CrossMark

Extending Cognitive Load Theory to Incorporate Working Memory Resource Depletion: Evidence from the Spacing Effect

Ouhao Chen [1,2] · Juan C. Castro-Alonso [3] · Fred Paas [4,5] ·
John Sweller [1]

Abstract Depletion of limited working memory resources may occur following extensive mental effort resulting in decreased performance compared to conditions requiring less extensive mental effort. This "depletion effect" can be incorporated into cognitive load theory that is concerned with using the properties of human cognitive architecture, especially working memory, when designing instruction. Two experiments were carried out on the spacing effect that occurs when learning that is spaced by temporal gaps between learning episodes is superior to identical, massed learning with no gaps between learning episodes. Using primary school students learning mathematics, it was found that students obtained lower scores on a working memory capacity test (Experiments 1 and 2) and higher ratings of cognitive load (Experiment 2) after massed than after spaced practice. The reduction in working memory capacity may be attributed to working memory resource depletion following the relatively

---

✉ Ouhao Chen
  chenouhao_lucky@126.com

✉ John Sweller
  j.sweller@unsw.edu.au

  Juan C. Castro-Alonso
  jccastro@ciae.uchile.cl

  Fred Paas
  paas@fsw.eur.nl

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

[2]   Present address: National Institute of Education, Nanyang Technological University, Singapore, Singapore

[3]   Center for Advanced Research in Education (CIAE), Universidad de Chile, Santiago, Chile

[4]   Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands

[5]   Early Start Research Institute, University of Wollongong, Wollongong, Australia

🖄 Springer

prolonged mental effort associated with massed compared to spaced practice. An expansion of cognitive load theory to incorporate working memory resource depletion along with instructional design implications, including the spacing effect, is discussed.

Resource depletion occurs when cognitive effort on one task depresses performance on a later task due to the first task depleting the resources available to complete the second task. In particular, working memory resource depletion occurs when the two tasks have similar cognitive components, and the depressed performance on the second task is due to a reduced working memory capacity. Depleted resources can be restored after rest periods (e.g., Tyler and Burns 2008). Our theoretical aim is to suggest that cognitive load theory can be substantially extended by including working memory resource depletion. Our empirical aim is to apply this extension to the spacing effect, thus explaining the spacing effect and incorporating it as a cognitive load theory effect. The spacing effect occurs when information processing that is spaced over longer periods (spaced presentation) results in superior test performance compared to the same information processed over shorter periods (massed presentation). In both cases, the amount of information presented and the total time processing the information are identical. The only difference is that spaced conditions include one or more temporal gaps between segments of information processed, while massed conditions have all the information presented consecutively without gaps. We will begin by discussing working memory resource depletion followed by its consequences for cognitive load theory and how the theory, enhanced by including resource depletion, can be used to explain the spacing effect. Two experiments were designed to test the suggested explanation.

Working Memory Resource Depletion

We are aware of very little work that has explored directly the effects of cognitive effort on working memory depletion. Furthermore, the limited work that is available is not concerned with cognitive effort while learning. A relevant study that we could find was conducted by Schmeichel (2007). He ran four experiments, with Experiments 1, 2, and 4 investigating resource depletion by presenting undergraduate students with difficult tasks and assessing the subsequent effects on working memory capacity. Experiment 1 found that requesting people to ignore irrelevant words that appeared on a screen depicting a person speaking without audio reduced their working memory capacity, compared to people who were not requested to ignore the irrelevant words. Actively attempting to ignore the words, a difficult task, depleted resources. Experiment 2 asked people to write a story without using the letters *a* or *n* and compared them to people who were similarly asked to write a story but without the restriction of avoiding the two letters. Performance scores on the subsequent working memory capacity test revealed that writing a story with the restriction reduced working memory capacity, compared to writing a story without the restriction. The results indicate that the effort involved in writing a story in the restricted condition depleted working memory resources more than the effort involved in the unrestricted condition. Lastly, Experiment 4 compared people who were asked to exaggerate their emotional responses when watching emotional films compared to people who simply watched the films. The effort involved in exaggerating emotional responses

depleted working memory resources more than expressing normal emotions. These experiments provided convincing evidence that increasing cognitive effort depleted working memory resources. None of the tasks were learning tasks but it seems reasonable to hypothesize that learning conditions requiring an increase in cognitive effort or an extension of cognitive effort over time would have a similar effect on working memory resource depletion.

There is previous work that obtained results in accord with the study by Schmeichel (2007), but did not use a measure of working memory capacity as a dependent variable. Schmeichel et al. (2003) used similar independent variables to Schmeichel (2007), but instead of using working memory capacity as a dependent variable used performance on reasoning, problem solving, or reading comprehension tasks. The results indicated that under resource depletion conditions, performance on these tasks deteriorated. In light of the subsequent Schmeichel (2007) study, it may be appropriate to speculate that the deterioration in test performance after resource depleting activity was due to working memory capacity reductions.

Whereas the studies by Schmeichel and colleagues found general depletion effects by showing that engaging in self-control tasks can lower performance on subsequent tests, Healey et al. (2011) extended these findings by providing evidence for more specific working memory depletion effects. They varied the similarity among stimuli that were to be ignored in an initial task and those that were to be remembered in a subsequent working memory task. In four experiments, the authors observed that ignoring words impaired performance on a subsequent working memory test based on words (Experiment 1) but not on arrows (Experiment 2), whereas ignoring arrows impaired performance on a working memory test based on arrows (Experiment 3) but not on words (Experiment 4). Thus, Healey et al. (2011) showed that depletion effects only occurred when there was a match between the to-be-ignored stimuli in the first task and the to-be-remembered stimuli in the working memory task. Again, while these depletion effects were due to cognitive processing, none of the processing tasks included learning.

