



Problem-solving or Explicit Instruction: Which Should Go First When Element Interactivity Is High?

Greg Ashman¹ · Slava Kalyuga¹ · John Sweller¹

Published online: 5 August 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

The concept of productive failure posits that a problem-solving phase prior to explicit instruction is more effective than explicit instruction followed by problem-solving. This prediction was tested with Year 5 primary school students learning about light energy efficiency. Two, fully randomised, controlled experiments were conducted. In the first experiment ($N = 64$), explicit instruction followed by problem-solving was found to be superior to the reverse order for performance on problems similar to those used during instruction, with no difference on transfer problems. In the second experiment, where element interactivity was increased ($N = 71$), explicit instruction followed by problem-solving was found to be superior to the reverse order for performance on both similar and transfer problems. The contradictory predictions and results of a productive failure approach and cognitive load theory are discussed using the concept of element interactivity. Specifically, for learning where element interactivity is high, explicit instruction should precede problem-solving.

Keywords Productive failure · Cognitive load theory · Expertise reversal effect · Element interactivity

Across multiple domains ranging from mathematics to reading comprehension, researchers have repeatedly demonstrated that fully guided forms of instruction are more effective for novice learners than unguided or partially guided forms of instruction (see Mayer 2004; Kirschner et al. 2006). This evidence comes from a range of sources such as randomised controlled trials investigating the use of worked examples (Sweller et al. 2011; Schwonke et al. 2009) and teacher effectiveness research (Rosenshine 2009). Furthermore, survey evidence from the Programme for International Student Assessment (PISA) suggests a negative correlation between the use of instructional strategies that are less teacher-directed and scores for mathematical and scientific literacy (Hwang et al. 2018).

✉ Greg Ashman
greg.ashman@gmail.com

¹ School of Education, University of New South Wales, Sydney, New South Wales 2052, Australia

However, while there are considerable data indicating the importance of full guidance, the effects of the sequence of providing more and less guidance are far less clear. Assuming that full guidance is given at some point, it remains a question as to whether it is optimal to provide this guidance at the start of an instructional sequence or to allow learners to first experiment with possible solution methods with little or no guidance.

An approach known as ‘productive failure’ has been developed in which learners first struggle to solve a problem on their own before being given full guidance in the canonical method (Kapur and Bielaczyc 2012). However, cognitive load theory (Sweller et al. 2011) predicts that problem-solving first should be less effective than an approach involving full guidance from the outset, for all but the simplest learning objectives. The main aim of the current study was to conduct fully randomised, controlled experiments investigating the effectiveness of a problem-solving first approach compared with an alternative model where full guidance is provided at the start of the sequence of instruction. Fully randomised experiments altering only one variable at a time are the exception rather than the rule in this area. We specifically used typical middle school science materials that were high in element interactivity because it is precisely for such materials presented to novice learners that cognitive load theory predicts initial, explicit instruction is important (see below).

Productive Failure

Productive failure has been proposed as an effective approach when learning how to solve problems. It consists of two phases. The generation and exploration phase requires learners to solve problems without explicit teacher guidance. The consolidation phase then involves responding to learner-generated solution strategies and instruction in the canonical method (Kapur and Bielaczyc 2012; Kapur 2016).

There are clear criteria that problems must meet in order to be accessible during the generation and exploration phase (see Kapur 2016). Problems must not be framed in such a way as to be unintelligible to a learner who has not been instructed in the canonical solution by, for example, using unfamiliar or technical terminology. Instead, problems must draw on prior formal and informal knowledge, must allow learners to attempt a number of solution strategies which a teacher may then build upon in the instructional phase, and must represent an appropriate level of challenge.

Kapur (2016) proposed a number of reasons why problem-solving first may be more effective than an approach that begins with explicit instruction. Problem-solving first may activate and differentiate prior knowledge, and such activation may make learners more aware of the gaps in their prior knowledge. When presented with the canonical solution method, learners who have already attempted to solve the problem are able to compare their solutions with the canonical one, better enabling them to attend to critical features of the canonical solution. Finally, learners involved in problem-solving first may be more motivated and engaged.

In addition, we might hypothesise that requiring learners to generate their own problem solutions prior to explicit guidance may strengthen the stimulus-response relation in memory in a similar way as has been proposed in order to account for the ‘generation effect’ (Slamecka and Graf 1978; Hirshman and Bjork 1988; Schwartz et al. 2009). This strengthening should lead to superior retention. Early problem-solving may also be superior because explicit guidance may interfere with implicit learning (Reber 1989) causing learners to focus on

procedures rather than the situational structures that make the procedures useful (Schwartz et al. 2009). Productive failure may therefore lead to superior transfer to new problems with a similar deep structure that are set in different contexts.

A number of studies directly support the relative effectiveness of problem-solving first when compared with an explicit instruction approach (e.g. Kapur 2012; Loibl and Rummel 2014a, b; Kapur 2014; Jacobson et al. 2017; Lai et al. 2017; Weaver et al. 2018). In addition, studies have been conducted that do not directly reference an attempt to meet the productive failure criteria, but nonetheless suggest the relative effectiveness of an exploratory phase prior to direct instruction, when compared with direct instruction from the outset (e.g. Schwartz and Bransford 1998; Schwartz and Martin 2004; DeCaro and Rittle-Johnson 2012; Schwartz et al. 2011).

However, only a subset of these studies varies the order of instruction while varying nothing else (Loibl and Rummel 2014b; Kapur 2014; Lai et al. 2017; DeCaro and Rittle-Johnson 2012; Weaver et al. 2018). For instance, in the Schwartz et al. study (2011), while learners in both conditions received the same lecture at different times, the other tasks they completed were different in nature, and in the Jacobson et al. (2017) study using a quasi-experimental design, different teachers taught the productive failure and explicit instruction conditions. Therefore, a factor varied in addition to the order of instruction, and it may be this factor, or a combined effect of this factor and the order of instruction, that caused the outcome.

Where attempts are made to vary only the order of instruction, this may result in creating comparison conditions that lack ecological validity. For example, in Kapur's (2014) study, learners in the productive failure condition were compared with learners in a direct instruction condition. Learners in the direct instruction condition were first given instruction in the canonical solution method before being asked to spend a substantial amount of time solving a single problem in a number of different ways. This enabled a match to the problem-solving task given to the learners in the productive failure condition, yet it seems unlikely that a teacher would choose to follow such an approach. For instance, Rosenshine (2009) argued from the perspective of teacher effectiveness research that the most effective forms of explicit instruction guide learner practice and are interactive.

