Effects of worked examples on step performance in solving complex problems

Ouhao Chen, Endah Retnowati & Slava Kalyuga

Published online: 13 Oct 2018.

Submit your article to this journal ⇗

View Crossmark data ⇗

Routledge
Taylor & Francis Group

Check for updates

Effects of worked examples on step performance in solving complex problems

Ouhao Chen[a] , Endah Retnowati[b] and Slava Kalyuga[c]

[a]National Institute of Education, Nanyang Technological University, Singapore, Singapore; [b]Faculty of Mathematics and Sciences, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; [c]School of Education, University of New South Wales, Sydney, Australia

ABSTRACT

The instructional effect of worked examples has been investigated in many research studies. However, most of them evaluated the overall performance of the participants in solving post-intervention problems, rather than individual step performance in multi-step problems. The two experiments reported in this article investigated the relations between using worked examples and individual step performance in solving isomorphic problems. In Experiment 1, the effect of worked examples was found for overall performance for novice learners, whereas this effect was gradually reduced from Step 1 (the most difficult one) at which the effect was the strongest, to Step 3 (the easiest one) at which the effect was the weakest or even disappeared. In Experiment 2, relatively more knowledgeable participants learned the same sets of materials, and no effect of worked examples was found for either overall performance or individual step performance. Learner levels of expertise and levels of element interactivity were used to explain the results.

Introduction

The use of worked examples is an instructional tool, which provides the professional solution to a problem for a learner to study. There is no specific definition for a worked example, but the typical components of a worked example include a problem statement followed by the sequential steps of the procedure for solving this problem (Atkinson, Derry, Renkl, & Wortham, 2000). The traditional design paradigms for testing the worked example effect compare worked examples only or worked example-problem solving pairs (each pair composed of a worked example followed by a similar problem to solve) with the equivalent number of problem solving only exercises using the overall post-test performance of participants to evaluate the effect (Van Gog, Kester, & Paas, 2011). In this article, the experiments were designed to compare worked examples only with the problem-solving only, however, in addition to the traditional overall post-test performance, learners' individual step performance in

CONTACT Endah Retnowati ✉ e.retno@uny.ac.id 🖳 Faculty of Mathematics and Sciences, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

multi-step problems was used to evaluate the effect. The following section explains why this innovation could be important for advancing our knowledge of conditions under which worked examples are instructionally effective.

Previous research on the effect of worked examples

A large body of research studies has indicated positive effects of worked examples on students' learning, especially for learners who are new to a specific task domain (Renkl, 2014; Sweller & Cooper, 1985; Van Gog & Rummel, 2010). Cognitive load theory (Sweller, Ayres, & Kalyuga, 2011, for an overview) has explained this effectiveness by reducing the unnecessary load on novice learners' working memory compared to alternative instructional approaches (such as problem-solving exercises or problem exploration as instructional methods). The theory is based on our contemporary knowledge of human cognitive architecture that consists of working memory with limited capacity and duration when dealing with novel information (Cowan, 2001; Miller, 1956; Peterson & Peterson, 1959), and long-term memory as a permanent repository of acquired organized knowledge structures.

Worked examples were successfully used in the domains of algebra (Sweller & Cooper, 1985), statistics (Paas, 1992), geometry (Schwonke et al., 2009; Van Gog & Rummel, 1994) and physics (Reisslein, Atkinson, Seeling, & Reisslein, 2006; Van Gog, Paas, & van Merriënboer, 2006; Van Gog et al., 2011). For example, Sweller and Cooper (1985) demonstrated the effectiveness of worked example-problem solving pairs in algebra task domain. In their experiment, worked examples facilitated students' relevant schema acquisition and so improved their performance compared to problem-solving only. Van Gog et al. (2011) compared worked examples only, example-problem pairs and problem-example pairs with problem-solving only. In their experiment, participants were novices in applying Ohm's law to determine potential problems in electrical circuits. The results indicated that the invested mental effort for training tasks was lower for the examples only and example-problem pairs conditions as compared to the problem-example pairs and problems only conditions; the example-problem pairs required the lowest level of mental effort. The performance on post-test tasks indicated the higher test performance in the examples only and example-problem pairs conditions. No differences were detected in the effectiveness of these two conditions.

There are two major interconnected factors influencing the effectiveness of worked examples: levels of element interactivity and levels of learners' expertise. In the next section, the relation between the worked examples and element interactivity is discussed.

Worked examples and element interactivity

Element interactivity is an index showing the difficulty of learning materials. Interactive elements are defined as elements that must be processed simultaneously in working memory as they are logically related (Sweller et al., 2011). The levels of

element interactivity are determined by the nature of learning materials as well as the expertise of learners.

