2021

# Towards an evidence-informed differentiated learning consolidation process to support classroom instruction

Nicola Carr-White
*Edith Cowan University*

# Towards an Evidence-Informed Differentiated Learning Consolidation Process to Support Classroom Instruction

This thesis is presented in partial fulfilment of the degree of

**Master of Education by Research**

Nicola Carr-White

Supervisors: Dr. Pauline Roberts and Dr. Gillian Kirk

School of Education,

Edith Cowan University

2021

**Abstract**

Despite many years of teaching experience, the differentiation and consolidation of classroom learning presented challenges for the researcher. In response, a Differentiated Learning Consolidation Process (DLCP) was developed through informal classroom-based action research over several years. Using low cost and accessible resources, it developed into a manageable supplementary intervention to support individual student needs and the retention of classroom instruction. Increasing interest from colleagues led the researcher to provide professional development on the instructional design and implementation of the DLCP. Through this experience, it became apparent that the DLCP theoretical assumptions were largely unknown. The current study was pursued to identify the theoretical components of the DLCP and determine if and how they could be aligned with evidence informed research. A simplified realist review was employed as it provided the opportunity to triangulate theory, the researcher's contextual experience, and the investigation of the DLCP instructional design. The study determined that the DLCP was situated within the field of cognitive psychology, aligning with cognitive load theory and the new theory of disuse. Within the context of the DLCP, spaced practice, retrieval practice, interleaved practice and strategies associated with metacognitive development were investigated to identify maintenance or modification of the instructional design. The findings of this analysis may support teachers to differentiate and consolidate classroom instruction. Additionally, the DLCP may hold potential as an instrument for classroom-based research on variables related to its theoretical constructs.

## Declaration

I certify that this thesis does not, to the best of my knowledge and belief:

i. incorporate without acknowledgement any material previously submitted for a degree or diploma in any institution of higher education;

ii. contain any material previously published or written by another person except where due reference is made in the text of the thesis; or

iii. contain defamatory material.

[redacted]

## Disclosure Statement

Prior to the current study, the researcher developed a hardcover commercial version of the Mastery Learning Folder tool to support the Differentiated Learning Consolidation Process. This version is not essential to the process and teachers can make their own hand-made tools.

## Tables

## Figures

# Towards an Evidence-Informed Differentiated Learning Consolidation Process to Support Classroom Instruction

## Chapter 1

Education deals in the currency of knowledge. Hobbes (1839, p. 7) wrote that "the end of knowledge is power", a reflection that encompasses one of the highest purposes of education. The equalising effects of education have been explored and promoted by the works of Hirsch (2016) who believes that the attainment of knowledge in the cognitive domain, including cultural literacy, gives students regardless of background, equity and agency in their societies (Hirsch, 2003). The current study investigates a process, originally developed in the classroom, to assist teacher management of knowledge acquisition according to individual student needs. Knowledge acquisition through our education system is currently under scrutiny within Australia.

Australian educational standards as reflected by the Program for International Student Assessment (PISA) are of concern according to Donnelly (2019). When compared with 2003, Australian PISA rankings have dropped from 10th in mathematics to 25th, fourth in reading to 16th and sixth in science to 14th (PISA, 2018). Donnelly's (2019) commentary on the 2018 PISA results cites multiple reasons for Australia's performance, including the failure to ensure that teaching and learning programmes are both evidence-informed and relevant to the complexities of the classroom. Whilst national, state-based and school authorities guide the sector, the responsibility for facilitating educational achievement rests predominately with classroom teachers who are guided by seven criteria within the Australian Professional Standards for Teachers (Australian Institute for Teaching and School Leadership, 2011).

The professional standards of most relevance to the current study include

- **Standard 1**: Know students and how they learn, which incorporates differentiation.

- **Standard 2**: Know the content and how to teach it, for example, effective learning strategies.
- **Standard 3**: Plan for and implement effective teaching and learning, which incorporates challenging learning goals and sequenced learning.

Together with the remaining standards, these high expectations are placed upon teachers within a challenging context which includes the time constraints of expanding curriculums and multifaceted demands related to administration, supervision and regulatory requirements amongst other expectations (Dinham, 2014).

The need for teachers to address widening learning gaps of up to six years within Australian classrooms testifies to the difficulties associated with differentiating learning (Goss et al., 2015; McNamara & Moreton, 1997). Despite over 25 years of experience, the management and attainment of individualised student learning was a challenge for the researcher and is recognised as such by many teachers (McNamara & Moreton, 1997; Tomlinson, 2015).

Learning differentiation is a strategy based on the belief that the differences in student achievement “are significant enough to make a major impact on what students need to learn, the pace at which they need to learn it, and the support they need from teachers and others to learn it well” (Tomlinson, 2000, p. 6). Learning it well, or consolidated learning, is defined in this thesis as knowledge held in long-term memory (Kirschner et al., 2006). The goal of differentiation and consolidation is therefore, learning that is durable. This standard of learning was not being achieved in the researcher’s classroom.

The lesson environment offered neither time nor support to satisfactorily address individualised student progress. Lacking, was a manageable process based on the content of classroom instruction but delivered outside of the restrictive lesson context. In response, the researcher began reflecting on her experiences and that of others, to remember what had worked and what had not, why, and what might be applied within the classroom context.

Reflection led to an informal action research journey to respond to the challenge of managing and attaining individualised student progress. The resulting intervention evolved within the classroom between 2010 and 2016. It consisted of human resources, a tool and a combination of learning strategies to facilitate individualised practice of the content of classroom instruction. Within the broader intervention, the process is described as the Differentiated Learning Consolidation Process (DLCP). Over time, the researcher experienced improved management of learning, and students of different ability levels demonstrated self-paced progress.

In discussions with colleagues, it became evident that beyond some fundamental understandings of Bloom's revised taxonomy (Anderson & Krathwohl, 2001; Bloom et al., 1956) and mastery learning (Bloom, 1968) the intervention had developed in the absence of theoretical understandings. If the intervention were to assist other teachers to facilitate student progress, the rudimentary DLCP needed to be adjudicated and revised by evidence-informed research, beginning with the identification of the theoretical foundation. Essentially, 'It works in practice, but does it work in theory?' This led to the current Master of Education thesis which is arranged in two sections. Section 1: Theory Identification, describes the researcher's classroom experience, annotates the DLCP and introduces a potential theoretical foundation. Section 2: Theory Application, seeks to apply evidence-informed research to the DLCP in the form of process maintenance or modification.

## Section 1: Theory Identification

Section 1 lays the foundation for the remainder of the thesis. The research design was based on a simplified realist synthesis logic of enquiry (Pawson, et al., 2004) which necessitated non-traditional thesis organisation. The methodology is therefore presented in Chapter 3 to explain and justify this departure.

Results data are presented across multiple chapters, including

- an autoethnographical account of the researcher's classroom experience (Chapter 2),
- the annotation of the original DLCP to provide a baseline with which to compare new data (Chapter 4), and
- the identification of the DLCP theoretical foundation (Chapter 5).

Following theory identification, Section 2: Theory Application, investigates research potentially applicable to the DLCP: the cognitive psychology learning strategies of spaced retrieval practice (Chapter 6), interleaved practice (Chapter 7) and strategies of metacognitive development (Chapter 8). Results in these chapters are presented in terms of maintenance or modification of the DLCP. Chapter 9 summarises and concludes the thesis.

# Chapter 2: The Researcher's Experience

The current study was motivated by a desire to understand and potentially improve a practice process developed in the classroom. This chapter commences with a first-person autoethnographic account deemed appropriate to share the researcher's personal journey in addressing the challenges of differentiating and consolidating student learning. A positionality statement follows, highlighting the potential for bias in researching a self-developed intervention and the ways in which the researcher has attempted to address this issue.

## Consolidation Challenges

Prior to the development of the DLCP, my most recent professional experience had been coordinating and teaching in a district level gifted and talented programme where the emphasis was on higher order knowledge and thinking skills. In my roles as programme coordinator, teacher and trainer, I received and delivered professional development in the 'analysing', 'evaluating' and 'creating' learning objectives of Bloom's revised taxonomy (Anderson & Krathwohl, 2001; Bloom et al., 1956). The anecdotally described 'lower order' thinking skills of remembering, understanding and applying, were not addressed in my professional development, received or delivered, and I perceived these skills as having far less importance to learning.

In 2009, I returned to the mainstream classroom and was allocated a composite Kindergarten/Pre-primary class (4 – 6-year-olds), followed by a Year 1 class (6 – 7-year-olds) in 2010. Having most recently taught the Year 5 – 7 extension programme (10 – 12-year-olds), addressing the learning needs of younger students was a considerable change. At that time, I would describe my pedagogy as eclectic and pragmatic based on my teacher training, professional development and the education system milieu. Like many of my colleagues, my approach included a mix of traditional and experiential learning, for example, explicit instruction followed by 'hands on' learning activities. Within my new teaching role, there were situations that raised concerns which ultimately led to my reflections on managing differentiation and consolidation.

One experience was related to teaching the Year 1 mental mathematics learning objective of number bonds to ten (2 + 8, 6 + 4 etcetera). Classroom activities included instruction, rotating group work, games, worksheets, software activities and drilling over the course of a week. Frequent class and individual questioning during these learning sessions suggested that students were both understanding and remembering the various number bond combinations. Unfortunately, this impression was recalibrated when, three weeks later, an end-of-term assessment revealed that less than 30% of the students had instant recall of the number bonds. I was confused by the results as most students had recalled them proficiently during the lessons. Adding to my concern, automaticity with number partitioning was required for the forthcoming mental mathematics curriculum and I was not sure how to facilitate students' mastery of the failed objective. Contrasting my experience in teaching the higher order skills, the learning objective of 'remembering' (Anderson & Krathwohl, 2001) was not as straight-forward as I had anticipated.

**Differentiation Challenges**

A second pivotal experience leading to the development of the intervention was the opportunity to follow the pre-primary students from the Kindergarten/Pre-primary composite class into Year 1. These students had received instruction and assessment on synthetic phonics throughout their pre-primary year. The end of year performance indicators had shown that most students were proficient with their phonics sounds and I had expected to begin the Year 1 instruction near to where I had left off the year before. Re-assessment at the start of the new year, however, indicated that many students had forgotten some essential learning over the summer holidays, creating a broad spectrum of individual mastery. Had I not followed these students through to the following academic year, I would have been unaware of the substantial degree and diversity of forgetting that had occurred. The Year 1 students needed mastery of the foundational phonics material to continue the development of their early literacy skills. The scope of remediation required during the first term of the new year motivated me to find different ways to address individual student learning needs.

### Seeking a Solution

In response to the challenges of differentiation and consolidation, I began to question how I could address these needs and ensure that the learning would be durable. I reflected upon my years of teaching for any tools or strategies that particularly addressed these concerns. I recalled a tool that I had observed, but not used, many years before called a mastery learning folder. I was aware that teachers had used this tool to assist students to remember phonics sounds and that they had described it as effective. The tool was a manila file which contained a series of library card pockets (Appendix A). I knew that the process involved placing phonics sound flashcards into a 'Start' pocket and that the flashcards moved forward through the pockets when they tested correctly, or back to the start when they were incorrect, however, beyond this, I did not know how the folders were used, nor were any teachers in my school still using the strategy. In 2010, I began an informal action research project to ascertain if the mastery learning folder tool could assist my Year 1 students to remember classroom instruction. Over several years, through trial and error, the process evolved.

### A Hardcover Tool

By 2013, the intervention was successfully supporting my classroom instruction and providing me with a manageable process to address the previous differentiation and consolidation concerns, however, one issue remained; the handmade tool was not fit for purpose.

Initially, the DLCP was a homework strategy and the cardboard folder went back and forth from home to school each week day. This intensity of use meant that students required a minimum of two or three handmade folders each year which involved ongoing teacher and/or education assistant time. In response, I had a small quantity of hardcover folders manufactured and this resource increased the time efficiency of the intervention substantially. During 2014, I began sharing the strategy with interested schools and teachers. Some used their own handmade folders, others chose to purchase the more durable version I had created.

**Motivation to Study**

Whilst sharing with colleagues, I increasingly realised that, beyond some fundamental understandings of Bloom's revised taxonomy (Anderson & Krathwohl, 2001; Bloom et al., 1956) and mastery learning (Bloom, 1968), I did not fully understand why the DLCP produced the observed results. I wondered if the process was a form of rote learning. To simply state that the intervention was achieving its objectives was insufficient and unlikely to benefit teachers and students outside of my sphere of influence. Having a science background, I began to wonder about the theoretical foundations of the DLCP, and whether it could be improved beyond my informal research. At the end of 2016, I made the decision to begin a research Master of Education degree to investigate the DLCP.

**Positionality Statement**

The researcher was aware of the potential bias, limitations and positionality of researching a self-developed intervention, particularly in the context of commercial interest. However, as an experienced classroom teacher who had faced difficulties managing differentiation and consolidation, the researcher felt that further investigation may facilitate an alignment of the process with evidence-informed research and deemed it a worthwhile project which may improve future student outcomes. Additionally, there was no necessity for teachers to use the commercial hardcover tool as they could create the hand-made version. An appropriate research design was sought to address this context.

