# General Discussion

I will first reiterate the aims of this thesis, then I will summarise the key findings from each chapter. Following this, I will discuss key themes that persisted throughout the thesis with a focus on retrieval during practice, individual differences, interference, and task complexity. Finally, I will provide an overall conclusion to the thesis.

The aim of this thesis was to investigate whether task complexity interacted with the spacing effect. I chose to investigate massed versus spaced practice as distributing practice is one of the top recommendations for educators to improve learners’ retention over a longer period. It is frequently recommended by cognitive psychologists ([Carpenter et al., 2012](#ref-carpenter2012); [Latimier et al., 2021](#ref-latimier2021)), governments ([Ofsted, 2021](#ref-ofsted2021)), and used in educational technology ([M.S, 2023](#ref-m.s2023); [*Sparx Maths - Home*, n.d.](#ref-sparxma)). For verbal facts and lists of words, there is considerable evidence for optimal spacing routines and accurate predictions of forgetting for remembering a series of isolated facts ([Cepeda et al., 2008](#ref-cepeda2008); [Pavlik & Anderson, 2008](#ref-pavlik2008); [Walsh et al., 2022](#ref-walsh2022)), however, none of these models account for differences in the complexity of the material. As a practical example, when a student is planning their revision should a new times table fact, such as “12 x 11 = 132”, be practised again as soon as learning a new rule in calculus such as the chain rule? As current verbal theories and models of the spacing effect do not make clear predictions as to how task complexity may affect the efficacy of spacing, it is not currently possible to make such a recommendation.

Before a school can confidently, and purposefully, space practice across a curriculum, or an edtech platform can create an effective personalised study schedule for each student, four questions need to be answered:

1. Is spacing effective for complex material?
2. What aspects of the task or material increase complexity in a way which affects spacing?
3. If spacing is still effective, how does complexity affect the optimal spacing of material?
4. When spacing multiple learning objectives. How does spacing affect material that has links between the elements to be learned?

In this thesis I aimed to answer the first two questions. First, the experimental results alongside the meta-analysis provide strong evidence that spacing can be effective for learning mathematics, but more research needs to be done to show that testing is effective for mathematics relative to restudy. Across the four experiments presented in this thesis, I found no evidence that changes to procedural complexity or element interactivity interacted with the spacing effect.

## Chapter Two Summary: Meta-analysis

Chapter two presented a meta-analysis (published in Educational Psychology REview) assessing whether spaced practice, retrieval practice, or the combination of the two learning techniques are effective for learning mathematics. After running the initial search, it was apparent that there was insufficient evidence to address whether the combination of spacing and retrieval was effective for mathematics learning. While completing the meta-analysis I endeavoured to follow current best practices including pre-registration, I followed PRISMA gundefineduidelines ([Page et al., 2021](#ref-page2021)), used Covidence for reproducibility ([*Covidence Systematic Review Software*, 2023](#ref-covidenc2023)), reported heterogeneity, and I accounted for the fact that most studies provided multiple effects using Robust Variance Estimation ([Pustejovsky & Tipton, 2022](#ref-pustejovsky2022)). These steps helped to ensure that the results and conclusions were as valid as possible and should help with any future replications of th

There was limited evidence for the efficacy of retrieval practice rather than restudying materials in mathematics learning. The search revealed seven studies for mathematics learning that compared restudy versus testing and while there was a positive mean effect it was not robust, and the 95% confidence interval crossed zero. Given the substantial number of successful testing effect studies in other domains such as verbal learning, I did not conclude that this result suggested testing was not effective for mathematics learning. Instead, I suggest that this result highlights the need for future studies involving testing and mathematics to confirm that it is a robust effect. Another meta-analysis, with broader search parameters found 12 other papers not included in this study with different control conditions other than restudy (including testing versus no activity, testing versus elaborative strategies or more versus fewer test questions), suggesting that alternate research questions have been of greater interest to prior researchers ([Yang et al., 2021](#ref-yang2021)). Perhaps prior researchers assumed the testing effect would hold for mathematics material and moved to questions of more or less testing or when testing occurs, before checking the robustness of the testing effect for mathematics material. My meta-analysis shows, however, that there is still a need for fundamental research into the testing effect and mathematics learning and it would be valuable to provide sufficient evidence for its adoption on a larger scale. For these future studies, the testing versus studying worked examples approach, as used by Yeo and Fazio ([2019](#ref-yeo2019)), would be a good approach to fill this gap.