Let's assume novice learners are asked to solve a math equation, such as $2x + 3 = 5$, for x. This problem, for novices, is high in element interactivity. Novices, without relevant prior knowledge, have to process interconnected elements, such as numerals, symbols for operations simultaneously rather than individually in their working memory in order to fully understand and successfully solve the equation. In the above example, the number of such interconnected elements (6) determines the level of element interactivity. However, with the increase in levels of learner expertise in this task domain, this question could become low in element interactivity, as acquired organized knowledge structures for this type of problems (schemas) would allow learners to treat several of the above-interconnected elements as a single entity in their working memory (thus, the level of element interactivity in the above example could be potentially reduced down to 1 for highly skilled learners). Therefore, the increased level of learner expertise decreases the number of interconnected elements that must be simultaneously processed in working memory, so reduces the level of element interactivity of learning materials.

A number of studies have investigated relations between using worked examples and levels of element interactivity (Ayres, 2006; Blayney, Kalyuga, & Sweller, 2010; Chandler & Sweller, 1991; Chen, Kalyuga, & Sweller, 2015, 2016a, 2016b). Ayres (2006) found that bracket expansion tasks such as $4(3x - 6) - 5(7 - 2x)$ were high in element interactivity and difficult for novices. With this kind of tasks, learners needed to consider numbers and mathematical symbols simultaneously. Ayres used the isolated-interactive elements method (Pollock, Chandler, & Sweller, 2002) which required learners initially to do one calculation at a time, for example, $4 \times 3x$ only, before proceeding to the complete, fully interactive-elements task. This method reduced the level of element interactivity and benefited less experienced learners, whereas more experienced learners benefited more from the fully interacting-elements instruction (fully worked example-problem pairs). Chen et al. (2015, 2016a, 2016b) investigated the relations between levels of guidance and levels of element interactivity more directly. Two types of learning materials – low and high in element interactivity – were designed in the mathematics task domain. Students were randomly assigned to conditions of low guidance (generating problem solutions) or high guidance (studying worked examples). Results indicated that the effectiveness of worked examples (high guidance) was higher for materials high in element interactivity, while the effectiveness of generating (low guidance) was higher for materials low in element interactivity. Therefore, whether the worked example effect is obtainable may depend on the levels of element interactivity of specific materials. If materials are low in element interactivity for given learners, the use of worked examples may be ineffective.

Interaction between learners' expertise and element interactivity when using worked examples

As mentioned above, the effectiveness of worked examples also depends on levels of learner expertise. The expertise reversal effect in cognitive load theory could be used

to explain the interaction between the levels of learner expertise and levels of element interactivity in the effectiveness of worked examples (Sweller et al., 2011). According to the expertise reversal effect, the information that is beneficial to novices may become redundant to more experienced learners. The more experienced learners have already acquired relevant procedural schemas for problem-solving, and if worked examples are presented to them again, they have to reconcile the presented information with their stored schemas. This cognitive processing is redundant, but since it requires additional working memory resources, it may result in high levels of cognitive load. Therefore, the expertise reversal effect focuses on the interaction between the characteristics of learners (levels of expertise) and the characteristics of learning tasks. Kalyuga and Renkl (2010) mentioned that the studies in the expertise reversal effect mainly focused on the role of learner knowledge and discussed the relationship between learners' expertise and the effectiveness of instruction: effective instructional methods that reduce extraneous load for novices may become ineffective or even hinder learning for more knowledgeable learners (Kalyuga, 2007). Two forms of this effect have been established: an ordinal interaction (the instruction is effective for novices, but has no effects on more experienced learners) and a dis-ordinal interaction (the instruction is effective for novices, but has negative effects on more experienced learners) (Nievelstein & Boshuizen, 2013).

Chen, Kalyuga, and Sweller (2017) suggested a close relation between the expertise reversal effect and levels of element interactivity. Novice learners may benefit more from using worked examples, as they do not need to generate moves on their own that may cause high levels of cognitive load, compared to problem-solving; whereas, with the increase in levels of learner expertise, this benefit may be reduced or even reversed, due to reduced complexity of materials that become redundant for these learners. Therefore, with the changes in learners' expertise, the levels of element interactivity are changed, which then influences the effectiveness of worked examples, namely, the effect of worked examples may be only obtained for materials that are high in element interactivity or for novices.

