



# Kent Academic Repository

Pyke, Wesley, Lunau, Johan and Javadi, Amir-Homayoun (2024) *Does difficulty moderate learning? A comparative analysis of the desirable Difficulties Framework and Cognitive Load theory*. *Quarterly Journal of Experimental Psychology*, 78 (10). pp. 2181-2195. ISSN 1747-0218.

## Downloaded from

<https://kar.kent.ac.uk/107794/> The University of Kent's Academic Repository KAR

## The version of record is available from

<https://doi.org/10.1177/17470218241308143>

## This document version

Publisher pdf

## DOI for this version

## Licence for this version

CC BY-NC (Attribution-NonCommercial)

## Additional information

## Versions of research works

### Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

### Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

## Enquiries

If you have questions about this document contact [ResearchSupport@kent.ac.uk](mailto:ResearchSupport@kent.ac.uk). Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

# Does difficulty moderate learning? A comparative analysis of the desirable difficulties framework and cognitive load theory

Quarterly Journal of Experimental Psychology  
1–15

© Experimental Psychology Society 2024




Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/17470218241308143

qjep.sagepub.com



Wesley Pyke<sup>1</sup> , Johan Lunau<sup>1</sup> and Amir-Homayoun Javadi<sup>1,2</sup>

## Abstract

There is evidence to suggest that variations in difficulty during learning can moderate long-term retention. However, the direction of this effect is under contention throughout the literature. According to both the Desirable Difficulties Framework (DDF) and the Retrieval Effort Hypothesis (REH), *increasing* difficulty (thus relative effort) during retrieval-based learning can help achieve superior long-term retention. One reason for this is due to improved schema formation following a deeper encoding strategy, allowing for more efficient retrieval techniques. A conflicting theory discussed in this review is the Cognitive Load Theory (CLT). The CLT states that conditions for learning are best when extraneous load is reduced, and intrinsic load is optimised. By doing this, germane resources can focus on schema formation. While both theories consider schema formation key to successful retention, the way in which it is best achieved is conflicting. To date, both theories have yet to be compared despite their commonalities. This review evaluates the aforementioned theories, before proposing a new model of difficulty in learning. The proposed model integrates principles from the DDF, REH, and CLT, incorporating insights from Perceptual Load Theory (PLT). It suggests that task difficulty should be adjusted based on the material's complexity and the learner's expertise. Increasing difficulty benefits low-element-interactivity tasks by enhancing focus and retention, while reducing difficulty in high-element-interactivity tasks prevents cognitive overload.

## Keywords

Desirable difficulties framework; cognitive load theory; retrieval effort hypothesis; perceptual load theory; retrieval-based learning

Received: 27 March 2024; revised: 6 November 2024; accepted: 8 November 2024

## Highlights

- Desirable difficulties framework and cognitive load theory oppose one another, prompting comparisons between the two.
- Both agree memory is modulated by successful schema formation which requires a degree of effort.
- A new model is put forward that emphasises the need to tailor instructional design to individual learners and circumstances.

## Introduction

When learning something new, there are various strategies we can employ to enhance memory retention, such as association with familiar objects or repetitive recitation. In the educational context, it is essential to identify

the most effective learning methods for increasing long-term retention and optimising educational programmes. This review aims to investigate the impact of difficulty during learning on long-term retention, specifically contrasting two opposing frameworks: the Desirable Difficulties Framework (DDF) and the Cognitive Load Theory (CLT). The DDF suggests that introducing certain difficulties during learning can enhance long-term

<sup>1</sup>School of Psychology, University of Kent, Canterbury, UK

<sup>2</sup>School of Rehabilitation, Tehran University of Medical Sciences, Tehran, Iran

### Corresponding author:

Amir-Homayoun Javadi, School of Psychology, Keynes College, University of Kent, Canterbury CT2 7NP, UK.

Email: a.h.javadi@kent.ac.uk

retention (Bjork, 1994; Bjork & Bjork, 2020). This idea is also supported by the Retrieval Effort Hypothesis (REH), which was developed from the DDF with a focus on retrieval learning. The CLT, on the contrary, posits that excessive difficulty (cognitive load) inhibits retention (Howard et al., 2015; Örün & Akbulut, 2019; Sweller, 1988). To date, the two conflicting frameworks are yet to be compared, despite sharing commonalities. After reviewing the current literature, this article aims to provide an updated framework on difficulty in learning.

## DDF

The DDF, developed by Bjork (1994), suggests that an effective way to improve long-term retention is to introduce a desirable amount of difficulty (effort) while learning. That is, during encoding, the task used should strike a balance between difficulty and achievability. According to this idea, while the initial performance on a difficult task may be poorer, greater improvements can be seen in the long term, compared to more simple tasks. There are several ways in which difficulty can be induced during learning, one of which being spacing, where study sessions are spread out over time, as opposed to cramming the information in a short period of time. Spacing requires more effort to remember the information over a longer period of time, which helps to strengthen the memory trace. Another method of learning that induces an increased amount of difficulty is retrieval-based learning (RBL), which will be discussed in more depth below.

## RBL

The concept of learning via retrieval has had its place in the literature since the early 20th century (Abbott, 1909; Gates, 1917; Spitzer, 1939). There has been much research on RBL, the fundamental basis of which relies on repeated testing of the participant on the subject material (Roediger & Butler, 2011). Typically, the participant is first shown the target stimuli (known as a study phase) before they move on to the next stage, the testing phase. Following this, there may be various patterns of study and test phases before a final test. In an experimental situation, this final test often measures the efficacy of RBL compared to say, study alone. An example of this pattern in an experimental group may be: Study, Test, Study, Test (STST); compared to a control Study alone group (SSSS). Learning via retrieval has consistently been shown to be more effective than studying the material alone (Carpenter et al., 2008; de Lima et al., 2020; Fazio & Marsh, 2019; Karpicke & Aue, 2015; Karpicke & Grimaldi, 2012; Karpicke & Roediger, 2008; Kornell et al., 2011; Roediger & Karpicke, 2006). Furthermore, as well as healthy populations, RBL has also been shown to be effective in language-impaired populations, compared to repetitive study (de Lima et al., 2021)

and when spacing is implemented (increasing difficulty), improved results are shown compared to massed study (Middleton et al., 2016, 2019). In an experimental setting, the difference in performance between the retrieval group and the study-only group is known as the “testing effect.”

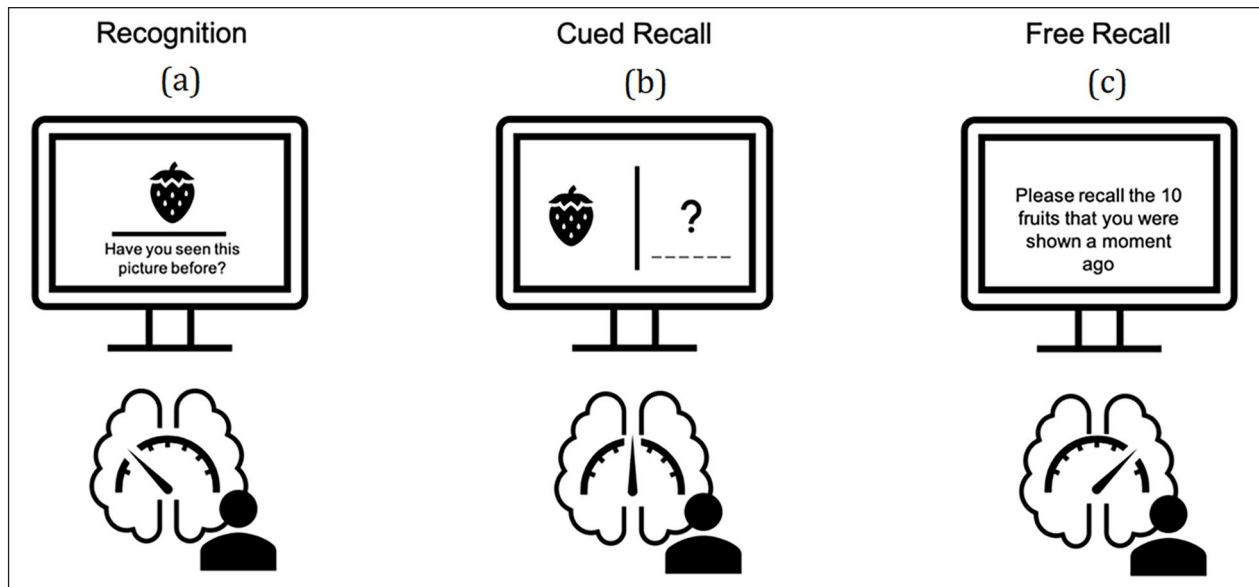
## REH

Consistent with Bjork’s (1994) DDF, the REH states that the more difficult a retrieval is, the more effort it requires from the learner; therefore, increasing the probability that the material will be consolidated in memory and retrieved at a later date. There are several studies that have directly tested the REH, each finding evidence in support of this notion (Carpenter & DeLosh, 2006; Karpicke & Roediger, 2007b; Pyc & Rawson, 2009).