Overall, spaced practice was effective for mathematics learning, with a small to medium effect size, however, the search revealed two distinct paradigms that researchers employ when conducting research into distributed practice and mathematics learning. Participants either learnt the required information in isolation or as part of a course alongside other material. When material was taught in isolation the effect was larger, possibly due to increased control under lab conditions, while the weighted mean effect was lower for material presented as part of a course, perhaps due to interference between items. Due to the sample size, the potential moderating factors could not inform why isolated material benefits more from spaced practice or help to account for the large spread of effect sizes. I concluded that spacing was effective for mathematics learning.

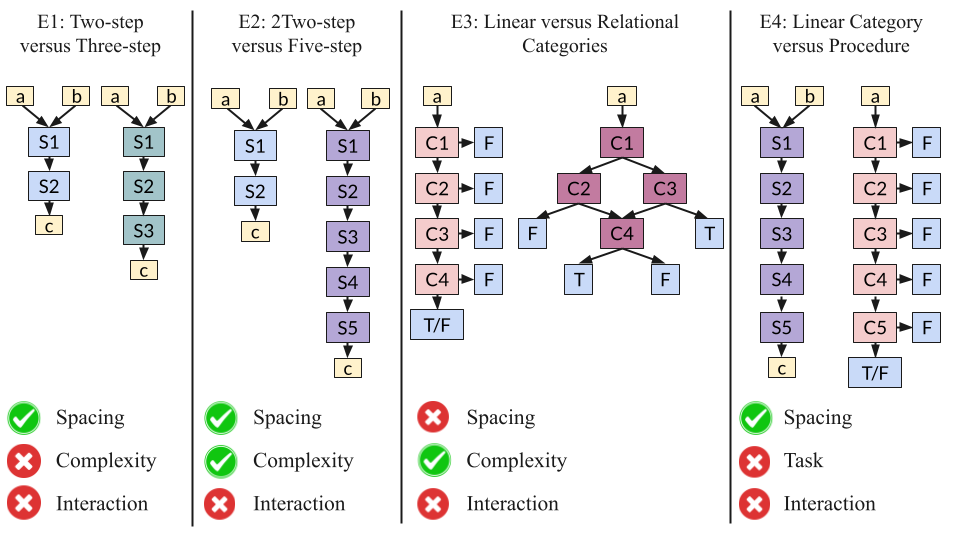
## Experimental work: Spacing Effect and Task Complexity

Chapter three and four describe the results of four experiments designed to investigate how the spacing effect changes when different aspects of task complexity are manipulated. Across all experiments I employed the same spacing schedule. I compared massed practice, where participants completed all the practice trials in a single session, with spaced practice, where participants completed the practice trials distributed evenly over three consecutive days. There was then a final session, after a seven-day delay, during which participants completed a post-test. Previous studies had issues where participants in the spaced condition had greater attrition which could have introduced a bias into the study ([Nazari & Ebersbach, 2018](#ref-nazari2018)). To minimise this potential bias all participants completed the four sessions and working memory, arithmetic fluency, and maths anxiety measures were organised to pad the days when massed participants did not complete the main task.

Across all experiments the main analysis was designed to detect a main effect of spacing (i.e., was there a significant difference between spaced and massed practice?), a main effect of Complexity/Task (i.e., was there a significant difference between the complexity or task conditions?) and finally whether there was a significant interaction between Spacing and complexity (i.e., was spacing more or less effective in one complexity condition or the other?).