Most of the previous studies obtained the worked example and expertise reversal effects by analyzing students' overall performance in solving post-test problems. No research studies, as far as the authors are aware, have directly investigated these effects when performing individual steps in solving complex multi-step problems and considered differential levels of element interactivity involved in those steps. The only relevant area of research within a cognitive load framework that has taken into account individual steps in problem solutions are studies of fading worked examples.

Fading worked examples

According to the expertise reversal effect, as learners acquire more experience in a task domain, worked examples should be replaced with problem-solving tasks. As a way to do it gradually, completion tasks were suggested (Van Merriënboer, 1990; Van Merriënboer, Kirschner, & Kester, 2003). A completion task provides example-style guidance for some solution steps but asks learners to complete a number of remaining steps on their own. A series of completion tasks with gradually diminishing levels

of guidance (e.g. the number of steps for learners to complete is gradually increased as the number of work-out steps is reduced) was suggested as fading worked examples (Renkl, Atkinson, & Maier, 2000). By using such a guidance fading strategy, the transition from a full worked example to a conventional problem may avoid providing redundant information to relatively more experienced learners and thus enhance the efficiency of instruction (a guidance fading effect in cognitive load theory).

Renkl, Atkinson, Maier, and Staley (2002) showed the superiority of the fading strategy compared to the traditionally worked example-problem solving pairs in both classroom and lab settings. Three different paces of fading, immediate, fast and slow, were compared (Reisslein et al., 2006). The immediate fading group, participants gave instruction followed directly by problem-solving; the fast fading group, a full worked example was presented initially, followed by task with one step omitted, then task with two steps omitted and so on; the slow fading group, one step was omitted for every two worked examples. Results showed that experts learnt more from immediate and fast fading design, whereas, novices benefited more from the slow pacing design.

However, the fading worked example research that focuses on individual steps neither directly evaluated levels of learner expertise to determine the exact point of fading, nor did it apply the concept of element interactivity to evaluate the complexity of different steps for the same purpose. Using these concepts to predict and explain the learner performance at individual steps is the main focus of this study.

Present study

This study examines the relations between the effectiveness of worked examples, element interactivity and learners' expertise by considering not only the overall performance of students in post-intervention problem-solving tasks but also their performance in individual problem steps. Investigating the effect of worked examples on individual steps may be important for our understanding of the processes involved in example-based learning: firstly, students may have different levels of expertise in dealing with different steps; secondly, they may gain differential experiences in different steps while going through a training session. This understanding may also allow better balancing of guided vs. unguided step performance in fading worked examples.

It is assumed that in solving multiple-step problems, the levels of element interactivity may change from one step to another (e.g. be higher for the initial steps with more novelty caused by a new problem than for final steps that could be entirely based on the previously acquired knowledge). Therefore, an effect of worked examples may be found for the initial steps, but not for the final steps (Hypothesis 1). Also, for the same reason, the strength of the effect of worked examples might be gradually reduced from the initial to final steps (Hypothesis 2) due to the gradually reduced levels of element interactivity. Of course, the traditional effect of worked examples should be replicated by using overall performance scores (Hypothesis 3). Finally, with the increase in levels of learner expertise, all effects of worked examples might be eliminated for both individual step performance and overall performance, indicating a traditional expertise reversal effect (Hypothesis 4). Two experiments that used students

with different levels of expertise, were conducted to test these hypotheses in this study. The study was approved by a relevant ethics committee.

Experiment 1

Participants

Forty-two Year 7 students (between 12 and 13 years old) were recruited from a secondary school in Yogyakarta, Indonesia. The participants were randomly assigned into two experimental conditions, with 20 students in the worked example only condition and 22 students in the problem solving only condition. The topic of multiplication of two polynomials, normally taught in Year 8, was chosen for students' learning. Therefore, in this experiment, all students were considered as novices in multiplying two polynomials, although from prior studies in Year 7, they have acquired some experience in performing the relevant lower-level operations such as simple multiplications of numbers and/or numerals or addition of two like terms.

Materials

Two sets of learning materials and a post-test were designed for the study. The first set of learning materials was used for the worked example only condition (Figure 1).

It included four worked examples, with some arrows added to avoid unnecessary searching processes between the lines of the solutions. All questions could be solved in three steps and were isomorphic. The second set of learning material was designed for the problem solving only condition. The four questions used in the worked example only condition were used for the problem-solving condition. The only difference was that students were required to generate solutions of each question by themselves without the worked examples provided.

