A Systematic Review of Interleaving as a Concept Learning Strategy: A Study Protocol

Jonathan Firth[1*], Ian Rivers[1] and James Boyle[2]

[1]School of Education, University of Strathclyde
[2]School of Psychological Sciences and Health, University of Strathclyde

ABSTRACT

Background: Education Scotland's (2018) framework for interventions for equity supporting the Scottish Attainment Challenge highlights the promotion of high quality learning and the effective use of evidence and data. This study protocol outlines the methodology of a systematic review of the literature into the use of interleaving to facilitate the effective learning and teaching of new concepts.

Methods: The systematic review has been pre-registered with PROSPERO, an international database of prospectively registered systematic reviews. The review will investigate whether presenting examples of to-be-learned concepts in an interleaved order is a more effective learning strategy than presenting examples blocked by topic, in terms of learners' ability to remember examples and to transfer learning to novel examples.

Discussion: Interleaving is widely recommended as an evidence-based approach to teaching with considerable potential as a strategy for learners experiencing difficulties in working memory functioning and conceptual learning, but to date there has not been a comprehensive review of the evidence base. The review will address this gap. It will synthesize primary research studies from the past decade, investigate boundary conditions and variables that interact with interleaving, and will include a meta-analysis of recent studies. This protocol provides the details of the rationale of the review, and details the inclusion criteria and approaches to data extraction.

Keywords: Interleaving, memory, transfer, concept learning, education, spacing, attainment challenge, working memory

1. Background

The term interleaving refers to variability within a set of tasks or example items such that each item is immediately followed and preceded by an example of a different type or concept rather than appearing in blocks of the same type of item repeatedly; the latter is termed a 'blocked' sequence. For example, in Figure 1, presentations of item types '1', '2' and '3' are shown blocked (Example A), or with the three item types interleaved (Example B):

* Corresponding author: Jonathan Firth, School of Education, University of Strathclyde, 141 St James Road, Glasgow, Scotland, G4 0LT. Email: jonathan.firth@strath.ac.uk

Figure 1. Comparison of interleaved and blocked sequences.

Example A: blocked sequence: 111112222233333

Example B: interleaved sequence: 123123123123123

The benefit of interleaving is sometimes termed the interleaving effect, and has been investigated in numerous contexts, many of which are of direct relevance to education. It has been found to benefit maths learning (e.g., Rohrer et al, 2015), the conceptual learning of science categories and examples (Eglington & Kang, 2017; Rawson et al, 2015), and the inductive learning of images of animal species and modern art paintings (Birnbaum, Kornell, Bjork & Bjork, 2013; Kornell & Bjork, 2008). If learners were being taught about species of bird, for example, blocking would involve showing multiple examples of the same species of bird consecutively, while interleaving would involve showing examples of one species of bird, then another, then yet another, and so on.

Interleaving has been described as a desirable difficulty, in that it can make learning slower at first but more durable over the long term (Yan, Bjork & Bjork, 2016). Its benefits were discovered by William Battig in the 1960s via his research into the learning of word pairs, and he believed that the interference caused by interleaving can make learning more resilient — an idea consistent with the role of varied environmental contexts in learning (e.g., see Smith, Glenberg & Bjork, 1978). As such, Battig (1979) tended to describe the effect in terms of 'contextual interference', a term which emphasises the role of the broader context rather than just the interleaved items themselves (Battig, 1979; Magill & Hall, 1990). This term is still widely used in the domain of motor learning. However, recent research into concept learning has suggested that it may be the relationship between interleaved items that leads to the interleaving benefit rather than simply the interference which results from the format, with interleaving making it easier for learners to compare and contrast items and thereby notice subtle conceptual differences between them (Birnbaum et al, 2013). In one study which helped to revive recent interest in applying interleaving to education, Kornell and Bjork (2008) found that interleaving sets of artwork by different artists led to more effective learning than spending the same amount of time looking at blocked examples of paintings by the same artist. Viewing paintings by different artists in an interleaved order improved learners' ability to later identify the style of these artists as indicated by their ability to correctly identify novel example paintings, perhaps due to the greater ease of learning when contrasting items are seen consecutively — an idea known as the discrimination-contrast hypothesis (Birnbaum et al, 2013). This fits with a body of research into categorisation which tends to show that highlighting differences (i.e. making discriminative contrast more salient) has a more beneficial effect on category learning than highlighting similarities (Higgins & Ross, 2011).