The important factor of positionality is addressed in multiple sections of the thesis, and initially here, in addressing the goals of the research. The first goal of the research was explanatory. This included (a) the mapping of the intervention, (b) the defining of operational definitions and (c) through research literature, identification of the DLCP theoretical assumptions. The second goal was to synthesise research evidence on the identified theories to adjudicate and potentially revise the DLCP according to the theoretical evidence.

A measure of triangulation was afforded by (a) the researcher's anecdotal experiences, (b) a research focus on theory and (c) a systematised literature review process that sought to be transparent and replicable (Grant & Booth, 2009). The juxtaposition of the researcher's experience and theory held the potential to "support fresh thinking to revise policy and launch it in new circumstances" (Pawson, 2006, p. 2). It is hoped that the theoretical foundation resulting from the Master of Education will form the basis of future empirical investigations.

# Chapter 3: Methodology

The researcher's challenge to differentiate and consolidate learning according to the observation and assessment of classroom instruction led to an informal action research project based upon a traditional mastery learning folder tool (Appendix A). An intervention emerged through the development of a process (the DLCP), an appropriate tool and accessible resources. The goal of the DLCP was to facilitate the transfer of classroom instruction to students' long-term memory, and over time, the researcher observed positive results. The current study sought to identify the causal factors of the undocumented process, commencing with the identification of the theoretical foundation followed by an assessment of the DLCP against selected theories for potential revisions.

The methodology chapter is positioned near the start of the thesis to explain the features of the realist synthesis logic of enquiry (Pawson et al., 2004) which guided the study and thesis organisation. Methodological searching revealed academic papers based on the approach, but not theses. Consequently, discussions with supervisors and academic writing consultants, assisted in clarifying the arrangement and content of chapters.

The methodology chapter begins with the researcher's theoretical perspectives. This is followed by a description of the challenges involved in identifying an appropriate research design that would theoretically adjudicate and extend the previous informal action research. Justifications for the use of the realist synthesis approach, its limitations and necessary accommodations are then explained. The methods including autoethnography (the researcher's experience), systematised literature review (spaced retrieval and interleaved practice) and narrative literature review (theoretical foundation and metacognitive development strategies) and are then discussed. The central place of theory within the project is clarified, followed by a description of the scope adaptations appropriate to solo postgraduate study. The guiding elements of the realist synthesis logic of enquiry are then elaborated.

**Theoretical Perspectives**

Proceeding from the researcher’s previous informal action research, the ontology rests on a pragmatic foundation which is known as the paradigm concerned with applications, in this case, within the context of the classroom (Cresswell, 1998). Research design decisions were based upon a methodological orientation of realism and the principle of “fitness for purpose” (Cohen et al., 2011, p. 1). Realism can be described as a logic of inquiry in pursuit of what works for whom and in what circumstances (Pawson et al., 2004). Realism and pragmatism accommodate the use of predominantly quantitative research literature, and the interpretative nature of qualitative synthesis.

Realist investigations are associated with theory-driven methodologies which seek to make explicit the assumptions of how interventions work by identifying programme theory, or “mechanisms-of-action” (Pawson et al., 2004, p. 3). Methods and results can be diverse with research questions focussing the collection of requisite data (Bell, 2005; Cohen et al., 2011). Data in this study includes the researcher’s classroom experience, the intervention presented as a logic model, the identification of theoretical assumptions and the application of these theories to the DLCP.

Foundational to this data collection was the challenge experienced by the researcher to facilitate the durable learning of the content of classroom instruction. Kirschner et al. (2006, p. 76) state that “if nothing has changed in long-term memory, nothing has been learned”. Essential to this task is an understanding of each student’s prior knowledge. To build upon existing schema, learning objectives need to be appropriately differentiated; this was the second challenge experienced by the researcher. The study is therefore situated within a cognitivist theoretical framework which “focuses on understanding human perception, thought, and memory [portraying] learners as active processors of information [and assigning] critical roles to the knowledge and perspective that students bring to their learning" (Bruning et al., 1999, p. 2). Further defined, the current study sits within the field of cognitive science and cognitive psychology, which includes behaviours such as "perceiving, attending, remembering, thinking and decision making' (Agarwal & Bain, 2019, p. 19).

## Research Design

### *Selection*

In seeking assistance to find a research design, the researcher was frequently advised of the unusual nature of the project. The development of interventions is often the result of applying research, however, the research problem was a functioning intervention without a fully identified theoretical foundation. Consequently, the search for an appropriate research design was extensive.

The research goals informed the search which began with an exploration of the methods of document analysis (based on the tool), autoethnography (based on the researcher's experience) and theory-driven evaluation to facilitate theoretically based validity (Chen, 1990, 2012). Drawing on the work of Booth et al. (2016) these diverse methods were followed by an exploration of various types of literature review to facilitate a secondary research synthesis. These included an integrative review (Cooper, 1984), rapid review, the constructive research approach (Lehtiranta et al., 2016), meta-framework synthesis (A. Booth, personal communication, August 1, 2018) and realist synthesis (Pawson, 2006). A realist synthesis was assessed in depth and selected.

Pawson and Tilly (1997) first introduced the concept of realist evaluation, with a later derivation called realist synthesis (Pawson et al., 2004). It is described by Wong et al. (2013) as an increasingly popular theory-driven approach. A realist synthesis consists of a "review and synthesis, which focuses on understanding the mechanisms by which an intervention works (or not) … identifying underlying causal mechanisms and exploring how they work under what conditions” (Rycroft-Malone et al., 2012, p. 1). It is described by Pawson et al. (2004) as a *logic of enquiry*, rather than a research design and, although frequently applied in health contexts, it appeared well suited to the goals of researching an educational intervention. Additionally, it was thought that a focus on theory may help to counter potential bias.

According to Pawson (2006), theory building is of central importance in a realist synthesis. An intervention is considered a theory because interventions are “implemented on a hypothesis of if we do X in this way, then it will bring about an outcome” (Rycroft-Malone et al., 2012, p. 3). In constructing theory, “concepts are captured; links are explored, created and tested; ideas are documented and systematically reworked, in textual memos, models, and diagrams” (Denzin & Lincoln, 1994, p. 447). Additionally, Booth and Carroll (2015) propose that theory assists with the collection, organisation, analysis and evaluation of improvement programmes. These undertakings were consistent with the goal of identifying, comparing and revising theoretical elements and their application within the DLCP.

Related goals of realist synthesis are to address not only theory, but also context and outcomes to provide explanations rather than judgements (Rycroft-Malone et al., 2012).

> The purpose is to articulate underlying programme theories and then to interrogate the existing evidence to find out whether and where these theories are pertinent and productive. Primary research is examined for its contribution to the developing theory. The overall intention is to create an abstract model of how and why programmes work, which then can be used to provide advice on the implementation and targeting of any novel incarnation of the intervention (Pawson, 2006, p. 3).

As previously described in Chapter 2, it is the classroom—and specifically the lesson context—that primarily creates the management difficulties associated with the differentiation and consolidation of student learning. Therefore, realist synthesis is an "intuitively appealing approach to those trying to expose and unpack the complexities of contexts and interrelated mechanisms underlying implementation activity" (Rycroft-Malone et al., 2012, p. 2). The realist synthesis approach provided the opportunity to juxtapose the researcher's classroom experience, theory, the intervention as a whole and the DLCP, to identify and refine selected mechanisms according to evidenced-informed research.

An overview of realist synthesis logic of enquiry (Pawson et al., 2004) is displayed in Table 1. Given the multitude of contextual variations, Pawson et al. (2004) caution against prescriptive use of these guidelines. The current study represents one interpretation within the described constraints.

Table 1
*The Realist Synthesis Logic of Enquiry*

| Stages and Action | Activity |
|---|---|
| 1. Define the scope | |
| (a) Identify the question | What are the objectives of the intervention?<br>What is the nature and content of the intervention?<br>What are the nature and form of its outcomes or impacts?<br>What are the circumstances or context of its use?<br>What were the initial theoretical assumptions? |
| (b) Clarify the purpose(s) of the review | Selected purpose: Theory integrity<br>Could the intervention work according to the identified theory? |
| (c) Find and articulate the programme theories | Search for relevant theories in the literature.<br>Annotate and categorise identified programme theories.<br>Develop a theoretically-based evaluation strategy.<br>Design a bespoke data extraction form. |
| 2. Search for and appraise the evidence | |
| (a) Search for evidence | Decide and define purposive sampling strategy.<br>Define search sources, terms, methods and limits.<br>Set the thresholds for search saturation. |
| (b) Test of relevance | Relevance: Does the research address the operational definition of the learning strategy within the DLCP?<br>Rigour: Is the selected research conducted by leaders in the field? |
| 3. Extract and synthesise findings | |
| (a) Extract the results | Seek confirmatory and contradictory findings.<br>Extract theory data for comparison with the DLCP. |
| (b) Synthesise findings | Use findings to address the review purposes.<br>Refine DLCP programme theories. |
| 4. Develop narrative | Use expert framing of the DLCP and classroom context in the review, application and discussion of findings. |

*Note.* Adapted from “Realist Synthesis: Illustrating the Method for Implementation Research,” By J. Rycroft-Malone, B. McCormack, A. M. Hutchinson, K. DeCorby, T. K. Bucknall, B. Kent, A. Schultz, E. Snelgrove-Clarke, C.B. Stetler, M. Titler, L. Wallin, and V. Wilson, 2012, *Implementation Science*, 7(1), p. 3 (https://doi.org.10.1186/1748-5908-7-33)

### ***Limitations***

Realist syntheses typically investigate and synthesise evidence on complex interventions sitting within complex social systems, in the present case, a school environment (Pawson et al., 2004). Interventions are ‘chains’ that include multiple components, mechanisms-of-action, stakeholders, participants and theories, all of which influence each other and the whole. At all points in the chain, the participants, theories and mechanisms are fallible. An intervention’s programme theory is described by Pawson et al. (2004) as a theory of theories, therefore, the potential scope of such research is vast.

Investigating an intervention brings together an extensive selection of research literature comparable to a systematic review (Booth et al., 2016) and it is usually conducted by a team of experienced researchers. Even in this context, limitations are essential in terms of scope, processes, theories, settings, the nature and quality of information and the applicability of recommendations (Pawson et al., 2004). For these reasons and others that follow, the current study is a *simplified* realist synthesis, a logic of enquiry used to guide and structure the investigation within scope, timeline and researcher experience limitations. Using the work of Pawson et al. (2004), described below are the ways in which the realist synthesis approach was ‘fit for purpose’ for the current study.

#### **Study Data.**

De Bruyckere (2018) identifies that many studies related to memory may be consistent in the laboratory, without translating to the complex setting of the classroom. Some studies included in the thesis are not classroom-based and therefore it is acknowledged that the applications of laboratory studies—and those with adult populations—will require further empirical research in an authentic classroom setting.

**Justification.**

Realist synthesis ascribes value to personal experience and expertise (Pawson et al., 2004). Although a novice in post-graduate study, the researcher has spent several years developing and using the intervention and over 25 years as a teacher in the primary school setting. Additionally, the researcher volunteers to assist schools as they navigate on-going DLCP implementation within their specific contexts. The realist synthesis logic of enquiry recognises and gives a voice to this experience. Stakeholder participation and expert framing are essential components of realist synthesis, roles to which the researcher may contribute through