One of the simplest ways of varying the order of instruction while maintaining full and valid experimental control along with ecological validity is to compare studying worked examples followed by problem-solving with exactly the same worked example and problem-solving phases but in reverse order. Thus, an example–problem sequence can be compared with a problem–example sequence with no other difference between groups. That comparison frequently has been made both in order to test the productive failure hypothesis (Hsu et al. 2015) and for other, unrelated reasons (Leppink et al. 2014; Van Gog et al. 2011). In all cases, the example–problem sequence has proved superior to the problem–example sequence, in contradiction of the assumption that problem-solving first is advantageous.

Other experimental studies also have looked for an advantage to learning from an initial exploratory phase prior to instruction and have either found a null result or an effect in the opposite direction (e.g. Fyfe et al. 2014; Rittle-Johnson et al. 2016). In addition, Glogger-Frey et al. (2015) compared an exploratory phase with studying worked examples, prior to instruction in the domains of education and physics, and found that transfer was better supported by studying worked examples in both domains. Similarly, Cook's (2017) experimental study found evidence that studying worked examples prior to explicit instruction was superior to a productive failure condition for undergraduate biology students learning statistical methods. Worked examples are a form of explicit instruction and so these studies do not support the predictions of productive failure.

In contrast, Glogger-Frey et al. (2017) found the opposite effect, potentially adding support to the predictions of productive failure. In this case, invention activities prior to a lecture led to superior transfer than studying worked examples prior to a lecture. The domain examined, the concept of density and ratio indices, and the invention activities that were used were similar to those used in the Schwartz et al. (2011) study. A key difference between Glogger-Frey et al. (2015) and Glogger-Frey et al. (2017) is that, in the later study, learners were provided with additional practice activities. The consequences of that difference will be discussed below.

Cognitive Load Theory

Cognitive load theory has the aim of developing effective instructional procedures. It is based on a combination of evolutionary psychology and human cognitive architecture (Sweller et al. 2011, 2019). The theory assumes that most instructional information results in the acquisition of biologically secondary rather than primary knowledge (Geary 2008; Geary and Berch 2016) where primary knowledge consists of information that we have evolved to acquire such as learning to listen and speak a native language while secondary knowledge consists of information we need to acquire for cultural reasons such as learning to read and write. Primary knowledge tends to be acquired easily and unconsciously and so does not need to be explicitly taught while secondary knowledge requires conscious effort and is assisted by explicit instruction. Education and training institutions were largely developed to impart biologically secondary information that tends not to be acquired without explicit instruction rather than biologically primary information that is routinely acquired without any explicit instruction.

There is a specific cognitive architecture associated with the acquisition and processing of biologically secondary information (Sweller and Sweller 2006). That architecture can be described by five basic principles. (1) The vast bulk of secondary information is obtained by either reading what other people write or listening to what they say. (2) When information cannot be acquired from others, it can be generated using a random generation and test process. (3) Acquired information can be stored in a long-term memory that has no known capacity limits. (4) When novel information is acquired, it first must be processed in a limited capacity, limited duration working memory before being transferred to long-term memory. (5) Once information is processed by working memory and stored in long-term memory, it can be transferred back to working memory to govern action appropriate to the extant circumstances. Working memory has no known capacity or temporal restraints when dealing with information transferred from long-term memory.

This architecture explains the transformative effects of education. The purpose of education is to allow learners to store information in long-term memory. Once stored, we are transformed into our ability to act. Nevertheless, before being stored, novel information must be processed by our limited capacity, limited duration working memory. Information differs in the extent to which it imposes a heavy working memory load. The concept of element interactivity accounts for those differences and is a central concept of cognitive load theory.

Consider the process of learning about a group of organisms in a forest. Learning the names of each of the organisms could be quite taxing. However, the name of one organism does not depend on another and so all of these names may be learned and recalled independently, requiring few items to be processed simultaneously in working memory. The intrinsic cognitive load of such material is low. In contrast, if the task is to learn how to determine what will happen to the hawk population if the slug population declines, learners have to also consider

the feeding relationships between the different organisms. In this case, learners not only have to consider items but also the relationships between items and these relationships will also consume working memory resources. The intrinsic cognitive load is higher than just needing to learn the names of the organisms (Sweller and Chandler 1994).

Initially, rules governing relations between items are external to the learner and the purpose of instruction is, in part, to enable the learner to map these rules to knowledge held in long-term memory. However, once the rules governing relations have been mapped in long-term memory in this way, they no longer need to be manipulated in working memory. Initially, the learner must consider how the reduction in the slug population may affect the population of thrushes that feed on the slugs and how, in turn, this will affect the hawk population that feeds on the thrushes. Once the knowledge has been acquired, the learner will effortlessly be able to state how the reduction in the slug population will affect the hawks. Thus, as the learner's relative expertise increases, the effective element interactivity of a given problem decreases (Chen et al. 2015, 2017; Sweller 2010).

It is important to note that element interactivity is therefore not a simple measure of the complexity of the learning materials but a measure of how many elements must be processed in working memory and this will change as the learner gains expertise and can retrieve more from long-term memory. Element interactivity therefore depends upon *both* the complexity of the learning materials and the expertise of the learner.

Cognitive load theory predicts that when learners *first* learn concepts with a fairly large number of interacting elements, problem-solving first would overload working memory and lead to the retention of little in the subsequent instructional phase. We contend that high element interactivity learning events are common in schools and so verifying this prediction has practical relevance. Results favouring a productive failure approach may be explained by lower levels of element interactivity. For instance, the differential results of Glogger-Frey et al. (2015) who obtained a worked example–problem-solving superiority and Glogger-Frey et al. (2017) who obtained the reverse result have been explained by Sweller and Paas (2017) using the concept of element interactivity. The additional practice given to learners in the 2017 paper should increase expertise and so decrease element interactivity resulting in a reversal of the results obtained in 2015.