Figure 1

Figure providing a simplified overview of the four experiments presented in the thesis. Experiment one and two are presented in chapter three and experiments three and four are presented in chapter four. Within the rectangles S\_ represents a step to be performed in a procedure while C\_ represents a constraint required to be checked in a category. The inputs and outputs “a”, “b” and “c” represent the numbers given to, and required from, participants (i.e., participants were provided with two integers a and b and were required to provide the integer c as an answer). For the category tasks, the required outputs were T (True) or F (False). The ticks and crosses along the bottom of the figure show whether there was a significant main effect of Spacing, Complexity/Task or an interaction between the two terms.



## Chapter Three summary:

Chapter three describes the results of two experiments where I varied the number of steps in mathematics procedures and compared a massed versus spaced practice schedule. These experiments were the first spacing studies to attempt to vary complexity within a mathematics learning task. I used procedures consisting of a series of arithmetic steps (i.e., “add the two numbers together” or “divide the answer by three”). The artificial nature of the procedures made the experiment less ecologically valid, but minimised the effect of prior knowledge, intending that the only change was the procedural complexity (the number of steps in the procedures).

I initially compared a two- versus three-step procedure in experiment one, before comparing a two- versus five-step procedure in experiment two. In experiment one, I found a significant main effect of spacing, but I did not find a significant effect of complexity, nor a significant interaction effect between the two terms. This suggested that an additional step was not sufficient for a difference in complexity to be detected, so I added two extra steps to the complex condition for the next experiment. In experiment two, comparing a two- versus a five-step procedure, I found both a significant main effect of spacing and complexity, but no significant interaction effect between spacing and complexity. These results suggest that spacing is robust to changes in procedural complexity and another measure of complexity may be more important for educators to consider.

## Chapter Four summary:

Chapter four drew on research that was published during data collection of experiments one and two. Chen et al. ([2023](#ref-chen2023)) published an article which provided an in depth definition for element interactivity, which had previously been critiqued for not being sufficiently well operationalised ([Karpicke & Aue, 2015](#ref-karpicke2015)). Furthermore, Chen et al. ([2024](#ref-chen2024)) published a series of experiments exploring how element interactivity interacted with a version of the spacing effect they dubbed the resting effect. They found mixed results that showed a spacing effect when element interactivity was high (which they suggested was due to participants being able to recover working memory resources during rest) and when it was low (which they suggested was due to participants being able to rehearse the information during rest). In their third and fourth experiment, where they compared spaced and massed practice with high and low element interactivity, they used the same material and instead manipulated the expertise of participants. They had one expert group (low element interactivity) and a novice group (high element interactivity). I aimed to further this research in two ways. First, I wanted to find a task where I could manipulate element interactivity directly instead of manipulating participants’ prior knowledge. I did this by having participants learn two categories, a linear category, where each constraint could be learnt in isolation (low element interactivity) and a relational category, where each constraint was only useful when other elements were taken into consideration (high element interactivity). This was a useful next step as in Chen et al. ([2024](#ref-chen2024)) ’s experiment the expert and novice groups may have had a large variation in prior knowledge, while this material was designed so that each participant should have equal knowledge of the individual elements and only the structure of how elements are linked together was varied. Second, I used longer inter-session intervals and retrieval intervals. This is useful as shorter retrieval intervals can be a boundary condition of the spacing effect and at short intervals massed practice may be superior to spaced practice ([Rohrer & Taylor, 2007](#ref-rohrer2007)).

I hypothesised that the linear category would benefit from spacing while the relational category would not because of the prior research showing that higher complexity leads to reduced efficacy of spacing ([Donovan & Radosevich, 1999](#ref-donovan1999)). While the working memory resource depletion hypothesis ([Chen et al., 2024](#ref-chen2024)) would instead suggest that the relational category would benefit more from spacing as the intervals between the sessions would allow for working memory resources to be restored and lead to greater subsequent learning. In contrast to either hypothesis, neither the linear nor relational category benefited from spaced practice relative to massed practice. I suggested two reasons for this, either both categories were too high in element interactivity to benefit from spacing, in comparison to the procedures task used in chapter three, or there were high levels of interference between the two category learning tasks that hindered learning.