Using the concept of element interactivity, the number of interactive elements for each step was evaluated. For example, for the question $(3x + 1)(x - 2)$, in Step 1, students needed to open across the brackets by calculating four components $3x \cdot x$, $3x \cdot (-2)$, $(+1) \cdot x$, and $(+1) \cdot (-2)$. In order to calculate each component correctly, students had to process all the elements (numerals, symbols, signs) simultaneously, rendering the total number of interactive elements for Step 1 to be 22 (4 elements for $3x \cdot x$; 6 elements for $3x \cdot (-2)$ considering the negative sign and bracket; 5 elements for $(+1) \cdot x$ considering the positive sign and bracket; 7 elements for $(+1) \cdot (-2)$

Task 1: Calculate $(x+4)(3x-2)$

$$(x + 4)(3x - 2)$$
$$= x \cdot 3x + x \cdot (-2) + 4 \cdot (3x) + 4 \cdot (-2)$$
$$= 3x^2 - 2x + 12x - 8$$
$$= 3x^2 + 10x - 8$$

Figure 1. Sample material used for worked example condition.

considering the positive and negative signs and bracket). For Step 2, the calculation involved simple multiplications between numbers or numerals, such as $x \cdot x$ or $3 \cdot (-2)$, with the total number of interactive elements (8) reduced compared to Step 1. Finally, for Step 3, only the addition of two like terms was needed, with only two interactive elements. Therefore, the number of interactive elements was gradually decreased with three consecutive steps.

A post-test, including five questions (isomorphic to the problems used in learning materials) on polynomial multiplication such as $(x + 4)(x - 9)$, was designed for all students. The participants were required to write down their student number on all materials.

Procedure

The whole experiment lasted for 40 minutes, a normal class time (Figure 2).

Before the intervention, students were randomly assigned to worked example only and problem-solving only conditions (10 minutes). Then students in different conditions studied four worked examples or solved four problems in two learning phases, each of which (10 minutes) had two worked examples or two problems. After 10 minutes, learning materials for the first learning phase were collected, and then the learning materials for the second learning phase were distributed for the second 10-minutes study (two learning phases took 20 minutes). After collecting the materials for the second learning phase, the final 10 minutes were used for the post-test. No students handed in their test solutions until the allocated time elapsed.

Scoring

Each question could be solved in three steps, with the first two steps consisting of four terms, and the last step involving one term (the addition of the like items).
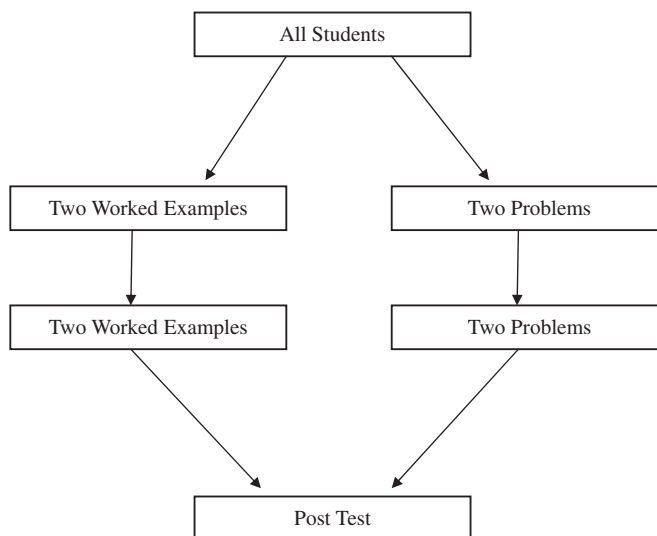

Figure 2. Experimental procedure.

Table 1. Means and standard deviations for percentage correct scores for each step and overall performance in Experiment 1.

Conditions	Step 1	Step 2	Step 3	Overall score
worked Example ($N = 20$)	77.50 (34.36)	48.50 (39.41)	14.00 (21.62)	57.56 (31.80)
Problem Solving ($N = 22$)	34.77 (38.31)	17.05 (33.44)	5.45 (18.70)	23.64 (31.66)

Awarding 1 point for each correct term (where terms are separated by addition/subtraction), resulted in 4 full marks for one correctly performed Step 1 or Step 2, and 1 mark for Step 3. For each of the first two steps, the maximum possible (all correct) score across all five test questions was 20. For the third step, the maximum possible (all correct) score across all five test questions was 5. Scores for each step and the total score of post-test were converted to percentage correct scores and were recorded separately. As the math problems used in the post-test were well-defined, the test was fully objective, so no alternative scorer was required. The Cronbach's alpha for the post-test was .84.

Results

A two-factor 2 (conditions: worked example only vs. problem-solving only) × 3 (steps: 1, 2 and 3) ANOVA was used with the second factor repeatedly measured. Means and standard deviations of percentage correct scores are presented in Table 1.