The Present Study

According to the above overview, the phenomenon of productive failure initially appears at odds with both the predictions of cognitive load theory and a considerable body of data, and yet this contradiction may be reconciled if element interactivity is taken into account. Productive failure should not occur if element interactivity is sufficiently high to exceed working memory limits as in the present study. The specific hypothesis tested in the current experiments was that explicit guidance first is superior to problem-solving first using high element interactivity information. That hypothesis was tested using fully randomised, controlled experiments altering a single variable at a time with that variable being order of presentation.

Two experiments were conducted to test this hypothesis. The task areas of energy efficiency and the law of conservation of energy were selected. In the first experiment, while element interactivity was high, it was lower in comparison with the second experiment, in which an additional step was added to problem-solving in order to increase element interactivity.

Experiment 1

The purpose of this experiment was to investigate the hypothesis that a productive failure effect would not be observed with high element interactivity problems. The participants received explicit instruction followed by problem-solving or the same instructional episodes in the reverse sequence.

Method

Participants

The participants were 64 Year 5 students from an independent school in Victoria, Australia. They were approximately 10 years old and had not previously received instruction in the conservation of energy or the related concept of energy efficiency. An entire cohort of Year 5 students were invited to participate and all students who had returned consenting ethics approval forms and who were present on both days of the experiment were included in the sample. The students were randomly assigned to either the group that received explicit instruction first or the group that received problem-solving first. Prior to the study, approval was obtained from the Human Research Ethics Advisory Panel of the lead author's institution.

Materials

Learners were asked to solve problems in which they were given data on the energy taken in per second, and the light energy given out per second, by various light globes. Light globes take in electrical energy and give out heat and light energy. By computing the useful light energy given out as a proportion of the electrical energy taken in, learners could find the efficiency of each type of globe and hence decide which globe was the most 'energy saving'.

To correctly solve each problem, learners needed to identify the input energy; identify the output light energy; divide the output light energy by the input energy (three elements—'divide', 'output', 'input'); multiply by 100 to complete the percentage calculation for each globe from these two numbers; repeat for each globe; and identify the lowest and/or highest percentages, resulting in eight interacting elements, a number considerably above the current assumptions of a working memory limit of four or fewer elements when processing information (Cowan 2001). The learners used a simple calculator to complete each calculation. The questions were designed so that they would be meaningful and make intuitive sense to learners without any prior instruction in the area and the solution method. Accordingly, terms like 'power' and 'efficiency' were avoided in the wording of the questions. The questions also allowed learners to attempt different solution methods (see Table 1).

Various iterations of these questions were compiled into a booklet and a set of PowerPoint slides, with the latter to be used in the interactive explicit instruction phase of the experiment. The problem-solving booklet was compiled so that there were multiple problems to complete involving increasing numbers of light globes—four questions involved three globes, one question involved five globes, and one question involved six globes. In addition, the PowerPoint slides addressed a common, incorrect solution method that learners were observed to deploy in previous exploratory work—many students indicate that the globe giving out the most light energy is the most efficient. Taken together, these conditions were consistent with key design features that are typically considered to enable a productive failure effect (Kapur 2016).

Table 1 Example question data. Students were asked to determine which light globe was the most efficient and/or which was the least efficient

Globe	Electrical energy used by the globe per second	Light energy given out per second
A	30 joules	15 joules
B	20 joules	12 joules
C	30 joules	12 joules

In addition to the experimental materials, a booklet of reading materials was also prepared. These materials were related to the topic of study (energy) but were not directly related to the experimental materials. One reading discussed the reasons why humans, unlike plants, cannot directly use sunlight as an energy resource and so it drew on concepts of photosynthesis that are unrelated to efficiency calculations. The second reading explained how some deep-sea organisms are able to make use of sulphur from deep-sea hydrothermal vents in a process similar to photosynthesis. Again, this was unrelated to efficiency calculations. These materials were used for the reading filler task described below.

For the post-test, two sets of questions were prepared. The first set included questions that were similar to the items in the problem-solving booklet and which will be referred to below as ‘similar questions’. They used the same context of light globes but the values used in the questions were different (three questions involved three globes and then one question each involved four, five, and six globes). The second set of questions varied in comparison to the problem-solving questions and will be referred to below as ‘transfer questions’. Two questions were set in the different context of an electric fan (see Table 2). Two questions involved light globes but presented additional redundant information in the table about the heat energy given out by these globes, requiring learners to select the useful energy. A final question required learners to use the principle of conservation of energy to complete an additional step and compute the light energy given out when given data on the electrical energy used and heat energy given out (see Table 3).

Procedure

The experiment took place in a 200-seat lecture theatre that was available for use in the learners’ school. All stages of the experiment took place during the time allocated for the learners’ regular science lessons. Learners were randomly assigned to one of two conditions: problem-solving–lecture (30 learners) and lecture–problem-solving (34 learners). Learners in each condition were randomly placed in alternate rows of the lecture theatre. Each learner was issued with a basic calculator.

Table 2 Example of a transfer question. Students were asked to determine which fan was the most efficient and/or which was the least efficient. Electric fans take in electrical energy and give out heat, sound, and movement energy (which is also known as ‘kinetic energy’)

Fan	Electrical energy used by the fan per second	Movement energy given out per second
A	30 joules	15 joules
B	20 joules	12 joules
C	30 joules	12 joules

Table 3 Example of transfer questions. Students were asked to determine which light globe was the most efficient and/or which was the least efficient

Globe	Electrical energy used by the globe per second	Heat energy given out per second	Light energy given out per second
A	50 joules	25 joules	
B	60 joules	30 joules	
C	50 joules	24 joules	
D	60 joules	33 joules	

The instruction proceeded in three stages. In the first stage, learners in the problem-solving–lecture condition were given the booklet of problems to solve, with the following instructions: ‘This booklet contains some problems to try to solve. They are set in everyday situations so think how you would solve the problem in real life. You are not expected to solve all of the problems. Just do what you can.’ They were given 15 min to work on these problems. During this time, learners in the lecture–problem-solving condition completed the reading task. After 15 min, both tasks were halted and materials were collected.

In the second stage, all learners simultaneously received 25 min of interactive explicit instruction in the canonical method for solving the light globe problems, which involves calculating efficiency by dividing useful energy output by total energy input and then comparing the efficiencies of the different globes. This stage included a discussion of the common incorrect solution approach and why it was incorrect. In this stage, the PowerPoint slides were displayed on a screen to the learners. The lecture was interactive in that as the teacher performed relevant calculations, learners were also asked to perform the same calculations and hold their calculators aloft once they had an answer. The teacher scanned these calculator responses but did not offer any feedback to the learners.