Interestingly, the deterioration in performance on complex tasks following mental effort found by Schmeichel et al. (2003) did not extend to simple tasks. Unlike the complex tasks, there was no effect on a nonsense syllable memorization task. Given the centrality of element interactivity to cognitive load theory, obtaining an effect using a high element interactivity but not a low element interactivity task, a result repeatedly obtained by experiments using cognitive load theory (see next sections), suggests a linkage between working memory resource depletion and cognitive load theory. That proposed linkage provides the major impetus for the current work. Thus, our aim is to extend cognitive load theory by adding resource depletion to the theory. That addition may have general instructional consequences, but in the present paper, we will only use the spacing effect to test the advisability of adding resource depletion to cognitive load theory.

Cognitive Load Theory

Cognitive load theory is an instructional theory based on our knowledge of evolutionary educational psychology and its relation to human cognitive architecture (Sweller et al. 2011). The theory uses Geary's division of biologically primary and biologically secondary information (Geary 2008, 2012; Geary and Berch 2016) to determine categories of knowledge that are amenable to instructional design (Paas and Sweller 2012; Sweller 2016a). Primary knowledge is knowledge leading, as examples, to our ability to solve problems, self-regulate our thoughts,

or learn to listen to and speak our native language that we have evolved to acquire over many generations. Such knowledge tends to be generic-cognitive rather than domain-specific (Sweller 2015) and is critically important to our basic cognitive psychology. Because of its evolutionary importance, we have evolved to acquire it automatically and unconsciously without explicit tuition and so it tends not to be amenable to instruction. In contrast, biologically secondary knowledge includes information that a particular culture has deemed to be important. We are able to acquire secondary knowledge through conscious effort on the part of the learner and explicit instruction by teachers or other instructors (Kirschner et al. 2006; Sweller et al. 2007). Unlike the generic-cognitive skills that commonly are biologically primary, secondary skills are commonly domain-specific (Tricot and Sweller 2014). Virtually, every topic that is taught in educational and training institutions provides an example of biologically secondary knowledge and skills.

In large part, cognitive load theory is concerned with the acquisition of the domain-specific, biologically secondary information that is the subject of educational and training curricula. The information processes relevant to cognitive load theory mimic the information processes of evolution by natural selection (Kirschner et al. 2006; Sweller et al. 2007; Sweller and Sweller 2006) and provide the cognitive architecture used by the theory (Sweller 2016b). That architecture can be described by the following five basic principles. These principles are biologically primary and so cannot be taught because they are learned automatically.

- Information store principle. All learned information, both primary and secondary, is stored in long-term memory. The long-term memory store is very large with no known limits.
- Borrowing and re-organizing principle. The bulk of information stored in the long-term memory store is borrowed from other people by copying what they do, listening to what they say, and reading what they write. Once obtained, information is re-organized by combining it with previously stored information.
- Randomness as genesis principle. Information that cannot be obtained from others can instead be obtained during problem solving using random generation and test processes. Information can be retained in long-term memory if it is useful or jettisoned if it is not useful.
- Narrow limits of change principle. In order to preserve the contents of long-term memory, only very limited amounts of novel information can be processed at any given time. Accordingly, novel information from the external environment is processed by a working memory that is severely limited both in capacity and the duration over which it can hold information.
- Environmental organizing and linking principle. Based on environmental signals, unlimited amounts of information can be transferred from long-term to working memory, in order to generate action that is appropriate to the prevailing environment. When dealing with information previously organized and stored in long-term memory, there are no known working memory limits.

This cognitive architecture, based on biologically primary information, is used to acquire, organize, and store biologically secondary information for subsequent use. As such, it provides the cognitive base for instructional design. The purpose of instruction is to assist learners to acquire novel, domain-specific, biologically secondary knowledge intended to be stored in long-term memory. To be stored in long-term memory, novel information first must be processed in working memory which is severely limited in capacity when processing novel information.

Element Interactivity and Cognitive Load

Novel information varies along an element interactivity continuum where element interactivity refers to the number of elements that must be processed simultaneously (Sweller 2010). Interacting elements must be processed simultaneously in working memory. For example, if learning to deal with a mathematical equation, all the symbols that constitute that equation must be processed simultaneously since most changes affect the entire equation. When processing a mathematical equation, element interactivity is high and so working memory load is high. In contrast, if a new vocabulary is being learned, each item can be learned without reference to any other item. Element interactivity and working memory load are low, even though the task may be difficult if there are many vocabulary items. Such element interactivity caused by the intrinsic properties of the information imposes an *intrinsic cognitive load*. Similarly, element interactivity associated with the same information can vary depending on the instructional procedures used, thus varying *extraneous cognitive load*. Different instructional procedures can increase or decrease the number of elements that must be simultaneously processed (Sweller 2010). A major aim of cognitive load theory has been to devise instruction intended to reduce the levels of element interactivity imposed by instructional procedures. Where element interactivity is reduced by changing instructional procedures dealing with the same information, extraneous cognitive load is reduced.

An implicit assumption of cognitive load theory, based on the narrow limits of change principle, has been that working memory capacity is relatively constant with the only major factor influencing capacity being the content of long-term memory. As indicated by the environmental organizing and linking principle, the limitations of working memory when dealing with novel information can be eliminated if the same information has been stored in long-term memory. High element interactivity information, once stored in long-term memory can be transferred to working memory as a single element that imposes a minimal working memory load.

Based on the working memory resource depletion hypothesis investigated in the present work, the assumption that the content of long-term memory provides the only major determinant of working memory characteristics may be untenable. It is possible that extensive cognitive effort can substantially deplete working memory resources and if so, that factor may be important when designing instruction. That possible extension of cognitive load theory provides the rationale for the current experiments. In those experiments, the spacing effect was used to investigate possible instructional consequences of working memory resource depletion.