In experiment four, participants were taught a five-step procedure (adapted from experiment two) and a five-constraint linear category (adapted from experiment three). This time I found a significant effect of spacing across both task conditions, but not a main effect of task or an interaction between spacing and task (although it was the closest to significance of all the other experiments). This suggests that in experiment three the linear category in experiment one was not too high in element interactivity to benefit from the spacing effect, but instead it was the fact that it was learnt alongside the relational category. Following these two experiments I can make no strong statements about whether element interactivity interacts with the spacing effect. Furthermore, future experiments should avoid learning material where there is likely to be a lot of interference, manipulating the structure of the material while changing and counterbalancing the surface features of the task. Alternatively, researchers could explicitly manipulate the material so that there is a high interference condition and a low interference condition to see if this a potential boundary condition for the spacing effect in mathematics learning.

### Theoretical and methodological considerations

After summarising the results of the meta-analysis and two experimental chapters I will now focus on four key themes that persisted across the thesis. First, I will discuss the role of retrieval during practice in mathematics learning and whether I saw the trade off between lower early performance and greater long-term retention typical in spacing studies. Then, I will discuss the individual difference measures I tracked across the experiments. Next, I will expand on interference as the explanation for why I did not see a spacing effect in experiment four before a final discussion of task complexity and presenting overall conclusions.

#### Retrieval Accuracy on the Practice Trials.

During data extraction of the meta-analysis, it was difficult to understand whether the included studies required participants to actively retrieve the material from memory or allowed them to use external aids, such as notes, examples, or prior work. As participants’ retrieval of the material after initial acquisition is key to the study-phase account of the spacing effect ([Thios & D’Agostino, 1976](#ref-thios1976)), this seemed important and was one of the key ways the meta-analysis informed the design of the tasks used in chapters two and three. Across all the experiments presented in this thesis, I wanted to ensure that after the initial learning phase participants would be required to retrieve the information. I asked participants not to write anything down and, after the initial learning session, participants could not see the procedure/category they were trying to learn on screen. However, participants were given feedback if incorrect, so after the first question participants may have seen the procedure/category again. In which case, they only had to rehearse the procedure/category in working memory, rather than recall from long-term memory, but this is still more explicit than many prior studies and something that should be included in mathematics learning and spacing research going forward. Furthermore, it would be difficult to rehearse the whole procedure or category at once. One way to test whether feedback reduces the efficacy of ensuring retrival during practice could be to add a distractor between each practice trial so that participants cannot maintain the procedure or category in working memory, however this was not feasible in my experiments due to time constraints. I think that being explicit about what participants had to retrieve and when they were able to follow an example was an important methodological improvement.

Spacing did not routinely lead to poorer practice performance when compared to massed practice in three out of the four experiments presented in this thesis. Spacing practice has been described as a desirable difficulty, a learning technique that makes practice more difficult to the benefit of future learning and retention ([Bjork & Bjork, 2020](#ref-bjork2020); [Lyle et al., 2022](#ref-lyle2022)). It is often a trade off between poorer initial performance and greater future performance. In experiment four I found this to be the case, retrieval accuracy during practice was significantly higher in the massed condition than in the spaced condition. This creates a potential danger when spacing practice in the classroom as students find it more difficult and may perform worse, reducing their confidence. In contrast, in experiments one, two, and three there was no significant difference in retrieval accuracy between spacing conditions. This is positive as participants benefited from spaced practice in experiments one and two while not suffering reduced performance during practice. However, as participants in the spaced condition in experiment four did find their practice more difficult, this highlights the importance of ensuring that students understand why they are being asked to space their practice, that it will be more difficult, but that it will help them to learn.