The third stage proceeded in exactly the same manner as the first stage except that the tasks were reversed between the two groups. Learners in the lecture–problem-solving condition were now given the booklet of problems to solve, whereas learners in the problem-solving–lecture condition completed the reading task.

The purpose of the reading task was purely to act as a filler activity so that the explicit instruction phase in both conditions would take place at the same time, allowing all students to receive this explicit instruction together (see Fig. 1).

By structuring the experiment in this way, only strictly one experimental factor—the instructional sequence—was manipulated with all other possible influencing factors equalised between the experimental conditions. Therefore, the outcomes could be directly compared for learners who had interactive explicit instruction prior to problem-solving with learners who solved problems prior to interactive explicit instruction.

Six days later, learners in both conditions completed the post-test which consisted of two components. The first component included 6 similar questions and the second component consisted of 5 transfer questions. Both components were timed and lasted for 15 min each. The procedure is illustrated in Fig. 1.

Scoring

In the similar questions, learners needed to decide which bulb was the most efficient and/or which bulb was the least efficient by the canonical method. In order to do this, learners needed

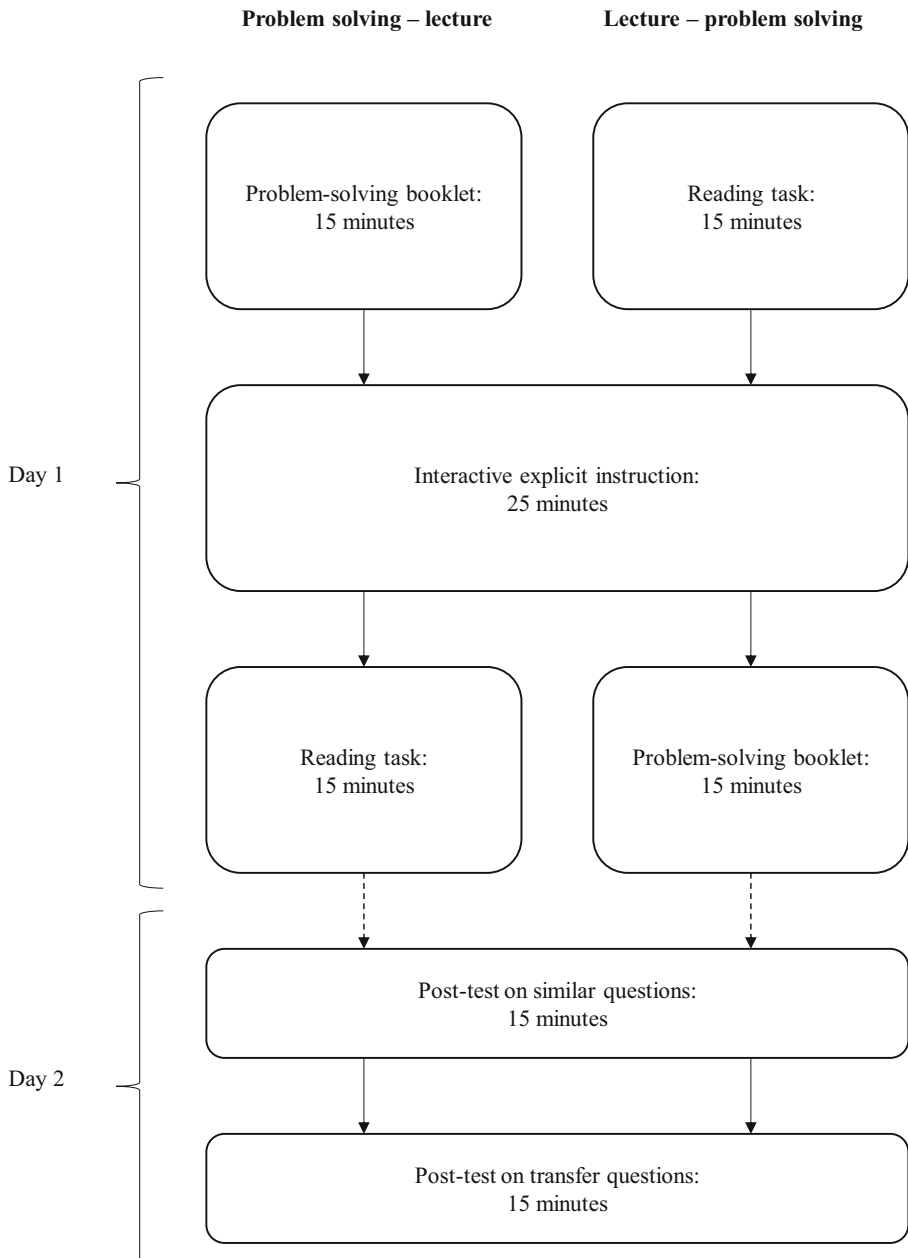


Fig. 1 Procedure for conducting Experiment 1

to calculate the efficiency of each bulb separately. Each of these multiple calculations was therefore scored as 1 if correct and the correct decisions of the most efficient bulbs (letter choice/s) were also scored as 1. The maximum possible score was 35. It was possible for the learners to guess the correct letter choice, but in this case, they would not have the supporting

multiple calculations and so would not score fully for the question. Items on the similar post-test component were highly reliable with Cronbach's $\alpha = 0.94$.

In order to vary questions for the transfer component and make them more complex, usually, an additional step was required to be added to the solution procedure. The transfer questions were scored similarly, with correct calculations and correct answers each being scored with 1 mark. The maximum possible score was 28. Items on the transfer post-test component were reliable with Cronbach's $\alpha = 0.75$.

Only one scorer was used to score the tests because there was no subjectivity in scoring. Either a calculated number was correct or it was not, and either a selected globe letter was correct or it was not. The scorer did not have knowledge of the group to which each learner had been allocated.

Results and Discussion

Means and standard deviations for the post-test scores are presented in Table 4. For the similar post-test questions, learners in the lecture–problem-solving condition, who received explicit instruction first, scored significantly higher than learners in the problem-solving–lecture condition who received problem-solving first ($t(62) = 2.25$, $p = 0.03$, Cohen's $d = 0.56$). A visual representation of the data is presented in Fig. 2 (Ho et al. 2018).