The Spacing Effect

The spacing effect occurs when information that is presented over a longer period with spaces between presentation episodes results in superior learning compared to identical information presented over a shorter period with no interruptions. The effect is sometimes referred to as the massed vs. spaced effect. It is well established with a very large number of replications over many decades dating back to the beginnings of experimental psychology (Ebbinghaus 1885/1964). For two recent educationally relevant papers, see Gluckman et al. (2014) and Kapler et al. (2015).

There are many theories used to explain the spacing effect with no consensus supporting any particular theory (see Küpper-Tetzel 2014), although Delaney et al. (2010), in an extensive review, favor the study-phase retrieval theory. This theory suggests that the gaps between learning events during spaced practice increase forgetting compared to massed practice. As a consequence, more effortful retrieval is required at the next learning event for spaced as opposed to massed learning, thus assisting in memory by supporting more active retrieval. This theory is plausible although it is difficult to provide direct evidence for it. It is, of course, entirely possible that the spacing effect has multiple causes.

We suggest that working memory resource depletion provides a strong candidate explanation for at least some versions of the spacing effect. Where massed practice is continuous, we might expect working memory resource depletion to result in reduced learning during the practice sessions compared to spaced learning, resulting in the spacing effect. Direct evidence for this hypothesis could be obtained by testing learners' working memory capacity after massed or spaced learning and immediately prior to providing a content test. We tested the hypothesis in two experiments with primary school students learning mathematics. It was predicted that working memory capacity would be reduced by massed compared to spaced practice and that this reduction in working memory capacity would be associated with the spacing effect. Higher working-memory resource depletion after massed practice was expected to result both in lower performance scores (Experiments 1 and 2) and higher perceived difficulty ratings on the working memory capacity test (Experiment 2). Considering the study by Healey et al. (2011), which only showed depletions with similar stimuli, in both experiments, we used similar mathematical depictions for the practice tasks, the working memory tests, and the content tests.

Experiment 1

This study was conducted as an initial investigation into the relation between the spacing effect and working memory depletion. The goal was to test the hypothesis that the spacing effect is caused by working memory depletion following massed compared to spaced practice. Accordingly, we predicted that massed learning would result in lower content test scores than spaced learning and that a working memory test immediately preceding the content test would reveal more working memory depletion for the massed than for the spaced group.

Method

Participants Two classes totaling 85 Year 4 students with a mean age of approximately 10 years were chosen from a primary school in the urban area of Chengdu, China. The two classes were taught by the same mathematics teacher. Using a quasi-experimental design, one class was assigned to the massed and the other to the spaced learning condition. Thirteen students from the massed learning condition and 18 from the spaced learning condition were excluded from the final analyses because they recorded items on paper for later use rather than memorized information during the working memory test. Therefore, data from the remaining 54 students (30 in the massed and 24 in the spaced condition) were used for the analyses. Students and teachers from the two classes did not know the purpose of this experiment. The lesson content required students to learn how to add together two positive fractions with different denominators, an area that had not previously been taught in class.

Materials The materials comprised slides to present the fraction addition information that needed to be learned, a working memory test with an answer sheet for this test, and a content post-test of fraction addition knowledge. The learning slides consisted of three worked example–problem solving pairs in the domain of algebra, specifically focused on how to add up two positive fractions with different denominators. An example of a worked example–problem solving pair is shown in Fig. 1.

For each pair, students were presented an example showing all solution steps, followed by a similar problem, in which they were asked "Please calculate the following algebraic expression", that had to be solved. All three worked example–problem solving pairs were presented as Microsoft™ PowerPoint™ 2016 slides on a screen in front of the classroom. For each pair, the worked example and problem solving tasks were presented on different slides, and so students could not refer to the worked example when solving its paired problem. For both conditions, after students had attempted to solve a problem, the final answer without the steps to the answer was provided. Each of the resulting six slides was presented for 150 s, leading to a total duration of 900 s.

Based on the study of Conlin et al. (2005), a complex working memory test for children was developed with PowerPoint slides. This working memory test consisted of a memory task interrupted by a processing task. In this case, the memory task was to remember the first digit of a given equation, while the processing task was to answer if that equation showed a correct or an incorrect solution (see example in Fig. 2). The test presented four difficulty levels with three trials each, thus totaling 12 trials. The levels are ranged from three to six (Level 3 included three equations per trial and Level 6 included six equations per trial). As shown in the Level 3 example of Fig. 2, after observing Equation 1, the participants indicated if the solution

Fig. 1 An example of (a) a worked example – (b) a problem solving pair in Experiment 1

of the shown equation was correct or incorrect (by marking a happy or a sad face, respectively), while memorizing the first digit of the equation. Then, Equation 2 was shown; students answered whether it was correct or not, while memorizing that first digit, and so on. At the end of each trial, the participants had to write down from memory all the first digits of the equations, in the order they were shown. A 27-page booklet (A4 size, printed one-sided) was designed for students to record their answers. For each trial, separate pages were used for the students to indicate the accuracy of the equations and the memorized first digits of the equations.

As suggested by Unsworth et al. (2005), we measured performance in both the memory and processing tasks. Thus, one point was given per accurately recalled first digit (in the correct order), and one point was awarded per accurately judged equation. In other words, we followed the recommendation of Conway et al. (2005) against span scoring and counted every single correct digit accurately memorized. For each trial, whether each equation presented to learners was correct or incorrect was determined randomly with the restrictions that at least one third of the equations were correct or incorrect for a given trial and that half of the equations were correct for the total test. The internal consistency, as estimated with Cronbach's coefficient alpha, was .96 for memorized digits and .88 for processed equations.

A content post-test with five problems was administered after the learning phase, for example, please calculate $\frac{1}{2} + \frac{1}{2}$. All problems required the students to add up two positive fractions with different denominators. Four steps were required to solve a problem. Each correct step was marked with one point, resulting in a maximum score for each problem of 4, and 20 for the entire post-test of five problems. All raw scores were converted to percentage correct scores for the analysis. The Cronbach's alpha for this content post-test was .98.