The main effect of condition was significant, $F(1, 40) = 10.71$, $MSE = 2232.39$, $p = .002$, $\eta_P^2 = .211$, indicating that the worked example condition was superior to the problem solving condition overall. The main effect of step was also significant, $F(2, 80) = 40.58$, $MSE = 414.59$, $p < .001$, $\eta_P^2 = .675$. According to Bonferroni-corrected post-hoc tests, the students performed significantly better in Step 1 than in Step 2, $p = .01$, in Step 2 than in Step 3, $p = .012$, and in Step 1 than in Step 3, $p < .001$. The interaction between condition and step was significant, $F(2, 80) = 5.70$, $MSE = 414.59$, $p = .007$, $\eta_P^2 = .226$. Following the significant interaction, simple effect analyses were conducted for the condition factor. For Step 1, the effect of condition was significant, $t(40) = 3.79$, $SE_{diff} = 11.27$, $p < .001$, $d = 1.02$, indicating that the worked example condition was superior to the problem solving condition on Step 1. Similarly, for Step 2, the effect of condition was also significant, $t(40) = 2.80$, $SE_{diff} = 11.25$, $p = .008$, $d = 0.80$, indicating that the worked example condition was again superior to the problem solving condition on Step 2. However, for Step 3, the effect of condition was not significant, $t(40) = 1.37$, $SE_{diff} = 6.22$, $p = .177$, $d = 0.42$. The results indicate that although the worked example condition was superior to the problem solving condition for overall performance, this superiority, based on the decreased effect size, was gradually reduced as steps proceeded.

The examination of the learners' post-test solutions showed that on eight occasions (out of $42 \times 5 = 210$ events in total for both conditions, or less than 4%), participants had possibly made some errors in Step 2 because of errors in the immediately preceding Step 1 for the task, and on eight occasions (out of 210, or less than 4%), students had possibly made errors in Step 3 because of errors in Step 2. In order to account for a possible influence of such carry-over errors on the results for Steps 2 and 3, additional analyses were conducted. The participants' performance at Steps 2 and 3 were

Table 2. Means and standard deviations for percentage correct scores with carry-over errors for each step and overall performance in Experiment 1.

Conditions	Step 1	Step 2	Step 3	Overall score
worked Example ($N = 20$)	77.50 (34.36)	51.00 (38.07)	17.00 (21.79)	59.00 (30.14)
Problem Solving ($N = 22$)	34.77 (38.31)	17.50 (33.97)	6.36 (18.91)	23.94 (32.12)

re-scored so that the marks for actions at Steps 2 and 3 that were correct given the values obtained in the immediately preceding step (even if those values were incorrect by themselves) were counted as correct (i.e., step correctness was considered not in absolute, but in relative terms – relative to the immediately preceding step). Means and standard deviations of percentage correct scores with carry-over errors are presented in Table 2.

With re-scored performance results, another two-factor 2 (conditions: worked example only vs. problem solving only) × 3 (steps: 1, 2 and 3) ANOVA indicated similar outcomes to the original analyses. The main effect of condition was significant, $F(1, 40) = 12.25$, $MSE = 2150.87$, $p = .001$, $\eta_P^2 = .234$, with the worked example only condition superior to the problem solving only condition. The main effect of step was also significant, $F(2, 80) = 46.64$, $MSE = 443.88$, $p < .001$, $\eta_P^2 = .538$. According to Bonferroni-corrected post-hoc tests, students performed significantly better in Step 1 than Step 2 ($p = .017$), in Step 2 than Step 3 ($p = .015$), and in Step 1 than in Step 3 ($p < .001$). The interaction between condition and step was also significant, $F(2, 80) = 6.44$, $MSE = 443.88$, $p = .012$, $\eta_P^2 = .202$. Simple effect analyses indicated significant effects of condition for the first two steps: $t(40) = 3.79$, $SE_{diff} = 11.27$, $p < .001$, $d = 1.02$, for Step 1; $t(40) = 3.01$, $SE_{diff} = 11.11$, $p = .004$, $d = 0.85$, for Step 2; and $t(40) = 1.69$, $SE_{diff} = 6.28$, $p = .098$, $d = 0.51$, for Step 3. The worked example condition was superior to problem solving condition although with gradually reduced effect size from Step 1 to Step 3. Thus, the pattern of results remained the same – based on the decreased effect size, the superiority of the worked example condition over the problem solving condition was reduced as steps proceeded.