For the transfer post-test questions, there was no significant difference between the conditions ($t(62) = 1.89$, $p = 0.06$, Cohen's $d = 0.47$). A visual representation of the data is presented in Fig. 3 (Ho et al. 2018).

As expected for high element interactivity information, Experiment 1 did not lead to a superiority of the problem-solving–lecture sequence. Instead, there is evidence that the lecture–problem-solving sequence resulted in higher test scores.

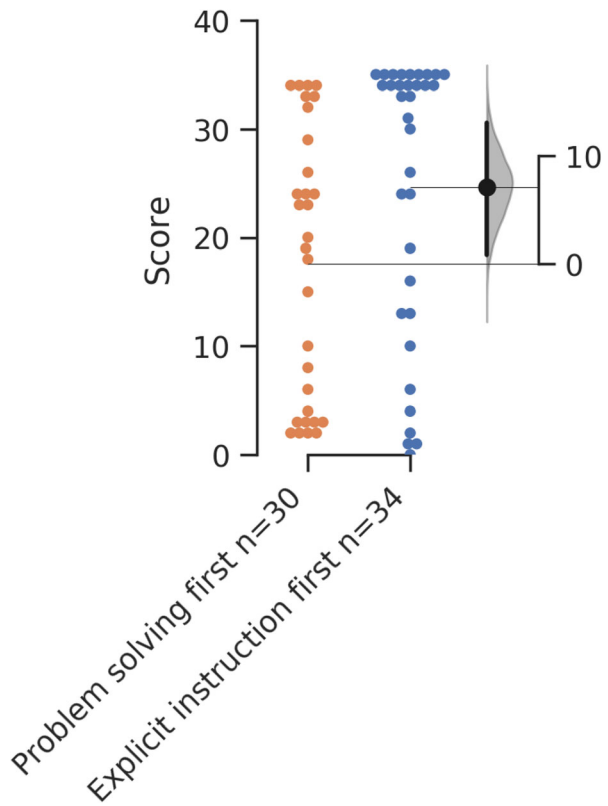
Experiment 2

Experiment 1 demonstrated the superiority of a lecture–problem-solving sequence over the reverse sequence for similar problems but not transfer problems. Since increases in element interactivity may increase the effect size, the intent of Experiment 2 was to replicate the approach of Experiment 1 while increasing the element interactivity. It was hypothesised that higher element interactivity may lead to an observed effect on transfer problems as well as similar problems. As in Experiment 1, the participants received interactive explicit instruction followed by problem-solving or the same instructional episodes in the reverse sequence.

Table 4 Means and standard deviations (in brackets) for the similar and transfer questions of Experiment 1

	Problem-solving–lecture ($n = 30$)	Lecture–problem-solving ($n = 34$)
Similar questions (total 35 marks)	17.57 (12.39)	24.68 (12.78)
Transfer questions (total 28 marks)	8.33 (7.60)	11.41 (5.33)

Fig. 2 An estimation plot for similar questions in Experiment 1. The filled curve indicates the complete distribution for the difference in means, given the observed data. The bold line shows the 95% confidence interval (Ho et al. 2018)



Method

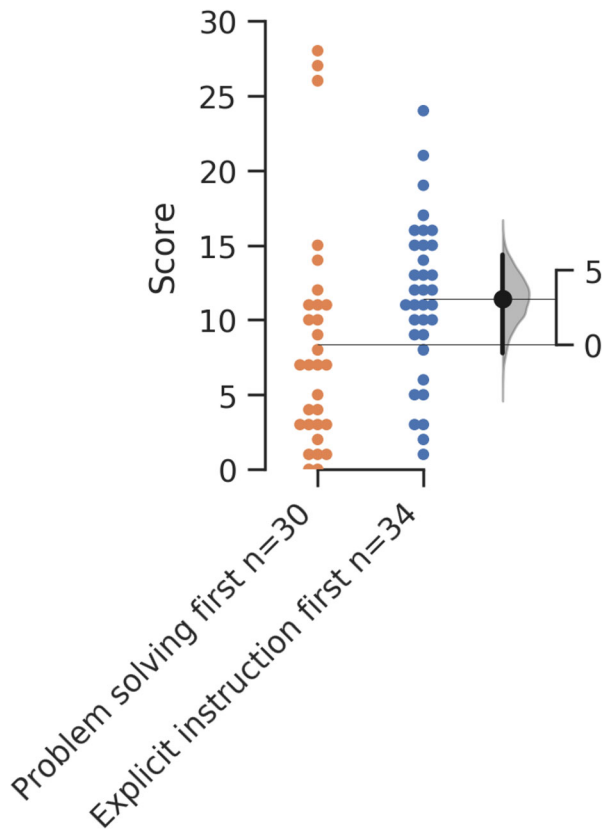
Participants

The participants were 71 Year 5 students from an independent school in Victoria, Australia (a different group of students to those who participated in Experiment 1). They were approximately 10 years old and had not previously received instruction in the conservation of energy or the related concept of efficiency. An entire cohort of Year 5 students were invited to participate and all students who had returned consenting ethics approval forms and who were present on both days of the experiment were included in the sample. The students were randomly assigned to either the group that received explicit instruction first or the group that received problem-solving first. Prior to the study, approval was obtained from the Human Research Ethics Advisory Panel of the lead author's institution.

Materials

The reading task was identical to Experiment 1. The problem-solving booklet was very similar except for two key differences. Learners were given the energy taken in by each light globe per second and the energy given out as heat per second. Therefore, they were

Fig. 3 An estimation plot for transfer questions in Experiment 1. The filled curve indicates the complete distribution for the difference in means, given the observed data. The bold line shows the 95% confidence interval (Ho et al. 2018)



required to use this information and the law of conservation of energy, to compute the light energy given out by each globe prior to calculating the efficiency in the same way as the final transfer question in Experiment 1. This increased the number of procedural steps and therefore the element interactivity (see Table 5). To the eight interacting elements identified for the problems of Experiment 1, an additional four elements need to be added—subtract (1) heat energy given out (2) from electrical energy used (3) to obtain light energy (4) resulting in a total of twelve interacting elements. In other respects, the problems remained the same, with learners being asked to identify the most and least ‘energy saving’ light globes. Again, there were three questions involving three globes, one question involving five globes and one question involving six globes. The structure of post-test was the same as in Experiment 1 and included a set of 7 similar questions and a set of 5 transfer questions. The similar questions involved 3, 3, 3, 4, 4, 5, and 6 globes respectively. Two transfer questions involved fans where learners had to calculate the relevant movement energy, two questions involved globes, one with the irrelevant heat energy omitted, and one with the electrical energy supplied omitted, requiring learners to compute this. The final question involved determining the truth of two statements given about two leaf blowers based upon data presented on the electrical energy used and the heat and sound energy produced by each blower.