Procedure The general procedure included three phases: learning, working memory test, and post-test (see Fig. 3). The only difference between the spaced and massed conditions was in the learning phase. In the massed condition (Class A in Fig. 3), students were presented the learning materials for 900 s, followed by a working memory test for around another 900 s. Finally, a post-test was administered for the final 600 s. The whole intervention for the massed condition lasted approximately 40 min, and was done within the teaching time of a regular class. In the spaced condition (Class B in Fig. 3), the intervention was spread over 4 days. On each of the first 3 days, students were required to both study one worked example and then solve a similar problem in 300 s. Consequently, the total time for learning was 900 s. On the fourth day, the working memory test (900 s) and post-test (600 s) were administered.

For the working memory test, a general instruction slide was presented for 60 s to provide instruction on how to perform this test, followed by a practice task for another 60 s. In the test, each equation of each trial was shown for 5 s followed by another 5 s for students to immediately circle a happy or a sad face to indicate whether each equation was correct or incorrect, respectively. For each trial, immediately after the accuracy of the last equation
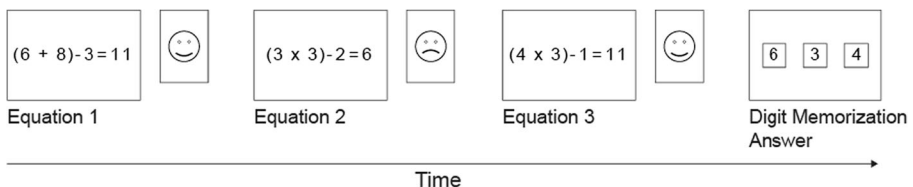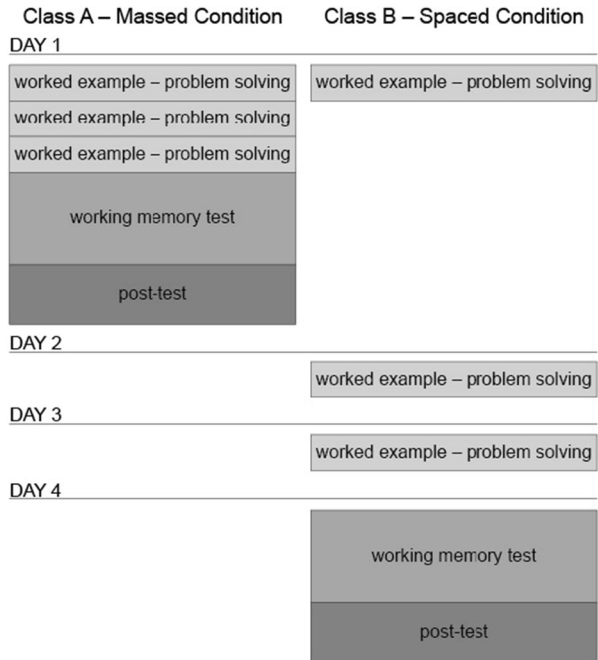

Fig. 2 Example of a correctly solved Level 3 trial in the working memory test of Experiment 1

Fig. 3 The general procedure of Experiment 1, comparing massed (Class A) to spaced (Class B) conditions

was indicated, students were required to write down the memorized first digits in order, with 2.5 s provided for each digit. The total time for the working memory test was 900 s for both conditions.

Results and Discussion

Three ANOVAs were used to analyze the data, in which the condition (massed vs. spaced) was the independent variable using a between-subject designed. The dependent variables were (a) scores on the memorized digits of the working memory test, (b) number of correctly classified equations on the working memory test, and (c) content post-test scores. Mean percentages and standard deviations for these dependent variables, as a function of the two conditions, are displayed in Table 1.

The effect of condition was significant for the memorized digits of the working memory test, $F(1, 52) = 11.029$, MSE = 342.91, $p = .002$, $\eta_P^2 = .175$. Similarly, the effect of condition for the classified equations of the working memory test was significant, $F(1, 52) = 8.314$, MSE = 200.20, $p = .006$, $\eta_P^2 = .138$. Again, the effect of condition was significant for the post-test, $F(1, 52) = 6.305$, MSE = 1793.71, $p = .015$, $\eta_P^2 = .108$. These results indicated that the

Table 1 Mean percentages (and SD) for memorized digits, classified equations, and post-tests, for Experiment 1

Condition	n	Digits	Equations	Post-test
Massed	30	47.28 (39.00)	80.19 (14.82)	45.67 (47.74)
Spaced	24	78.47 (27.22)	91.36 (13.26)	74.79 (34.38)

students in the spaced condition memorized more digits, classified more equations correctly, and achieved higher scores on the post-test than the students in the massed condition.

The aim of Experiment 1 was to test the hypothesis that resource depletion after cognitive effort was influenced by working memory depletion and that the spacing effect in turn, was due to differential effects of working memory depletion following spaced rather than massed learning. The results provided support for this hypothesis. We replicated the spacing effect and, more importantly, indicated that it was associated with a reduction in working memory resources for learners presented information under massed rather than spaced conditions. Immediately prior to their content test, the massed condition had significantly fewer working memory resources than the spaced condition.

However, there were two procedural concerns associated with this experiment. The first concern was that the experiment used a quasi-experimental design rather than a fully randomized, controlled experiment. Intact classes were used as experimental groups. In order to maintain ecological validity with children studying real curriculum materials, it was not possible to use randomization. We ameliorated this problem in Experiment 2 by using a counterbalanced design that could equate for any differences between classes.

The second issue was the failure of 31 students to follow the required procedure on the working memory test. When indicating whether an equation was correct or incorrect, these students also wrote down the first number of the equation, eliminating any need to remember it, as the memory task required. Those students were eliminated from the data analysis. That issue also was rectified by a change in the procedure of Experiment 2.