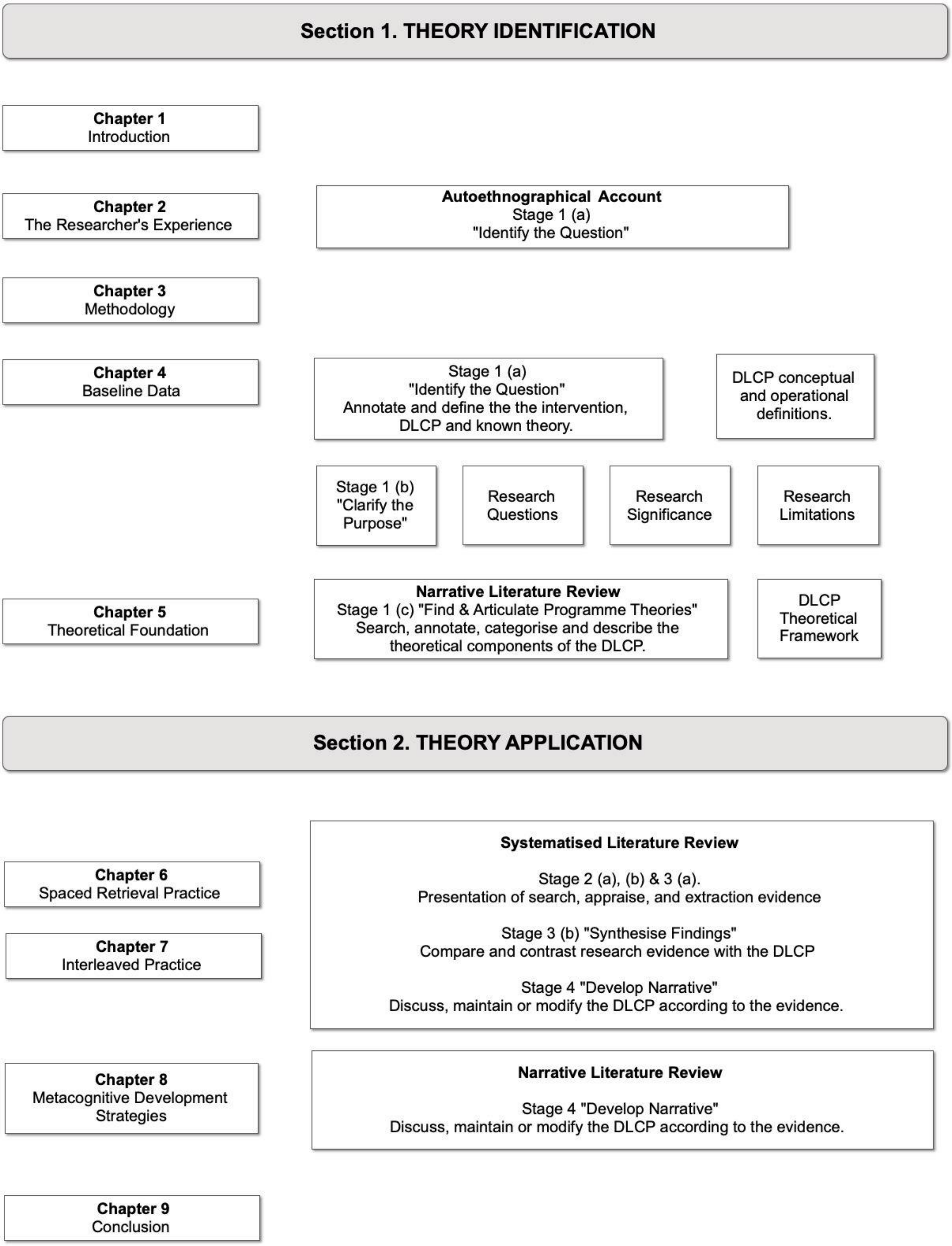
- the lived experience (Ch. 2 The Researcher's Experience),
- understanding of the intervention and some theoretical aspects (Ch. 4 Baseline Data),
- the potential to adjudicate between theories and potential key mechanisms (Ch. 5 Theoretical Foundation), which combine to contribute to
- decisions on the applicability of research study findings to the DLCP within the classroom context (Section 2 Theory Application).

The complete thesis chapter organisation is displayed in Figure 1.

The approach prioritises the recognition of context, an essential recurring theme of this thesis. An educational intervention must be able to recognise and adapt to the complexities of the classroom. This includes the ability to address student learning needs with minimal physical, financial and human resources. These factors underlie the analysis of the intervention and, more specifically, the DLCP.

Realist synthesis facilitated the annotation the intervention in its entirety, a process that resulted in many ideas for revision beyond the scope of the current study. It provided the means to fill a theoretical knowledge gap and to address the research priority, the identification and refinement of selected theories within the DLCP. It is hoped that the approach may facilitate a potentially evidence-informed option for educators seeking a manageable process to support the differentiation and consolidation of classroom instruction. The realist synthesis approach is non-prescriptive and recognises a variety of methods, procedures and data (Pawson et al., 2004).

*Figure 1*. The simplified realist synthesis approach and thesis organisation.



#### ***Accommodations.***

To accommodate timelines and scope, a systematised literature review was selected for the refinement of the DLCP by spaced retrieval and interleaving theory. Grant and Booth (2009, p. 102) describe the approach as one seeking to "include one or more elements of the systematic review process". A systematised literature review has been described as typically conducted by a solo postgraduate student when time constraints limit the comprehensiveness of the review process. In the current study, this involved purposive sampling of studies judged by the researcher as highly relevant to the DLCP. A narrative review was used to identify established metacognitive theory that was potentially applicable to the DLCP.

The selected methods required expert framing in making data judgements of relevance and applicability. Pawson et al. (2004) promote validity in the form of clarity on the part of the researcher in explaining selections and reasoning. It is acknowledged, therefore, that the findings of realist syntheses are fallible. Reader judgement may follow in the form of refutation. Such feedback is encouraged as "exposure to scrutiny and critique is thus the engine for the revision and refinement of programme theories" (Pawson et al., 2004, p. 38).

## Method

The goal of realist synthesis Stage 1 was to define the scope of the investigation. Stage 1 (a) included the researcher's autoethnographical account and the annotation of the intervention's components.

### *Stage 1 Define the Scope*

#### Stage 1 (a) Identify the Question.

Realist synthesis begins by "identifying its subject matter [and] the construction of an embryonic theory of how [it] may work" (Pawson, 2006, p. 3). Foundational to this purpose was the description of the researcher's management issues with the differentiation and consolidation of student learning found in Chapter 2. A personal narrative, known as an autoethnography, was selected as the method to recount the classroom experience.

Autoethnography begins with life experiences and seeks to relate them to a context or culture (Denzin & Lincoln, 2011). In this study, it provided the opportunity for the researcher to reflect upon and understand the self as a teacher, whose educational goals for her students were not being satisfactorily realised. The autoethnography facilitated the description of the experience within the complex context of a primary school classroom. It was through these experiences that the current study emerged. Throughout the research journey, the researcher as *teacher,* sought to continually assess the relevance and the applicability of solutions to the classroom context.

To address issues of teacher management, it was essential to situate the DLCP research focus within the context of the broader intervention. A logic model was employed to fulfil this purpose. Logic models are descriptive tools that pictorially represent the systematic thinking behind an intervention (W.K. Kellogg Foundation, 1998). The relevant terms of reference include intervention, stakeholders, programme theory, instructional design, process, learning strategies and mechanisms. Within the thesis, these terms are defined as follows:

- **Intervention**: the broadest conception of the programme implementation within the classroom. It includes pedagogy problems and needs, resources and inputs, theory, DLCP instructional design, as well as the anecdotal outcomes and impact.
- **Stakeholders**: the school administration, teachers, education assistants, tutors, parents and students associated with the intervention. Additionally, experts in the theoretical constructs utilised in the process.
- **Programme theory**: the underlying assumptions of how the intervention is understood to work (Pawson et al., 2004). Within this study, it refers specifically to the theoretical components of the DLCP.

- **Instructional design**: a description of the DLCP components, how they are related and expressed within the process.
- **Mechanisms**: the conceptual and operational definitions of the components (theory) by which the DLCP seeks to achieve student outcomes.

The logic model provides a framework for baseline data discussion in Chapter 4 and assists with further discussions throughout the thesis. The data from Stage 1 (a) informed the research purpose of Stage 1 (b).

**Stage 1 (b) Clarify the Purposes of the Review.**

Pawson (2006, p. 25) describes three different potential purposes of realist synthesis: (a) “to question programme theory integrity”, (b) “to adjudicate between rival programme theories” and / or (c) “to consider the same theory in comparative settings”. The purpose of this investigation relates to (a) which was to

- explore the integrity of the DLCP theory,
- identify inconsistencies with established theory and
- facilitate process maintenance or modification of selected key learning strategies.

Exploring the integrity of the DLCP began with a narrative literature review to identify the potential theoretical foundation of the process (Booth et al., 2016).

**Stage 1 (c) Find and Articulate the Programme Theories.**

Conceptual and operational definitions of the DLCP components were identified. Conceptual definitions describe the general recognisable characteristics, while operational definitions identify measurable outcomes in the related source data (Cooper, 1998). The literature search commenced with broad definitions of the DLCP components to “err on the overly inclusive side” (Cooper, 1998, p. 26), allowing relevance to be revealed over time. The identified definitions were expanded according to the availability of studies and the definitions used within them. Controlled vocabulary was sourced from the subject terms used in relevant literature. During the initial theory investigations, two types of data were sought: the programme theory underlying each DLCP component and the abstraction of these theories into related theoretical categories.

Bartholomew et al. (1998) suggest three classifications when searching for programme theory: issue, concept and general theory. Theory searching based on *issues*, situates the learning strategy within its problem setting, identifying theories seeking solutions. Searching for *concept* related theory presumes some knowledge of learning strategy effects and may reveal related evidence. Searching for *general theory* conceptually expands the view and may identify overarching theory constructs for multiple learning strategies, a goal of realist synthesis.

The search for theory and the abstraction of theory into categories was initially trialed for one identified DLCP learning strategy, later followed by the remaining components. The overall DLCP programme theory was reworked iteratively over the course of the research and is displayed as a DLCP theoretical model in Chapter 5: Theoretical Foundation. Subsequently, specific theories and key mechanisms were selected for closer examination in Section 2: Theory Application.

***Theoretical Focus.***

Realist synthesis suggests prioritising a mid-range theory to narrow the scope of the research and focus the evidence synthesis (Pawson, 2006). In making this decision, Pawson et al. (2004, p. 16) recommend expert framing of the “hunches, the expectations, the rationales and rationalisations for why the intervention might work”, a role fulfilled by the researcher. Also of relevance, were the theories that could contribute to maintenance or modification decisions.

Cognitive load theory (Sweller et al., 2019) and the new theory of disuse (Bjork & Bjork, 1992), and the desirable difficulty framework (Bjork & Bjork, 2011), emerged as key theoretical constructs. The scope was reduced by the selection of the relevant desirable difficulty strategies of spaced retrieval practice (the spacing and testing effects) and interleaved practice (interleaving theory) for focussed DLCP programme theory investigation and refinement. Factors related to student metacognition emerged through the study of these mechanisms. A narrative review of the associated metacognitive theories and development strategies was then included.

### *Stage 2. Search for and Appraise the Evidence*

Operational definitions are described as *units of analysis* (Palmberger & Gingrich, 2014). For spaced retrieval practice and interleaved practice, the units of analysis facilitated data collection through a systematised literature review. The review sought to explore the theories associated with each strategy, and identify a theoretical consensus, if one emerged.

#### Search for the Evidence.

The search for data within the systematised literature review consisted of purposive sampling for the evidence-informed operation of each learning strategy. An initial search trial, based on interleaving theory, was conducted on multiple databases to inform future search procedures. To conclude database searching, Scopus was used with similar search terms.

The Edith Cowan University library database describes Scopus as a large abstract and citation database of multidisciplinary peer-reviewed literature. The *Analyse Search Results* function provided a list of prominent authors, journals and articles which concurred with the list gathered through the previous, much longer search procedure. With a broad range of topics for investigation and limited time, Scopus was selected as the initial search procedure (for an example see Appendix B).

As recommended for realist synthesis, the search focused on primary research articles (Wong et al., 2013), particularly seminal papers or those based upon student populations in recent publications. Synthesised research in review articles and textbooks were also explored. Database alerts (Google Scholar and journal-based) were used to identify newly released research.

Citation searching proceeded article searching in the form of pearling and snowballing. Pearling is the identification of a "highly relevant article (the 'pearl') to identify terms … on which a search can subsequently be based" (Booth et al., 2016, p. 115). Snowballing involves the use of a relevant article "as a starting point for either working back from its references or for conducting additional citation searches" (Booth et al., 2016, p. 315). Papers containing potential evidence data for each learning strategy were tallied in flow charts (Appendix C displays the format). Articles identified through the third screenings were then summarised in tables (a section is displayed in Appendix D). Purposive sampling followed being facilitated by the research questions, operational definitions of each theory and researcher judgement (expert framing) to identify the most relevant articles for further data extraction.

**Test of Relevance, Rigour and Saturation.**

Quality appraisal of research articles was first assessed by relevance. Primary research needed to be based upon the relevant theory and 'fit for purpose' within the DLCP. Second, to ensure rigour, the prominent researchers in each learning strategy area were identified, assisted by the Scopus *Analyse Search Results* (Appendix B) function. For each author this included an assessment of

- the quantity of published articles,
- their appearance in prominent publications,
- their presence with other noted authors on articles, and
- citations and references to these researchers in books and academic textbooks.

For each learning strategy, searching finished when no new theories emerged. New studies by prominent authors were also explored across the duration of the study.

### ***Stages 3 and 4. Extract, Synthesise and Discuss Findings***

#### **Extract the Results.**

The purpose of Stage 3 and 4 of the investigation was to investigate and discuss the integrity of selected DLCP learning strategies in relation to established theories to inform potential revisions. Data was assessed through operational definitions, the expression of the strategy within the DLCP and researcher judgement. For each strategy, evidence data was identified from the articles in the third screening. To organise and annotate the primary source materials, the findings of relevance were tabulated on data extraction forms (Appendix E displays an example).

