an answer through the borrowing and reorganizing principle.

Posttest materials

The posttest consisted of retention, near and far transfer problems. The retention test problem was identical to one of the four problems in the learning phase:

> Folktales and fairytales show universally familiar female characters. What do you think are the characteristics of the female figures, in relation to their male characters? Are fairytale females typically "feminine" or do they demonstrate different characteristics from "traditional feminine" characters?

The near transfer problem was similar to that in the learning phase:

> In contemporary times, picture books are a very popular genre of the children's book market and children are encouraged to access many of those books particularly from an early stage. What do you think of this classification of picture books as children's books? Do you think it is valid or useful? Do you think "picture books" are mainly for children? If yes, why do you think so? If not, what are your reasons?

The far transfer problem was different from the problems in the learning phase:

> There are two arguments, regarding the teaching of literature to the poor in a society. One party argues that the poor cannot afford to study literature or philosophy, since what they need is a job to earn income and to manage their life. This first party says that literature is simply too luxurious an item for the poor. The other party argues that access to literature and philosophy education is as necessary (or even more necessary) for the poor as it is for the middle and the upper class. They argue that in the long term, instruction in humanity subjects is the very way for the poor to escape their poverty, since what they ultimately need is to change their mindset and to be aware of purpose in their life. What do you think of these two arguments? Discuss your opinion.

It was important to use not only a near transfer problem (similar to those used in the learning phase) but also a far transfer problem in order to evaluate the ability of students to apply the acquired knowledge in a domain different from children's English literature. The posttest questions were also produced by the professor who teaches this subject. These questions were actual problems used in a quiz or a mid-term exam.

Procedure

This experiment included four phases conducted during the second and third week of a semester in teaching children's English literature to university students. In Phase 1 (pretest), students were allocated 15 min to complete the pretest questionnaire on their age and prior academic history. In allocating students to groups, we considered students' English score (English proficiency) and the number of previous college courses taken in English literature including children's English literature. We initially organized students' English ability level on the basis of an English conversion score table between the Computer-Based Test of the Test of English as a Foreign Language, the Internet-Based Test of the Test of English as a Foreign Language, and the Test of English for International Communication and examined whether or not students had experience in the area of children's English literature in previous semesters and the credit they obtained. Considering these two factors, each student was paired with a student of a similar level based on their sequential ranks in the sample in terms of English ability and previous experience in the area of English literature and then members of each pair were randomly assigned to one of two conditions, a worked-example group (32 participants) and problem-solving group (31 participants).

In Phase 2 (the lecture phase), the two groups were provided a lecture by the same lecturer on an "Historical introduction to children's English literature." The lecturer provided examples of many selections of children's books, their historical meanings and status. Also, students were asked to pay attention to how the history of children's literature in western society has changed and what kind of ideas, cultures and social phenomena have been involved. The lecture lasted for approximately 50 min.

In Phase 3 (the learning phase) 2 days later, students in the worked example and problem-solving groups were presented their learning materials as described earlier. All students had 60 min to complete this phase (15 min per problem). At the end of this phase, the subjective self-rating scale was administered to students.

In Phase 4 (the posttest) 5 days later, both groups were provided with the same retention, near and far transfer problems and were given 45 min to complete these problems (15 min per

problem). Every 15 min during the learning and posttest phases, the instructor informed students of the time and asked them to proceed to the next problem.

Scoring procedure

The student answers to two essay questions during the learning phase (corresponding to Questions 2 and 4 in the worked example format) and their answers to the retention, near and far transfer questions during the posttest phase were scored by two professional lecturers in the area of English literature. The International English Language Testing System academic writing test rubrics (Appendix B), which is offered by University of Cambridge English for Speakers of Other Languages Examinations, were used as a marking standard. The academic writing task of International English Language Testing System is designed in terms of the format of a university essay that can be allocated to the genre "written argument" (see Shaw & Falvey, 2008, p. 202). In the present study, students were expected to argue their perspectives to the given questions in the area of children's English literature. Because this characteristic of writing was considered a similar characteristic of International English Language Testing System academic writing, the rubrics have been chosen as the assessment tool to assess the results of students' writing.

According to the International English Language Testing System Academic writing test rubrics, each essay was marked in terms of four categories—*task response*, *coherence and cohesion*, *lexical resource*, and *grammatical range and accuracy*—and was differentiated into 10 levels (from 0 to 9) for each subcategory. Therefore, the maximum score was 36 points for each marker. Each student's score was based on the sum of the two markers, therefore the maximum possible score was 72. Pearson product-moment correlation coefficients between the two raters' scores for Tests 1 and 2 during the learning phase were .69 and .77, respectively; for the retention, near and far transfer tests during the posttest phase, the coefficients were .74, .69, and .82 respectively.

RESULTS AND DISCUSSION

Out of 63 participants, 4 students who did not complete all phases of the experiment were excluded from the analysis. For the remaining 59 participants, Table 1 shows the mean scores and standard deviations for each of the two groups for the dependent variables. A one-way analysis of variance (ANOVA) was conducted with instructional method (worked example vs. problem solving) as the between-subject factor.

There were significant differences between groups for the two similar questions during the learning phase with $F(1, 58) = 8.26$, MSE $= 57.20$, $p < .01$, partial $\eta^2 = .13$, for Question 1, and $F(1, 58) = 8.9$, MSE $= 52.69$, $p < .01$ partial $\eta^2 = .14$, for Question 2, both favoring the worked-example group.

There were no significant differences on any of the posttest problems: $F(1, 58) = 1.73$, MSE $= 45.70$, $p = .19$, partial $\eta^2 = .03$, for the retention question; $F(1, 58) = .17$, MSE $= 38.32$, $p = .69$, partial $\eta^2 = 0$, for the near transfer question; and $F(1, 58) = .68$, MSE $= 73.26$, $p = .41$, partial $\eta^2 = 0$, for the far transfer question.