Supporting this idea, Hausmann and Kornell (2014) mixed the study of Indonesian vocabulary with the learning of biology terms, and did not find a benefit of interleaving, presumably because the two sets of material were too conceptually distant to productively be compared or organised. Similar items do tend to show a benefit of interleaving, especially when the difference between them is subtle (Carvalho & Goldstone, 2014), possibly because subtle differences would be very hard for learners to notice if viewed during separate study sessions. Interestingly, though, Eglington and Kang (2017) did not find that explicitly highlighting differences during the learning phase impacted on the benefits of interleaving compared to blocking.

There is an unavoidable connection between interleaving and the spacing effect (i.e. distributed practice), and the two are often conflated. This is because interleaving items inevitably increases the gaps between one example item and the next (as can be seen in Figure 1). In Kornell and Bjork's (2008) study of interleaving and modern art paintings, the researchers initially attributed their findings to spacing, an effect which was already well established at the time (e.g., see Dempster, 1996). However, Kang and Pashler (2012) carried out a replication where spacing was held constant, using filler images to increase the temporal space between one blocked item and the next, and found that the interleaved condition was nevertheless superior to the spaced or blocked conditions. In a similar study which used trivia questions as filler items, Birnbaum et al (2013) concluded that the benefit to inductive learning of visual items was largely due to interleaving rather than spacing, while Taylor and Rohrer (2010) found that for mathematics practice, both spacing and interleaving have separate beneficial effects. It could be argued that failing to control for spacing (as was typical in studies prior to 2012) leads to a confound between the two variables.

Interleaving and spacing therefore tend to be seen as separate phenomena today, and both are widely recommended by sources which advocate applying cognitive science to classroom teaching, for example the UK's Chartered College of Teaching (https://chartered.college), the Learning Scientists (learningscientists.org), as well as numerous recent books on teaching practice.

A practical difference between spacing and interleaving is that rather than re-studying the same material on separate occasions, the procedure of interleaving studies tends to involve presenting several different examples of a category during a learning phase, and then presenting novel items during a test phase (instead of or as well as testing memory for the original items). This has educational implications; spacing is likely to be beneficial when items studied are exactly the same, such as for the practice of foreign language vocabulary. Interleaving, in contrast, stands to benefit knowledge transfer following the learning of varied prior examples. It is likely to be useful in situations where learners may have to identify novel instances of previously studied concepts, such as identifying signs of glaciation in a previously unseen landscape, or recognising social psychology phenomena during an everyday encounter.

Although the potential of interleaving to help learners to compare and contrast exemplars has been noted, another theoretical explanation of the benefit is that learners tend to pay more attention to interleaved items. This is known as the attention-attenuation hypothesis (Wahlheim, Dunlosky & Jacoby, 2011). It is an idea which fits with evidence that blocked presentations tend to lead to more mindwandering (Metcalfe & Xu, 2016). It also fits with recent findings that working memory capacity does not play a major role in the interleaving effect (Guzman-Munoz, 2017; Sana, Yan, Kim, Bjork, & Bjork, 2018) — if contrast between current and previous items is the key factor in the effect then a larger working memory capacity should increase the benefit, but this does not appear to be the case. This finding suggests that interleaving could generalise to younger learners whose working memory is still developing, or to pupils with additional support needs or adults with impaired working memory.

The present study is a systematic review and meta-analysis of research into interleaving, focusing on the work over the ten years since Kornell and Bjork's seminal study of interleaved learning of art paintings. It aims to provide a much-needed overview of the evidence base for interleaving as an educational technique, as well as an indication of the effect size (if any) of the technique as an intervention, and any important interactions that may emerge. Given that many of the studies cited thus far were conducted on undergraduate populations and with specific tasks (art, maths, etc), the review will also aim to shed light on

whether the evidence base – such as it is – can support recommendations for applying interleaving to other educational domains. It may also give an indication of what kind of school tasks are most likely to benefit from this intervention, and an idea of where further research is needed. The review question is as follows: For a population of learners in mainstream education, is presenting examples of concepts in an interleaved order a more successful learning strategy than presenting examples blocked by type, in terms of learners' ability to remember examples and transfer to novel examples?