#### Individual differences.

Across my four experimental studies I measured participants’ working memory capacity (WMC), measured using forwards and backwards digit span tasks, arithmetic fluency, and mathematics anxiety (excluding experiment three where mathematics anxiety was omitted). For each experiment, there were no significant differences in working memory capacity or arithmetic fluency between the massed and spaced groups. This suggests that the differences in final performance were due to the spacing manipulation, not a bias introduced through our sampling of participants. For each experiment I ran exploratory analyses to detect whether any of the individual difference measures were significant covariates, in each case they were not significant. As these analyses were exploratory, they provided no potential insights, and I did not have sufficient power to detect small effects, I did not include them in the thesis, but the data and code to run them is available in the supplementary materials.

While it did not appear to have an effect in the studies presented in this thesis, there is some evidence to suggest that spacing may benefit participants with higher WMC more than those with lower WMC ([Bui et al., 2013](#ref-bui2013)), however, other results suggest that participants with higher WMC simply perform better on memory tests and higher WMC has an additive effect on performance and does not interact with the spacing effect ([Delaney et al., 2018](#ref-delaney2018)). Overall, while it is reasonable to expect participants with higher WMC to perform better on more complex material, I found no evidence that this is the case. If a future study finds a manipulation of task complexity that significantly affects the efficacy of spacing, then an additional high-powered study with multiple measures of working memory capacity would provide a better understanding of the relationship between WMC and the spacing effect.

In each experiment, I decided to test participants’ mathematics anxiety at the end of the study, after the final test, so that taking the test itself did not alter participants’ anxiety during the task. This presents a limitation as participation in any of the experiments could have either increased anxiety, if participants found the task difficult, or reduced it, perhaps due to increased fluency during massed practice leading to overconfidence ([Emeny et al., 2021](#ref-emeny2021)). I did not have the resources to test mathematics anxiety twice during the experiment, however, if mathematics anxiety is a focus of a future experiment, then an improvement would be to test participants at the start and end of the experiment. This would allow future researchers to better distinguish between a baseline mathematics anxiety of the participant and the change in mathematics anxiety during the experiment.

Due to a change in recruitment platform, participants were older in experiment four, than the first three experiments. In experiment one, two and three I recruited my participants using the the University of York’s human participant pool SONA and these participants were similar in age (around 20 years old). However, in experiment four I recruited participants via prolific and the participants were older (around 27 years old). Before experiment four, our criteria were that participants needed to be over 18 and an undergraduate student and I used the same criteria on prolific. Either prolific has a larger proportion of mature students, who signed up to our experiment quickly, or perhaps participants on prolific are incentivised to claim that they are a university student in order to gain access to more studies. However, I do not think that this caused a major issue as participants had high performance on the task and post-test, and previous research suggests the spacing effect is robust across many age ranges ([Delaney et al., 2010](#ref-delaney2010); [Kornell & Bjork, 2008](#ref-kornell2008); [MacLean et al., 2017](#ref-maclean2017)).

#### Interference.

Additionally, further research should be undertaken to see how different items affect the spacing schedules of other items learnt alongside each other in practice. The meta-analysis presented in chapter two suggested that a smaller spacing effect was found for items presented together in a course and the material in experiment three did not benefit from spacing when presented together, while an adapted version of the same item in experiment four did benefit when learnt alongside less similar material. This is not a ubiquitous finding, however, as the two procedures used in experiments one and two were visually and conceptually similar and they consistently benefited from spaced practice relative to massed practice. Future studies that manipulate the structure of categories, but vary the surface features, would be a valuable next step. As an example, one task could require participants to categorise a number, while the other requires participants to categorise a shape. Another future experiment could include a direct manipulation of spaced versus massed practice, high versus low element interactivity categories and similar versus dissimilar surface features. This 2 x 2 x 2 factorial design could test whether the high levels of interference across tasks may be a boundary condition for the spacing effect.