Experiment 2

As was the case in Experiment 1, the ultimate concern of Experiment 2 was whether cognitive load theory should be modified to include resource depletion as a factor when considering working memory capacity, rather than assuming a fixed working memory capacity for each individual. The spacing effect was used again to investigate this issue by considering the relation between spaced learning and working memory resource depletion with the same hypothesis as in Experiment 1. In Experiment 2, Year 5 students learning algebra were tested with two intact classes. There were two discrete areas of algebra taught using counterbalancing to reduce the effects of possible class differences. For the first area taught in week 1, Class A was presented the material in massed form while Class B was presented the material in spaced form, replicating the experimental design of Experiment 1. For the second area taught in week 2, the presentation modes were reversed: Class A was presented the material in spaced form while Class B was presented the material in massed form. The data then could be analyzed using a 2 (Condition: massed vs. spaced) × 2 (Test Phase: week 1 vs. week 2) ANOVA with repeated measures on the second factor. In this manner, any differences between the classes were eliminated by counterbalancing.

A second change to the procedure was to prevent students from writing during the presentation of the memory equations. Rather than indicating whether an equation was correct or incorrect immediately after it was seen during a trial, students had to wait until all of the equations for a trial had been presented and then recall whether the equations had been correct or not immediately prior to recalling the first digits of each equation. Students had to process each equation and remember the results of that processing. By preventing them from writing anything while the equation was present, they were unable to write the first digit. When tested,

they had to recall the correct and incorrect sequences for that trial (e.g., correct, incorrect, incorrect, correct) and the sequence of initial numbers for each equation.

A third variation from Experiment 1 was that we collected subjective ratings of task difficulty for the working memory test. These measures provided information concerning cognitive load in addition to the working memory measures (see Paas 1992; Paas et al. 2003).

Method

Participants Two classes totaling 82 Year 5 students with a mean age of approximately 11 years of the same primary school used for Experiment 1, were chosen for this experiment. As in Experiment 1, in week 1, a class was assigned to the massed learning condition, while another class was assigned to the spaced learning condition. In week 2, using different algebra curriculum materials, the two classes were reversed with the spaced learning condition in week 1 becoming the massed learning condition in week 2, and vice versa, resulting in counterbalancing (see Fig. 4). The number of participants from each class in each week may be found in Table 2. Note that the numbers of participants in Table 2 reflect only those that took part in both weeks. Whereas there were 82 participants in week 1, only 61 of them could eventually be used in the statistical analyses. This dropout was caused by the fact that some students had to participate in training for a national mathematics competition, and some others omitted writing down their personal details.

Materials As in Experiment 1, slides for the learning phase, a working memory test with an answer sheet for this test, and a post-test of algebra content were used. There were two separate versions of the learning phase slides for weeks 1 and 2. The week 1 slides taught students how to calculate with negative numbers. Three pairs of worked example–problems solving pairs were presented. For each problem of the pair, students were asked "Please calculate the following algebraic expression". In week 2, slides taught students how to solve fractional
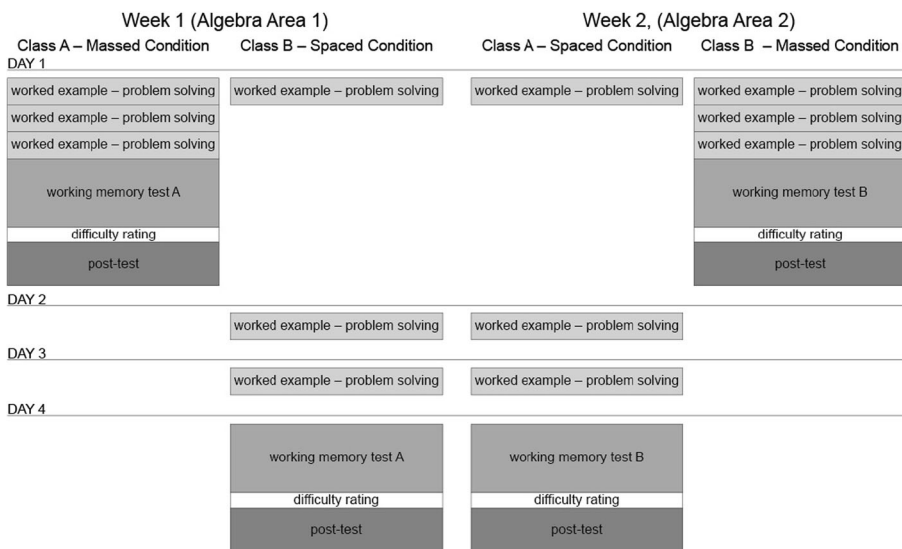

Fig. 4 The general procedure of Experiment 2 in which Class A followed a massed–spaced order, while Class B followed a spaced–massed order

Table 2 Mean percentages (and SDs) for memorized digits, classified equations, and post-tests, for both weeks of Experiment 2

Condition	Week 1				Week 2			
	n	Digits	Equations	Post-test	n	Digits	Equations	Post-test
Massed	28	62.22 (30.76)	65.86 (19.19)	63.33 (44.12)	33	77.10 (30.32)	76.83 (19.03)	21.36 (23.18)
Spaced	33	51.03 (36.13)	79.73 (15.50)	83.57 (33.77)	28	85.56 (21.74)	83.65 (10.23)	30.00 (28.42)

equations. Again, we designed three worked example–problem solving pairs. All worked example–problem solving pairs were again computer-based using the same presentation time as in Experiment 1. Examples of worked example–problem solving pairs for weeks 1 and 2 are shown in Fig. 5.