When testing the REH, it is important to understand how effort, a particularly subjective concept, can be manipulated by the experimenter. These variations affect the cognitive processes occurring simultaneously with learning. In retrieval learning literature, there are differences in how both task (the way in which the learning material is presented; (Carpenter & DeLosh, 2006; Kang et al., 2007; Pyc & Rawson, 2009; Stenlund et al., 2016) and item (the learning material itself; de Lima et al., 2020; Minear et al., 2018; Pyke et al., 2023; Vaughn et al., 2013) difficulty can affect final performance on a memory task. It is yet to be clearly established whether the increase in effort elicited by these variables is a direct manipulator or whether effort itself indirectly affects performance on a task. Indeed, one criticism of the REH is that it is a purely descriptive account (Karpicke et al., 2014) and fails to explain reasons why an increase in effort may produce memory benefits. Below, we outline how difficulty can be increased during learning (which increases relative effort) and studies that provide support for the DDF and REH.

## Task difficulty

The majority of studies investigating how difficulty can moderate later memory performance have manipulated the tasks that are used during both initial testing (learning phase) and final testing sessions (to assess the efficacy of the learning task). Most commonly, the variance is in the type of task administered. These typically employ either recognition, cued recall or free recall, during either the learning or final test phase. Both cued and free recall are said to be more difficult (and thus induce more effort) because they require the participant to *produce* a relevant answer (content generation), rather than just stating whether they have seen an item before, as in recognition (yes/no; old/new tasks; see Figure 1 for an example). When cued or free recall is employed during the learning phase, many studies have found improved subsequent memory performance on a final test compared to when



**Figure 1.** Variations in effort induced by different types of learning tasks: (a) recognition task, (b) cued recall, and (c) free recall.

recognition is used (Butler & Roediger, 2007; Kang et al., 2007; Stenlund et al., 2016); for reviews, see (McDermott, 2021; Rowland, 2014). This holds even when there is a mismatch in task type (i.e., recall during learning and recognition during retrieval; (Carpenter & DeLosh, 2006; Endres & Renkl, 2015; Rowland, 2014; Stenlund et al., 2016), highlighting the positive influence of *recall* during the learning phase.

As well as the type of test used during learning, it is also possible to manipulate task difficulty by increasing or decreasing intervals between retrieval trials, known as *interstimulus intervals* (Pyc & Rawson, 2009). By carrying out retrieval practice with progressively expanding spaced intervals, difficulty for the learner is increased, thus inducing more effort, which has been shown to subsequently improve long-term retention (Kang et al., 2014; Middleton et al., 2016). This notion however has received conflicting views, with (Karpicke & Roediger, 2007a) finding repeated expanding intervals to only provide short-term benefits, with *equally* spaced intervals providing superior long-term benefits (Cull, 2000).

Regardless of the spacing, there is evidence to suggest that *repeated* retrieval sessions with corrective feedback improves long-term retention (Butler, 2010; Karpicke & Roediger, 2007b; Tse et al., 2010); for a review, see (Binks, 2018). Corrective feedback is important, as it ensures the learner is correctly recalling the material, thus meeting the requirement of the DDF. Without corrective feedback, incorrect answers could be learnt without the learner realising until the final test (Bifurcation theory; Kornell et al., 2011).

The reasoning behind why the task type can induce greater retention may be explained by the Perceptual Load Theory (PLT; Lavie & Dalton, 2014). The PLT suggests

that performance on a task is greater when the perceptual load related to that task is higher (i.e., more attentional resources are dedicated towards it). This theory seems to align with the notion of the REH, in that the higher the degree of difficulty, the higher the degree of attention focused on it.

Practical examples of PLT related to retrieval learning can be found in research investigating how divided attention interacts with learning. This research typically involves applying a secondary cognitive task concurrently with a primary learning task. Studies that have carried this out with typical memory tasks (study-only) with separate encoding and retrieval phases, found that performance on the primary learning task was only affected by the secondary task during encoding, but not retrieval (N. D. Anderson et al., 1998; Craik et al., 1996). This implies that the encoding phase during a study-only intervention is susceptible to interference from a secondary task, due to lack of attentional focus. However, these studies do not indicate whether encoding via retrieval would also suffer from interference.

Mulligan and Picklesimer (2016) investigated this question by comparing an RBL task (cued recall) with a study-only group under either full attention or divided attention. They found that the testing effect was greater under divided attention than full attention (to reiterate, the "testing effect" is the difference in performance between the RBL group and the study-only group). Therefore, this suggests that the RBL group was more resilient to divided attention than the study-only group. The same was shown in two experiments using free recall (Buchin & Mulligan, 2017, 2019) and this held with both shorter and longer word lists, the latter of which required more effort.

These findings support the use of the PLT to explain why RBL is resistant to divided attention, as it requires considerable perceptual load, and thus perceptual capacity is exhausted. This then reduces the likelihood that irrelevant distractors will interfere with the primary task, potentially increasing the likelihood of successful retention. The PLT may also more broadly explain why an increase in difficulty during RBL tasks, can provide superior long-term retention, compared to easier tasks. As perceptual load is at capacity, full focus is given to the learning task, allowing for an uninterrupted consolidation process.