Regarding the mental effort invested during learning acquisition, there was no significant difference between groups, $F(1, 58) = 1.21$, MSE $= 2.78$, $p = .28$, partial $\eta^2 = .02$.

TABLE 1
Mean Scores and Standard Deviations in Experiment 1

	Worked example group (n = *31*)		*Problem-solving group* (n = *28*)	
	M	SD	M	SD
Learning phase				
Similar question 1	46.25	6.92	40.79	8.22
Similar question 2	45.16	7.24	39.50	7.23
Mental effort	5.71	1.15	6.14	1.46
Posttest phase				
Retention test	43.71	7.03	41.39	6.44
Near transfer test	43.45	6.66	44.11	5.63
Far transfer test	42.52	6.90	40.68	10.10
Efficiency score				
Similar question 1	9.14	4.58	7.20	2.94
Similar question 2	8.84	4.27	6.92	2.50
Retention test	8.61	4.38	7.26	2.67
Near transfer test	8.54	4.17	7.73	2.70
Far transfer test	8.39	4.48	7.15	3.03

Students' performance and mental effort scores were combined using the likelihood model of Hoffman and Schraw (2010) to calculate efficiency. For each student, the mental effort rating (M) was divided into the performance measure (P) to provide an indicator of cognitive efficiency (E), using the formula, $E = P/M$ (Kalyuga & Sweller, 2005). A higher performance level with less mental effort required for learning was considered as evidence of a higher level of efficiency. It should be noted that here and in all subsequent measures of efficiency using the likelihood ratio, with one exception, the same pattern of significant effects was obtained using the Paas and van Merriënboer (1993) formula, $E = (Zp - Zm)/2^{1/2}$, where Zp is a performance score and Zm is a mental effort rating, both expressed as Z scores.

An ANOVA performed on these data provided a marginally significant effect on the first similar problem during the learning phase, $F(1, 58) = 3.64$, MSE = 15.15, $p = .06$, partial $\eta^2 = .06$ (this difference was significant using the Paas and van Merriënboer (1993) formula, $p = .017$, partial $\eta^2 = .10$) and a significant effect on the second problem, $F(1, 58) = 4.31$, MSE = 15.57, $p < .05$, partial $\eta^2 = .07$. There were no significant effects using the retention, near and far transfer tests during the posttest phase: $F(1, 58) = 1.99$, MSE = 13.46, $p = .16$, partial $\eta^2 = .03$; $F(1, 58) = .76$, MSE = 12.59; $p = .39$, partial $\eta^2 = .01$ and $F(1, 58) = 1.53$, MSE = 14.91, $p = .22$, partial $\eta^2 = .02$, respectively.

In Experiment 1, the students' levels of knowledge of literature, including children's English literature were high, which may have rendered the problems too easy for many students. We know from previous research on the expertise reversal effect that the worked example effect may be unobtainable using high knowledge learners (Kalyuga et al., 2001). Worked examples can be redundant for such learners. Although the present results during the learning phase clearly indicate that the worked examples were not redundant, they may not have been as effective as might be obtained with less knowledgeable learners. Learners were influenced by worked examples on the immediate similar questions, but not affected on the retention and transfer tests. Students who

were highly knowledgeable with respect to these literature problems could use their knowledge to successfully respond to the test problems regardless of the format used in the learning phase. The failure to find a significant difference using the subjective rating scale suggests that these students did not gain a great deal from the presentation of the worked examples. Less knowledgeable students for whom the worked-example procedure was originally devised might need the worked examples of the learning phase sufficiently to use them during the test phase and so yield test effects. Experiment 2 tested this hypothesis by using less knowledgeable learners.

EXPERIMENT 2

Experiment 2 used a relatively less advanced group of students in the area of children's English literature. Although competent in English, they had not previously taken the subject of children's English literature as a college course although most of them were students majoring in English literature and Korean literature. The purpose of this experiment was to investigate whether a stronger worked example effect than that obtained using relatively more knowledgeable students could be obtained using a less knowledgeable group.

METHOD

Participants

The participants were 62 (60 Korean and 2 Chinese) female university students (ages ranged from 19 to 29 years) enrolled in the course titled "20th Century American Literature" at the Sookmyung Women's University, Seoul, Korea. More than half of the students were from the College of English Language and Literature; however, all of them lacked formal education in the domain of children's English literature. Of the 62 students, 37 were from the College of English Language and Literature, and 25 students were from other schools such as Korean Language and Literature and Education located in the College of Arts and Social Sciences. Of the 62 students, 27 were in their third and fourth university year and were preparing for graduation. In this experiment, all students were required to complete a modified version of the Test of English for International Communication during the pretest phase as well as a survey on their basic background information. Students' average test score was 45.19 ($SD = 8.17$, possible maximum score = 60), indicating a high level of competence in English. There were no students who had experience in the area of children's English literature. Students were randomly assigned to one of two conditions based on their English proficiency score from the modified version of the Test of English for International Communication. Half of the students served in the worked-example group ($M = 45.1$, $SD = 8.60$) and half served in the problem-solving group ($M = 45.3$, $SD = 7.86$).