The working memory test had two equivalent versions for Experiment 2: Version A for week 1 and Version B for week 2. Each version included PowerPoint slides that showed mathematics equations for 5 s again. For each equation, the processing task was to determine if the equation was correct or incorrect but unlike Experiment 1, to also remember the correct/ incorrect sequences. The digit memory task was identical to Experiment 1. For the whole test, half of the equations were randomly assigned to be correct and half incorrect. The Cronbach's alphas for memorized digits in weeks 1 2 were .93 for both weeks, whereas, for processing equations in weeks 1 and 2 the Cronbach's alphas were .83 and .72, respectively. In Experiment 2, all students were required to write down the answers only after the end of each trial. They had to write down all the first digits in the order they were shown and then indicate from memory the accuracy of the mathematics equations in that trial. As in Experiment 1, we measured performance on both the memory of the first number of each equation in each trial and the accuracy of the processing tasks except that memory of the accuracy of the sequence of the equations for each trial was included. One point was given per accurately recalled digit (in the correct order), and one point was awarded per accurately classified and remembered equation. With the change of testing procedure of the working memory test, all students in this experiment followed instructions. In both versions of the working memory tests, there were five difficulty levels with three trials each, thus totaling 15 trials. The levels ranged from one to five (Level 1 included one equation per trial, and Level 5 included five equations per trial). The instructions and two practice tasks of two equations for each practice task were given before the tests.

A 3-page booklet (A4 size, printed one-sided) was designed for students to record their answers of the working memory tests. This booklet was used for weeks 1 and 2. The first page was for students to provide their name and class number. It also had some important notes for taking the working memory test, such as "you can only record your answers for memorized digits when you see the slide saying *please start writing*" and "you cannot record your answers

Week 1

a)

$(-5)-(-8)+(-15)$

$= (-5) + 8 - 15$

$= 8 - 5 - 15$

$= 3 - 15$

$= -(15-3)$

$= -12$

b)

$(-4)-(-6)+(-12)$

Week 2

c)

$\dfrac{x+2}{4} - \dfrac{2x-3}{7} = 1$

$\left(\dfrac{x+2}{4} - \dfrac{2x-3}{7}\right) \times 4 \times 7 = 1 \times 4 \times 7$

$\dfrac{x+2}{4} \times 4 \times 7 - \dfrac{2x-3}{7} \times 4 \times 7 = 1 \times 4 \times 7$

$7(x+2) - 4(2x-3) = 28$

$7x + 14 - 8x + 12 = 28$

$7x - 8x + 14 + 12 = 28$

$-x + 26 = 28$

$-x = 28 - 26$

$-x = 2$

$x = -2$

d)

$\dfrac{x+2}{3} - \dfrac{2x-3}{5} = 1$

Fig. 5 Examples of (**a**) a worked example – (**b**) a problem solving pair for week 1 and of (**c**) a worked example – (**d**) a problem solving pair for week 2 in Experiment 2

anywhere else on this answer sheet except in the printed empty boxes". The practice section and the sections from Levels 1 to 5 followed. For each section, students had to record their memorization of the accuracy of each equation in a trial after the empty boxes for recording the memorized digits in order. Thus, students had to write down the memorized digits in order first and then indicate, from memory, the accuracy of the series of equations. Answers to all sections were recorded on the second and third pages. At the end of page 3, a 9-point symmetrical category scale (Paas 1992; Paas et al. 2003) was provided for students to rate their perceived difficulty of the working memory test. The numerical values and labels assigned to the categories ranged from *very, very easy* (1) to *very, very difficult* (9). The scale was explained to the students during the instruction phase.

A post-test with five algebra questions was presented after the learning phase in weeks 1 and 2. In week 1, all questions were calculations with negative numbers, for example, please calculate $(-5) - (-8) + (-9)$, whereas in week 2, all questions were concerned with how to solve fractional equations, for example, please calculate $\frac{x+2}{2} - \frac{x+2}{2} = 2$. The maximum score for each question was 4. Answers for week 1 required five steps for solution but only four key steps (excluding step 3) were scored with each correct step allocated one mark giving a total of 20 marks. For week 2, the key steps 1, 3, 6, and 9 were scored, again resulting in a full mark of 20. All raw scores were converted to percentage correct scores for analysis. For the post-test scores, Cronbach's alphas of .98 in week 1, and .94 in week 2 were obtained.

Procedure Other than as indicated above, the general procedure was the same as in Experiment 1.

Results and Discussion

A 2 (Condition: massed vs. spaced) × 2 (Test Phase: week 1 vs. week 2) design with repeated measures on the second factor was used in this experiment. As in Experiment 1, the dependent variables were the number of memorized digits, correctly classified equations, and post-test performance. Mean percentages and standard deviations for these dependent variables, as a function of the two factors, are shown in Table 2.

An ANOVA performed on the memorized digits of the working memory test yielded no significant main effect of condition, ($F < 1$, *ns*). The main effect of week was significant, $F(1, 59) = 17.081$, MSE $= 219.17$, $p < .001$, $\eta_P^2 = .225$, indicating that the scores of memorized digits in week 2 were significantly higher than those in week 1. The interaction between condition and test phase was not significant, $F(1, 59) = 2.700$, MSE $= 219.17$, $p = .106$, $\eta_P^2 = .044$.

An ANOVA performed on the means of processing equations yielded a significant main effect of condition, $F(1, 59) = 5.307$, MSE $= 64.04$, $p = .025$, $\eta_P^2 = .083$, indicating a lower number of correctly classified equations in the massed compared to the spaced learning conditions. The main effect of week was significant, $F(1, 59) = 14.262$, MSE $= 46.07$, $p < .001$, $\eta_P^2 = .195$, showing that the number of correctly processed equations in week 2 was significantly higher than in week 1. The interaction between condition and test phase was not significant, $F(1, 59) = 1.653$, MSE $= 46.07$, $p = .204$, $\eta_P^2 = .027$.