### Item difficulty

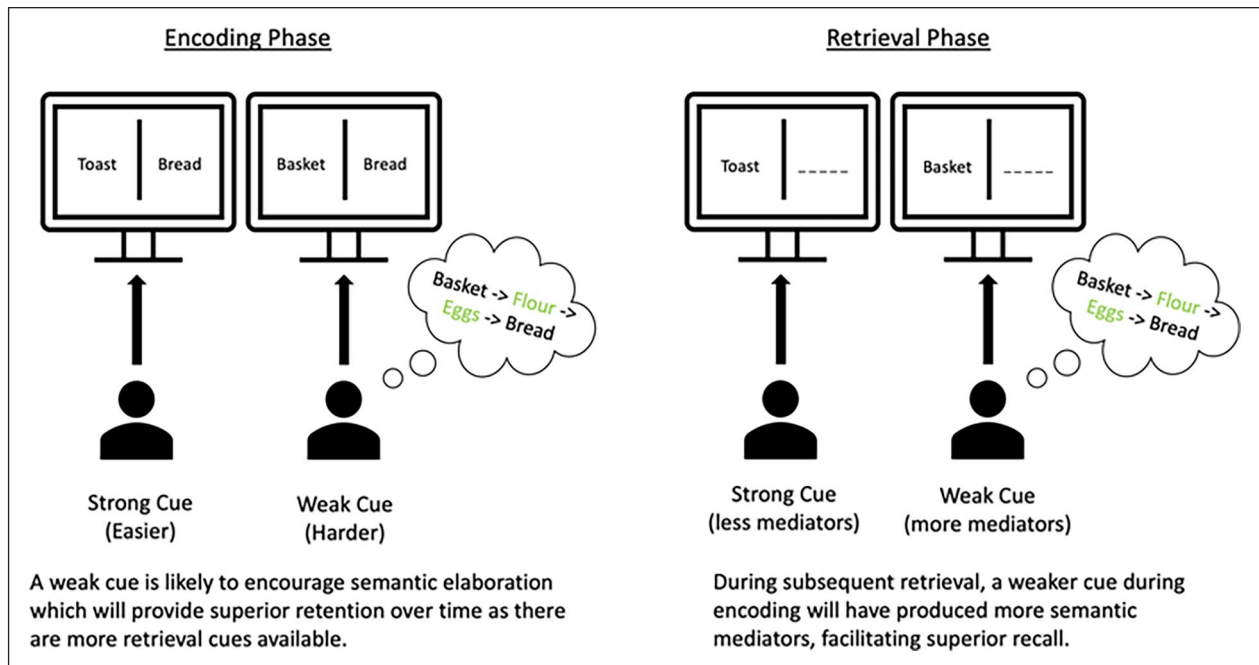
The majority of the learning material used to assess the DDF and REH is simple word pairs or sentences. In any memory task utilising stimuli word pairs, there is likely to be a difference in how each item is encoded and recalled. This notion can be quantified as item difficulty. In the literature, several previous studies have rigorously tested each item. For example, Cho et al. (2020) collected normative data for Chinese-English word pairs, requiring participants to partake in three study-test cycles for 160-word pairs. They found that Chinese characters with a higher number of strokes (visual complexity) were less likely to be recalled in a test phase. This type of normative study, among others (de Lima & Buratto, 2021; Grimaldi et al., 2010; Nelson & Dunlosky, 1994; Pyke et al., 2023), allows future research to utilise individual items differently. This can be useful to ensure an equal spread of difficulty between groups, reducing the likelihood that one condition will contain easier items than the other. This normative data have also been used in studies relating to the REH. Proponents of the DDF/REH state that difficult items are more likely to benefit from retrieval learning than easier items as they require more in-depth processing and attentional resources, similar to the idea put forward for cued and free recall over recognition. This was tested by (de Lima et al., 2020) who, over two experiments compared study versus test conditions and easy versus difficult items, with a follow-up cued recall test 48 hours later. Their findings in Experiment 1 showed a strong retrieval practice effect, with easier items recalled more than difficult items. The authors put this down to the lack of correct retrievals during the learning phase, something that is required under the DDF. In the second experiment, additional learning sessions were employed for difficult items (six compared to four for easy items). The final test results displayed a non-significant trend towards a greater retrieval practice effect for difficult items. Therefore, this suggests that item difficulty and perceived effort may contribute towards final retention, however, this is inconclusive. Future studies that seek to replicate the study above could consider employing further learning sessions for difficult items, perhaps employing a performance-based ending threshold. While there has been less research on the REH's

explanation of how item difficulty can affect long-term retention, the individual variability of participants, such as educational background or language, as well as previous semantic representations developed throughout the lifespan, will undoubtedly influence how difficult an item is perceived. Therefore, this highlights that the REH perspective on item difficulty, without the consideration of the potential covariates highlighted above, requires further exploration before a conclusion can be made.

Relating to item difficulty, the Elaborative Retrieval Hypothesis, states that during RBL, the individual activates elaborative information to help with retrieval of the target response. Using paired-word associates, Carpenter (2009) tested this theory by presenting participants with a cued recall test containing either strongly (e.g., Toast: Bread) or weakly (Basket: Bread) associated cues, or a restudy opportunity. While strongly associated pairs were easier to learn, scoring higher on an initial test, weakly associated pairs were better remembered on a final free-recall test due to activation of elaborative information during learning (see Figure 2). This notion is in line with the REH as it suggests that as more effort is required for the weakly associated pairs, these items benefit more from the retrieval practice effect.

Support for the Elaborative Retrieval Hypothesis has also been shown in a previous study by Carpenter & DeLosh (2006), even when controlling for item difficulty (Experiment 3) and in a more recent study (Endres & Renkl, 2015), where interestingly the authors stated that the testing effect disappeared when statistically controlling for mental effort. Consistent with the Elaborative Retrieval Hypothesis, they claimed that increased mental effort, as measured subjectively on a sliding scale, leads to spreading activation, which in turn is an indicator of semantic elaboration. Spreading activation can be defined as the implicit creation of new cues in memory, which can then be utilised during retrieval (J. R. Anderson, 1983). This semantic elaboration can aid future retrievals and can even spread to material that has not been initially tested, a phenomenon known as retrieval-induced facilitation (Chan, 2009; Chan et al., 2006; Oliva & Storm, 2023; Rowland & DeLosh, 2014). A difficult task will lead to more mental effort and thus a greater amount of spreading activation occurs, which in turn will strengthen the memory trace, allowing for multiple cue points (Carpenter, 2009). On the contrary, a comprehensive study (7 experiments) by (Lehman & Karpicke, 2016) tested whether semantic mediators (cues linking to target information) related to the target word (e.g., *Mother-Child-Father*) were exclusively produced during RBL, or whether they are also activated during restudy. They found that generation of mediators was *not* more likely during retrieval learning compared to study-only conditions. Furthermore, they also found that the activation of mediators was unrelated to subsequent free recall of targets. This may suggest that although the Elaborative Retrieval Hypothesis was a





**Figure 2.** A visual example of how harder cues during encoding can facilitate the creation of more semantic mediators to aid in subsequent retrieval.

theory developed predominantly with RBL in mind, these results highlight that the theory may also explain mechanisms behind traditional encoding techniques (i.e., visual presentations followed by a final test).

So far, this review has presented theories that, for the most part, are consistent with the DDF. These theories suggest that by increasing difficulty, either by manipulating the type of task or difficulty of items during the encoding phase of RBL, one can improve subsequent performance on a final test. While task type (recognition, cued or free recall) employed during learning seems to have an undisputed effect on final test performance, differences in item difficulty and the generation of semantic mediators is still open to debate.

In the next section, this review will address a different type of load, the CLT. This theory is often discussed in the context of PLT, with elements of the PLT said to be related to external aspects and CLT to be related to internal mechanisms of attention and learning. The CLT proposes that an increase in mental effort (comparable to the previously discussed difficulty) can be detrimental to learning. We provide a description of the key components of the CLT, and how it claims optimum learning is achieved while also drawing contrasts and comparisons to the DDF and REH.

## CLT

CLT details another explanation for the role of difficulty in learning. More specifically, CLT aims to explain the link between the processing load (i.e., cognitive load) induced

by learning tasks and students' ability to manage novel information to subsequently build knowledge in the form of long-term memory (Sweller, 1988; Sweller et al., 1998). The theory rests on three key assumptions. First, that working memory has a limited capacity and consists of multiple partially independent subsystems. Second, that long-term memory has an unlimited capacity and consists of schemas that categorise information based on how it will be used (Chi et al., 1982). These two assumptions form a third, which is that learning is most effective when instructional procedures *limit the load* imposed on working memory while simultaneously encouraging schema formation.

Beyond these foundational assumptions, while cognitive load can generally be defined as the number of working memory resources employed to perform a task, CLT distinguishes between three different types (Sweller et al., 1998, 2019) each of which are reviewed below.

### Intrinsic load

Intrinsic load refers to the inherent difficulty of the learning material, which is determined in large part by *element interactivity*. When components of the material have *low element* interactivity (e.g., learning the elements of the periodic table), they are unrelated and can be learned both in isolation and sequentially, reducing the intrinsic load. Conversely, a mathematical equation involves interactive and related elements that must be learned and considered together, thus imposing a higher intrinsic load (Sweller,

2010, 2011). The level of intrinsic load induced by the material will depend on the ability of the learner. A novice student would find a complex mathematical equation difficult, and therefore intrinsic load would be higher than for an expert mathematician (Chen et al., 2016b). A higher intrinsic load is considered to be more difficult and therefore requires more effort (Ayres, 2006; Beckmann, 2010; Sweller & Chandler, 1994; Wirzberger et al., 2016). Intrinsic load can be compared to previously described item difficulty; a variable that is commonly manipulated in tests of the REH. However, while both reference the role of the nature of the learning material regarding encoding, recall and memorisation, intrinsic load refers to the significance of the relationships between components of the learning material, whereas item difficulty in DDF/REH points to the complexity of isolated and single items, (e.g., de Lima et al., 2020). This is an important consideration, given that the learning material typically used to test the DDF/REH are paired associates (dog-chien), which are very low in element interactivity.