#### **Synthesise and Discuss Findings.**

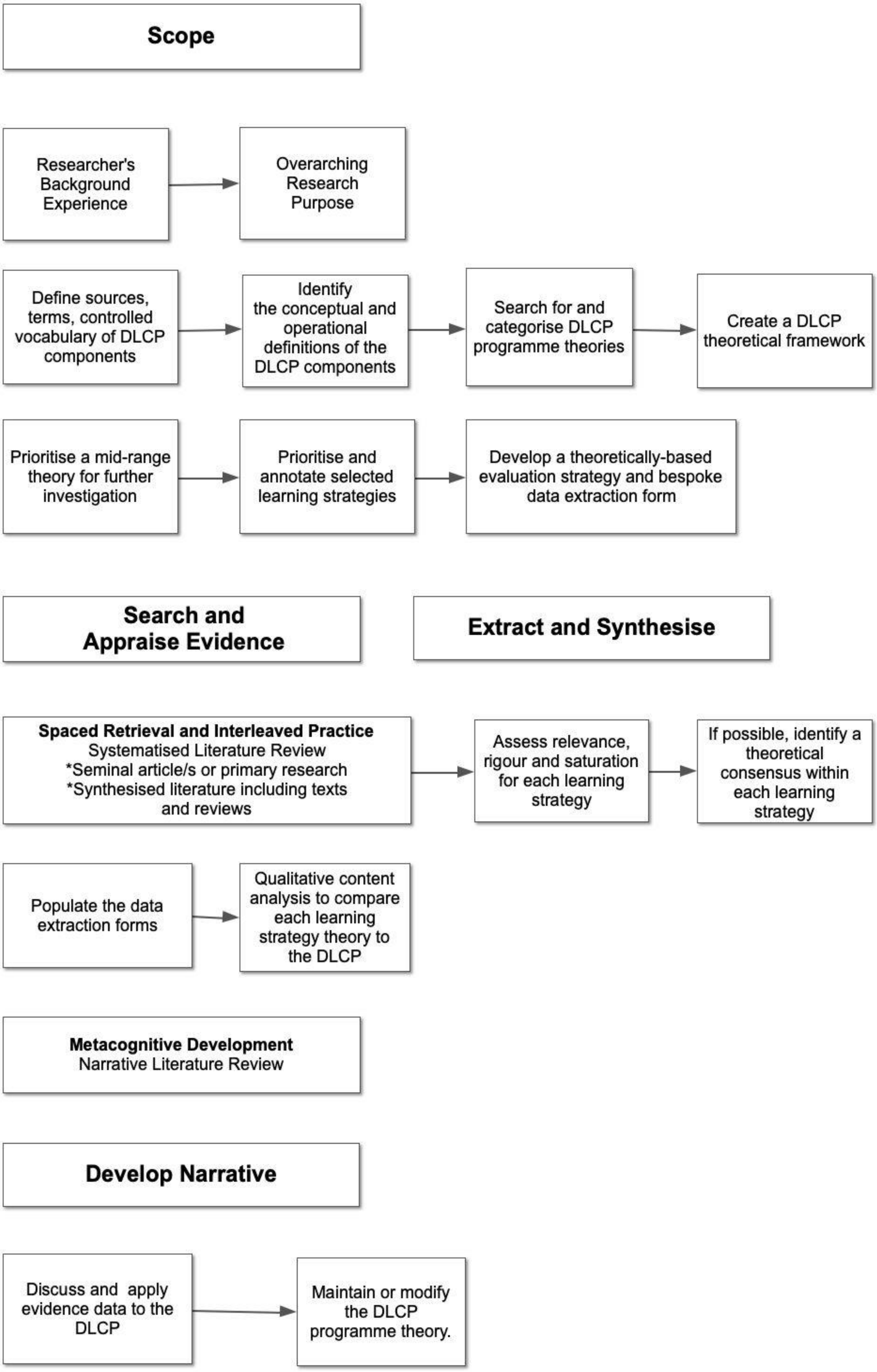
A realist review interprets synthesis as the refining of programme theory (Pawson et al., 2004) to further progress understanding of the mechanisms of action within the area of focus. For each learning strategy, the goal was to discover its defining characteristics, and the conditions under which it has been found to effect learning outcomes. To achieve this objective, the method of qualitative content analysis was employed.

Hsieh and Shannon (2005) classify the method into three approaches: conventional, direct and summative. All approaches are used to “interpret meaning from the content of text data”, however, for this study, the directed approach was most applicable as the analysis was to be guided by theory (Hsieh & Shannon, 2005, p. 1277). Hsieh and Shannon (2005) describe the method as a “subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes and patterns” (p. 1278). The focus was a comparative content analysis to assess the similarities and differences between the theory evidence data and the expression of operational definitions for each learning strategy. This type of analysis is “well equipped to tackle questions that require complex and combinatorial explanations” (Palmberger & Gingrich, 2013, p. 3).

The intervention brings together a novel combination of teaching and learning issues, resources and instructional design, therefore, subjective, but transparent and accountable abstractions were required. To ensure relevance, realist synthesis suggests expert framing through input from practitioners, a role fulfilled by the researcher. With the underlying programme theory identified and the alignment of theory to operational definitions determined, contextual judgements informed DLCP maintenance or modification. Validity was sought through the discussion of studies in sufficient depth to allow reader assessment of the reasoning behind revision decisions (Pawson et al., 2004). The synthesis results and discussion are presented in Section 2, Ch. 6 Spaced Retrieval Practice, Ch. 7 Interleaved Practice and Ch. 8 Metacognitive Development Strategies.

In conclusion, Figure 2 summarises and contextualises the realist synthesis approach within the thesis. Whilst the flow chart appears linear, the process is highly iterative within and between stages (Pawson et al., 2004). The forthcoming chapter annotates the baseline data of the original intervention and DLCP for later comparison with the results of synthesised studies in Section 2.

*Figure 2.* Methodological summary of the simplified realist synthesis.



# Chapter 4: Baseline Data

The overarching objective of the intervention, and the DLCP within it, was to provide teachers with a manageable strategy to address the differentiation and consolidation of the content of classroom instruction. The learning goal was to facilitate individualised student progress along established learning progressions. The classroom development of the intervention used an informal action research approach over several years to refine the strategy, ultimately assisting the researcher to address these objectives.

According to Pawson et al. (2004), Stage 1 of realist synthesis broadly defines the scope of the review through (a) the description of the intervention within its context which leads to (b) the clarification of the review's purpose.

Stage 1(a) includes the

- nature and content of the intervention,
- circumstances and context,
- available resources,
- components of the DLCP instructional design and
- anecdotal observations of the outcomes and impacts.

These results provide a baseline for Stage 1 (c) the identification of theory, and the analysis and application of synthesis data in Stages 2- 4.

Within the current study, research on the traditional mastery learning folder approach yielded virtually no information, however, a document from a Western Australian education department psychologist describing a version of the process was discovered (Appendix F). Informally, colleagues described their own variations of the procedure described. Early in the research, an instructional design of a similar nature was identified.

German journalist Sebastian Leitner, popularised science through his writings during the 1970s. His book *So Lernt Man Lernen* (How to Learn to Learn) was written in German and, unfortunately, an English translation was unable to be sourced. The work described a spaced retrieval system referred to as a Leitner box ("Leitner system," 2020). Given the similarities, it may have been the inspiration for the traditional mastery learning folder approach. Later in the thesis, the revised DLCP will be compared with these approaches. Information on the traditional mastery learning folder and Leitner box were not available during classroom development —the current chapter will explain how the DLCP evolved differently.

As an introduction to the intervention, the original DLCP is demonstrated through an animation accessed via the private video link[1] below and supported by Figure 3. Supplementary information includes background information on DLCP development from 2010 to 2016 (Appendix G), and the original DLCP instructions (Appendix H).

**Stage 1 (a): The Nature and Content of the Intervention**

Realist synthesis begins with the annotation of how an intervention works (Pawson, 2006, p. 3). Interventions consist of multiple mechanisms of action which depend on the cumulative success of the complete sequence (Pawson et al., 2004). In describing all facets of an intervention, logic models can assist those involved to plan, evaluate and revise programmes (W.K. Kellogg Foundation, 1998). The creation of a logic model on the intervention (Figure 4) increased the researcher's knowledge and made explicit the "flows, blockages and potential points of contention" (Pawson et al., 2004, p.3). Sectional views are provided throughout the chapter to assist with the discussion. Whilst the focus of the current

[1] Original DLCP Demonstration: https://youtu.be/wReNJRuYemE

study is the DLCP in a school environment, it should be noted that the process may not be limited by age or context.

*Figure 3.* A schematic of the original DLCP (2016).

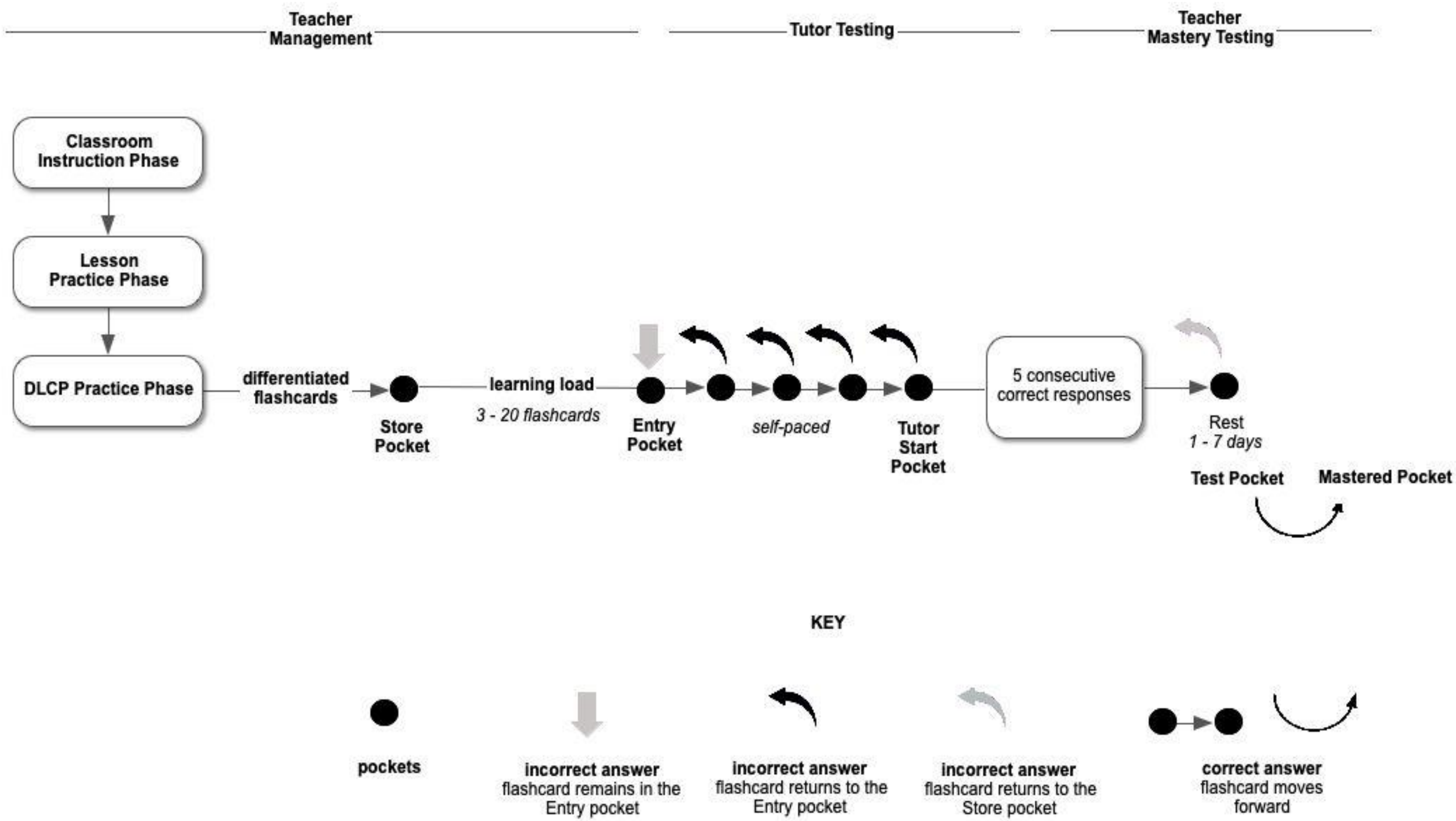
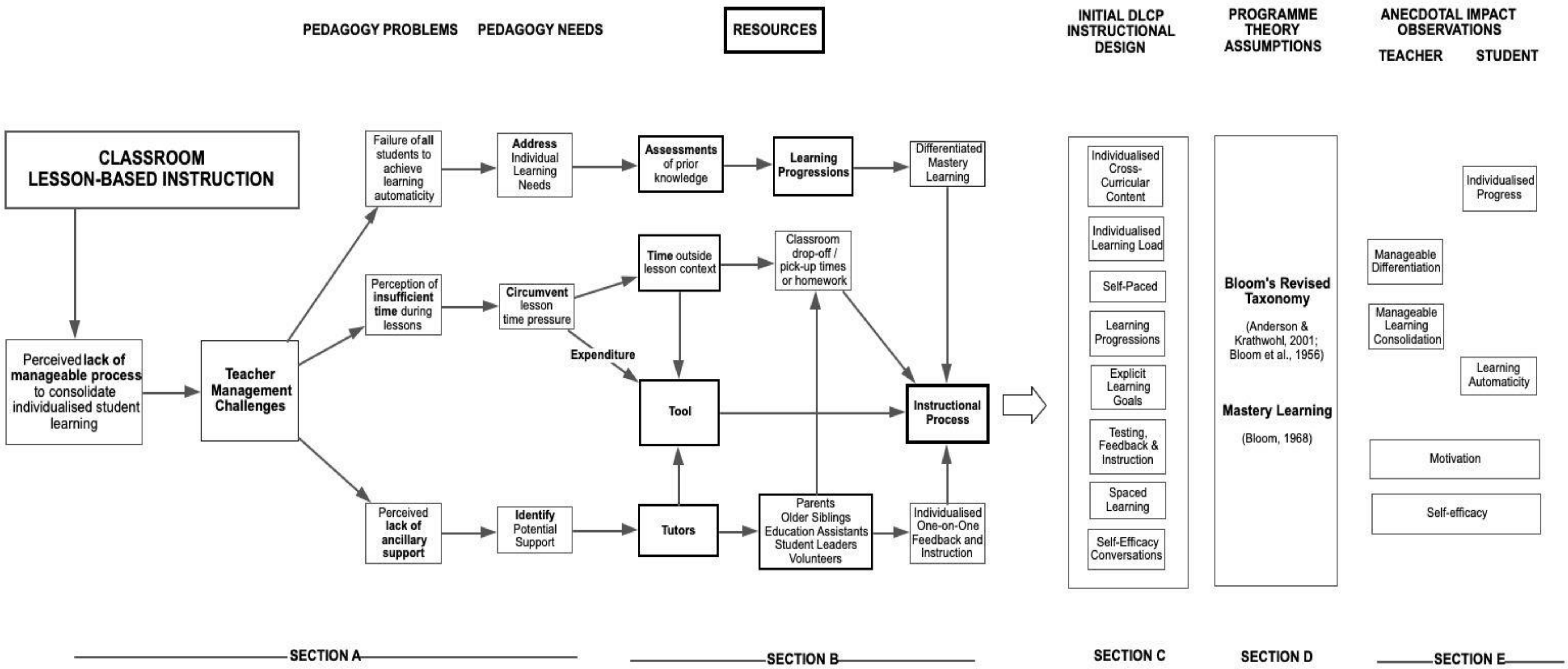
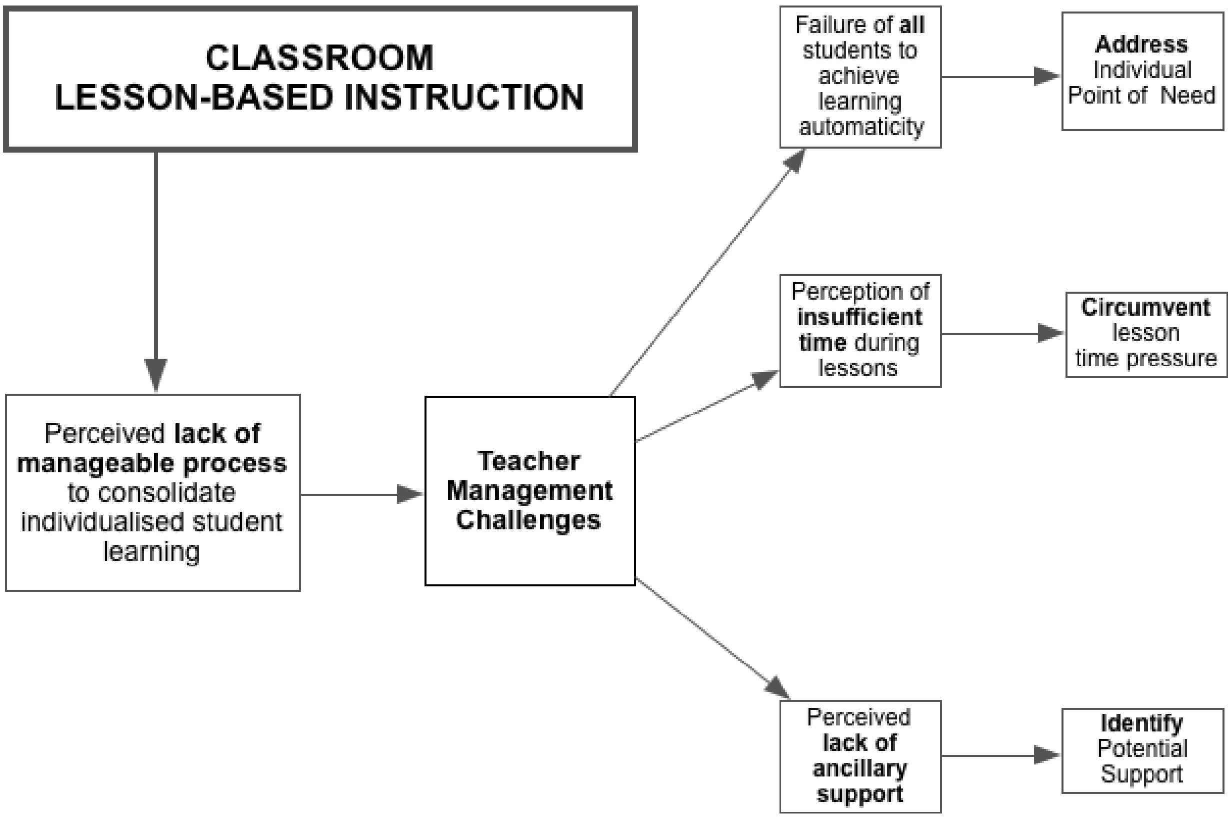