#### Task Complexity.

The most difficult part of research into task complexity is that it is unique to each participant. There is evidence that spacing benefits “complex” material, such as learning calculus ([Hopkins et al., 2016](#ref-hopkins2016); [Lyle et al., 2020](#ref-lyleHowAmountSpacing2020), [2022](#ref-lyle2022)) and some people may argue that spacing has already been shown for complex material. After all, the brain has a remarkable ability to chunk information together allowing us to gain expertise and make the complex simple ([Gobet et al., 2001](#ref-gobet2001); [Miller, 1956](#ref-miller1956)). However, if participants had high prerequisite prior knowledge and had portions of the “complex” material chunked into a single element then that material was not complex to that participant. Across four experiments I aimed to manipulate complexity and practice schedule while reducing the effect of prior knowledge. My key contribution was the use of artificial elements (i.e a step to perform “multiply by five” or a constraint to check “is the number a multiple of five?”) to build up novel learning tasks in a modular fashion. Importantly, I expected that each individual element would be known to the participants (and checked this during piloting) and that what they are learning is the structure of the procedure or category. By varying the structure, I sought to manipulate first procedural complexity in experiments 1 and 2, and then element interactivity in experiments 3 and 4.

When measuring procedural complexity as the number of steps participants were required to learn, we found large main effects of spacing. Further evidence, with larger sample sizes, would be required to make a strong reccomendation to educators that it can be ignored when scheduling practice, however in my experiments it did not affect spacing. My manipulation of element interactivity found no effects of spacing, but this was likely to not be due to the material itself, but instead to how the task was set up. The linear category learning task did not benefit from spacing when learnt alongside a similar relational category learning task. However, the linear category did benefit from spacing when paired with a procedure learning task. If element interactivity does matter, then one methodological concern future experiments should consider is to avoid interference between learning objectives. Overall, the experiments add to the consensus that the spacing effect is robust across different tasks.

One explanation for the spacing effects robustness to changes in task complexity could be due to multiple complimentary mechanisms underlying the spacing effect. Most modern overviews of the spacing effect argue that no single mechanism that can account for, and model, all the key phenomena of the spacing effect ([Delaney et al., 2010](#ref-delaney2010); [Küpper-Tetzel, 2014](#ref-küpper-tetzel2014); [Walsh et al., 2018](#ref-walsh2018)). For example, if more complex material with more elements increases the difficulty of retrieval then the study-phase retrieval account would suggest a reduced effectiveness of spacing ([Thios & D’Agostino, 1976](#ref-thios1976)), however perhaps if participants realise this during practice, either consciously or subconsciously, they may attend to the information more closely. If this occurs, then in the massed condition participants may feel fluent in the material and process it less deeply, as suggested in the deficient processing account ([Hintzman, 1974](#ref-hintzman1974)), while they may attend more closely to the complex material in the spaced condition when unable to recall the material.

I was unable to test this theory of multiple complementary mechanisms in the experiments presented in this thesis; however, future researchers could test this theory by running a series of experiments manipulating inter-session intervals and amount of practice trials. One study found that increasing the amount of practice beyond the minimum needed was found not to moderate the spacing effect ([Lyle et al., 2020](#ref-lyleHowAmountSpacing2020); [Rohrer & Taylor, 2006](#ref-rohrer2006)), however, this may have been due to the use of a simple learning objective. Increasing or reducing the number of practice sessions can enable a researcher to see whether deficient processing is occurring as at some point the less complex material will begin to no longer benefit from increased practice while the more complex may benefit from additional practice. Additionally, if the critical point where deficient processing occurs is moderated by changes to the inter-session interval then that may be evidence that participants are using the difficulty of retrieval as a sign to process the material at a deeper level. If this is indeed the case, then this would make spacing an even more valuable tool to improve learning.