### Extraneous load

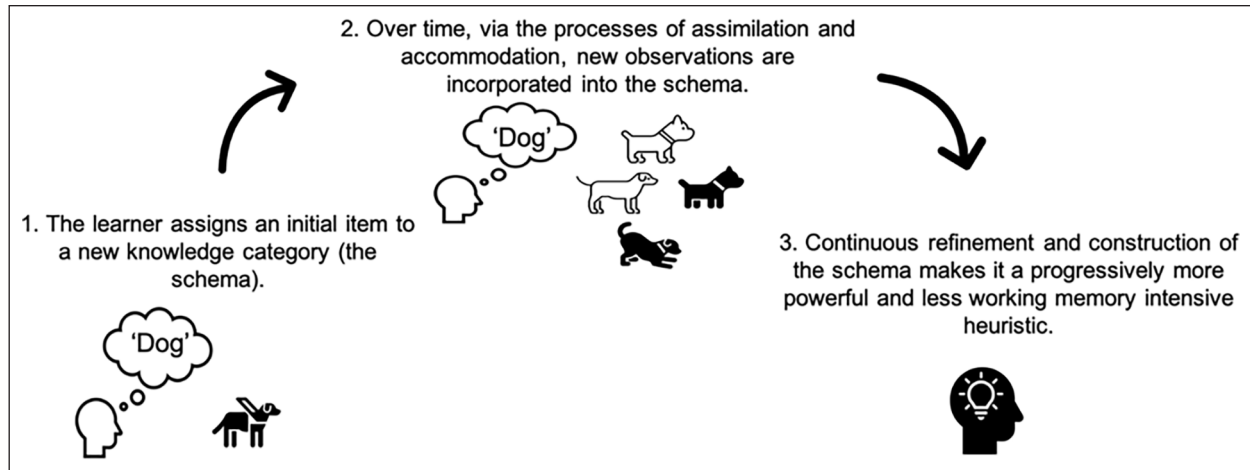
Extraneous load is related to instructional design, which concerns how a task is formatted and presented to the learner. As with intrinsic load, *element interactivity* is also relevant to extraneous load (Sweller et al., 2019), in the sense that a poor instructional design increases *element interactivity*, imposing a greater cognitive load (and therefore having a detrimental effect on memorisation (Sweller, 2010). A quality instructional design would decrease *element interactivity*, reducing cognitive load, and supporting memorisation. Reducing extraneous load can be achieved through manipulating the instructional design to provide a well-organised solution for the learner (Chen et al., 2016b; Sweller, 1988). Empirical research has highlighted that by increasing extraneous load (by increasing task difficulty and time pressure), lower recall performance is achieved (Galy et al., 2012). A contrast exists between extraneous load, as defined under the CLT, and the manipulation of the task difficulty variable in DDF/REH research. Whereas the principle of extraneous load in CLT encourages the creation of tailored instructional designs that *reduce* cognitive load (i.e., place less strain on working memory), proponents of the DDF/REH state that *increasing* task difficulty (e.g., through cued and free recall, as opposed to simple recognition), can provide superior recall (McDermott, 2021; Rowland, 2014).

### Germane load

Finally, *germane load* can be defined as cognitive load that is directly associated with schema construction, a process that is essential for long-term retention (Sweller et al., 1998). Whereas intrinsic and extraneous load are

problematic factors that instructional designers should aim to optimise or limit, germane load is the mental effort that is devoted to promoting learning and should be increased as much as possible (Sweller, 1988). Germane load can be induced by encouraging a learner to engage in conscious cognitive processing which is closely related to schema construction (Sweller, 2010; Sweller et al., 1998). Ideally, this should be done when the learner is engaging in a task that is inducing only minimal intrinsic and extraneous load, leaving some working memory capacity underutilised. Variations in the levels of mental resources devoted to schema construction have been shown to depend on intrinsic and extraneous load in empirical studies (Wirzberger et al., 2017, 2020). An example of encouraging schema construction through increasing germane load would be to present the learner with incomplete examples of how to solve a problem (requiring completion of the example) or ask the learner specific questions about said examples. Schema construction is a fundamental concept within germane load but has also been referenced in REH literature. Indeed, Zaromb & Roediger (2010) employed categorised lists and free recall tasks to investigate whether the established recall-enhancing testing effect also improved the participant's conceptual (i.e., schematic) organisation of the learning material, relative to studying alone. They found that testing improved three aspects of categorised list recall: recall organisation, access to higher-order units, and access to their contents. This led them to conclude that the testing effect, in terms of free recall, resulted from the construction of retrieval schemas that guide subsequent recall (see Figure 3). This joint acknowledgement of the importance of schema construction highlights a similarity between the REH and CLT and strengthens the concept of germane load.

More recently, the concept of germane “load” has been challenged in the literature. Initially, germane load was conceptualised as the mental effort dedicated to schema construction, considered beneficial for learning. However, this approach created ambiguity, as it overlapped with intrinsic load and suggested that increasing germane load would always enhance learning. Both intrinsic and germane loads are involved in processing the essential elements of learning tasks, making it difficult to clearly distinguish between the two. For instance, when learners engage deeply with complex material, the cognitive effort they expend can be seen as contributing to both the intrinsic and germane loads, blurring the lines between them (Greenberg & Zheng, 2023; Kalyuga, 2011). To address these issues, the concept of germane resources was introduced, emphasising the allocation of cognitive resources towards effective learning processes rather than viewing it as an additional load (Sweller, 2023). This shift aligns better with the finite capacity of working memory, focusing on optimising instructional design to maximise the use of cognitive resources for productive learning activities



**Figure 3.** A graphical representation of schema construction. This notion is acknowledged in both DDF/REH and CLT literature as a crucial process for learning.

(Kalyuga, 2011; Sweller et al., 2019). As a result, the emphasis has moved from simply increasing germane load to ensuring that learners' cognitive efforts are directed towards the most meaningful and supportive tasks in the learning process (Sweller et al., 2019).

### CLT in practice

Empirical evidence for CLT originates largely from studies that provide support for various effects set forward by the theory (Sweller et al., 1998), some of which will be discussed here. The *goal-free effect* is one example. Typically, when learners are faced with a novel problem for which they have no schema available, they engage in means-ends analysis (MEA), which requires the learner to identify a *problem state* and a *goal state*, and to reconcile the differences between them using a *problem-solving operator* (Sweller, 1988). While MEA could solve the problem, it is incompatible with learning because it is working memory intensive and does not lead to schema construction (Sweller, 1988; Sweller et al., 1998). However, if a *goal state* is not defined in the problem, the learner is effectively forced to identify only the *problem states* and apply an *operator* (such as an equation) to these. Theoretically, this method reduces working memory load and drives schema construction, leading to improved memorisation. In practice, this assumption is supported by empirical research in a variety of experimental contexts (Ayres, 1993; Bobis et al., 1994; Owen & Sweller, 1985; Vollmeyer et al., 1996).

The *worked example effect* is closely related to the *goal-free effect*, as it is based on the same premise of eliminating MEA and thereby reducing extraneous cognitive load while simultaneously promoting schema construction and memorisation. Worked examples focus the learner's attention on the *problem states* and their respective

*operators*, instead of the ultimate *goal state*. As with the *goal-free effect*, a wealth of studies lend support to the effectiveness of worked examples in terms of improving memorisation (Carroll, 1994; Cooper & Sweller, 1987; Paas & Van Merriënboer, 1994; Sweller & Cooper, 1985; Trafton & Reiser, 1993; Zhu & Simon, 1987). However, some limitations of worked examples have been identified. First, it has been suggested that the excessive use of worked examples could lead to the formation of a rigid problem-solving template that possibly inhibits more creative and novel problem-solving solutions (Smith et al., 1993). Second, the design of high-quality worked examples is challenging, particularly if these require integration of multiple information sources, which could lead to a higher extraneous cognitive load (Sweller et al., 1990, 1998; Tarmizi & Sweller, 1988; Ward & Sweller, 1990).

The *generation effect* contrasts with the *worked example effect* in that instead of providing the learner with a structure, they are encouraged to actively generate the response themselves. For material with low element interactivity, engaging in active generation of the answer alone has been shown to provide superior recall compared to when a structure is provided (Chen et al., 2016b). This notion can be compared to the DDF/REH in that tasks requiring a generated response (cued and free recall) consistently yield higher long-term retention (Butler & Roediger, 2007; Carpenter & DeLosh, 2006; Endres & Renkl, 2015; Kang et al., 2007; Stenlund et al., 2016); for reviews, see McDermott (2021) and Rowland (2014). Indeed, the literature surrounding the REH typically uses tasks that are low in element interactivity, such as paired associates. While the two theories agree that material with low element interactivity is best learnt using active generation, the CLT posits that for more complex learning material, with higher element interactivity, this will be less effective than a worked example (Chen et al., 2016a).