*Figure 4.* Intervention logic model of the original DLCP.



Section A of the logic model (Figure 5) highlights that classroom lesson-based instruction is the initial source of requisite knowledge and skills and the learning foundation upon which the intervention is based. This includes the provision of all types of instruction for broad, cross-curricular educational purposes, teacher directed or facilitated by ancillary staff such as education assistants. Primarily, the differentiation and consolidation of classroom instruction was problematic due to the lack of a manageable process within the lesson context.

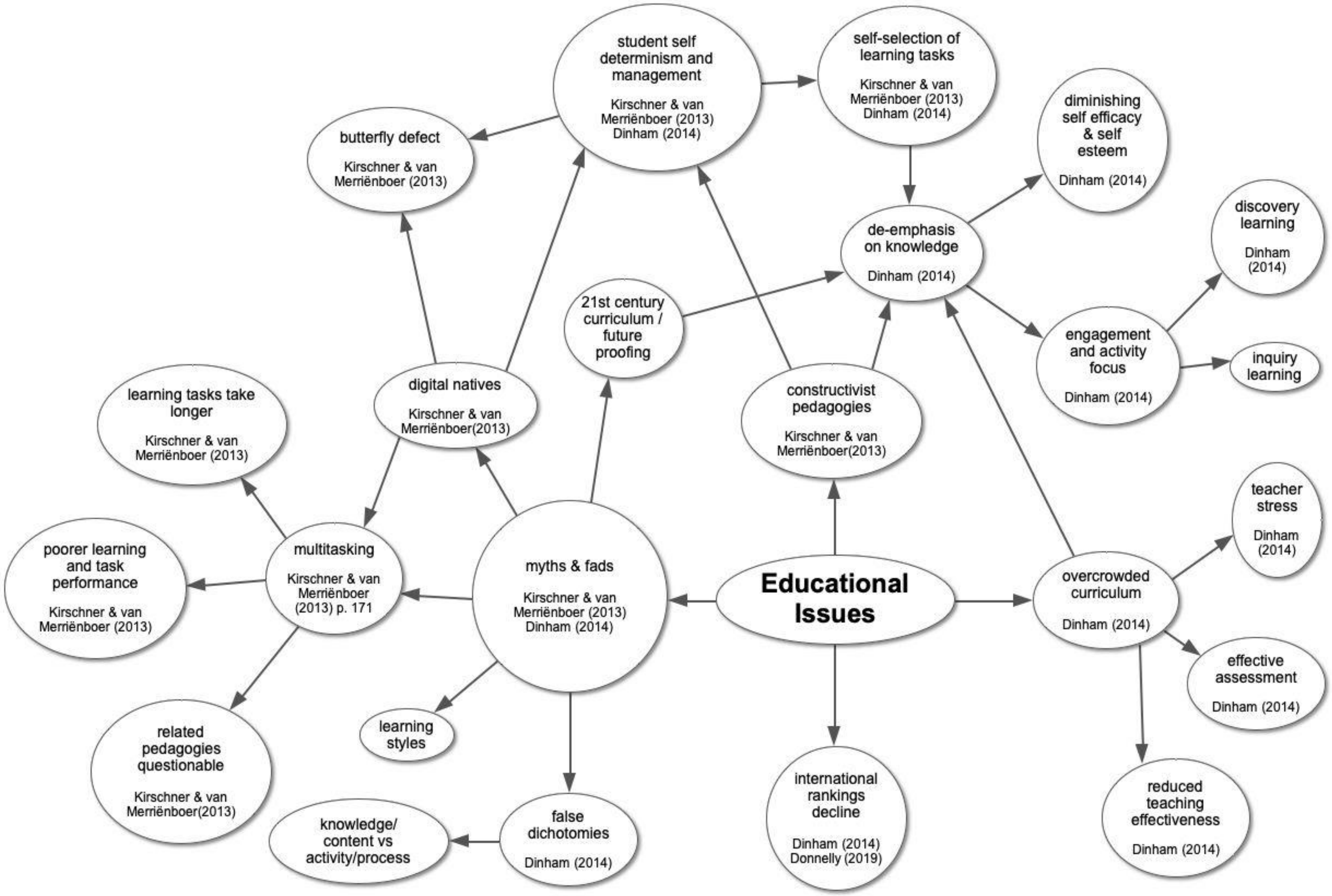
*Figure 5.* Teacher management challenges associated with differentiating and consolidating student learning according to individual needs.



### ***Circumstances and Context***

The identified pedagogical problems and needs were situated within the broader context of policy and educational philosophies within Australia. Based on articles by Dinham (2014) and Kirschner and van Merriënboer (2013), relevant examples of the challenges that impacted consolidation, differentiation, available time and support are illustrated in Figure 6. This milieu contributed to the researcher's confusion and challenge in trying to address individualised student learning needs. The challenges most relevant to the researcher's experience are elaborated below, supported by literature and brief introductions to the cognitive psychology perspectives which are elaborated in further detail in the forthcoming chapter.

*Figure 6.* Examples of issues effecting teaching and learning in Australia.



**Consolidation Challenges.**

The importance placed on knowledge as endorsed by cognitive psychology, is not shared by all educational philosophies and pedagogies. Philosophy, educational trends and time have influenced the status, means and effectiveness of knowledge acquisition, all of which influenced the researcher in the classroom. Observations of student performance during lessons were interpreted as a demonstration of learning. As the previous example described, it was not until the recall of the number bonds to ten was tested a few weeks after instruction, that it became obvious to the researcher that the learning had not endured. Multiple factors influenced the researcher's pedagogy which was situated within historical attitudes on the value of knowledge and its acquisition.

***History.***

The relevance and value of knowledge has been debated over time with knowledge frequently represented as isolated facts (Christodoulou, 2014a). Jean-Jacques Rousseau proposed the questionable relevance of learning facts in the 18$^{th}$ century, followed by John Dewey in the late 19$^{th}$ century and Paulo Freire amongst others in the 1960s (Christodoulou, 2014a). Central to their philosophical argument was the belief that personal experience brings true understanding and that facts are not only irrelevant but may also hinder learning. "To instruct someone… is not a matter of getting him to commit results to mind. Rather, it is to teach him to participate in the process that makes possible the establishment of knowledge … knowing is a process not a product" (Bruner, 1966 as cited in Wilson & Murdoch, 2009, p. 63). Cognitive psychology, however, claims that domain-specific knowledge informs reasoning (Hirsch, 2010) and higher order thinking skills (Hattie & Yates, 2014; Krathwohl, 2002).

In the Australian context, the shift from state-based curriculum and a focus on content knowledge in the 1960s, to school-based curriculum development in the 1970s, "recognised and privileged skills and values acquisition, diversity, experiential learning, cross-curricular thematic approaches, cooperative learning and 'group work', problem solving, critical thinking and more personalised learning" (Dinham, 2014, p. 2). Standardised curriculum was replaced by content that reflected school-based priorities (Dinham, 2014). Over time, explicit knowledge acquisition came to be viewed pejoratively and the presentation of factual subject-based content seen as dogmatic (Christodoulou, 2014a; Zhang, 2016). The devaluation of knowledge reduced the perceived importance of domain-specific learning. The researcher's training and pedagogy was developed within this context.

Presently, a derisive view of knowledge is implied in some readings advocating 21st century skills (Didau, 2019). The term *fact* is frequently paired with the term *mere*, to question the relevance of knowledge in an information rich new century. Proponents may presume that the learning objective of 'remembering' (Anderson & Krathwohl, 2001; Bloom et al., 1956) involves learning facts in isolation or without regard to prior knowledge or the benefit to further learning (Christodoulou, 2014a; Didau, 2019). The researcher's attitude to literacy and numeracy facts changed when it became clear that their automaticity was foundational to future learning and essential for higher level processing (Bloom et al., 1956, Willingham, 2009; Hattie & Yates, 2014). Cognitive psychology theorists seek to revise negative teacher views on the value of knowledge and claim their view is evidence-based.

Over many years of teaching, the researcher received professional development that may not have been based on a firm foundation of evidence. Biesta (2007), however, challenges the very notion of evidence-based practice by questioning the relevance of its key assumptions within complex systems such as education. Historically, many cognitive psychology investigations, particularly in memory, were conducted in a laboratory environment, the results of which may not apply directly to the classroom. De Bruyckere (2018), an educational scientist within cognitive psychology, agrees that there are several challenges when applying scientific results to educational practice. He advises, “Knowledge can be reliable without being universal” (De Bruyckere, 2018, p. 18) as learning interventions need to work within a variety of educationally contextual constraints. Therefore, research which identifies a means to learning are better termed *evidence-informed* rather than evidence-based and laboratory experiments must proceed to investigations within the classroom context. Together with opinions on evidence, beliefs regarding the importance and acquisition of knowledge have as their source different ontological and epistemological viewpoints which in turn inform pedagogy (McMullen & Madelaine, 2014).