2. Methods/Design

2.1 Searches

One researcher will conduct data extraction, using the PsycINFO, Web of Science, BEI, AEI and ERIC databases. Search terms will focus on the research variable interleaving (interleav*, with possible synonyms "contextual interference", shuffl*, intermix*) and on the outcome variable (learning/“conceptual knowledge”, inducti*), for records from 2008-present. Where necessary, database journal categories will be used to exclude items from irrelevant domains or on the basis of exclusion criteria below. Other search methods will include hand searching of relevant journals, and reference chasing from existing narrative reviews by Rohrer (2012), Carvalho & Goldstone (2015), and Kang (2016).

2.2 Domain and context

The focus will be on learning and memory. More specifically, the review will focus on the application of interleaving as an intervention for learning tasks, such that it could be applied to schools or other educational contexts. The review will focus on studies where participants are in formal education, but studies on equivalent adults from the general population (e.g., recruited via Amazon Mechanical Turk) need not be excluded.

2.3 Participants

As a first step, the review will look at studies of mainstream education populations, representative of those in formal education at school or university. This can comprise adolescents or adults or both, but will exclude older adults (65+) on the basis that this population may have memory issues that differ from those experienced by the bulk of students in education.

2.4 Intervention and comparator

The review will search for studies of interleaving versus blocking, where interleaving refers to immediate variation within a set of tasks or example items, whereby each item is immediately followed and preceded by an example of a different category/concept, such that examples of concept ‘1’, concept ‘2’ and concept ‘3’ would be presented in an order such as 123123123. Interleaving could arise either due to a randomisation or ‘shuffling’ of the order of items, or a more deliberate alternation of items.

Blocked presentation involves concepts being presented via multiple examples of the same concept (e.g., 111, 222, 333 for the concepts labelled above). It is defined as presentations where studied items appear in blocks of the same type of item repeatedly.

Other than the order of presentation, control conditions should be identical to the intervention condition to avoid confounding variables; studies where spacing is a potential confound will be included, but the potential effect of this issue will be considered when reviewing the evidence.

2.5 Types of study to be included

Inclusion criteria are as follows:

- Age 13-65, must be a typically (or assumed typically) developing sample without known memory problems.
- Experimental or quasi-experimental designs only.
- Studies must have collected primary data.
- One of the primary research variables must be interleaving as it relates to learning/memory/understanding.

In addition to excluding studies that don't meet all four inclusion criteria, two further exclusion criteria will be applied:

- Exclude neurological/fMRI-based studies.
- Exclude studies with outcome variables that don't directly demonstrate concept learning (e.g., studies of attention, mind-wandering, visual perception, motor learning, language learning).

Inclusions will be cross-checked within the research team. Any discrepancies will be solved via discussion. Domain (visual v's verbal, and relevance to educational contexts) will be assessed.

2.6 Main outcome(s)

Research studies must measure learning via memory recall (correctly identifying trained examples), or transfer (correctly identifying novel examples), or both.

It is notable that such results do not exist in isolation in the real world, but instead link to other domains. One is the mental health of school/university pupils at exam time, where it has been shown that cognitive processes interact with anxiety levels. The other is the psychological wellbeing of learners who are experiencing memory difficulties. Relevant studies on these health outcomes will be eligible for inclusion in the review, provided that other criteria are also met.

2.7 Data extraction

Key data to be included are name and year of study, experiment number (where there are multiple experiments in the same publication), design (within or between participants), outcome measure (memory v's transfer), mean score and SD for each condition, F values or t values, population, sample size for each condition, domain tested (visual v's verbal v's other), relevance to specific subject discipline (e.g., maths, science). These will be assessed again by the first researcher, and coding will be cross-checked among the research team.

2.8 Risk of bias (quality) assessment

Possible sources of bias include participant demand characteristics, incomplete or selective reporting of outcomes, lack of sample diversity, or conflicts of interest. These will be assessed again by the first researcher, and judgements will be cross-checked among the research team. Funnel plots, fail-safe N, Begg and Mazumdar's Test and Eggers' Test (see Borenstein et al., 2014) will be used to assess publication bias.

2.9 Strategy for data synthesis

For meta-analysis, we will use we will calculate Hedge's *g* standardised effect size, which minimises bias from studies with small sample size. Representation of tasks which link to specific-subject education disciplines will also be assessed.