Furthermore, the *generation effect* is reduced in novice learners, as there are no previous schemas to work with (Leahy et al., 2015; van Gog & Sweller, 2015).

The review of the CLT presented above has indicated how the different types of load may be either of detriment (extraneous and intrinsic) or of benefit (germane) to learning. There appears to be considerable support for the practical applications of CLT; however, these are dated and would therefore benefit from more recent evidence.

It is evident that the DDF (including REH) and CLT offer contrasting perspectives on how difficulty affects learning. While the DDF/REH argue that increasing difficulty during learning enhances long-term retention, the CLT suggests that, in many cases, reducing difficulty leads to better learning outcomes and, consequently, improved retention. However, these theories are often applied in different contexts, utilising different types of learning materials. In addition, CLT emphasises the importance of considering learner expertise and adapting instructional design to maximise efficiency. In the next section, a model is proposed that integrates these perspectives, aiming to optimise learning by strategically adjusting difficulty based on the nature of the material and the learner's level of expertise.

## Working towards a new model of difficulty in learning

This review sought to evaluate the role of difficulty in learning by examining the DDF and CLT, with a focus on how task difficulty impacts learning outcomes. Research shows that task difficulty moderates the effects seen in studies on DDF and the REH (see Table 1). Free and cued recall tasks yield the highest testing effect when implemented during encoding, even when there is a mismatch in initial and final test type (e.g., free recall followed by recognition; (Endres & Renkl, 2015)). This is likely due to the development of stronger memory traces and retrieval schemas during encoding, which is then able to facilitate easier recall when required (Zaromb & Roediger, 2010). In CLT, the *generation effect* shares similarities with the DDF's *testing effect*, both emphasising the value of actively generating responses to enhance retention. However, CLT suggests that this benefit is contingent upon the material being low in element interactivity (Chen et al., 2015, 2023; Sweller, 2023). Studies related to REH typically involve simple, low-element-interactivity material, such as paired associates (Pyc & Rawson, 2009; Vaughn et al., 2013) and word lists (Karpicke & Roediger, 2007b; Lehman et al., 2014). While generating responses is beneficial for such material (Chen et al., 2016b), research indicates that the generation effect diminishes or even reverses when dealing with high-element-interactivity tasks, such as complex problem-solving (Leahy et al., 2015; van Gog et al., 2015), for a review, see van Gog and Sweller (2015). For these

more complex tasks, providing a worked example can be more beneficial, as it frees up working memory resources, allowing for more effective schema formation (Paas & van Merriënboer, 2020; Sweller & Sweller, 2006). This implies that the type of task used (or instructional design) must rely on the type of learning material and the level of element interactivity (see Table 1 for example studies).

The DDF posits that difficulty must strike a balance between challenge and achievability for it to be desirable (Bjork & Bjork, 2020). If a task is too difficult due to a learner's lack of ability, the difficulty becomes undesirable. In such cases, corrective feedback can reduce difficulty and prevent disengagement (Binks, 2018; Kornell et al., 2011). Similarly, CLT suggests that instructional design should consider the learner's prior knowledge, with more guidance needed for novices than for experts (Chen et al., 2023). Both theories, therefore, agree that individual differences should inform the design of learning tasks.

Based on these insights, a new model of difficulty in learning is proposed that integrates DDF (including REH) and CLT (see Figure 4).

The proposed model of difficulty in learning seeks to integrate and extend the principles of the DDF (including REH) and CLT, by incorporating insights from PLT. This model emphasises that the effectiveness of learning tasks depends not just on the inherent difficulty of the material but also on the strategic modulation of difficulty to optimise cognitive resources and attention.

When dealing with learning material low in element interactivity, such as reading text or memorising word pairs, the perceptual load is naturally low, making the learner more susceptible to distractions like mind wandering. In such cases, increasing task difficulty through methods like testing (e.g., cued or free recall) can help to allocate more attentional resources to the task, thereby reducing interference and enhancing learning outcomes. This approach not only heightens attentional focus but also strengthens memory traces and promotes the formation of schemas, leading to better retention compared to passive study methods. This principle aligns with the Elaborative Retrieval Hypothesis, which suggests that more effortful retrieval of weakly associated pairs improves long-term memory (Carpenter, 2009; Carpenter & DeLosh, 2006; Endres & Renkl, 2015; Rawson et al., 2015).

Conversely, when the learning material is high in element interactivity, such as in complex problem-solving or mathematical tasks, the cognitive and perceptual loads are significantly higher. In these scenarios, increasing difficulty could overwhelm the learner's working memory, leading to ineffective learning or even cognitive overload. To avoid this, it would be most effective to decrease difficulty by providing worked examples and scaffolding, which help to reduce cognitive load and allow working memory to be used more efficiently (Chen et al., 2016a). This approach facilitates schema formation through the

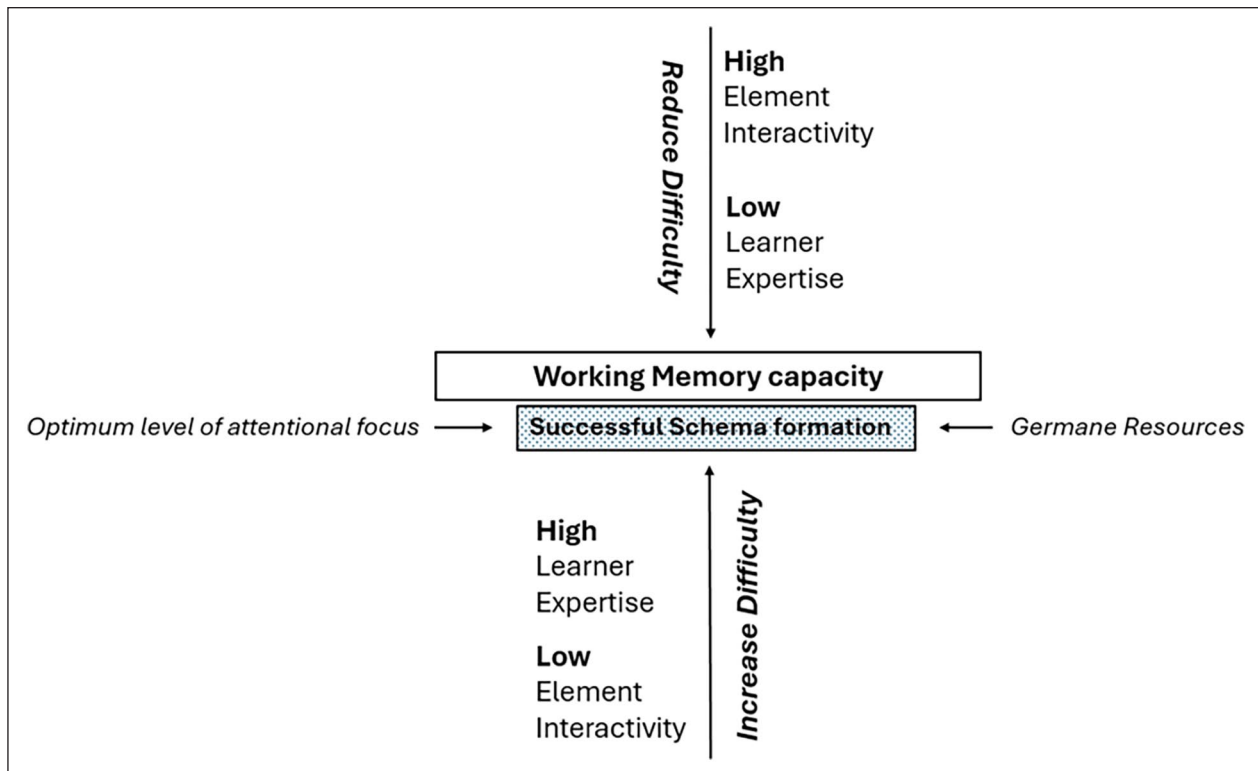
**Table 1.** Studies supporting increasing difficulty during learning.