***Pedagogy.***

In an absence of evidence-informed understandings, the researcher’s instructional approaches were based on the professional development received and presumptions of how knowledge is acquired. Cognitive psychology supports knowledge-based approaches such as explicit and direct instruction, yet McMullen and Madelaine (2014) claim that such interventions are maligned by some educators due to ontological and epistemological points of view which instead, promote interest-based, student self-determination in learning. These pedagogies are described by Kirschner and van Merriënboer (2013) as minimally guided. Learners become self-educators with learning to learn given priority (Dinham, 2014; Wilson & Murdoch, 2012).

Kirschner et al. (2006, p. 75) assert that minimally guided approaches such as discovery, problem-based, inquiry, experiential and constructivist learning strategies are equivalent in their approach to the acquisition of knowledge; “Rather than being presented with essential information [learners] must discover or construct essential information for themselves”. Constructivist pedagogies state that appropriately designed learning environments and activities will enable students to meaningfully construct the essential knowledge component (Wilson & Murdoch, 2012). The acquired knowledge may therefore differ between learners (Duchesne & McMaugh, 2018). A more scaffolded approach, guided inquiry, uses a learning team to guide students through “the flow of discovery in the process of learning from a variety of sources of information” (Kuhlthau, et al., 2015, p. 4). Motivation and engagement are recognised as key priorities in these approaches. The quality of learning derived from minimally guided approaches is described by Hirsch (2010) as less effective and efficient when compared to the explicit communication of concepts and skills.

In contrast to constructivist education theory, cognitive psychology advocates a separation of Piaget’s definition of schema construction (the *what*) from the pedagogical conclusion that knowledge is best acquired through an experiential approach (the *how*) (Hirsch, 2010; Kirschner et al., 2006). Kirschner et al. (2006) question how the presentation of partial information can enhance the construction of schemas *more* than the provision of all the essential information. They claim that the way schemas are mentally constructed should not dictate the way in which essential knowledge is presented through instructional design.

Minimally guided approaches raise cognitive psychology concerns over students’ ability to “determine what they do not know and what they, therefore, need to learn” (Kirschner & van Merriënboer, 2013, p. 177). The degree of intrinsic cognitive load (Sweller et al., 2019) required for learning self-determination is of concern as student attention is necessarily divided between tasks, instructions, materials, investigations and the ability to recognise and retain key learning elements (Kirschner et al., 2006; Kirschner & van Merriënboer, 2013). The consensus among cognitive psychologists is that without the presentation of essential learning content, these strategies do not reflect the “structures, functions, and characteristics of working and long-term memory; the relations between them; and [the] consequences for learning and problem solving” (Kirschner et al., 2006, p. 77-78). The researcher had provided a variety of activities to support the learning of number bonds to ten, as previously described, however, they were not sufficient to produce durable learning.

The researcher's understandings were also impacted by informal collegiate conversations and beliefs, media reports and programmes with social currency.

***Edu-Myths.***

In recent years, there has been an attempt to identify evidence-informed research for some widely-held beliefs within the teaching profession. The researcher subscribed to some of these beliefs, however, many have been identified as having a negative impact on the acquisition of knowledge (De Bruyckere et al., 2015; Christodoulou, 2014b). Questioning the validity of these claims, authors describe these beliefs as education myths. Christodoulou (2014b) lists seven myths:

"Myth 1: facts prevent understanding

Myth 2: teacher-led instruction is passive

Myth 3: the twenty-first century fundamentally changes everything

Myth 4: you can always just look it up

Myth 5: we should teach transferable skills

Myth 6: projects and activities are the best way to learn

Myth 7: teaching knowledge is indoctrination" (p. viii).

Some of these beliefs devalue knowledge and contrast with the views held by researchers in cognitive psychology (Kirschner et al., 2006; Kirschner & van Merriënboer, 2013; Krathwohl, 2002; Sweller et al., 2019). Practical issues related to the classroom context also influenced the ability of the researcher to support the consolidation of classroom instruction.

***Time Pressure.***

A major limiting factor in the classroom context is sufficient time. The researcher felt opposing pressures to fulfil curriculum responsibilities *and* provide enough time within lessons for student practice and therefore, the consolidation of learning. In Australia, the previously content defined curriculum of the 1960s moved on to incorporate cumulative ‘issues’-based topics such as the environment and multiculturalism. Dinham (2014, p. 6) claims “rarely is anything taken away to balance what is imposed … as the breadth of teaching increases, inevitably, depth and effectiveness decreases”. In addition, social welfare, mandatory reporting and administration further add to teacher workloads and time constraints (Scott et al., 2001 as cited by Dinham, 2014).

***Learning Support.***

Staffing and access to support is a key concern in attempting to meet individualised student needs. The researcher had two hours’ access to education assistant time which was utilised for multiple purposes within the classroom, including the DLCP. To provide one-on-one delivery of the DLCP, a variety of human resource support was needed.

***Summary.***

Attitudes towards the value and acquisition of knowledge are diverse. Beliefs frequently reflect different ontological and epistemological viewpoints which in turn inform pedagogy (McMullen & Madelaine, 2014). The pressure to cover broad curriculum goals and the influence of trends, impacts the time available for knowledge acquisition and consolidation. Until the failed learning objective of Year 1 students’ number bonds to ten, the researcher had underestimated the importance of knowledge held in long-term memory and was without a thorough understanding of the means to achieve it. It is now understood that if knowledge is interpreted as irrelevant isolated facts, or if time pressures prevent learning mastery, or if educators do not understand the mechanism, nature of schema and the benefit of assimilated information in long-term memory (Sweller et al., 1998), then exposure to knowledge rather than its acquisition may result. If, however, the importance of knowledge is recognised, then a determining factor of its acquisition is prior knowledge (Ausubel et al., 1968) and therefore the need for differentiation.

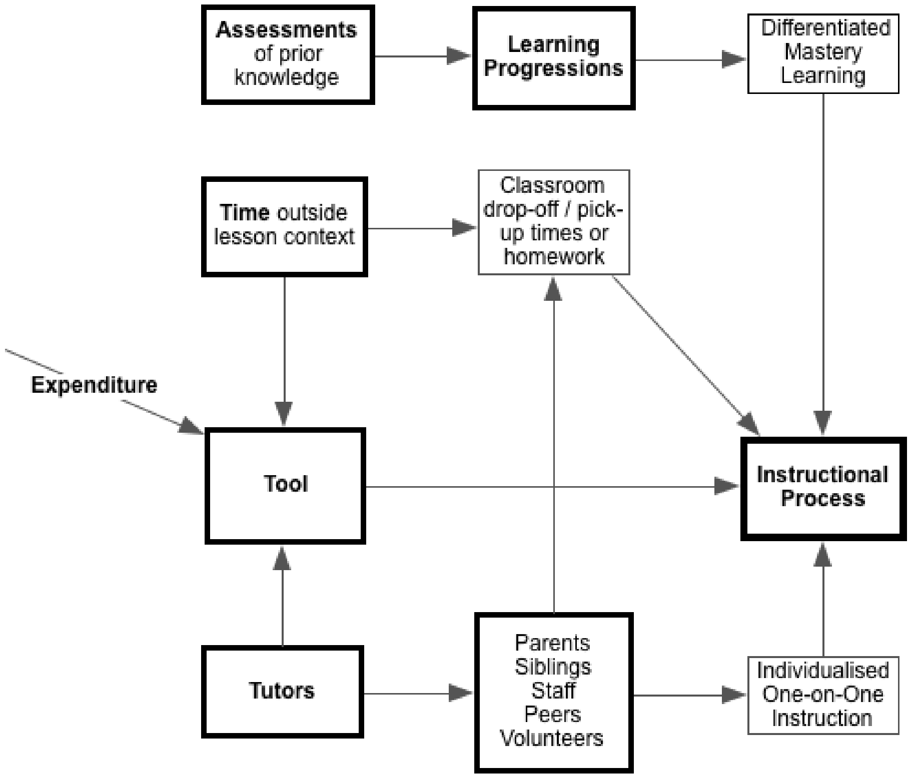
### Differentiation Challenges.

Despite multiple differentiation theories, McNamara and Moreton (1997) claim that these theories have had limited impact in classroom practice at their time of writing. Eighteen years later, Tomlinson (2015) makes a similar assertion. She states that learning differentiation "calls on teachers to be thoughtful about what they teach, responsive to what they teach, and resourceful in how they connect … it is likely that classrooms in which teachers routinely exhibit [these] characteristics are currently in short supply" (p. 206).

The need to differentiate was obvious when the researcher moved with the pre-primary students into Year 1. The students had demonstrated sound performance with synthetic phonics in pre-primary, yet forgetting over the summer holidays had created a great diversity of individual content knowledge. The scope of student abilities and achievement contributes to the challenge of differentiation.

Research identifies that Australian teachers can expect to cater for an average five to six-year gap between the highest and lowest achieving students in their classrooms (Goss et al., 2014). According to McNamara and Moreton (1997, p. 1), these and other factors result in teachers feeling "overwhelmed and frustrated at the amount of work and relative lack of success they have experienced as a result of struggling to differentiate". The lack of an identified process for manageably and effectively addressing *individual* students' prior knowledge and learning progress, was the researcher's experience. During the classroom development of the intervention, accessible resources were identified which assisted the researcher to address these challenges (Figure 7).

*Figure 7*. Intervention resources (in bold).




### *Resources*

Pawson and Tilly (1997, p. 6) identify that resourcing is of central importance to interventions stating that "it is not programmes that work but the resources they offer to enable their subjects to make them work". The subjects, in this case, are teachers. Difficulties related to teachers' ability to manage the differentiation and consolidation of classroom instruction are multi-faceted and contribute to work-related stress. The intervention attempted to address these stressors through the following resources:

- pedagogies recognised as effective (for example, mastery learning),
- a mastery tool,
- sufficient and flexible time,
- adaptable support and
- very low expenditure (as defined by Higgins, et al., 2016).

**Pedagogy.**

During intervention development, four pedagogical resources were identified to assist in the differentiation and consolidation of student learning: prior knowledge, observation and assessment, learning progressions and the instructional method of mastery learning. In 1968, Ausubel et al. defined prior knowledge as the most important factor in the acquisition of new knowledge. Prior knowledge is arranged in schemas which facilitate the integration of new learning (Kirschner & Neelen, 2019). A student's prior knowledge is the foundational resource for future learning and the reason that learning differentiation is necessary.

Assessment and observation are methods by which teachers can identify prior knowledge and the understandings gained through classroom instruction. Tomlinson et al. (2008, p. 6) state that "the most powerful differentiation is based on pre-assessment and ongoing assessment of student progress toward key goals". These assessments may be facilitated through established teacher routines and documentation such as formative and summative assessments, observation and subject-based checklists. They may reveal the need for a combination of lesson-based extension, remediation or consolidation, and include strategies such as re-teaching and small group work. Beyond verbal feedback, however, the provision of significant individual differentiation is extremely difficult within a whole class lesson context due to time, support and classroom management constraints. An objective of the intervention was to facilitate this level of differentiation of classroom instruction, outside of this context.

Following the assessment of a student's knowledge or skills the third pedagogical resource within the intervention was the understanding and application of learning progressions. Learning progressions are described as sequential learning content arranged from "simple to complex so that a meaningful context is created to integrate subsequent ideas" (Kirschner & van Merriënboer, 2013, p. 179). They provide teachers with a scope and sequence to facilitate progress. Tomlinson (2015) describes classroom differentiation as scaffolding learner progress from their prior knowledge to a defined expectation.

During the development of the intervention, the researcher's overarching theoretical understandings were based on the revised Taxonomy of Educational Objectives: The Cognitive Domain (Anderson & Krathwohl, 2001; Bloom et al., 1956) and Bloom's (1968) Learning for Mastery framework (which incorporated Carroll's (1963) Model of School Learning). As displayed in Table 2, the revised Bloom's taxonomy recognises the foundational importance of knowledge by assigning it as a separate dimension against which the cognitive processes of the remaining learning objectives can be classified (De Bruyckere et al., 2020). Whilst less rigid in hierarchy when compared to the original taxonomy (Krathwohl, 2002), knowledge informs and contributes to the development of the remaining learning objectives (Didau, 2019). It should be noted that the familiar Bloom's hierarchy pyramid was not devised by the original or revised taxonomies (De Bruyckere et al., 2020) however, it also implies the foundational importance of knowledge.

Table 2

*The Revised Bloom's Taxonomy Matrix*

| Knowledge Dimension | Cognitive Process Dimension | | | | | |
|---|---|---|---|---|---|---|
| | Remember | Understand | Apply | Analyse | Evaluate | Create |
| Factual | | | | | | |
| Conceptual | | | | | | |
| Procedural | | | | | | |
| Metacognitive | | | | | | |

*Note.* Adapted from "A revision of Bloom's taxonomy: An overview," By D. Krathwohl, 2002, *Theory into Practice, 42*(4), p. 216.