Materials and Procedure

The materials and experimental procedure were identical to those used in Experiment 1, except for the use of the modified version of the Test of English for International Communication to measure students' English ability directly. This modified version consisted of grammar and vocabulary (40 questions), and reading comprehension (20 questions) that were extracted from the original

TABLE 2
Mean Scores and Standard Deviations in Experiment 2

	Worked example group (n = *28*)		*Problem-solving group* (n = *27*)	
	M	SD	M	SD
Learning phase				
Similar question 1	41.93	7.00	36.48	10.68
Similar question 2	43.68	8.03	37.37	8.26
Mental effort	5.18	1.57	5.59	1.47
Posttest phase				
Retention test	41.96	6.63	36.89	9.61
Near transfer test	40.43	5.63	38.93	0.91
Far transfer test	35.89	11.38	36.52	12.12
Efficiency score				
Similar question 1	9.22	4.56	7.04	3.05
Similar question 2	9.70	5.08	7.27	2.77
Retention test	9.23	4.68	7.07	2.99
Near transfer test	8.87	4.36	7.48	2.27
Far transfer test	7.80	4.73	6.95	3.16

test. The maximum possible score was 60 (1 point per question). Students were given 25 min to complete the 60 problems.

Scoring procedure

The data analyses were identical to those used in Experiment 1. The correlation coefficients between the two markers indicated Pearson product-moment correlation coefficients for Tests 1 and 2 during the learning phase were .78 and .80, respectively. For the retention, near and far transfer tests during the posttest phase, the coefficients were .83, .77, and .94, respectively. (The inferential statistics reported in the next section provide the ultimate test of reliability.)

RESULTS AND DISCUSSION

Of the 62 participants, 7 did not complete all phases of the experiment and were excluded from the analysis: For the remaining 55 participants, Table 2 shows the mean scores and standard deviations for each of the two groups for the dependent variables. An ANOVA was conducted with instructional method (worked example vs. problem solving) as the between-subject factor.

There were significant differences for the two similar questions during the learning phase: $F(1, 54) = 5.04$, MSE = 80.88, $p < .05$, partial $\eta^2 = .09$, for Question 1, and $F(1, 54) = 8.25$, MSE = 66.31, $p < .01$, partial $\eta^2 = .13$, for Question 2, both favoring the worked-example group.

There was a significant difference for the retention posttest problem, $F(1, 54) = 5.23$, MSE = 67.65, $p < .05$, partial $\eta^2 = .09$, favoring the worked-example group. However, there were no significant differences on either the near transfer question, $F(1, 54) = 1.14$, MSE =

27.22, $p = .29$, partial $\eta^2 = .02$, nor the far transfer question, $F(1, 54) = .39$, MSE = 138.07, $p = .84$, partial $\eta^2 = 0$.

Regarding the mental effort invested during learning acquisition, there was no significant difference on mental effort ratings between the two groups, $F(1, 54) = .90$, MSE = 2.31, $p = .35$, partial $\eta^2 = .02$.

An ANOVA performed on the efficiency measure data using the likelihood model provided significant effects on the two similar problem tests during the learning phase, $F(1, 54) = 4.31$, MSE = 15.16, $p < .05$, partial $\eta^2 = .08$ and $F(1, 54) = 4.79$, MSE = 16.91, $p < .05$, partial $\eta^2 = .08$ respectively, and on the retention test during the posttest phase, $F(1, 54) = 4.12$, MSE = 15.52, $p < .05$, partial $\eta^2 = .07$. There were no significant effects on the near and far transfer tests during the posttest phase, $F(1, 54) = 2.18$, MSE = 12.21, $p = .15$, partial $\eta^2 = .04$ and $F(1, 54) = .594$, MSE = 16.31, $p = .44$, partial $\eta^2 = .01$, respectively.

In Experiment 2, a significant difference was found between the two groups on the retention posttest. Also, similar to Experiment 1, the students who received worked examples performed significantly better than those who were required only to solve problems on the two questions during the training session. With less knowledgeable students, worked examples had an effect on learner performance on tasks similar to those used in instruction. However there was no such effect on either near or far transfer problems. Furthermore, as was the case for Experiment 1, there was no difference between conditions using the subjective ratings, suggesting that for these learners, the provision of worked examples did not substantially reduce cognitive load. Experiment 3 used less knowledgeable learners than Experiment 2.

EXPERIMENT 3

Experiment 2 found that the worked example condition was beneficial for less advanced students in the area of children's English literature, but only on problems similar to those studied. The aim of Experiment 3 was to investigate if the effect of worked examples could be further increased by using less knowledgeable students in the area of children's English literature than those who participated in the previous two experiments. In addition, the number of sets of learning materials was reduced in half with one pair used instead of two and instructional time was doubled from 15 to 30 min per problem. It was expected that this more extensive instruction with less knowledgeable students would yield a stronger worked example effect.

METHOD

Participants

Participants were 129 female university students (ages ranged from 19 to 22 years) enrolled in a general psychology course at the Sookmyung Women's University, Seoul, Korea. Students had various backgrounds. Most of the students were from the College of Natural Sciences (School of Mathematics, School of Statistics, School of Physics, School of Chemistry, or School of Computer Science) and the remaining students were from Journalism, Economics, or the College of Law. Of 129 students, 116 were in their first or second university year who had recently

graduated from high school. All of them were enrolled in general psychology as a liberal arts subject and lacked formal education in the domain of children's English literature: They had not previously taken any subject of literature, including children's English literature as a college course. In this experiment, all students were required to complete a modified version of the Test of English for International Communication during the pretest phase as well as a survey on their basic background. According to the results of the modified test (including tests of English grammar, vocabulary, and reading comprehension), almost 30% of students were excluded in analysis because of their very low English score. The remaining students' average test score was 44.5 ($SD = 6.21$; possible maximum score 60), indicating a sufficiently high level of competence in English to deal with the materials. Students were randomly assigned to one of two conditions based on their English proficiency score. We allocated 45 students to the worked-example group ($M = 44.5$, $SD = 6.55$) and 46 to the problem-solving group ($M = 44.5$, $SD = 6.00$).