Study	Task Used	Theory Supported	Findings
Ayres (1993)	General problem-solving tasks	Cognitive load theory	Removing specific goals during problem-solving allows learners to focus on understanding principles rather than being overwhelmed by goal-directed processing.
Bobis et al. (1994)	Geometric problem-solving tasks	Cognitive load theory	Found that geometric problem-solving tasks require significant cognitive resources, which can impede learning if not managed properly.
Butler (2010)	Prose passages	Testing effect	Highlighted the importance of repeated testing in promoting not just retention, but also transfer of learning to new contexts.
Carroll (1994)	Algebra equations	Cognitive load theory	Demonstrated that worked examples reduce cognitive load and enhance problem-solving abilities in algebra.
Carpenter (2009)	Word pairs	Testing effect, elaborative retrieval	Showed that stronger retrieval cues lead to better memory performance initially, with weaker cues benefitting long-term memory performance.
Carpenter & DeLosh (2006)	Word pairs	Testing effect, elaborative retrieval	Argues that less support during retrieval forces more elaborative processing, improving long-term retention.
Chan (2009)	Prose passages	Testing effect, retrieval-induced forgetting	Explored dual effects of retrieval on memory, showing that retrieval can lead to both forgetting and facilitation.
Chen et al. (2016b)	Geometric problem-solving tasks	Cognitive load theory	Highlights the importance of instructional guidance for novice learners, especially in complex learning environments.
Cooper & Sweller (1987)	Algebra equations	Cognitive load theory	Demonstrates how acquiring schemas and automating rules enhance problem-solving abilities, particularly in mathematics.
Endres & Renkl (2015)	Prose passages	Testing effect	Provides empirical evidence supporting retrieval practice as a key strategy for deep, meaningful learning.
Galy et al. (2012)	Basic memory task	Cognitive load theory	Explores how intrinsic, extraneous, and germane cognitive loads interact with mental workload, providing insights for task design.
Kang et al. (2014)	Word pairs	Spacing effect	Concludes that equally spaced retrieval practices are more beneficial for long-term retention than expanding intervals.
Karpicke & Roediger (2007a)	Word pairs	Spacing effect	Supports equally spaced retrieval for better long-term retention compared to expanding intervals.
Karpicke & Roediger (2008)	Word pairs	Retrieval-based learning	Underlines the importance of active retrieval in learning, impacting educational strategies.
Leahy et al. (2015)	Mathematical problem-solving tasks	Cognitive load theory	Suggests that high element interactivity can overwhelm cognitive resources, potentially failing to observe the testing effect.
Mulligan & Picklesimer (2016)	Word pairs	Testing effect	Highlights that the testing effect is robust under divided attention but better with full attention.
Paas & Van Merriënboer (1994)	Geometric problem-solving tasks	Cognitive load theory	Demonstrates that varying worked examples can enhance skill transfer by managing cognitive load.
Pyc & Rawson (2009)	Word pairs	Retrieval effort hypothesis	Finds that more difficult retrievals lead to stronger memory retention.
Rawson et al. (2015)	Word pairs	Elaborative retrieval hypothesis	Longer lags between learning and testing can enhance memory through elaborative retrieval processes.
Roediger & Karpicke (2006)	Prose passages	Testing effect	Highlights the effectiveness of testing over passive review in enhancing long-term learning.

(continued)

**Table 1. (continued)**

Study	Task Used	Theory Supported	Findings
Sweller & Cooper (1985)	Algebra equations	Cognitive load theory	Demonstrates that worked examples are more effective than problem-solving alone in reducing cognitive load and improving learning.
Sweller et al. (1990)	Geometric problem-solving tasks	Cognitive load theory	Highlights the importance of reducing cognitive load in technical subjects to enhance learning outcomes.
Tarmizi & Sweller (1988)	Geometric problem-solving tasks	Cognitive load theory	Finds that guidance during problem-solving reduces cognitive load and enhances learning.
van Gog & Sweller (2015)	General problem-solving tasks	Cognitive load theory	Shows that the testing effect diminishes with increasing complexity of learning materials.
Ward & Sweller (1990)	Geometric Physics problem-solving tasks	Cognitive load theory	Demonstrates that well-structured worked examples reduce cognitive load and improve learning outcomes.
Wirzberger et al. (2017)	Working memory updating task (basal letter-learning task)	Cognitive load theory	Finds that interruptions and task complexity affect the progression of cognitive load, impacting learning efficiency.
Zaromb & Roediger (2010)	Word pairs	Testing effect	Suggests that the testing effect enhances organisational processes, improving recall.
Zhu & Simon (1987)	Algebra equations	Cognitive load theory	Demonstrates that learning from worked examples is more effective for novices due to reduced cognitive load.



**Figure 4.** A model to incorporate the DDF and CLT. The top section of the diagram follows the CLT theory in that it is necessary to reduce difficulty so as not to exhaust working memory (WM) capacity. The bottom section of the diagram suggests a new model that utilises the PLT to explain why easier learning material with lower element interactivity will benefit from an increase in difficulty. Both theories agree that the overall aim is to encourage successful schema formation.

borrowing and reorganising principle, enabling learners to develop effective strategies without the strain of excessive cognitive load (Paas & van Merriënboer, 2020; Sweller & Sweller, 2006).

It is also important to take into account the expertise of the learner. For novice learners, increasing the difficulty of low-element-interactivity material can be beneficial, as long as the task remains achievable and is supported by corrective feedback to prevent frustration or disengagement. However, for expert learners with low-element interactivity, a ceiling effect may occur, where additional difficulty does not yield further benefits. For novices dealing with high-element-interactivity material, reducing difficulty through guided instruction and worked examples is essential to avoid overwhelming their cognitive capacities and facilitate effective learning. Expert learners facing high-element-interactivity material are more likely to benefit from tasks that require active problem-solving and critical thinking, as they have already developed schemas that allow them to handle complex information more efficiently. In such cases, increasing the difficulty would require more attentional resources (as per the PLT) and promote the refinement of their knowledge.

Building upon the proposed framework, future research should aim to empirically validate the model through controlled experimental studies that manipulate task difficulty, element interactivity, and learner expertise. For instance, experiments could investigate how increasing task difficulty in low-element-interactivity tasks affects retention in both novice and expert learners. Conversely, studies could examine how reducing difficulty in high-element-interactivity tasks impacts cognitive load and schema formation. Longitudinal research assessing retention over extended periods would provide valuable insights into the durability of learning outcomes associated with these manipulations.

Another promising avenue is the development of adaptive learning systems that adjust task difficulty in real time based on the learner's performance and the complexity of the material. Such systems could incorporate corrective feedback mechanisms to maintain optimal challenge levels, ensuring that cognitive and attentional resources are allocated effectively. Applying the model across various educational domains—such as mathematics, language learning, and science—would test its generalisability and practical utility. Integrating the framework into educational technology and e-learning platforms could facilitate personalised learning experiences, enhancing engagement and efficiency.

Finally, exploring individual differences and learner characteristics presents an opportunity to refine the model further. Research could examine how factors like working memory capacity, prior knowledge, motivation, and cognitive abilities interact with task difficulty and element interactivity. This approach would allow for the tailoring of instructional designs to meet the needs of diverse learner

populations, including those with special educational requirements. By addressing these areas, future research can strengthen the theoretical foundations of the framework and contribute to the development of effective educational practices.