Procedure

Learners were randomly assigned to one of two conditions, a problem-solving–lecture sequence and a lecture–problem-solving sequence. The instructional phase also addressed an additional incorrect solution method—many students indicated that the globe giving out the least heat energy was the most efficient. Due to the scheduling of science lessons, the post-test took place on the day following the instructional phase rather than 6 days later, as in Experiment 1. In all other respects, the procedure was identical to Experiment 1 and is illustrated in Fig. 1.

Scoring

Similar test questions were scored in the same way as in Experiment 1 except that there was an additional mark available per question for successfully computing all of the values for the light energy given out. Items on the similar post-test component were highly reliable with Cronbach's $\alpha = 0.96$. The transfer questions were scored similarly to the transfer questions in Experiment 1. The maximum score possible for the similar questions was 48 and the maximum score possible for the transfer questions was 27. Items on the transfer component were highly reliable with Cronbach's $\alpha = 0.88$. Again, there was no subjectivity in the scoring.

Results and Discussion

Means and standard deviations for the post-test question may be found in Table 6. Learners in the lecture–problem-solving condition, who received explicit instruction first, scored significantly higher on similar questions than learners in the problem-solving–lecture condition who received problem-solving first ($t(69) = 2.41, p = 0.02$, Cohen's $d = 0.57$). A visual representation of the data is presented in Fig. 4 (Ho et al. 2018).

Similarly, learners in the lecture–problem-solving condition, who received explicit instruction first, scored significantly higher on transfer questions than learners in the problem-solving–lecture condition who received problem-solving first ($t(69) = 2.35, p = 0.02$, Cohen's $d = 0.56$). A visual representation of the data is presented in Fig. 5 (Ho et al. 2018).

These results strongly support the lecture–problem-solving sequence. Mean test scores for both the similar and transfer problems were almost 50% higher using the lecture–problem-solving sequence compared with the problem-solving–lecture sequence.

General Discussion

Overall, by using fully randomised, controlled experiments only altering the single variable of sequence, the study found that explicit instruction prior to problem-solving was more effective

Table 5 Example of the problem data used in Experiment 2

Globe	Electrical energy used by the globe per second	Heat energy given out per second	Light energy given out per second
A	100 joules	80 joules	
B	150 joules	96 joules	
C	100 joules	70 joules	

Table 6 Means and standard deviations (in brackets) for the similar and transfer questions of Experiment 2

	Problem-solving–lecture ($n = 35$)	Lecture–problem-solving ($n = 36$)
Similar questions (total 48 marks)	21.40 (15.90)	31.06 (17.79)
Transfer questions (total 27 marks)	10.20 (8.87)	15.22 (9.15)

than problem-solving prior to explicit instruction for the acquisition of domain-specific schema. This was true for both experiments. The only difference was that, for Experiment 1, the difference between conditions was significant for similar questions but not for transfer questions, whereas for the higher element interactivity Experiment 2, this difference was significant for both similar and transfer questions with a slightly large effect size for the transfer questions of Experiment 2.

While we did not test a problem-solving first advantage in the current experiments by using low element interactivity information, the conflicting data in the literature (e.g. Loibl and Rummel 2014b; Kapur 2014; Fyfe et al. 2014; Rittle-Johnson et al. 2016) do require an explanation and the element interactivity explanation does appear consonant with previous findings, especially the more recent findings. While much of the early work on productive failure did not use valid experimental or quasi-experimental designs that either eliminated extraneous factors or attempted to equalise them, we believe more recent work is immune to this criticism. Based on these and other studies from other areas, there is evidence that element

Fig. 4 An estimation plot for similar questions in Experiment 2. The filled curve indicates the complete distribution for the difference in means, given the observed data. The bold line shows the 95% confidence interval (Ho et al. 2018)

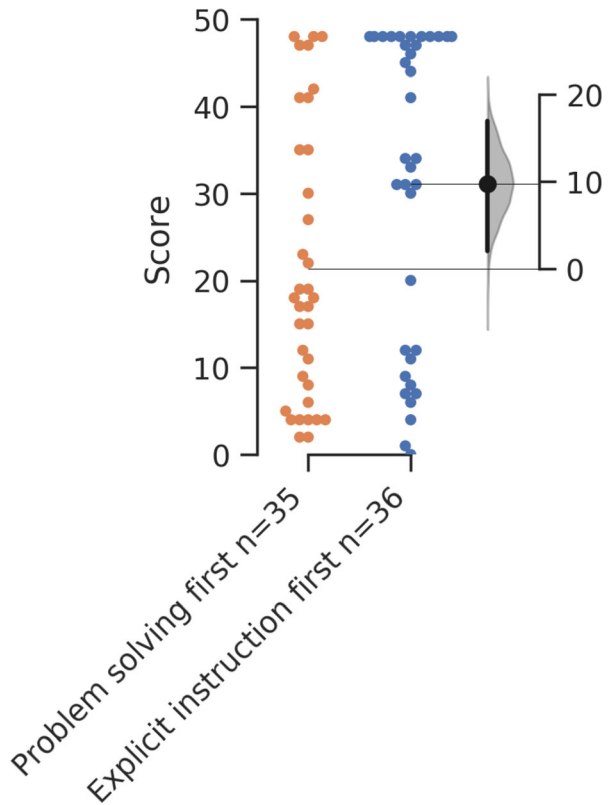
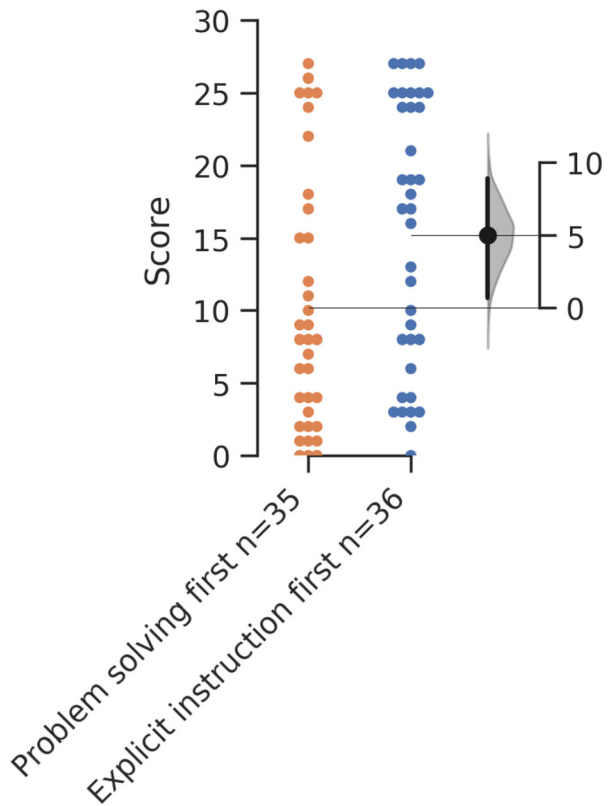