Materials and Procedure

The learning and test materials were identical to those used in Experiments 1 and 2, and the modified version of the Test of English for International Communication was identical to that used in Experiment 2. Regarding experimental procedures, there were two differences from the procedures used in Experiments 1 and 2. First, in the learning acquisition phase, the number of sets of learning materials that students had to study was reduced from two pairs to one pair and the time that students could spend on studying learning materials was increased from 15 to 30 min per problem. The reduced number of tasks and, accordingly, the increased time per task, provided students with more extensive instruction on each task. Second, the posttest was conducted twice. The first posttest was administered two days after the learning acquisition phase, and the second posttest was conducted 1 week later.

Scoring Procedure

The data analyses were identical to those used in Experiments 1 and 2. The correlation coefficients between the two markers indicated that Pearson product-moment correlation coefficients were .64 for the similar problem test during the learning phase, .67 and .47, respectively, for the retention and near transfer test during the first posttest phase, and .69, .52, and .61, respectively, for the retention, near and far transfer tests during the second posttest phase.

RESULTS AND DISCUSSION

Of 129 participants, 58 were excluded from the analysis because they either received very low English proficiency scores in the pretest (38 students) or did not complete all phases of the experiment (20 students). For the remaining 71 participants, Table 3 shows the mean scores and standard deviations for each of the two groups for the dependent variables. An ANOVA was conducted with instructional method (worked example vs. problem solving) as the between-subject factor.

There were significant differences for the similar question during the learning phase: $F(1, 70) = 14.51$, MSE $= 47.70$, $p < .001$, partial $\eta^2 = .17$, and for the retention question during the first

TABLE 3
Mean Scores and Standard Deviations in Experiment 3

	Worked example group ($n = 32$)		Problem-solving group ($n = 39$)	
	M	*SD*	*M*	*SD*
Learning phase				
Similar question	36.25	7.82	29.97	6.06
Mental effort	4.91	1.42	5.54	0.88
1st posttest phase				
Retention test	35.47	7.40	31.38	5.85
Near transfer test	34.81	6.77	33.64	5.30
2nd posttest phase				
Retention test	38.34	6.40	33.85	7.86
Near transfer test	38.47	7.08	35.82	5.50
Far transfer test	38.28	6.55	36.44	6.30
Efficiency score				
Similar question	8.45	4.42	5.54	1.39
1st retention test	8.16	3.77	5.85	1.59
1st near transfer test	8.05	3.88	6.21	1.36
2nd retention test	8.83	4.09	6.25	1.79
2nd near transfer test	8.74	3.83	6.67	1.64
2nd far transfer test	8.64	3.54	6.75	1.67

posttest phase, $F(1, 70) = 6.75$, MSE = 43.44, $p < .05$, partial $\eta^2 = .09$, favoring the worked-example group. However, there was no significant difference on the near transfer question, $F(1, 70) = .67$, MSE = 36.06, $p = .416$, partial $\eta^2 = .01$.

There was a significant difference for the retention test during the second posttest phase, $F(1, 70) = 6.78$, MSE = 52.44, $p < .05$, partial $\eta^2 = .09$. Also, there was a marginally significant difference for the near transfer test during the second posttest phase, $F(1, 70) = 3.15$, MSE = 39.12, $p = .08$, partial $\eta^2 = .04$, favoring the worked-example group. However, there was no significant difference on the far transfer test during the second posttest phase, $F(1, 70) = 1.46$, MSE = 41.10, $p = .23$, partial $\eta^2 = .02$.

Regarding the mental effort invested during learning acquisition, there was a significant difference between groups, $F(1, 70) = 5.25$, MSE = 1.34, $p < .05$, partial $\eta^2 = .07$, indicating that students who had studied worked examples invested a lower mental effort than the students who had practiced conventional problems.

An ANOVA performed on the efficiency measure data using the likelihood model provided significant effects on all test problems, $F(1, 70) = 15.18$, MSE = 9.84, $p < .001$, partial $\eta^2 = .18$, for the similar problem during the learning phase; $F(1, 70) = 12.04$, MSE = 7.78, $p < .01$, partial $\eta^2 = .15$ and $F(1, 70) = 7.64$, MSE = 7.76, $p < .01$, partial $\eta^2 = .10$, respectively, for the retention and near transfer test during the first posttest phase; $F(1, 70) = 12.53$, MSE = 9.29, $p < .01$, partial $\eta^2 = .15$, $F(1, 70) = 9.31$, MSE = 8.08, $p < .01$, partial $\eta^2 = .12$ and $F(1, 70) = 8.79$, MSE = 7.18, $p < .01$, partial $\eta^2 = .11$, respectively, for the retention, near and far transfer test during the second posttest phase. All of these results favored the worked-example group.

In Experiment 3, a marginally significant difference was found between the two groups on the near but not the far transfer posttest. The students who received worked examples performed significantly better than those who were required to solve problems on the similar problem during the training session. The students in the worked-example group performed significantly better than those in the conventional problem-solving group on both the first and second retention posttest. It is critical to note that, in this experiment, a significant difference was obtained between the two conditions on the subjective rating of task difficulty. That difference indicated that learners found that studying the worked examples reduced cognitive load compared to answering the equivalent questions. As a consequence, efficiency gains were obtained on all measures by the worked-example group compared to the problem-solving group.

GENERAL DISCUSSION

The purpose of the present study was to investigate the worked-example effect in the ill-structured domain of English literature. This study further extended the effects of worked examples from well-defined domains such as mathematics, science, and computer programming to an ill-defined domain.