An ANOVA performed on the mean percentage correct scores on the post-test revealed a main effect of condition, $F(1, 59) = 4.074$, MSE $= 1550.13$, $p = .048$, $\eta_P^2 = .065$, showing that

students in the spaced learning condition achieved higher scores than students in the massed learning condition, which is indicative of a spacing effect. The main effect of week was significant, $F(1, 59) = 100.45$, MSE $= 688.26$, $p < .001$, $\eta_P^2 = .630$, indicating that the percentage correct scores on the post-test in week 1 were significantly higher than those in week 2, suggesting that the material and test of week 2 was considerably more difficult than in week 1. The interaction between condition and test phase was not significant, $F(1, 59) = 1.481$, MSE $= 688.26$, $p = .228$, $\eta_P^2 = .024$.

With respect to the subjective rating scales during the working memory test, a 2 (Condition: massed vs. spaced) × 2 (Test Phase: week 1 vs. week 2) ANOVA with repeated measures on the second factor was performed to analyze these ratings of perceived difficulty. The means and standard deviations are given in Table 3.

For the subjective difficulty ratings, the main effects of condition, $F(1, 59) = 18.574$, MSE $= 5.03$, $p < .001$, $\eta_P^2 = .239$, and test phase, $F(1, 59) = 55.383$, MSE $= 4.51$, $p < .001$, $\eta_P^2 = .484$, were significant. However, these main effects were qualified by a significant interaction between condition and test phase, $F(1, 59) = 10.249$, MSE $= 4.51$, $p = .002$, $\eta_P^2 = .148$. The interaction indicated that the expected decrease in perceived difficulty as a function of practice from weeks 1 to 2 was significantly higher for students in the massed–spaced order than for students in the spaced–massed order. Massed practice after spaced practice was experienced as harder than spaced practice after massed practice. The fact that the working memory test was perceived as being particularly easy when it followed spaced practice in week 2 may be due to that condition being the only one influenced by a spacing effect, practice effect, and a relative complexity effect. The practice effect holds that participants were expected to perceive the second attempt on the working memory test as less difficult than the first attempt. The relative complexity effect holds that spaced practice would be perceived as less difficult after massed practice than spaced practice before massed practice.

The results of Experiment 2 confirm the findings of Experiment 1. A spacing effect is associated with more working memory resource depletion following massed presentation compared to spaced presentation. These results suggest a possible causal explanation of the spacing effect. More importantly, they suggest that cognitive resource depletion, at least under some circumstances, may be characterized as working memory resource depletion.

General Discussion

Several conclusions flow from the findings of the two experiments reported in this paper. The results support the general conclusion that the resource depletion that occurs after cognitive effort can be characterized as working memory resource depletion, at least under some circumstances. In both experiments, tests of working memory capacity decreased significantly immediately after cognitive exertion compared to the insertion of rest periods during learning

Table 3 Mean ratings (and SD) of perceived difficulty for the working memory tests of Experiment 2

Condition	n	Week 1	n	Week 2
Massed	28	7.48 (2.18)	33	5.85 (2.05)
Spaced	33	6.96 (2.27)	28	2.86 (2.26)

Potential range of ratings = 1–9

and prior to the working memory test. Additional evidence in support of the working memory resource depletion explanation was provided in Experiment 2 by the interaction showing higher ratings of perceived difficulty on the working memory test after massed practice (observed in the massed–spaced condition). Further work will be needed to establish whether working memory depletion is the sole cause of resource depletion or whether there are other possible causes as well.

The results also suggest that the spacing effect, which has been replicated on numerous occasions over many decades, can be directly attributed to working memory resource depletion. Currently, while the spacing effect is empirically well-established, there is no consensus on its cause (e.g., Benjamin and Tullis 2010; Delaney et al. 2010). Much research associated with the effect is concerned more with its empirical and, to a lesser extent, practical characteristics rather than its theoretical context (Küpper-Tetzel 2014). Our results provide both empirical evidence and a theoretical context for at least some versions of the effect. They do not, of course, eliminate the possibility that the effect has multiple causes under different conditions.

The demonstration of the spacing effect using students under real classroom conditions studying relevant curriculum materials, rather than under laboratory conditions using artificial materials, is rare, as pointed out by Kapler et al. (2015). The effect is usually assumed to have educational significance but is rarely tested in educational contexts and seems to have little practical impact. Additional demonstrations of the effect within classroom rather than laboratory contexts may be important.

An important issue concerns one of our experimental design decisions. For both experiments, both the working memory tests and the content tests for the spaced conditions were conducted 1 day after the last learning phase rather than immediately after the last learning phase, as they were for the massed group. We used this procedure because our major reason for running the experiments was to determine whether the fixed working memory capacity assumption of cognitive load theory should be retained. If it was to be jettisoned, we needed evidence that working memory capacity reduced after cognitive effort and increased after a rest. The procedure we used had the capacity to provide that evidence. Of course, because a delay in testing can result in increased forgetting leading to reduced performance by the spaced groups, that procedure potentially compromised evidence for the spacing effect. Nevertheless, the delay in testing for the spaced groups did not prevent a conventional spacing effect in either experiment despite the potential bias against the spacing groups.

As indicated above, the primary purpose in running the current experiments was to provide data to assist in determining whether the assumption of a fixed working memory capacity used by cognitive load theory should be retained or discarded in favor of a working memory depletion assumption following cognitive effort. The current results suggest that a fixed working memory assumption needs to be jettisoned. If so, such a change should have considerable consequences and result in a considerable extension of cognitive load theory. To this point, under the narrow limits of change principle, the theory has implicitly assumed a fixed working memory capacity for any individual. Of course, that capacity can be dramatically increased under the assumptions of the environmental organizing and linking principle. In an environment where information stored in long-term memory can be used, by transferring that information from long-term to working memory, vast increases in working memory capacity ensue. Nevertheless, in the absence of that stored, previously acquired information, it was assumed that for any given individual, working memory capacity was essentially fixed. Based on the current data, that assumption is untenable. Working memory capacity can be

variable depending not just on previous information stored via the information store, the borrowing and reorganizing, and the randomness as genesis principles, but also on working memory resource depletion due to cognitive effort.