2.10 Analysis of subgroups

We will conduct subgroup analyses in order to evaluate the impact of moderators on pooled effect sizes, including the key research question (transfer v's memory), type of category studied (which will be divided by science/factual v's art/images), and within v's between participant designs.

3. Discussion

This PROSPERO-registered protocol outlines the search and analysis strategy for a systematic review into interleaving as it applies to education. Conducting the review will provide a much clearer picture of the interleaving effect, including some of the variables with which interleaving may interact, and any relevant boundary conditions. It will help to clarify whether interleaving is useful as a memory intervention, or if it is beneficial for transfer of prior learning to new contexts, or both.

At a time when interleaving is increasingly being advocated as a strand of evidence-based teaching practice, it will be useful to gain an objective summary of the evidence for the effect, and also to investigate its potential as a strategy to support learners who struggle with concept learning, or who have reduced working memory. This will include finding out what type of school-based tasks it appears to be best suited to, with a view to supporting pupils with additional support needs.

Interleaving promotes the contrast of real-world examples and thereby facilitates learners' induction of new concepts. Given its potential for use with visual examples, interleaving could boost the development of conceptual knowledge learning among pupils with reading disabilities, thereby supporting their subsequent understanding of texts (Willingham, 2006). It could also be a suitable technique for helping pupils who struggle with traditional approaches to inductive learning such as discovery or problem-based learning, due to the working memory demands of such tasks (Kirschner et al, 2006).

References

Battig, W. F. (1979). The flexibility of human memory. In L. S. Cermak, & F. I. M. Craik (Eds), *Levels of processing and human memory.* Hove, UK: Lawrence Erlbaum.

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition, 41*(3), 392-402.

Borenstein, M., Hedges, L. V., Higgins, J. T., & Rothstein, H. R. (2014). *Comprehensive Meta Analysis Version 3*. Englewood, NJ: Biostat.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition, 42*(3), 481-495.

Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review, 22*(1), 281-288.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 318-339). San Diego, CA: Academic Press.

Education Scotland (2018). Interventions for equity. Retrieved from *https://education.gov.scot/improvement/self-evaluation/ Interventions%20for%20Equity*

Eglington, L. G., & Kang, S. H. (2017). Interleaved presentation benefits science category learning. *Journal of Applied Research in Memory and Cognition*, *6*(4), 475-485.

Guzman-Munoz, F. J. (2017) The advantage of mixing examples in inductive learning: a comparison of three hypotheses. *Educational Psychology, 37*(4), 421-437.
Hausman, H., & Kornell, N. (2014). Mixing topics while studying does not enhance learning. *Journal of Applied Research in memory and Cognition, 3*(3), 153-160.
Higgins, E. J., & Ross, B. H. (2011). Comparisons in category learning: How best to compare for what. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1388–1393). Austin, TX: Cognitive Science Society.
Kang, S. H., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26*(1), 97-103.
Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86.
Kornell, N. & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585-592.
Magill, R. A., & Hall, K. G. (1990). A review of the contextual interference effect in motor skill acquisition. *Human Movement Science*, *9*(3-5), 241-289.
Metcalfe, J., & Xu, J. (2015). People mind wander more during massed than spaced inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(6), 978-984.
Rawson, K. A., Thomas, R. C., & Jacoby, L. L. (2015). The power of examples: illustrative examples enhance conceptual learning of declarative concepts. *Educational Psychology Review, 27*(3), 483-504.
Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology, 107*, 900-908.
Sana, F., Yan, V. X., Kim, J. A., Bjork, E. L., & Bjork, R. A. (2018). Does working memory capacity moderate the interleaving benefit?. *Journal of Applied Research in Memory and Cognition, 7*(3), 361-369.
Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6*(4), 342-353.
Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24*(6), 837-848.
Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition, 39*(5), 750-763.
Willingham, D. T. (2006). How knowledge helps. *American Educator, Spring*, 30-37.
Yan, V. X., Clark, C. M., & Bjork, R. A. (2016). Memory and metamemory considerations in the instruction of human beings revisited: Implications for optimizing online learning. J. C. Horvath, J. Lodge, & J. A. C. Hattie (Eds). *From the laboratory to the classroom: Translating the learning sciences for teachers*. Abingdon, UK: Routledge.