The results obtained in this study convey theoretical and educational implications that can lead to effective instruction. We believe that providing learners with information that fully explains the concepts and procedures that students are required to learn (Kirschner, Sweller, & Clark, 2006) is necessary for problem-solving performance. This explicit instructional guidance can facilitate the acquisition of domain-specific knowledge which is the major determinant of skilled problem-solving performance. Students in language-based subjects such as social studies, humanities, history, or literature have traditionally been required to write essays as a form of practice. Often, they need to construct those essays with little guidance concerning what constitutes a good essay in a specific domain. This instructional technique was developed and became a dominant teaching paradigm before our current understanding of human cognitive architecture. The technique is heavily based on instructional processes associated with discovery learning and problem-based learning (Sweller, 2009).

The human cognitive architecture associated with biologically secondary knowledge suggests that the primary function of instruction is to assist in the acquisition of knowledge in long-term memory. That knowledge can be acquired through the randomness-as-genesis principle or the borrowing and reorganizing principle. Knowledge acquired through the randomness-as-genesis principle requires a generate and test procedure that imposes a heavy working memory load. Essay writing without guidance places a heavy emphasis on the randomness-as-genesis principle. In contrast, we have evolved to acquire biologically secondary knowledge from the borrowing and reorganizing principle. We are skilled at obtaining information from other people. The use of worked examples provides an ideal procedure for obtaining information from others.

The three experiments in this study support the suggestion that providing worked examples can be beneficial even in discursive domains such as literature. In all of the experiments, as measured by the questions in the training session, the students who received worked examples demonstrated that they learned significantly more than those who were required to construct their own answers without previous guidance. In Experiment 2, using less knowledgeable students, this superiority extended to the retention test in the posttest phase. In Experiment 3, using even

less knowledgeable students, there was an indication of this superiority on the near transfer test in the posttest phase. Furthermore, and critically, in Experiment 3 and only in Experiment 3, a significant difference was obtained between conditions on the subjective ratings of task difficulty. The worked-example group found their task of studying worked examples easier than the problem-solving group found answering the equivalent questions. In Experiments 1 and 2, using more knowledgeable learners this difference was not obtained.

While there were no significant differences on the near and far transfer problems through the three experiments and the effect sizes were small, it is noticeable that the effect sizes increased with decreasing learner expertise from Experiment 1 to the 3. Furthermore, using the instructional efficiency measure in which learners' performance and mental effort scores were combined using the likelihood model, there was an increasing effect size as the learners' prior knowledge decreased, along with significant differences on the both near and far transfer problems in Experiment 3.

The increased effectiveness of studying worked examples with decreased knowledge levels is predicted by cognitive load theory and is one of the bases for the expertise reversal effect. Only low knowledge learners require worked examples. For higher knowledge learners, worked examples are unnecessary and, depending on levels of expertise, learners may gain more from solving problems. An absolute measure of expertise is desirable to determine whether worked examples are likely to be beneficial. Some progress on such a measure is available for technical areas (Kalyuga & Sweller, 2004, 2005). A suitable measure is not yet available for the ill-defined curriculum area of the present experiments. In the absence of such a measure, the judgment of instructors must be used to determine whether worked examples are likely to be beneficial.

As far as we are aware, the present study is the first to attempt to study worked examples in the domain of literary analysis. To do so, we required a new type of worked examples very different from the classical worked examples used in technical domains. We have defined an essay question as an ill-defined problem and model answers to an essay question were classed as worked examples using the perspective of problem solving in ill-defined learning domains. In well-defined problem areas, a worked example is defined as a problem for which a solution is explicitly provided. We have used this definition to construct worked examples in the area of literature. An essay question provides the problem statement that is analogous to the problem statement found in mathematics or science problems. A model answer to the essay question provides a worked example that is analogous to a worked out problem solution in mathematics or science. Of course, depending on the learning domain, the representation of the worked examples is different because it must reflect the manner in which the discipline is organized. Accordingly, the worked examples in the present experiments differ from mathematics worked examples. This new type of worked example is influenced by the nature of the ill-defined learning domain of literature, in which the given, goal and problem solving operators are defined in more general terms resulting in a large number of acceptable solutions. Notwithstanding, the worked examples used in the present experiments still include the basic, definitional requirements of a worked example in that they provide learners with an explicit answer to a problem that students are required to study in order to improve subsequent problem-solving performance.

There are limitations to the present study that need to be addressed in subsequent work. Most of the limitations occur because the sole purpose of the experiments was to test the hypothesis, disputed by some, that worked examples could be effective in a literature-based area. The simple research design used in this study aimed to demonstrate that instruction that conformed to the

definition of worked examples could be constructed in the area of English literature and that those worked examples would prove to be superior to more conventional, problem-solving instruction. The experiments succeeded in this aim. In future research, a factorial design incorporating different levels of expertise to demonstrate the conditions under which the effect might be enhanced as well as testing for the worked example effect will need to be carried out.

In addition, the reported experiments were limited to female English-as-a-foreign-language participants only. In future studies, samples representing a variety of learners studying in their native language should be used. Given that the correlations coefficients between the markers used in this study were low in some cases, a more consistent essay scoring procedure (or, more generally, scoring procedures for problem solutions in ill-structured task domains) needs to be developed and used in future research.

It needs to be emphasized that we have not suggested that essay writing should be abandoned. Students are unlikely to gain skill in essay writing without actual practice. Our point is that practice should be carefully guided, and the use of worked examples provides an ideal vehicle for guidance. Our results suggest that the use of worked examples can provide a beneficial instructional procedure even in areas where they are rarely used. That benefit increases with decreased student knowledge levels.