Fig. 5 An estimation plot for transfer questions in Experiment 2. The filled curve indicates the complete distribution for the difference in means, given the observed data. The bold line shows the 95% confidence interval (Ho et al. 2018)



interactivity provides an explanatory variable for the presence or absence of a problem-solving first advantage.

The clearest evidence for a problem-solving first advantage comes from the expertise reversal effect which is a variant of the more general element interactivity effect (Chen et al. 2017). We know that for high element interactivity information, studying worked examples facilitates learning solution schemas compared with problem-solving. However, when expertise increases and element interactivity therefore decreases, the advantage of worked examples reduces and eventually reverses (Kalyuga et al. 2001). The reversal of the results obtained by Glogger-Frey et al. (2015), who obtained an explicit instruction first advantage to the results obtained by Glogger-Frey et al. (2017) who obtained a problem-solving first advantage following increased practice and hence presumably increased expertise and decreased element interactivity, accords with this possibility. The results of Chen et al. (2015); Chen et al. (2016a, b), who found a generation effect in which problem-solving was superior to worked examples for low element interactivity information but worked examples were superior to problem-solving for high element interactivity information, also strongly supports the suggestion that problem-solving first may be beneficial for low but not high element interactivity information.

It follows that we should begin teaching procedures for solving high element interactivity problems with explicit instruction before shifting to more problem-based instructional methods. Martin (2016) has referred to this as ‘load reduction instruction’. It also has been

described as the guidance fading effect (Sweller et al. 2011) for which there is considerable empirical evidence. It is possible that a problem-solving first strategy is effective, but only for relatively low element interactivity information. For novices dealing with high element interactivity information, the problem–example vs. example–problem literature unambiguously indicates that explicit instruction should come first (Hsu et al. 2015; Leppink et al. 2014; Van Gog et al. 2011). The current results that used lecture material instead of worked examples fully accord with these findings.

There also is evidence that studies within a productive failure context also support the element interactivity hypothesis. DeCaro and Rittle-Johnson (2012) observed a problem-solving first advantage for conceptual knowledge but not for procedural knowledge. In their case, conceptual knowledge involved understanding the principle of equivalence—that the equal sign in mathematical questions (i.e. ‘=’) means ‘the same as’ and not ‘put your answer here’. Although it is a fundamental concept, the tasks associated with this concept may have been relatively low in element interactivity compared with the tasks associated with procedural knowledge that usually involve a series of interrelated steps. For instance, one question involved recalling three equations after a 5-s delay, whereas another question required the recall of a definition of the equal sign. Other studies (e.g. Kapur 2014) demonstrated similar findings, replicating a difference favouring problem-solving first on conceptual but not procedural knowledge. Together, these findings may be explained by element interactivity that is likely to be relatively lower for conceptual than procedural knowledge.

Crooks and Alibali (2014) conducted a review of the construct of conceptual knowledge in the mathematics education literature. They noted that conceptual knowledge is often left undefined or is vaguely defined and that the tasks designed to measure conceptual knowledge do not always align with theoretical claims about mathematical understanding. For instance, their review found that the most common conceptual task in the literature on mathematical equivalence involved providing a definition of the equal sign, as in the DeCaro and Rittle-Johnson (2012) study. It is therefore not clear that we can draw firm conclusions from research showing a problem-solving first advantage for conceptual knowledge independently of element interactivity.

Limitations and Future Studies

A limitation of this study is that, although both conditions experienced exactly the same materials and explicit instruction, in order to accommodate this design, learners in the problem-solving–lecture condition completed the experimentally relevant tasks around 15 min earlier in the session than learners in the lecture–problem-solving condition. This would have made results of an immediate post-test difficult to interpret. However, given that the post-test was delayed in both experiments (6 days and 1 day correspondingly), we have assumed a negligible effect of this difference. It might be noted that despite the differential delay in post-tests, brought about by the scheduling requirements of a functioning school, similar results were obtained in the two experiments.

Also, it may be the case that learners in the problem-solving–lecture condition suffered from some form of interference from the subsequent reading task. An attempt was made to minimise this possibility by using a dissimilar set of concepts in the reading task, so the experience may have been similar to that of learners in the lecture–problem-solving condition attending their next lesson of the day. It may also be the case that learners in the lecture–problem-solving condition were more cognitively depleted during

the experimentally relevant tasks and this may have impaired their learning (Chen et al. 2018), although this possibility would only further strengthen the observed benefits of the lecture–problem-solving sequence.

The materials used in both of the experiments of this study were high in element interactivity. While they strongly support the hypothesis that for high element interactivity information, explicit guidance first is necessary, we have not directly tested the hypothesis that problem-solving first may be beneficial for low element interactivity information. Given the evidence for this hypothesis in the literature, it should be tested in future experiments.