## Overall conclusions

I found no evidence that procedural complexity or element interactivity affects the efficacy of spaced practice. Importantly, however, my manipulation of element interactivity was impeded by high interference between the two tasks and should not be considered evidence that element interactivity does not affect the efficacy of spacing. More evidence is needed to show procedural complexity has no effect on the efficacy of spacing, however, future research should focus on explicit manipulations of element interactivity that use different surface features, to reduce interference, while manipulating the structure of the materials. If spacing is affected by task complexity, then further studies should vary the inter-session and retrieval intervals to ascertain how task complexity affects the optimal spacing schedules for different tasks. Taken together, the results of the meta-analysis and robustness of the spacing effect across three out of four experiments presented in this thesis supports the inclusion of spaced practice in mathematics learning.

# References

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.jarmac.2020.09.003>

Bui, D. C., Maddox, G. B., & Balota, D. A. (2013). The roles of working memory and intervening task difficulty in determining the benefits of repetition. *Psychonomic Bulletin & Review*, *20*(2), 341–347. <https://doi.org/10.3758/s13423-012-0352-5>

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educ. Psychol. Rev.*, *24*(3), 369–378. <https://doi.org/10.1007/s10648-012-9205-z>

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention: Research article. *Psychol. Sci.*, *19*(11), 1095–1102. <https://doi.org/10.1111/J.1467-9280.2008.02209.X>

Chen, O., Kai Yin Chan, B., Anderson, E., O’sullivan, R., Jay, T., Ouwehand, K., Paas, F., & Sweller, J. (2024). The effect of element interactivity and mental rehearsal on working memory resource depletion and the spacing effect. *Contemporary Educational Psychology*, *77*, 102281. <https://doi.org/10.1016/j.cedpsych.2024.102281>

Chen, O., Paas, F., & Sweller, J. (2023). A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educ. Psychol. Rev.*, *35*(2), 63. <https://doi.org/10.1007/s10648-023-09782-w>

*Covidence systematic review software*. (2023). Veritas Health Innovation. [www.covidence.org.](https://www.covidence.org.)

Delaney, P. F., Godbole, N. R., Holden, L. R., & Chang, Y. (2018). Working memory capacity and the spacing effect in cued recall. *Memory*, *26*(6), 784–797. <https://doi.org/10.1080/09658211.2017.1408841>

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* (Vol. 53, pp. 63–147). <https://doi.org/10.1016/S0079-7421(10)53003-2>

Donovan, J. J., & Radosevich, D. J. (1999). *A meta-analytic review of the distribution of practice effect: Now you see it, now you don’t*.

Emeny, W. G., Hartwig, M. K., & Rohrer, D. (2021). Spaced mathematics practice improves test scores and reduces overconfidence. *Appl. Cogn. Psychol.*, *35*(4), 1082–1089. <https://doi.org/10.1002/acp.3814>

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*(6), 236–243. <https://doi.org/10.1016/s1364-6613(00)01662-4>

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. *Theories in Cognitive Psychology: The Loyola Symposium.*, *386*. <https://psycnet.apa.org/fulltext/1975-00291-001.pdf>

Hopkins, R. F., Lyle, K. B., Hieb, J. L., & Ralston, P. A. S. (2016). Spaced retrieval practice increases college students’ short- and long-term retention of mathematics knowledge. *Educ. Psychol. Rev.*, *28*(4), 853–873. <https://doi.org/10.1007/s10648-015-9349-8>

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educ. Psychol. Rev.*, *27*(2), 317–326. <https://doi.org/10.1007/s10648-015-9309-3>

Kornell, N., & Bjork, R. A. (2008). Learning Concepts and Categories: Is Spacing the “Enemy of Induction”? *Psychological Science*, *19*(6), 585–592. <https://doi.org/10.1111/j.1467-9280.2008.02127.x>

Küpper-Tetzel, C. E. (2014). Understanding the distributed practice effect. *Zeitschrift Für Psychologie*, *222*(2), 71–81. <https://doi.org/10.1027/2151-2604/a000168>