In conclusion, this new model does not reject existing theories but rather synthesises them into a more comprehensive framework. It reiterates the importance of considering both the nature of the learning material and the learner's prior knowledge when designing instructional tasks. By individually adapting task difficulty, learning can be optimised by ensuring that cognitive and attentional resources are allocated most effectively, which in turn will promote superior long-term retention.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

Wesley Pyke  <https://orcid.org/0000-0002-0975-1308>

### References

- Abbott, E. E. (1909). On the analysis of the factor of recall in the learning process. *The Psychological Review: Monograph Supplements*, 11(1), 159–177. <https://doi.org/10.1037/h0093018>
- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22(3), 261–295. [https://doi.org/10.1016/S0022-5371\(83\)90201-3](https://doi.org/10.1016/S0022-5371(83)90201-3)
- Anderson, N. D., Craik, F. I. M., & Naveh-Benjamin, M. (1998). The attentional demands of encoding and retrieval in younger and older adults: I. Evidence from divided attention costs. *Psychology and Aging*, 13(3), 405–423. <https://doi.org/10.1037//0882-7974.13.3.405>
- Ayres, P. L. (1993). Why goal-free problems can facilitate learning. *Contemporary Educational Psychology*, 18(3), 376–381. <https://doi.org/10.1006/ceps.1993.1027>
- Ayres, P. L. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, 16(5), 389–400. <https://doi.org/10.1016/j.learninstruc.2006.09.001>
- Beckmann, J. F. (2010). Taming a beast of burden: On some issues with the conceptualisation and operationalisation of cognitive load. *Learning and Instruction*, 20(3), 250–264. <https://doi.org/10.1016/j.learninstruc.2009.02.024>
- Binks, S. (2018). Testing enhances learning: A review of the literature. *Journal of Professional Nursing*, 34(3), 205–210. <https://doi.org/10.1016/j.profnurs.2017.08.008>
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A.



- Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.
- Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, 9(4), 475–479. <https://doi.org/10.1016/j.jar-mac.2020.09.003>
- Bobis, J., Sweller, J., & Cooper, M. (1994). Demands imposed on primary-school students by geometric models. *Contemporary Educational Psychology*, 19(1), 108–117. <https://doi.org/10.1006/ceps.1994.1010>
- Buchin, Z. L., & Mulligan, N. W. (2017). The testing effect under divided attention. *Journal of Experimental Psychology: Learning Memory and Cognition*, 43(12), 1934–1947. <https://doi.org/10.1037/xlm0000427>
- Buchin, Z. L., & Mulligan, N. W. (2019). Divided attention and the encoding effects of retrieval. *Quarterly Journal of Experimental Psychology*, 72(10), 2474–2494. <https://doi.org/10.1177/1747021819847141>
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning Memory and Cognition*, 36(5), 1118–1133. <https://doi.org/10.1037/A0019902>
- Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19(4–5), 514–527. <https://doi.org/10.1080/09541440701326097>
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1563–1569. <https://doi.org/10.1037/a0017021>
- Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory and Cognition*, 34(2), 268–276. <https://doi.org/10.3758/BF03193405>
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory and Cognition*, 36(2), 438–448. <https://doi.org/10.3758/MC.36.2.438>
- Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology*, 86(3), 360–367. <https://doi.org/10.1037/0022-0663.86.3.360>
- Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, 61(2), 153–170. <https://doi.org/10.1016/J.JML.2009.04.004>
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. In *Journal of Experimental Psychology: General*, 135(4), 553–571. <https://doi.org/10.1037/0096-3445.135.4.553>
- 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. <https://doi.org/10.1037/edu0000018>
- Chen, O., Kalyuga, S., & Sweller, J. (2016a). Relations between the worked example and generation effects on immediate and delayed tests. *Learning and Instruction*, 45, 20–30. <https://doi.org/10.1016/J.LEARNINSTRUC.2016.06.007>
- Chen, O., Kalyuga, S., & Sweller, J. (2016b). When instructional guidance is needed. *Educational and Developmental Psychologist*, 33(2), 149–162. <https://doi.org/10.1017/edp.2016.16>
- Chen, O., Paas, F., & Sweller, J. (2023). A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educational Psychology Review*, 35(2), 63. <https://doi.org/10.1007/s10648-023-09782-w>
- Chi, M., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7–75). Erlbaum.
- Cho, K. W., Tse, C. S., & Chan, Y. L. (2020). Normative data for Chinese-English paired associates. *Behavior Research Methods*, 52(1), 440–445. <https://doi.org/10.3758/s13428-019-01240-2>
- Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, 79(4), 347–362. <https://doi.org/10.1037/0022-0663.79.4.347>
- Craik, F. I. M., Naveh-Benjamin, M., Govoni, R., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, 125(2), 159–180. <https://doi.org/10.1037/0096-3445.125.2.159>
- Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, 14(3), 215–235. [https://doi.org/10.1002/\(SICI\)1099-0720\(200005/06\)14:3<215::AID-ACP640>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1)
- de Lima, M. F. R., & Buratto, L. G. (2021). Norms for familiarity, concreteness, valence, arousal, wordlikeness, and recall accuracy for Swahili–Portuguese word pairs. *SAGE Open*, 11(1), 215824402098852. <https://doi.org/10.1177/2158244020988524>
- de Lima, M. F. R., Cavendish, B. A., De Deus, J. S., & Buratto, L. G. (2021). Retrieval practice in memory–And language-impaired populations: A systematic review. *Archives of Clinical Neuropsychology*, 35(7), 1078–1093. <https://doi.org/10.1093/ARCLIN/ACAA035>
- de Lima, M. F. R., Venancio, S., Feminella, J., & Buratto, L. G. (2020). Does item difficulty affect the magnitude of the retrieval practice effect? An evaluation of the retrieval effort hypothesis. *Spanish Journal of Psychology*, 23, 1–22. <https://doi.org/10.1017/SJP.2020.33>
- Endres, T., & Renkl, A. (2015). Mechanisms behind the testing effect: An empirical investigation of retrieval practice in meaningful learning. *Frontiers in Psychology*, 6, Article 1054. <https://doi.org/10.3389/fpsyg.2015.01054>
- Fazio, L. K., & Marsh, E. J. (2019). Retrieval-based learning in children. *Current Directions in Psychological Science*, 28(2), 111–116. <https://doi.org/10.1177/0963721418806673>
- Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology*, 83, 269–275. <https://doi.org/10.1016/j.ijpsycho.2011.09.023>
- Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, 6(40), 104.
- Greenberg, K., & Zheng, R. (2023). Revisiting the debate on germane cognitive load versus germane resources. *Journal of Cognitive Psychology*, 35(3), 295–314. <https://doi.org/10.1080/20445911.2022.2159416>