AUTHOR NOTES

Suna Kyun received a PhD from the School of Education at the University of New South Wales and currently works as a postdoctoral research fellow in the College of Medicine at Yonsei University, South Korea. Her research interests are cognitive processes in learning and their application to instructional design of ill-defined domains such as medical domains. **Slava Kalyuga** is Professor of Educational Psychology at the School of Education, the University of New South Wales, where he received a PhD and has worked since 1995. His research interests are in cognitive processes in learning and the role of learner prior knowledge, adaptive multimedia learning, and diagnostic assessment methods. **John Sweller** is Emeritus Professor of Education in the School of Education at the University of New South Wales. His research is concerned with cognitive processes and instructional design.

REFERENCES

Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research*, *54*, 577–596.

Atkinson, R. K., & Renkl, A. (2007). Interactive example-based learning environment: Using interactive elements to encourage effective processing of worked examples. *Educational Psychology Review*, *19*, 375–386.

Beckmann, J. (2010). Taming a beast of burden: On some issues with the conceptualisation and operationalisation of cognitive load. *Learning and Instruction*, *20*, 250–264.

Best, J. B. (1989). *Cognitive psychology* (2nd ed.). St. Paul. MN: West.

Carroll, M. W. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology*, *86*, 360–367.

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

Crippen, K. J., & Earl, B. L. (2007). The impact of web-based worked examples and self-explanation on performance, problem solving, and self efficacy. *Computers & Education*, *49*, 809–821.

De Groot, A. (1965). *Thought and choice in chess*. The Hague, Netherlands: Mouton. (Original work published 1946)
Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication*, *28*, 122–128.
Geary, D. (2007). Educating the evolved mind: Conceptual foundations for an evolutionary educational psychology. In J. S. Carlson & J. R. Levin (Eds.), *Psychological perspectives on contemporary educational issues* (pp. 1–99). Greenwich, CT: Information Age.
Geary, D. (2008). An evolutionarily informed education science. *Educational Psychologist*, *43*, 179–195.
Geary, D. (2012). Evolutionary educational psychology. In K. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook* (Vol. 1, pp. 597–621). Washington, D.C.: American Psychological Association.
Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self explanation? *Learning and Instruction*, *16*, 104–121.
Goel, V. (1992a, July–August). *A comparison of well-structured and ill-structured task environments and problem spaces*. Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society, Hillsdale, New Jersey.
Goel, V. (1992b, July–August). *Ill-structured representations for ill-structured problems*. Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society, Hillsdale, New Jersey.
Greeno, J. (1976). Indefinite goals in well-structured problems. *Psychological Review*, *83*, 479–491.
Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, *17*, 612–634.
Hilbert, T. S., & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Computers in Human Behavior*, *25*, 267–274.
Hoffman, B., & Schraw, G. (2010). Conceptions of efficiency: Applications in learning and problem solving. *Educational Psychologist*, *45*, 1–14.
Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction*, *20*, 18–29.
Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior worked examples. *Journal of Educational Psychology*, *93*, 579–588.
Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, *96*, 558–568.
Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology, Research and Development*, *53*, 83–93.
Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*, 75–86.
Kyun, S., & Lee, H. (2009). The effects of worked examples in computer based instruction: Focus on the presentation format of worked examples and prior knowledge of learners. *Asia Pacific Education Review*, *10*, 495–503.
Owen, P., & Sweller, J. (2008). Cognitive load theory and music instruction. *Educational Psychology*, *28*, 29–45.
Paas, F., & van Merrienboer, J. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors*, *35*, 737–743.
Paas, F., & van Merriënboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive load approach. *Journal of Educational Psychology*, *86*, 122–133.
Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, *88*, 144–161.
Renkl, A., Atkinson, R. K., & Große, C. S. (2004). How fading worked solution steps works- A cognitive load perspective. *Instructional Science*, *32*, 59–82.
Rourke, A., & Sweller, J. (2009). The worked example effect using ill-defined problems: Learning to recognise designers' style. *Learning and Instruction*, *19*, 185–199.
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.
Schwonke, R., Renkl, A., Salden, R., & Aleven, V. (2011). Effects of different ratios of worked solution steps and problem solving opportunities on cognitive load and learning outcomes. *Computers in Human Behavior*, *27*, 58–62.
Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, *99*, 285–296.
Shaw, S., & Falvey, P. (2008). *The IELTS writing assessment revision project: Towards a revised rating scale* (Research reports No. 08.1). Cambridge, England: University of Cambridge ESOL Examinations.

Spiro, R. J., & DeSchryver, M. (2009). Constructivism: when it's the wrong idea and when it's the only idea. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: success or failure?* (pp. 106–123). New York, NY: Routledge.

Stark, R., Kopp, V., & Fischer, M. R. (2011). Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. *Learning and Instruction*, *21*, 22–33.

Stark, R., Mendl, H., Gruber, H., & Renkl, A. (2002). Conditions and effects of example elaboration. *Learning and Instruction*, *12*, 39–60.

Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 215–266). San Diego, CA: Academic Press.

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

Sweller, J. (2008). Instructional implications of David C. Geary's evolutionary educational psychology. *Educational Psychologist*, *43*, 214–216.

Sweller, J. (2009). What human cognitive architecture tells us about constructivism. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: success or failure?* (pp. 127–143). New York, NY: Routledge.

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

Sweller, J. (2011). Cognitive load theory. In J. Mestre & B. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (Vol. 55, pp. 37–76). Oxford, England: Academic Press.

Sweller, J. (2012). Human cognitive architecture: Why some instructional procedures work and others do not. In K. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook* (Vol. 1, pp. 295–325). Washington, DC: American Psychological Association.

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

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 progressing systems. *Evolutionary Psychology*, *4*, 434–458.

Torcasio, S., & Sweller, J. (2010). The use of illustrations when learning to read: A cognitive load theory approach. *Applied Cognitive Psychology*, *24*, 659–672.