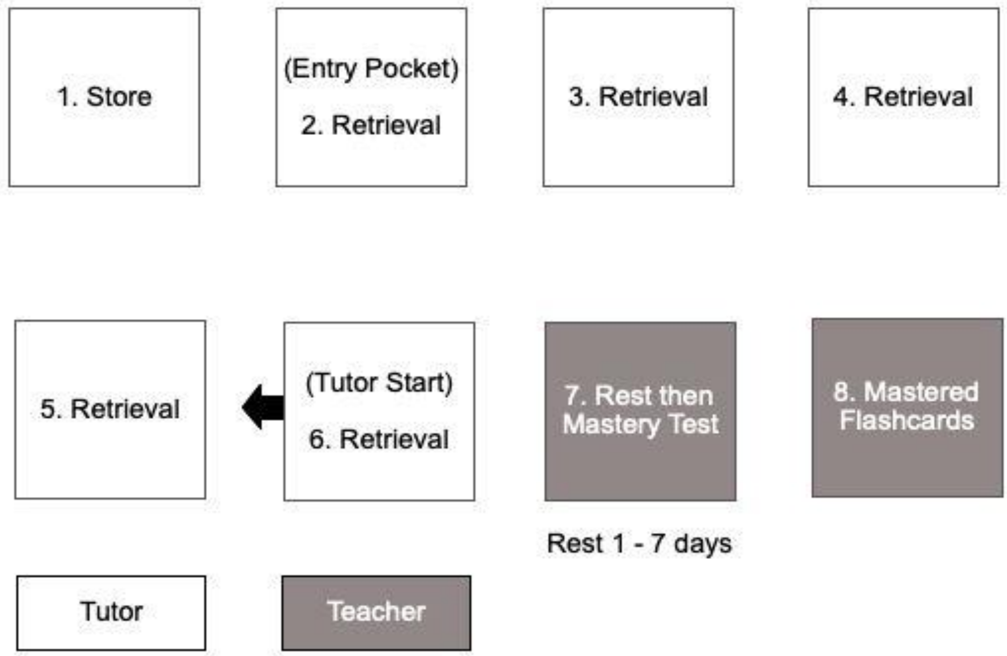
Bloom's Learning for Mastery is an instructional method based upon students' prior knowledge. "Bloom believed that all students could be helped to reach a higher criterion of learning if both the instructional methods and time were varied to match students' individual learning needs" which he termed *mastery learning* (Guskey, 2007, p. 9). Assessments identify the entry levels along established learning progressions with mastery criteria determining progress through advancing curriculum. To attain differentiation to the level of the individual, the DLCP required the combination of a mastery tool and learning content in flashcard format.

**Mastery Tool.**

Early versions of the handmade mastery learning folders (Appendix I, 2010 – 2012) were both time consuming to create and nondurable. As previously described, a hard-cover version was created to engineer out the inefficient use of time required to repeatedly make the folders by hand. Additionally, the pocket labels included a bee-theme which facilitated conversations with young students related to the movement of flashcards, particularly when content returned to the entry pocket (the "Hive").

The hardcover folder (2013 – 2016) improved classroom management efficiency but was not essential. Many schools continue to make the cardboard manila file version for use with the DLCP. Other formats are possible if they allow for the sequential movement of flashcards such as a divided box, a collection of zip-lock bags or envelopes, or alternative teacher innovations. Regardless of tool format, simplicity, practicality and the classroom context should guide the design. For research discussions, a generic version of the folder tool is displayed in Figure 8. Flashcards were used as they facilitated practice and mastery assessment of one concept at a time. The mastery tool required one-on-one administration which challenged the resources of time and support, however, several factors combined to facilitate this requirement.

*Figure 8.* The research version of the DLCP tool.




**Time and Support.**

The intervention addressed the management of time and support in the following ways:

- The folder content was teacher directed but the practice sessions were facilitated by parents, older siblings, education assistants, student leaders, or volunteers.
- Sessions occurred outside of the regular lesson context.
- The number of sessions per week was flexible.
- The DLCP was of short duration: approximately five, and not more than 10 minutes.
- The DLCP was self-paced due to the movement of content according to mastery criteria.

- The timing and frequency of the teacher facilitated mastery test was flexible as tutors were responsible for moving new flashcards from the Store pocket into the entry pocket as each flashcard arrived at the inactive Test pocket.
- An online bank of free printable flashcards was created for foundational literacy and numeracy concepts.

Three process delivery options were investigated for all or selected students: homework, classroom-based or special needs support. Initially practice sessions were delivered by parents or older siblings as a homework strategy. The folder was stored in the student's school bag so that it was accessible to the teacher for mastery testing and the provision of new content during the school day. When delivered as a homework strategy, the most noticeable gains were achieved by students with committed parent tutors. In response, the mode of delivery was changed to classroom-based so that all students could receive equal access to practice.

The classroom-based session was scheduled during the morning drop-off period prior to the start of class. In the researcher's context, the timeslot prior to the first bell was typically 15 – 20 minutes, a time when tutor support in the form of volunteer parent help was available. Given the short duration of the process, one tutor could conduct a practice session with three to four students in succession within the given time frame. This kept the overall number of tutors required to a manageable number. Whilst the tutor pool was usually a small group of parents, supplementary support options were also available when necessary. These included education assistants, student leaders/buddies, student teachers, work experience students or other volunteers within the school community. The intervention was used by some teachers for selected students only, in which case, practice sessions were usually conducted by an education assistant.

The use of tutors enabled classroom practice sessions to be scheduled several times a week. This frequency of one-on-one provision would otherwise have been unattainable given average ratios of one teaching staff member to approximately 14 students in 2016 (Australian Bureau of Statistics, 2017). Tutor training was essential and consisted of an explanation and demonstration of the DLCP including guidelines on feedback and instruction. Instructions and animations were displayed on the interactive whiteboard during the initial practice sessions. Additionally, the tool included printed instructions. Content guidelines could be provided for tutors, however these were not usually required for junior primary material. As a homework strategy, the expectations of tutors and students results may be reduced.

The process, tool and mastery criteria, enabled the consolidation of learning to be self-paced; content that required greater remediation time cycled back to the start, whilst learning content recalled or demonstrated advanced through the tool. This meant that the consolidation of active learning items was responsive to individual student needs without further teacher intervention.

Classroom management issues often make it difficult for teachers to have one-on-one time with students. The weekly mastery test, however, had a measure of flexibility. With only one pocket to test, the mastery test duration was short (a couple of minutes per student). When the DLCP was delivered at morning drop-off time, the researcher could conduct concurrent mastery testing, staggering the delivery to students across the week. This mode was possible through the combination of students arriving at different times, the presence of tutors working with several students, established routines and the availability of independent activities after testing. At other times, mastery testing occurred during independent learning activities or when education assistants or parent help was present in the classroom. Additionally, if due to contextual circumstances, a mastery test session was missed, tutor testing could continue and the larger accumulation of flashcards in the Mastery Test pocket could be tested at the next mastery test. After mastery testing, the researcher topped up the Store pocket with new learning content. Classroom management often includes a small budget. If school financial support is unavailable this is an important practical consideration.

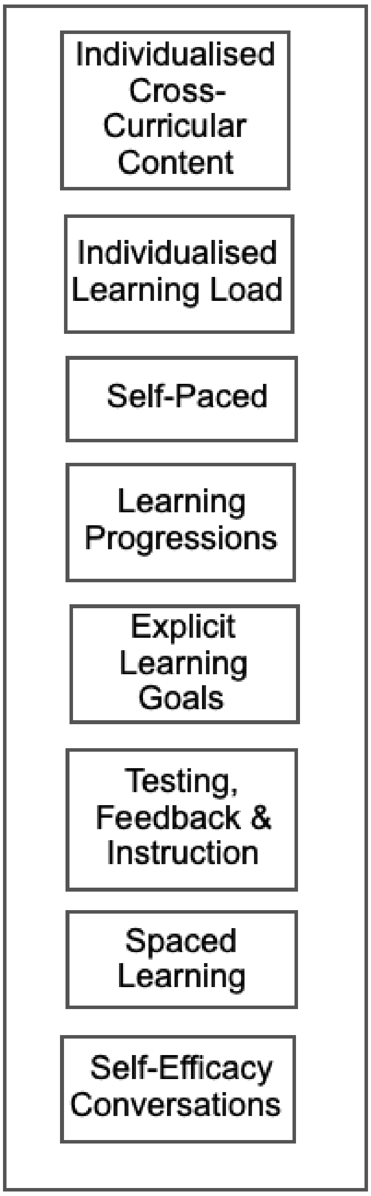
**Expenditure.**

The researcher developed the intervention within the financial context of a classroom budget of $200 which was provided to cover all discretionary spending during the school year. The use of volunteer tutors from within the school community was free. The approximate cost of the tool, initially cardboard mastery learning folders, was $4 per student (as two or three were required over the year). The EEF Teaching and Learning Toolkit (Higgins et al., 2016, p. 4) describes any learning intervention costing less than $140 Australian dollars per student per year as “very low”. By EEF criteria, the intervention was a very low-cost strategy, hence, the minimal cost is described as a *resource* that facilitates implementation. The strategy may be an economically feasible differentiation and consolidation option for schools. The final resource within the intervention was the instructional process.

***DLCP Instructional Design Components***

Section C (Figure 9) of the logic model displays the components of the process as understood prior to commencing the current study. These were the initial descriptions used to identify the conceptual and operational definitions.

*Figure 9.* Process understandings.




**Learning Content.**

The overarching objective of the intervention was to assist the researcher to manage and address the diversified outcomes of classroom instruction. The traditional mastery learning folder approach was colloquially used for remembering subject-specific information such as phonics and sight words (Appendix F). The DCLP had a broader scope, incorporating different levels of cognitive processing and cross-curricular content.

***Cognitive Processing.***

The traditional approach emphasised the cognitive process of *remembering* with content based on the recognition or recall of discrete items (Appendix F). The DCLP evolved to incorporate the learning objectives of *understand* and *apply* (Anderson & Krathwohl, 2001) which required student explanation or demonstration. Table 3 displays questions relating to cognitive processes beyond remembering for the number bonds to ten. *Understanding* utilises the recall of the number bonds to ten through situating their use within a more complex task such as addition or subtraction. *Applying* requires not only recall of the number bonds and the understanding of addition and subtraction, but also the selection of a strategy, for example, the reordering of the algorithm to take advantage of known facts. This level of processing can include cumulative application of multiple strategies such as mental maths doubles and more complex partitioning.

Table 3

*Examples of Cognitive Processes Using Number Partitioning to Ten*

| Knowledge Dimension | Cognitive Process Dimension | | |
|---|---|---|---|
| | Remember | Understand | Apply |
| Factual | Who is friends with 8? | | |
| | | 8 + __= 10<br>10 – 8 = __<br>2 + __ = 10<br>10 – 2 = __ | |
| Procedural | | | 8 + 3 + 7 + 2<br>6 + 1 + 6 + 9 |
| Metacognitive | | Can friends of ten help me to answer these questions? | Which strategies will I use to make adding these numbers easier? How will I order the numbers? |

*Note.* Adapted from "A revision of Bloom's taxonomy: An overview," By D. Krathwohl, 2002, *Theory into Practice, 42*(4), p. 216.

***Cross-Curricular Content.***

After student progress was demonstrated with subject-specific content during development (Appendix G), the researcher investigated the inclusion of mixed cross-curricular content. Additionally, later in development, the use of behaviour goals was trialled. Anecdotally, these goals assisted students to remember classroom expectations and routines, as well as coping strategies when faced with difficult situations. Tables 4 provides examples of cross-curricular learning content for different levels of cognitive processing as well as questions to assist students to remember their personal goals.

Table 4

*Example Questions*

| Subject | Remembering | Understanding | Applying |
|---|---|---|---|
| Maths | What is double 9? | Use doubles to solve 2 x 9. | What is the easiest way to add 5, 9, 5 & 1? |
| English | Decode *queen*. | Clap the syllables in *dinosaur*. | How do you spell *bright*? |
| Health | How should we wash our hands? | Why do we wash our hands? | Show me how you wash your hands. |
| Science | What are the characteristics of living things? | Why do pets need food and water? | What do the guinea pigs need today? |
| Routines | Did you remember to order your lunch? | | |
| Expectations | Are you remembering to put up your hand when answering a question? | | |
| Self-regulation | Did you breathe slowly when you felt upset today? | | |

### Differentiation.

#### *Learning Content.*

Learning content included remediation, consolidation or extension material based on the content of classroom instruction, individual student learning needs and learning progressions. Identification of individualised learning content occurred through student observation, assessment and checklist procedures, or the post-testing of related flashcards from the online flashcard bank. Flashcards were also handwritten in response to observations of student work during lessons, for example, incorrect spelling words across subject areas.

***Learning Load.***

The inclusion of an inactive Store pocket enabled a differentiated learning load to be applied. The quantity of active flashcards was thought to address the degree of difficulty and so was tailored to each student's perceived aptitude. The judgement presupposed that students of lesser ability may benefit from a reduced quantity of flashcards and vice versa, but this presumption was untested. With the learning load established, tutors transferred a new flashcard from the Store pocket into the Entry pocket as each flashcard moved from the last active pocket into the inactive Test pocket. In this way, the learning load was maintained. At the time of the mastery test, the teacher could also increase or decrease the learning load based on student results.