- Grimaldi, P. J., Pyc, M. A., & Rawson, K. A. (2010). Normative multitrial recall performance, metacognitive judgments, and retrieval latencies for Lithuanian-English paired associates. *Behavior Research Methods*, 42(3), 634–642. <https://doi.org/10.3758/BRM.42.3.634>
- Howard, S. J., Burianová, H., Ehrich, J., Kervin, L., Calleia, A., Barkus, E., Carmody, J., & Humphry, S. (2015). Behavioral and fMRI evidence of the differing cognitive load of domain-specific assessments. *Neuroscience*, 297, 38–46. <https://doi.org/10.1016/j.neuroscience.2015.03.047>
- Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, 23(1), 1–19. <https://doi.org/10.1007/S10648-010-9150-7/METRICS>
- Kang, S. H. K., Lindsey, R. V., Mozer, M. C., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin and Review*, 21(6), 1544–1550. <https://doi.org/10.3758/s13423-014-0636-z>
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19(4–5), 528–558. <https://doi.org/10.1080/09541440601056620>
- Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review*, 27(2), 317–326. <https://doi.org/10.1007/s10648-015-9309-3>
- Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review*, 24(3), 401–418. <https://doi.org/10.1007/s10648-012-9202-2>
- Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning. An episodic context account. *Psychology of Learning and Motivation: Advances in Research and Theory*, 61, 237–284. <https://doi.org/10.1016/B978-0-12-800283-4.00007-1>
- Karpicke, J. D., & Roediger, H. L. (2007a). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning Memory and Cognition*, 33(4), 704–719. <https://doi.org/10.1037/0278-7393.33.4.704>
- Karpicke, J. D., & Roediger, H. L. (2007b). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57(2), 151–162. <https://doi.org/10.1016/j.jml.2006.09.004>
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. <https://doi.org/10.1126/science.1152408>
- Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, 65(2), 85–97. <https://doi.org/10.1016/j.jml.2011.04.002>
- Lavie, N., & Dalton, P. (2014). Load theory of attention and cognitive control. In A. C. Nobre & S. Kastner (Eds.), *The Oxford handbook of attention* (pp. 56–75). Oxford University Press. <https://doi.org/10.1093/OXFORDHB/9780199675111.013.003>
- Leahy, W., Hanham, J., & Sweller, J. (2015). High element interactivity information during problem solving may lead to failure to obtain the testing effect. *Educational Psychology Review*, 27(2), 291–304. <https://doi.org/10.1007/S10648-015-9296-4>
- Lehman, M., & Karpicke, J. D. (2016). Elaborative retrieval: Do semantic mediators improve memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(10), 1573–1591. <https://doi.org/10.1037/xlm0000267>
- Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning Memory and Cognition*, 40(6), 1787–1794. <https://doi.org/10.1037/XLM0000012>
- McDermott, K. B. (2021). Practicing Retrieval Facilitates Learning. *Annual Review of Psychology*, 72(1), 1–25. <https://doi.org/10.1146/annurev-psych-010419-051019>
- Middleton, E. L., Rawson, K. A., & Verkuilen, J. (2019). Retrieval practice and spacing effects in multi-session treatment of naming impairment in aphasia. *Cortex*, 119, 386–400. <https://doi.org/10.1016/j.cortex.2019.07.003>
- Middleton, E. L., Schwartz, M. F., Rawson, K. A., Traut, H., & Verkuilen, J. (2016). Towards a theory of learning for naming rehabilitation: Retrieval practice and spacing effects. *Journal of Speech, Language, and Hearing Research*, 59(5), 1111–1122. [https://doi.org/10.1044/2016\\_JSLHR-L-15-0303](https://doi.org/10.1044/2016_JSLHR-L-15-0303)
- Miner, M., Coane, J. H., Boland, S. C., Cooney, L. H., & Albat, M. (2018). The benefits of retrieval practice depend on item difficulty and intelligence. *Journal of Experimental Psychology: Learning Memory and Cognition*, 44(9), 1474–1486. <https://doi.org/10.1037/xlm0000486>
- Mulligan, N. W., & Picklesimer, M. (2016). Attention and the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(6), 938–950. <https://doi.org/10.1037/xlm0000227>
- Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory*, 2(3), 325–335. <https://doi.org/10.1080/09658219408258951>
- Oliva, M. T., & Storm, B. C. (2023). Examining the effect size and duration of retrieval-induced facilitation. *Psychological Research*, 87(4), 1166–1179. <https://doi.org/10.1007/S00426-022-01729-0/TABLES/1>
- Örün, Ö., & Akbulut, Y. (2019). Effect of multitasking, physical environment and electroencephalography use on cognitive load and retention. *Computers in Human Behavior*, 92, 216–229. <https://doi.org/10.1016/j.chb.2018.11.027>
- Owen, E., & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Educational Psychology*, 77(3), 272–284. <https://doi.org/10.1037/0022-0663.77.3.272>
- 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(1), 122–133. <https://doi.org/10.1037/0022-0663.86.1.122>
- Paas, F., & Van Merriënboer, J. J. G. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science*, 29(4), 394–398. <https://doi.org/10.1177/0963721420922183>

- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60(4), 437–447. <https://doi.org/10.1016/j.jml.2009.01.004>
- Pyke, W., Lunau, J., & Javadi, A.-H. (2023). Normative data for logographic and lexical Japanese paired associates. *Social Sciences & Humanities Open*, 7(1), 100398. <https://doi.org/10.1016/j.ssaho.2023.100398>
- Rawson, K. A., Vaughn, K. E., & Carpenter, S. K. (2015). Does the benefit of testing depend on lag, and if so, why? Evaluating the elaborative retrieval hypothesis. *Memory and Cognition*, 43(4), 619–633. <https://doi.org/10.3758/s13421-014-0477-z>
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20–27. <https://doi.org/10.1016/j.tics.2010.09.003>
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463. <https://doi.org/10.1037/a0037559>
- Rowland, C. A., & DeLosh, E. L. (2014). Benefits of testing for nontested information: Retrieval-induced facilitation of episodically bound material. *Psychonomic Bulletin & Review*, 21(6), 1516–1523. <https://doi.org/10.3758/s13423-014-0625-2>
- Smith, S. M., Ward, T. B., & Schumacher, J. S. (1993). Constraining effects of examples in a creative generation task. *Memory & Cognition*, 21(6), 837–845. <https://doi.org/10.3758/BF03202751>
- Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, 30(9), 641–656. <https://doi.org/10.1037/h0063404>
- Stenlund, T., Sundström, A., & Jonsson, B. (2016). Effects of repeated testing on short- and long-term memory performance across different test formats. *Educational Psychology*, 36(10), 1710–1727. <https://doi.org/10.1080/01443410.2014.953037>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1207/s15516709cog1202\\_4](https://doi.org/10.1207/s15516709cog1202_4)
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22(2), 123–138. <https://doi.org/10.1007/s10648-010-9128-5>
- Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation: Advances in Research and Theory*, 55, 37–76. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
- Sweller, J. (2023). The development of cognitive load theory: Replication crises and incorporation of other theories can lead to theory expansion. *Educational Psychology Review*, 35(4), 95. <https://doi.org/10.1007/s10648-023-09817-2>
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12(3), 185–233. [https://doi.org/10.1207/s1532690xc1203\\_1](https://doi.org/10.1207/s1532690xc1203_1)
- Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General*, 119(2), 176–192. <https://doi.org/10.1037/0096-3445.119.2.176>
- Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, 2(1), 59–89. [https://doi.org/10.1207/s1532690xc10201\\_3](https://doi.org/10.1207/s1532690xc10201_3)
- Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology*, 4(1), 147470490600400. <https://doi.org/10.1177/147470490600400135>
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology*, 80(4), 424–436. <https://doi.org/10.1037/0022-0663.80.4.424>
- Trafton, J. G., & Reiser, B. J. (1993). Studying examples and solving problems: Contributions to skill acquisition. In *Proceedings of the 15th Conference of the Cognitive Science Society* (pp. 1017–1022). Lawrence Erlbaum Associates, Inc.
- Tse, C. S., Balota, D. A., & Roediger, H. L. (2010). The benefits and costs of repeated testing on the learning of face-name pairs in healthy older adults. *Psychology and Aging*, 25(4), 833–845. <https://doi.org/10.1037/a0019933>
- van Gog, T., Kester, L., Dirkx, K., Hoogerheide, V., Boerboom, J., & Verhoeven, P. P. J. L. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educational Psychology Review*, 27(2), 265–289. <https://doi.org/10.1007/S10648-015-9297-3/FIGURES/1>
- van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, 27(2), 247–264. <https://doi.org/10.1007/S10648-015-9310-X/TABLES/1>
- Vaughn, K. E., Rawson, K. A., & Pyc, M. A. (2013). Repeated retrieval practice and item difficulty: Does criterion learning eliminate item difficulty effects? *Psychonomic Bulletin and Review*, 20(6), 1239–1245. <https://doi.org/10.3758/s13423-013-0434-z>
- Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science*, 20(1), 75–100. [https://doi.org/10.1207/s15516709cog2001\\_3](https://doi.org/10.1207/s15516709cog2001_3)
- Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, 7(1), 1–39. [https://doi.org/10.1207/s1532690xc10701\\_1](https://doi.org/10.1207/s1532690xc10701_1)
- Wirzberger, M., Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2016). One for all?! Simultaneous examination of load-inducing factors for advancing media-related instructional research. *Computers and Education*, 100, 18–31. <https://doi.org/10.1016/j.compedu.2016.04.010>
- Wirzberger, M., Borst, J. P., Krems, J. F., & Rey, G. D. (2020). Memory-related cognitive load effects in an interrupted learning task: A model-based explanation. *Trends*

- in *Neuroscience and Education*, 20, 100139. <https://doi.org/10.1016/j.tine.2020.100139>
- Wirzberger, M., Esmaili Bijarsari, S., & Rey, G. D. (2017). Embedded interruptions and task complexity influence schema-related cognitive load progression in an abstract learning task. *Acta Psychologica*, 179, 30–41. <https://doi.org/10.1016/j.actpsy.2017.07.001>
- Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory and Cognition*, 38(8), 995–1008. <https://doi.org/10.3758/MC.38.8.995>
- Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction*, 4(3), 137–166. [https://doi.org/10.1207/s1532690xci0403\\_1](https://doi.org/10.1207/s1532690xci0403_1)