Tynjälä, P., Mason, L., & Lonka, K. (2001). Writing as a learning tool: An introduction. In P. Tynjälä, L. Mason, & K. Lonka (Eds.), *Writing as a learning tool: Integrating theory and practice* (pp. 7–22). Dordrecht, The Netherlands: Kluwer.

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

van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2008). Effect of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficacy. *Learning and Instruction*, *18*, 211–222.

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

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

APPENDIX A
Learning Materials for the Worked Example Group

Faculty:_____ **Student Number:**_____ **Name:**_____

(Questions 1–2) Questions 1 and 2 are closely related to each other. After studying Question 1 and its possible answers, answer Question 2. **When you answer Question 2, you can, of course, make use of answers to Question 1.** Please note the time limit for studying and answering Questions 1 and 2.

Question 1. In contemporary times, fairytales are mainly considered as stories for children and children are encouraged to read many fairytales from an early stage. Do you think that the fairytales are mainly children's stories? If yes, why do you think so? If not, why not?

<Answer 1> I think fairytales can be stories for children. First of all, the plot is very simple but the stories are very interesting. Despite being mostly short and simple, the stories contain vivid conflicts but the ending is always a happy one where the problems are solved. I think these aspects can encourage children to think more positively about the problems or conflicts they face: when there are conflicts and problems, there are solutions and rewards too. Also, many fairytales include child characters. Hansel and Gretel is a good example and in other fairytales, children grow up and get married to a prince or princess. Although the process of their growth is normally mentioned very briefly, the story normally starts the time when the main character was young, having particular problems such as a witch's curse, and in most cases, having lost their mother or both. So there are aspects, in which child readers can sympathize more with the characters too. There is no reason that adults cannot enjoy reading fairytales but I think child readers is the main reason for the simple plot but complex relationships and conflicts, and for happy ending with all problems solved, and child characters involved in fairytales.

<Answer 2> I don't think fairytales are mainly for children. Children can read fairytales of course but they were mainly for adults before the 17th century and even now for them too. It was mainly regarded as the Brothers Grimm who published fairytales for children and contributed to the trend that fairytales are mainly for children. They removed many scenes and elements, which were not considered proper for children such as the elements of sex and violence. I think this creates two problems. First, despite their omission of those sexual and violent elements from the stories, many fairytales published for children still contain hints of sex and violence. The way that a stepmother treats her stepdaughter cruelly or how Gretel kills the witch, or how a kind husband suddenly turns out to be a murderer of several wives in Blue Beard, all show aspects of cruelty and violence. So, children actually experience those violent and cruel aspects repeatedly through reading fairytales. The other problem is that when an editor and a publisher omit particular elements or scenes, which they consider are not appropriate for children's version, is it likely that those original fairytales also lose some important symbols or the integrity of the story. Most fairytales are passed down orally and there are several versions of a similar type of story, so we cannot say there is a definite original story. However, I think when people intend to make a particular story for children and omit sections of the original oral stories, it is unavoidable that those children's versions are somewhat artificial and lose some elements of the original stories that might be important for the integrity of the fairytale. Considering these aspects, I think fairytales are mainly for adults and when they are constructed as children's versions, they often lose particular integrity of symbols or meanings not only some specific elements of sex or violence.

<Answer 3> For reasons of space, the third answer has been omitted.

Question 2. There are different versions of fairytales in different countries even if they share similar motifs or symbols. Since the western industrialization, children's versions of fairytales began to appear and in the contemporary times, there are diverse modified versions of fairytales and creative stories, which adopt main motifs from familiar fairytales. Despite the universality and diversity of fairytales, fairytales still tend to be classified as children's stories. What do you think are the reasons for that?

Question 3–4) The following Questions 3 and 4 are closely related to each other. After studying Question 3 and the answers, answer Question 4. **When you answer Question 4, you can, of course, make use of the answers to Question 3.** Also, you need to note the time limit for studying and answering the question

<Answer 1> I think most of the fairytales familiar to me show a typical plot where an outstandingly beautiful but passive female character wins ultimate happiness by gaining a prince's love and getting married to him. She is not only beautiful but also good in her heart and endures many sufferings and hardships in her life until she meets the prince and get married to him. One of the most famous figures is perhaps Cinderella. She lost her real mother and her father has another wife with two daughters. Cinderella's stepmother treats her badly but she is very patient to obey her stepmother and endures all kinds of maltreatment and hard tasks at home. As most people know, a prince holds a big party to look for his "future" wife and invites all the young ladies in the land. Cinderella's stepmother does not allow her to go to the party but thanks to a fairy godmother's help, she is changed into a splendid princess and attends the party. The prince falls in love with Cinderella and eventually they get married to each other despite the problems her stepmother caused. In the story of the Sleeping Beauty, a princess is cursed to sleep for one hundred years but is saved by a prince, who has heard about her and has been looking for her. Cinderella is different from the Sleeping Beauty in that she has gone through difficulties in her life; however, they ultimately both achieve their happiness by with the help of princes who appear as saviors in their lives. Their characters are described as 'obedient and patient', and most of all, beautiful, which is typically required for an "ideal female character" in a traditional patriarchal society. In this way, I think the fairytales portray "typical feminine" characters and even reinforce these stereotypes, and female characters are portrayed as obtaining a happy and privileged status by a royal male character.