Although the gradual decrease in the reported percentage correct scores might reflect carry-over errors which could be a factor affecting the poor results of Step 3, considering that students presumably had sufficient prior knowledge (from Year 7 studies) to complete this simple and easy step involving only addition of two like terms, students' motivation may be another factor to be considered.

In Experiment 1, only novices in the task domain (multiplication of two polynomials) were recruited to test the first three hypotheses. In order to test our last hypothesis, Experiment 2 was conducted with more knowledgeable students involved.

Experiment 2

Participants

Forty-seven Year 8 students (between 13 and 14 years old) were recruited from one secondary school that was used in Experiment 1. The participants were randomly assigned into two experimental conditions, with 24 students in the worked example only condition and 23 students in the problem solving only condition. Multiplication

Table 3. Means and standard deviations for percentage correct scores for each step and overall performance in Experiment 2.

Conditions	Step 1	Step 2	Step 3	Overall score
worked Example ($N=24$)	89.79 (16.97)	81.46 (28.99)	63.33 (38.97)	83.24 (23.44)
Problem Solving ($N=23$)	98.26 (5.76)	92.17 (14.53)	73.04 (39.82)	92.75 (11.92)

of two polynomials was again used in Experiment 2. Since in this experiment, all students had been previously taught how to multiply two polynomials, they were considered to be more knowledgeable in this task area than the participants in Experiment 1.

Materials

All materials were the same as those used in Experiment 1. Similarly, using the question $(3x+1)(x-2)$ as an example for evaluating the number of interactive elements for each step, the total number of interactive elements should be considerably reduced for all steps compared to Experiment 1, as Year 8 students had relevant schemas for each step and the whole question which allowed them to treat multiple elements as a single chunk. For example, in Step 1, students needed to open across the brackets by calculating four components, but they didn't have to process simultaneously all numerals, symbols, and signs, as they already had schemas for handling all these elements as a single unit, rendering the total number of interactive elements for Step 1 to be 4 (one element for $3x \cdot x$; one element for $-6x$; one element for $+x$; and one element for -2). For Step 2, the calculation involved simple multiplication between numerals $3x \cdot x$ to produce $3x^2$ with one element; for Step 3, only the addition of like terms was needed which was counted as one element for these students.

Procedure and scoring

The procedure and scoring in Experiment 2 were the same as in Experiment 1. The Cronbach's alpha for the post-test was 0.79.

Results

A two-factor 2 (conditions: worked example only vs. problem-solving only) × 3 (steps: 1, 2 and 3) ANOVA was used again with the second factor repeatedly measured. Means and standard deviations for the percentage correct scores are presented in Table 3.

The main effect of condition was not significant, $F(1, 45)=2.11$, $MSE=1551.61$, $p=.154$, $\eta_p^2=.045$, indicating that the superiority of worked examples disappeared for overall performance. The main effect of step was significant, $F(2, 90)=13.31$, $MSE=347.97$, $p<.001$, $\eta_p^2=.377$. According to Bonferroni-corrected post-hoc tests, the students performed significantly better in Step 2 than in Step 3, $p=.004$, and in Step 1 than in Step 3, $p<.001$, with no other significant differences. The interaction between condition and step was not significant, $F(2, 90)=0.19$, $MSE=347.97$, $p=.829$, $\eta_p^2=0.009$. The results may reveal that with the disappearance of the worked

Table 4. Means and standard deviations for percentage correct scores with carry-over errors for each step and overall performance in Experiment 2.

Conditions	Step 1	Step 2	Step 3	Overall score
worked Example (N = 24)	89.79 (16.97)	81.67 (28.62)	66.67 (35.71)	83.61 (22.71)
Problem Solving (N = 23)	98.26 (5.76)	92.39 (14.21)	73.91 (38.82)	92.95 (11.64)

examples effect on overall performance, the superiority of using worked examples on each step, found in Experiment 1, was eliminated as well.

The analyses of the re-scored post-test performance results (similar to the procedure used in Experiment 1) to account for a possible influence of carry-over errors, did not change this pattern of results. Means and standard deviations of percentage correct scores with carry-over errors are presented in Table 4.

The results of these analyses indicated a non-significant main effect of condition ($p = .170$, $\eta_P^2 = 0.041$) and non-significant interaction between condition and step ($p = .738$, $\eta_P^2 = 0.014$). The main effect of step was still significant, $F(2, 90) = 21.13$, $MSE = 330.74$, $p < .001$, $\eta_P^2 = 0.358$. with Bonferroni-corrected post-hoc tests indicating that the students performed significantly better in Step 2 than in Step 3, $p = .008$, and in Step 1 than in Step 3, $p < .001$, with no other significant differences.