Based on working memory resource depletion, there may be considerable scope for extending cognitive load theory and the instructional effects it generates. The narrow limits of change principle assumes that in order to preserve the contents of long-term memory, only very limited amounts of novel information can be processed at any given time. We may need to add to that principle that the limit narrows further following cognitive effort but expands following rest. This expansion should lead to new instructional hypotheses.

The spacing effect provides the first installment of this extension of cognitive load theory. As far as we are aware, the effect has never previously been explained within a cognitive load theory framework and indeed, without the extension made available by the concept of working memory resource depletion, cognitive load theory was not capable of explaining the effect.

There are many issues that require investigation when considering working memory resource depletion, cognitive load theory, and the spacing effect. For example, how much cognitive effort and for how long must cognitive effort be exercised before significant depletion occurs? What is the speed with which the effects of resource depletion are reversed by cognitive rest, and are there conditions under which the reversal can be accelerated? Despite the large number of empirical studies on the spacing effect, there seems to be no consensus on these matters.

Another issue concerns the nature of working memory resource depletion. Is complete rest required or is rest only required from the type of cognitive activity that caused resource depletion, with changes in cognitive activity having the same effect as rest? There is some possible evidence (Rohrer and Taylor 2007) that presenting students with a variety of mathematics problems rather than grouping problems in the same category is beneficial. While Rohrer and Taylor (2007) relate their procedure and results to the spacing effect with grouping together similar problems categorized as massed practice and mixing, or to use their term, "shuffling" problems, categorized as spaced presentations, the experimental design and results probably more closely resemble a variability effect paradigm (Paas and van Merriënboer 1994).

In conclusion, characterizing resource depletion as working memory depletion, adding working memory depletion to cognitive load theory and using cognitive load theory to explain the spacing effect may have both theoretical and educational implications. While considerably more research is required, these initial empirical results are encouraging.

References

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology, 61*, 228–247. https://doi.org/10.1016/j.cogpsych.2010.05.004.

Conlin, J. A., Gathercole, S. E., & Adams, J. W. (2005). Children's working memory: investigating performance limitations in complex span tasks. *Journal of Experimental Child Psychology, 90*, 303–317. https://doi.org/10.1016/j.jecp.2004.12.001.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: a methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769–786. https://doi.org/10.3758/BF03196772.

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: a deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *The psychology of learning and motivation: advances in research and theory* (Vol. 53, pp. 63–147). New York: Academic. https://doi.org/10.1016/S0079-7421(10)53003-2.

Ebbinghaus, H. (1885/1964). *Memory: a contribution to experimental psychology*. Oxford: Dover.

Geary, D. C. (2008). An evolutionarily informed education science. *Educational Psychologist, 43*, 179–195. https://doi.org/10.1080/00461520802392133.

Geary, D. (2012). Evolutionary educational psychology. In K. Harris, S. Graham, & T. Urdan (Eds.), *APA Educational Psychology Handbook* (Vol. 1, pp. 597–621). Washington, D.C.: American Psychological Association.

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.

Gluckman, M., Vlach, H. A., & Sandhofer, C. M. (2014). Spacing simultaneously promotes multiple forms of learning in children's science curriculum. *Applied Cognitive Psychology, 28*, 266–273. https://doi.org/10.1002/acp.2997.

Healey, M. K., Hasher, L., & Danilova, E. (2011). The stability of working memory: do previous tasks influence complex span? *Journal of Experimental Psychology: General, 140*, 573–585. https://doi.org/10.1037/a0024587.

Kapler, I. V., Weston, T., & Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: long-term benefits for factual and higher-level learning. *Learning and Instruction, 36*, 38–45. https://doi.org/10.1016/j.learninstruc.2014.11.001.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86. https://doi.org/10.1207/s15326985ep4102_1.

Küpper-Tetzel, C. E. (2014). Understanding the distributed practice effect: strong effects on weak theoretical grounds. *Zeitschrift für Psychologie, 222*, 71–81. https://doi.org/10.1027/2151-2604/a000168.

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

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review, 24*, 27–45. https://doi.org/10.1007/s10648-011-9179-2.

Paas, F., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of Educational Psychology, 86*, 122–133. https://doi.org/10.1037/0022-0663.86.1.122.

Paas, F., Tuovinen, J. E., Tabbers, H. K., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63–71. https://doi.org/10.1207/S15326985EP3801_8.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science, 35*, 481–498. https://doi.org/10.1007/s11251-007-9015-8.

Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology: General, 136*, 241–255. https://doi.org/10.1037/0096-3445.136.2.241.

Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Intellectual performance and ego depletion: role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology, 85*, 33–46. https://doi.org/10.1037/0022-3514.85.1.33.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*, 123–138. https://doi.org/10.1007/s10648-010-9128-5.

Sweller, J. (2015). In academe, what is learned, and how is it learned? *Current Directions in Psychological Science, 24*, 190–194. https://doi.org/10.1177/0963721415569570.

Sweller, J. (2016a). Cognitive load theory, evolutionary educational psychology, and instructional design. In D. Geary & D. Berch (Eds.), *Evolutionary perspectives on child development and education* (pp. 291–306). Switzerland: Springer. https://doi.org/10.1007/978-3-319-29986-0.

Sweller, J. (2016b). Working memory, long-term memory, and instructional design. *Journal of Applied Research in Memory and Cognition, 5*, 360–367. https://doi.org/10.1016/j.jarmac.2015.12.002.

Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology, 4*, 434–458.

Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: a reply to commentaries. *Educational Psychologist, 42*, 115–121. https://doi.org/10.1080/00461520701263426.

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

Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review, 26*, 265–283. https://doi.org/10.1007/s10648-013-9243-1.

Tyler, J. M., & Burns, K. C. (2008). After depletion: the replenishment of the self's regulatory resources. *Self and Identity, 7*, 305–321. https://doi.org/10.1080/15298860701799997.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498–505. https://doi.org/10.3758/bf03192720.