<Answer 2> I think many fairytales show "typical feminine" characters, which can be described as passive, obedient and beautiful. Many female characters such as Cinderella seem to show those characteristics and ultimately gain the happiness by male "saviors." However, I think we cannot underestimate their positive attitudes in their life and some of the female characters show distinct leadership qualities. For example, Cinderella, who is remembered as beautiful but passive and obedient, has strength to endure the difficulties in her life and is able to communicate with "Nature." She is helped by small ants and a fairy godmother. This can be another example of her reliance on "somebody else." However, I think this aspect also proves her strength and her ability to harness the power of nature and to make the most of it when she really needs some help. Also, in the story of *Beauty and the Beast*, the Beauty is a typical female character in that she is beautiful, kind and obedient to her father. However, it is her decision to choose her "destiny" to get married to the Beast and to accept "the ugly creature" in her life. Her destiny does not appear bright and fantastic to other people but she "chooses" to accept it and make the most of her destiny. It is not that she is forced to stay with the Beast. Gretel, in the story of *Hansel and Gretel*, might be a character who changed from a seemingly helpless little girl to a leading character to save her own and her brother's life. When they face the witch's cunning plan, it is Gretel who dares to push the witch into the fire. So, female characters in fairytales are apparently passive but there are aspects and situations where they also show their own courage and independent decision making, which I think should be noted. Thus, female characters in fairytales do not always depend on male characters but can be said to play a role of cooperating male figures or even leading them.

<Answer 3> For reasons of space, the third answer has been omitted.

Question 4. Many fairytales are regarded as showing "typical" male figures such as a prince who saves a main female character. What do you think of male figures in fairytales in relation to female characters? Are they all typical "saviors" or do they show different aspects?

APPENDIX B
International English Language Testing System Task 2 Writing Band Descriptors (Public Version)

Band	*Task response*	*Coherence and cohesion*	*Lexical resource*	*Grammatical range and accuracy*
9	– fully addresses all parts of the task – presents a fully developed position in answer to the question with relevant, fully extended, and well-supported ideas	– uses cohesion in such a way that it attracts no attention – skillfully manages paragraphing	– uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as "slips"	– uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'
8	– sufficiently addresses all parts of the task – presents a well-developed response to the question with relevant, extended, and supported ideas	– sequences information and ideas logically – manages all aspects of cohesion well – uses paragraphing sufficiently and appropriately	– uses a wide range of vocabulary fluently and flexibly to convey precise meanings – skillfully uses uncommon lexical items, but there may be occasional inaccuracies in word choice and collocation – produces rare errors in spelling and/or word formation	– uses a wide range of structures – the majority of sentences are error-free – makes only very occasional errors
7	– addresses all parts of the task – presents a clear position throughout the response – presents, extends, and supports main ideas, but there may be a tendency to over generalize and/or supporting ideas may lack focus	– logically organizes information and ideas; there is clear progression throughout – uses a range of cohesive devices appropriately although there may be some under- or overuse – presents a clear central topic within each paragraph	– uses a sufficient range of vocabulary to allow some flexibility and precision – uses less common lexical items with some awareness of style and collocation – may produce occasional errors in word choice, spelling, and/or word formation	– uses a variety of complex structures – produces frequent error-free sentences – has good control of grammar and punctuation but may make a few errors
6	– addresses all parts of the task although some parts may be more fully covered than others – presents a relevant position, although the conclusions may become unclear or repetitive – presents relevant main ideas but some may be inadequately developed/unclear	– arranges information and ideas coherently, and there is a clear overall progression – uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical – may not always use referencing clearly or appropriately – uses paragraphing, but not always logically	– uses an adequate range of vocabulary for the task – attempts to use less common vocabulary but with some inaccuracy – makes some errors in spelling and/or word formation, but they do not impede communication	– uses a mix of simple and complex sentence forms – makes some errors in grammar and punctuation but they rarely reduce communication

(Continued on next page)

APPENDIX B
International English Language Testing System Task 2 Writing Band Descriptors (Public Version)
(Continued)

Band	*Task response*	*Coherence and cohesion*	*Lexical resource*	*Grammatical range and accuracy*
5	– addresses the task only partially; the format may be inappropriate in places – expresses a position but the development is not always clear and there may be no conclusions drawn – presents some main ideas but these are limited and not sufficiently developed; there may be irrelevant detail	– presents information with some organization but there may be a lack of overall progression – makes inadequate, inaccurate or overuse of cohesive devices – may be repetitive because of lack of referencing and substitution – may not write in paragraphs, or paragraphing may be inadequate	– uses a limited range of vocabulary, but this is minimally adequate for the task – may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader	– uses only a limited range of structures – attempts complex sentences but these tend to be less accurate than simple sentences – may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader
4	– responds to the task only in a minimal way or the answer is tangential; the format may be inappropriate – presents a position but this is unclear – presents some main ideas but these are difficult to identify and may be repetitive, irrelevant, or not well supported	– presents information and ideas but these are not arranged coherently and there is no clear progression in the response – uses some basic cohesive devices but these may be inaccurate or repetitive – may not write in paragraphs or their use may be confusing	– uses only basic vocabulary that may be used repetitively or that may be inappropriate for the task – has limited control of word formation and/or spelling; errors may cause strain for the reader	– uses only a very limited range of structures with only rare use of subordinate clauses – some structures are accurate but errors predominate, and punctuation is often faulty
3	– does not adequately address any part of the task – does not express a clear position – presents few ideas, which are largely undeveloped or irrelevant	– does not organize ideas logically – may use a very limited range of cohesive devices, and those used may not indicate a logical relation between ideas	– uses only a very limited range of words and expressions with very limited control of word formation and/or spelling – errors may severely distort the message	– attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning
2	– barely responds to the task – does not express a position – may attempt to present one or two ideas but there is no development	– has very little control of organizational features	– uses an extremely limited range of vocabulary; essentially no control of word formation and/or spelling	– cannot use sentence forms except in memorized phrases
1	– answer is completely unrelated to the task	– fails to communicate any message	– can only use a few isolated words	– cannot use sentence forms at all
0	– does not attend – does not attempt the task in any way – writes a totally memorized response			