Latimier, A., Peyre, H., & Ramus, F. (2021). A meta-analytic review of the benefit of spacing out retrieval practice episodes on retention. *Educ. Psychol. Rev.*, *33*(3), 959–987. <https://doi.org/10.1007/s10648-020-09572-8>

Lyle, K. B., Bego, C. R., Hopkins, R. F., Hieb, J. L., & Ralston, P. A. S. (2020). How the amount and spacing of retrieval practice affect the short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, *32*(1), 277295. <https://doi.org/10.1007/s10648-019-09489-x>

Lyle, K. B., Bego, C. R., Ralston, P. A. S., & Immekus, J. C. (2022). Spaced retrieval practice imposes desirable difficulty in calculus learning. *Educ. Psychol. Rev.*, *34*(3), 1799–1812. <https://doi.org/10.1007/s10648-022-09677-2>

MacLean, A. C., Bell, M. C., & Simone, P. M. (2017). Interference Effects, Age, and the Spacing Benefit. *The American Journal of Psychology*, *130*(3), 295–302. <https://doi.org/10.5406/amerjpsyc.130.3.0295>

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. <https://doi.org/10.1037/h0043158>

M.S, E. O. (2023). *Duolingo*. <https://blog.duolingo.com/spaced-repetition-for-learning/>

Nazari, K. B., & Ebersbach, M. (2018). Distributed practice: Rarely realized in self-regulated mathematical learning. *Front. Psychol.*, *9*(NOV), 2170. <https://doi.org/10.3389/fpsyg.2018.02170>

Ofsted. (2021). *Research review series: mathematics*. <https://www.gov.uk/government/publications/research-review-series-mathematics>

Page, M. J., Moher, D., Fidler, F. M., Higgins, J. P. T., Brennan, S. E., Haddaway, N. R., Hamilton, D. G., Kanukula, R., Karunananthan, S., Maxwell, L. J., McDonald, S., Nakagawa, S., Nunan, D., Tugwell, P., Welch, V. A., & McKenzie, J. E. (2021). The REPRISE project: Protocol for an evaluation of REProducibility and replicability in syntheses of evidence. *Syst. Rev.*, *10*(1). <https://doi.org/10.1186/s13643-021-01670-0>

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *J. Exp. Psychol. Appl.*, *14*(2), 101–117. <https://doi.org/10.1037/1076-898X.14.2.101>

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prev. Sci.*, *23*(3), 425–438. <https://doi.org/10.1007/s11121-021-01246-3>

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Appl. Cogn. Psychol.*, *20*(9), 1209–1224. <https://doi.org/10.1002/acp.1266>

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning | instructional science. *Instructional Science*, *35*(6), 481–489. <https://link.springer.com/article/10.1007/s11251-007-9015-8>

*Sparx maths - home*. (n.d.). <https://sparxmaths.com/>

Thios, S. J., & D’Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning & Verbal Behavior*, *15*(5), 529–536. <https://doi.org/10.1016/0022-5371(76)90047-5>

Walsh, M. M., Gluck, K. A., Gunzelmann, G., Jastrzembski, T., & Krusmark, M. (2018). Evaluating the theoretic adequacy and applied potential of computational models of the spacing effect. *Cogn. Sci.*, *42 Suppl 3*, 644–691. <https://doi.org/10.1111/cogs.12602>

Walsh, M. M., Krusmark, M. A., Jastrembski, T., Hansen, D. A., Honn, K. A., & Gunzelmann, G. (2022). Enhancing learning and retention through the distribution of practice repetitions across multiple sessions. *Mem. Cognit.* <https://doi.org/10.3758/s13421-022-01361-8>

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychol. Bull.*, *147*(4), 399–435. <https://doi.org/10.1037/bul0000309>

Yeo, D. J., & Fazio, L. K. (2019). The optimal learning strategy depends on learning goals and processes: Retrieval practice versus worked examples. *J. Educ. Psychol.*, *111*(1), 73–90. <https://doi.org/10.1037/edu0000268>