General discussion

This study was designed to investigate the influence of using worked examples on individual step performance in solving complex multi-step problems in an attempt to extend some findings of the traditionally worked example effect based on measuring overall performance (Chen et al., 2015, 2016a, 2016b; Sweller & Cooper, 1985). The results confirmed all four hypotheses. In Experiment 1, a worked example effect was found for overall performance, which was in line with the majority of studies on worked example effect and confirmed the Hypothesis 3. Furthermore, the positive effect of using worked examples on individual step performance was found for the initial step (with the highest level of element interactivity), as well as for the second step (with the intermediate level of element interactivity), although with a smaller effect size. No statistically significant worked example effect was found for the final step (with the lowest level of element interactivity) with a very small effect size. The results confirmed our Hypotheses 1 and 2.

The same pattern of results, based on the gradually decreased effect size as steps proceeded, was also found when carry-over errors in step performance were taken into account. It should be noted that if the traditional post-test measures based on the final test performance results (which are the same as the last step scores, i.e. Step 3 scores in this study) were used in this study, there would be no worked example effect found for novices in Experiment 1. However, using total scores as the sum of step scores (or taking into account carry-over errors at Step 3) revealed the worked example effect for novices as predicted by cognitive load theory. Therefore, it might be important for detecting actual effects to analyze the step-level performance and calculate the overall score as the sum of step scores rather than using the final-step score as the overall score.

The pattern of results found on individual step performance in Experiment 1 may correspond to a partial expertise reversal effect, namely, an ordinal interaction between learners' expertise and instructional formats (Kalyuga, 2007; Nievelstein & Boshuizen, 2013). For the initial step, which had the highest level of element interactivity (the most difficult step), novices had little or no knowledge on how to complete it, and using worked examples could facilitate their study. With the gradually reduced level of element interactivity from the initial step to Step 2, the novice learners might gradually increase knowledge on completing this step, therefore, the effect of worked example was reduced. Similarly, the final step, with its lowest level of element interactivity and prior learner experience available (the easiest step), made the learners even more knowledgeable in accomplishing it, rendering the effect of worked example further reduced (or eliminated, depending on the scoring procedure). Therefore, the effect of worked examples found for overall performance may be decomposed into gradually reduced effects on the level of individual steps. Also, the results indicate that the traditionally worked example effect for overall performance may be mostly due to the strong effects of the initial steps. Furthermore, the gradually reduced level of element interactivity may also indicate the increased number of better organized and learned underlying knowledge schemas the learners used for each step, which resulted in the gradually reduced effect of worked examples for each step.

Learners' performance on the easiest step (Step 3) with the lowest score may need further explanation. It could be expected that the mean score for Step 3 should be the highest as it had the lowest level of element interactivity and the required knowledge (addition of two like terms which is the only operation used in Step 3) was studied and practiced by Year 7 students prior to the experiment. This contradiction may possibly be explained by students' motivation when solving problems. It has been found that people who have already had the knowledge of to-be-taught information would show less attention and be complacent during learning. Therefore, they got worse performance than novices (Wood & Lynch, 2002). Therefore, more knowledgeable learners have schemas to get better scores, however, having expertise in the domain could also make them complacent leading to worse performance during learning. Further research is needed to investigate how students' motivation may affect the use of worked examples in both overall performance and individual step performance.

In Experiment 2, with the increased level of learners' expertise, the level of element interactivity for both the whole questions and for the individual steps was vastly reduced, therefore, the effect of worked examples were eliminated for both overall performance and for individual step performance. The results of Experiment 2 combined with the results of Experiment 1 demonstrate an ordinal interaction between learners' expertise and instruction (partial expertise reversal effect): a worked example effect found for novices was not obtained for more knowledgeable students, which confirmed our Hypothesis 4.

Educational implications

This study may have some important educational implications related to the element interactivity of instructional materials. Firstly, for novices, teachers should show explicit

solutions for the initial and key steps of the procedures (most difficult steps with most interactive elements included); after that, the use of worked examples (especially for less difficult steps with less interactive elements included) should be gradually reduced. Accordingly, if completion problems or fading worked examples are used to gradually reduce explicit instructional guidance (Renkl et al., 2000, 2002; Van Merrienboer, 1990), such less difficult, low element-interactivity steps might not include explicit instruction from (or almost from) the beginning and presented as problem-solving tasks. Secondly, with increasing levels of learner expertise, the same step of the same task may turn to be simpler (the level of element interactivity is reduced), therefore, teachers should encourage students to study without worked examples.