Conclusions

The reported study found no evidence to support a problem-solving first strategy as an effective instructional approach. Initial explicit guidance was superior in both experiments. Nevertheless, there may be sufficient evidence in the literature to indicate that problem-solving first is effective under some circumstances. The use of cognitive load theory and element interactivity may resolve the contradiction in the same way as it may have resolved the apparent contradictions between the generation and worked example effects, as well as other effects associated with the expertise reversal effect. For high element interactivity information, initial explicit instruction seems essential. For low element interactivity information, initial problem-solving may be beneficial. At this point, it would seem premature to advise a problem first strategy as a general approach to teaching either problem-solving procedures or curriculum content until the relevant conditions can be clearly defined and the positive effects reliably replicated.

Acknowledgements We would like to acknowledge the students, parents, staff, and leadership team of the Ballarat Clarendon College for their support with this research.

Compliance with Ethical Standards Prior to the study, approval was obtained from the Human Research Ethics Advisory Panel of the lead author's institution.

References

- Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology*, *107*(3), 689–704.
- Chen, O., Kalyuga, S., & Sweller, J. (2016a). Relations between the worked example and generation effects on immediate and delayed tests. *Learning and Instruction*, *45*, 20–30.
- Chen, O., Kalyuga, S., & Sweller, J. (2016b). When instructional guidance is needed. *Educational and Developmental Psychologist*, *33*(2), 149–162.
- 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*(2), 393–405. <https://doi.org/10.1007/s10648-016-9359-1>.
- Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2018). Extending cognitive load theory to incorporate working memory resource depletion: evidence from the spacing effect. *Educational Psychology Review*, *30*(2), 483–501. <https://doi.org/10.1007/s10648-017-9426-2>.
- Cook, M. A. (2017). A comparison of the effectiveness of worked examples and productive failure in learning procedural and conceptual knowledge related to statistics (Order No. 10666475). Available from ProQuest Dissertations & Theses Global. (1984948629).
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.

- Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review, 34*(4), 344–377.
- DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology, 113*(4), 552–568.
- Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2014). An alternative time for telling: when conceptual instruction prior to problem solving improves mathematical knowledge. *British Journal of Educational Psychology, 84*(3), 502–519.
- Geary, D. (2008). An evolutionarily informed education science. *Educational Psychologist, 43*(4), 179–195.
- Geary, D., & Berch, D. (2016). Evolution and children's cognitive and academic development. In D. Geary & D. Berch (Eds.), *Evolutionary perspectives on child development and education* (pp. 217–249). Switzerland: Springer.
- Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction, 39*, 72–87.
- Glogger-Frey, I., Gaus, K., & Renkl, A. (2017). Learning from direct instruction: best prepared by several self-regulated or guided invention activities? *Learning and Instruction, 51*, 26–35.
- Hirshman, E., & Bjork, R. A. (1988). The generation effect: support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 484.
- Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2018). Moving beyond P values: everyday data analysis with estimation plots. *bioRxiv, 377978*.
- Hsu, C.-Y., Kalyuga, S., & Sweller, J. (2015). When should guidance be presented in physics instruction? *Archives of Scientific Psychology, 3*(1), 37–53.
- Hwang, J., Choi, K. M., Bae, Y., Dong, & Shin, H. (2018). Do teachers' instructional practices moderate equity in mathematical and scientific literacy? An investigation of the PISA 2012 and 2015. *International Journal of Science and Mathematics Education*. Advance online publication. <https://doi.org/10.1007/s10763-018-9909-8>.
- Jacobson, M. J., Markauskaite, L., Portolese, A., Kapur, M., Lai, P. K., & Roberts, G. (2017). Designs for learning about climate change as a complex system. *Learning and Instruction, 52*, 1–14.
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*(3), 579–588.
- Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*(4), 651–672.
- Kapur, M. (2014). Productive failure in learning math. *Cognitive Science, 38*(5), 1008–1022.
- Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289–299.
- Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45–83. <https://doi.org/10.1080/10508406.2011.591717>.
- 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*(2), 75–86.
- Lai, P. K., Portolese, A., & Jacobson, M. J. (2017). Does sequence matter? Productive failure and designing online authentic learning for process engineering. *British Journal of Educational Technology, 48*(6), 1217–1227.
- Leppink, J., Paas, F., Van Gog, T., Van der Vleuten, C., & Van Merriënboer, J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32–42.
- Loibl, K., & Rummel, N. (2014a). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science, 42*(3), 305–326.
- Loibl, K., & Rummel, N. (2014b). Knowing what you don't know makes failure productive. *Learning and Instruction, 34*, 74–85.
- Martin, A. J. (2016). *Using load reduction instruction (LRI) to boost motivation and engagement*. Leicester: British Psychological Society.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14–19.
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219–235.
- Rittle-Johnson, B., Fyfe, E. R., & Loehr, A. M. (2016). Improving conceptual and procedural knowledge: the impact of instructional content within a mathematics lesson. *British Journal of Educational Psychology, 86*(4), 576–591.
- Rosenshine, B. (2009). The empirical support for direct instruction. In S. Tobias and T. Duffy (Eds.) *Constructivist instruction: success or failure?* (pp. 201–220). New York: Routledge. <https://doi.org/10.1037/0003-066X.59.1.14>, 59, 1.
- Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475–522.

- Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: the hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, 22(2), 129–184.
- Schwartz, D. L., Lindgren, R., & Lewis, S. (2009). Constructivism in an age of non-constructivist assessments. In *Constructivist Instruction* (pp. 46–73). Routledge.
- Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: the effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103(4), 759–775.
- Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Alevén, V., & Salden, R. (2009). The worked-example effect: not an artefact of lousy control conditions. *Computers in Human Behavior*, 25(2), 258–266.
- Slamecka, N. J., & Graf, P. (1978). The generation effect: delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), 592.
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review*, 22(2), 123–138.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12(3), 185–233.
- Sweller, J., & Paas, F. (2017). Should self-regulated learning be integrated with cognitive load theory? A commentary. *Learning and Instruction*, 51, 85–89.
- Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology*, 4, 434–458.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.
- Sweller, J., van Merriënboer, J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292.
- 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(3), 212–218. <https://doi.org/10.1016/j.cedpsych.2010.10.004>.
- Weaver, J. P., Chastain, R. J., DeCaro, D. A., & DeCaro, M. S. (2018). Reverse the routine: problem solving before instruction improves conceptual knowledge in undergraduate physics. *Contemporary Educational Psychology*, 52, 36–47.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.