Limitations of study

Firstly, the work of Wood and Lynch (2002) focused on those with prior knowledge, which may be used to explain the worst performance in Step 3 for knowledgeable learners in Experiment 2. However, the similar pattern obtained for novices in Experiment 1 may still be an open question for more research (even though it needs to be noted that Year 7 students in Experiment 1 were novices in multiplying polynomials, they were not novices in adding two like terms - the operation used in Step 3). Secondly, this research study only used the materials on multiplication of polynomials – which is a task area in which the later steps are usually based on the prior skills and experience (such as the addition of two like terms in Step 3 of this study) rather than involving learning new procedures. It is possible that different results may be found with different learning materials, especially with tasks in which learning previous steps act as a necessary stage for learning the next step as a new one (rather than relying on previously acquired skills as in this study). Thirdly, as the sample sizes in both experiments were relatively small, the study did not have the sufficiently strong statistical power needed to reliably detect medium-size effects. These limitations need to be addressed in future studies.

Disclosure statement

ORCID

Ouhao Chen http://orcid.org/0000-0001-9645-1946

References

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked example research. *Review of Educational Research*, *70*, 181–214. doi:10.3102/00346543070002181

Ayres, P. (2006). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology*, *20*, 287–298. doi:10.1002/acp.1245

Blayney, P., Kalyuga, S., & Sweller, J. (2010). Interactions between the isolated–interactive elements effect and levels of learner expertise: Experimental evidence from an accountancy class. *Instructional Science*, *38*, 277–287. doi:10.1007/s11251-009-9105-x

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

Chandler, P. & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*, 293–332. doi:10.1207/s1532690xci0804_2

Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology*, *107*, 689–704. doi:10.1037/edu0000018

Chen, O., Kalyuga, S., & Sweller, J. (2016a). Relations between the worked example and generation effects on immediate and delayed tests. *Learning and Instruction*, *45*, 20–30. doi:10.1016/j.learninstruc.2016.06.007

Chen, O., Kalyuga, S., & Sweller, J. (2016b). When instructional guidance is needed. *The Educational and Developmental Psychologist*, *33*, 149–162. doi:10.1017/edp.2016.16

Chen, O., Kalyuga, S., & Sweller, J. (2017). The expertise reversal effect is a variant of the more general element interactivity effect. *Educational Psychology Review*, *29*, 393–405. doi:10.1007/s10648-016-9359-1

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*, 509–539. doi:10.1007/s10648-007-9054-3

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: introduction to the special issue. *Instructional Science*, *38*, 209–215. doi:10.1007/s11251-009-9102-0

Miller, G. A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information . *Psychological Review*, *63*, 81–97. doi:10.1037/h0043158

Nievelstein, V. G., Boshuizen V. D. (2013). Instructional support for novice law students: Reducing search processes and explaining concepts in cases. *Applied Cognitive Psychology*, *25*, 408–413.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*, 429–434. doi:10.1037/0022-0663.84.4.429

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*, 193–198. doi:10.1037/h0049234

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, *12*, 61–86. doi:10.1016/S0959-4752(01)00016-0

Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and Instruction*, *16*, 92–103. doi:10.1016/j.learninstruc.2006.02.008

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*, 1–37. doi:10.1111/cogs.12086

Renkl, A., Atkinson, R. K., & Maier, U. H. (2000). *From studying examples to solving problem: fading worked-out solution steps helps learning*. Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Renkl, A., Atkinson, R., Maier, U., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *The Journal of Experimental Education*, *70*, 293–315. doi:10.1080/00220970209599510

Schwonke, R., Renkl, A., Kreig, 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. doi:10.1016/j.chb.2008.12.011

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

Sweller, J., & Cooper, G. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*, 59–89. doi:10.1207/s1532690xci0201_3

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, *22*, 155–174. doi:10.1007/s10648-010-9134-7

Van Gog, T., Paas, F., & van Merriënboer, J. J. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction*, *16*, 154–164. doi:10.1016/j.learninstruc.2006.02.003

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, *36*, 212–218. doi:10.1016/j.cedpsych.2010.10.004

Van Merriënboer, J. J. (1990). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of Educational Computing Research*, *6*, 265–285. doi:10.2190/4NK5-17L7-TWQV-1EHL

Van Merriënboer, J. J., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist*, *38*, 5–13. doi:10.1207/S15326985EP3801_2

Wood, S. L., & Lynch, J. G. (2002). Prior knowledge and complacency in new product learning. *Journal of Consumer Research*, *29*, 416–426. doi:10.1086/344425