***Response Time.***

Response time is the maximum amount of time allocated to retrieve an answer and demonstrate mastery of a flashcard learning item. Approximately $\leq 4$ seconds was provided for students to respond by recall, recognition, demonstration or explanation.

***Learning Time.***

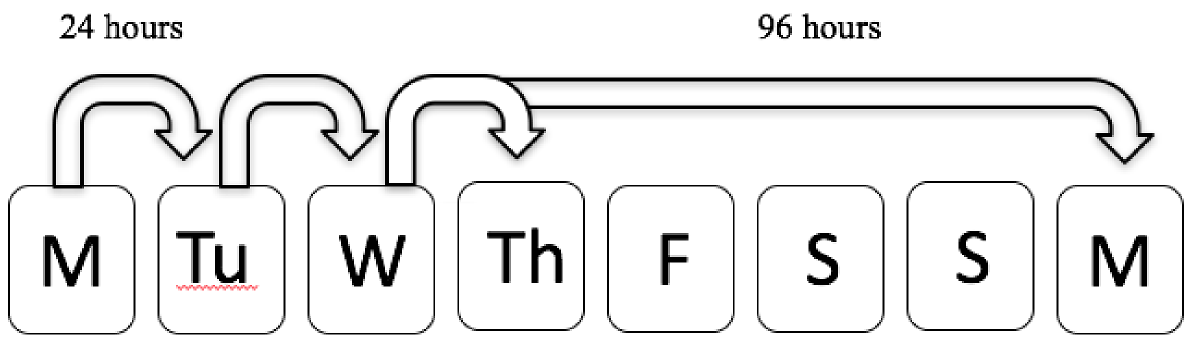
The DLCP was self-paced. When a flashcard met the mastery criteria, it was moved forward to the next pocket. If the criterion was not met, it returned to the Entry pocket (Figure 3). Correct content advanced at the rate of one pocket per practice session, ultimately arriving at the Test pocket where it awaited the teacher directed mastery test, usually conducted weekly. If remembered, the flashcard was moved temporarily to the Mastered pocket and ultimately removed from the tool. If forgotten, it returned to the Store pocket to be reintroduced over forthcoming days, extending the remediation time available and facilitating a self-paced process.

***Learning Intervals.***

According to the educational psychologist document (Appendix F), the traditional process consisted of the repetitive testing—also known as drilling—of flashcards, and their movement through multiple pockets according to accuracy, *within* each practice session as well as between. In the DLCP, however, the testing of each flashcard occurred only once per session moving forward just one pocket when remembered, or back to the Entry pocket when forgotten.

The researcher had a basic general knowledge of the benefits of spaced learning, however, this knowledge was incidental to the development of the process. The practice sessions (Figure 10) were scheduled according to classroom practicalities. As the school assembly was held first period on a Friday, there was insufficient time available to get organised and conduct a practice session on that day. Therefore, there were three 24-hour intervals and one 96-hour interval over the weekend.

*Figure 10.* Spaced interval practice sessions.




***Instruction and Consolidation.***

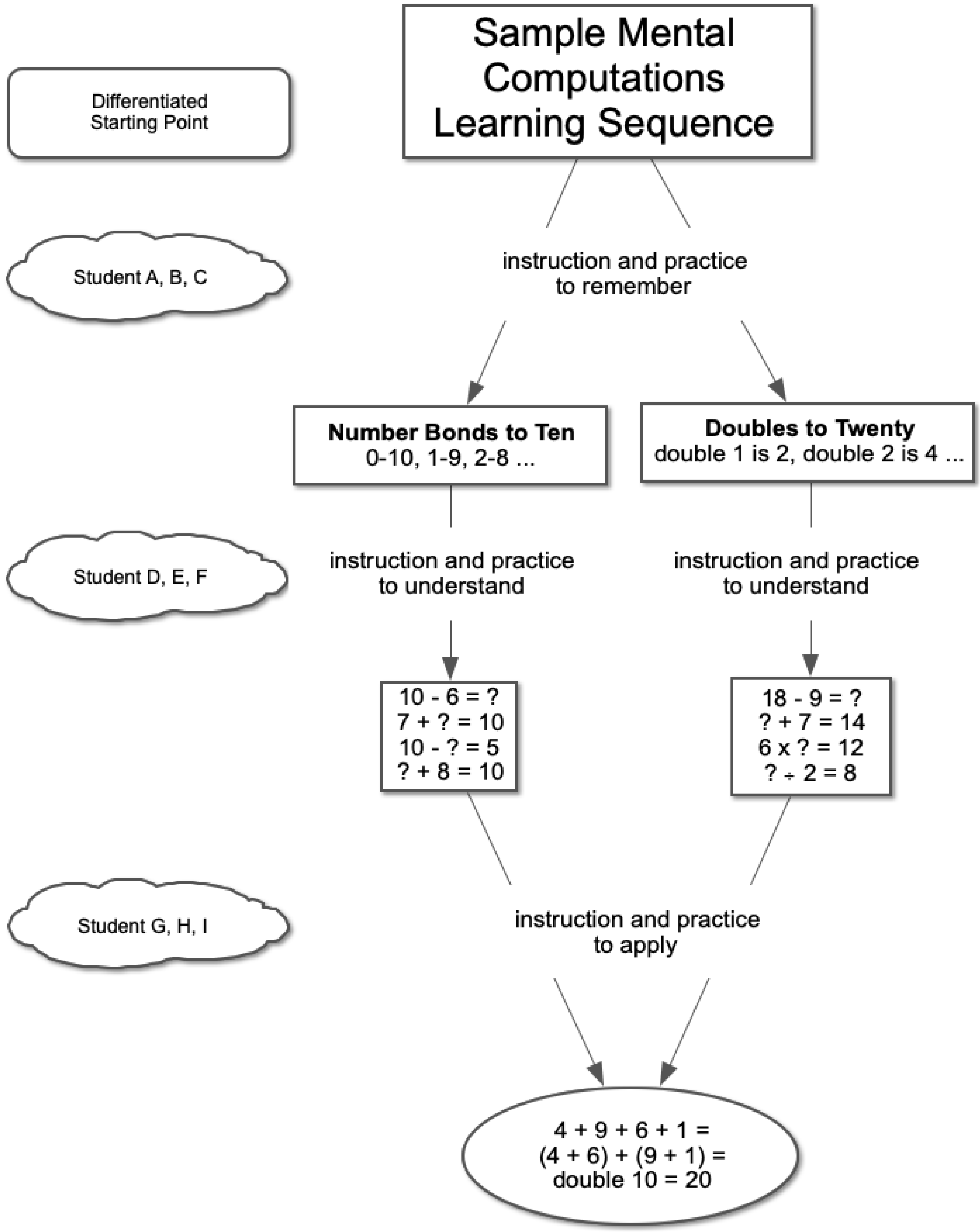
During practice sessions, immediate feedback on mastery criteria was provided through the movement of flashcards. After testing, tutors were asked to check for student understanding of the incorrect content. Any individual difficulties with concepts or skills were then addressed through instruction. The researcher and tutors used conversation during practice sessions to encourage the development of student self-efficacy and self-regulation. To complete the session, the tutor guided the student in repeated retrieval of all the active flashcards, which remained in their allocated pockets.

**Integrated Example.**

Mastery learning theory addresses the differentiation of content and time as well as the consolidation of learning along learning progressions, according to mastery criteria. Figure 11 provides a Year 1 mental maths example showing how the DLCP could be used to support differentiated consolidation for number bonds to ten and doubles. Consolidation practice begins with ‘remembering’, followed by the objectives of ‘understanding’ and ‘applying’. Remembering requires the recall of discrete content (Anderson & Krathwohl, 2001; Hopkins & Bayliss, 2017) which is then available for more complex cognitive processes (Anderson & Sosniak, 1994; Brown et al., 2014; Didau, 2019). Practice to ‘understand’ uses this knowledge in the context of simple arithmetic. In practising the ‘application’ of this knowledge, students need to self-select the most appropriate combination of partitioning, using number bonds and doubles to solve the more complex algorithms.

Figure 11 also describes the incorporation of differentiation. After lesson-based instruction, assessments and remediation, students labelled A to I begin their remediation, consolidation or extension at different points along the learning progressions. Students whose assessment indicates that they can use the strategies independently, could practice using more advanced material or not use the intervention for this content. The researcher found that the DLCP and tool provided a manageable means of addressing the differentiation and consolidation of mental mathematics computations.

*Figure 11.* Remembering, understanding and applying two mental computation strategies.



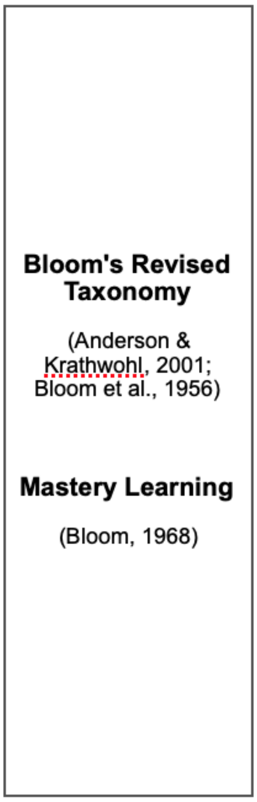
### *Theoretical Assumptions*

Bloom's revised taxonomy (Anderson & Krathwohl, 2001; Bloom et al., 1956), and mastery learning (Bloom, 1968; Carroll, 1963) were the major theoretical constructs understood in the original DLCP (Figure 12). The developers of Bloom's revised taxonomy assumed a "cumulative hierarchy; that is each simpler category was prerequisite to mastery of the next more complex one" (Krathwohl, 2002, p. 213). The DLCP focused on the practice of content related to the *remembering*, *understanding* and *applying* levels of the taxonomy.

Carroll's (1963) Model of School Learning documented the factors pertinent to student success in school learning and the relationship between them (Airasian, 1971). During Carroll's research in foreign language learning, he identified that "a student's aptitude for a language predicted not only the level to which he learned in a given time, but also the amount of time he required to learn to a given level" (cited in Airasian, 1971, p. 5). He subsequently defined aptitude as the *amount of time* required for a student to learn content in optimum conditions according to given criteria. Bloom used Carroll's concepts in developing his Learning for Mastery (1968) model.

Instruction based upon student learning needs and differentiated learning time are known as mastery learning principles. Learning progressions are an essential component of mastery learning as student progress is assessed against mastery criteria before instruction moves on to more advanced levels. Bloom (1968) and Carroll (1963) proposed that by accommodating the time required for a student to understand a concept, and varying the instructional strategies as necessary, most students should be enabled to achieve a higher criterion of learning (Guskey, 2007). As a self-paced learning 'safety-net' beneath classroom instruction, the DLCP could be described as instructional method variation. Prior to the results of the current study, practice sessions and the mastery test were understood only as mastery procedures. These elementary understandings of theory were used to begin the search for baseline conceptual and operational definitions of the DLCP components.

*Figure 12.* The DLCP theoretical assumptions.




**Conceptual and Operational Definitions.**

Conceptual and operational definitions of the DLCP components and strategies were annotated to assist with the identification of underlying programme theory. Conceptual definitions describe the general recognisable characteristics, while operational definitions identify measurable outcomes in the related source data (Cooper, 1998). Theoretical assumptions were limited and based predominately on Bloom's revised taxonomy (Anderson & Krathwohl, 2001; Bloom et al., 1956) and mastery learning principles (Bloom, 1968). Table 5 displays the conceptual and operational definitions used to begin the theory identification.

Table 5

*DLCP Learning Strategies, Definitions and Theoretical Assumptions*

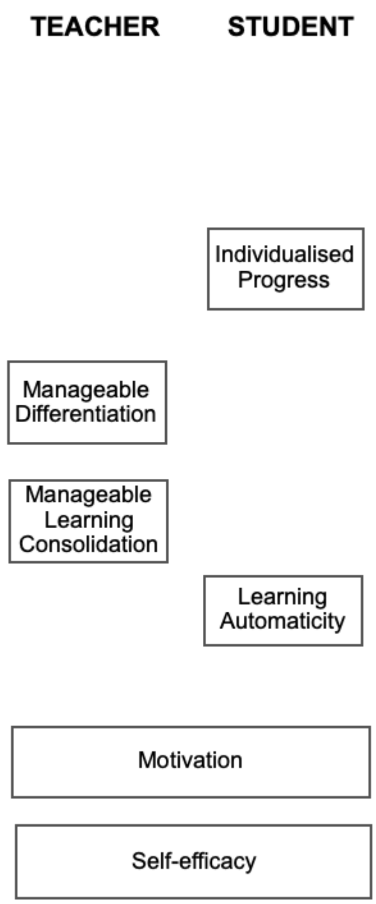
| Learning Strategy | Conceptual Definition | Operational Definition | Theoretical Assumption |
|---|---|---|---|
| Cognitive Strategies: | | Theories related to: | |
| differentiated content | The selection of learning material according to student learning needs. | Student prior knowledge and learning content progressions to address individual learning needs. | Bloom's revised Taxonomy of Learning Objectives: Cognitive Domain (Anderson & Krathwohl, 2001; Bloom et al., 1956)<br>Mastery Learning<br>Bloom (1968) |
| differentiated learning time | Provision of the amount of time of time required for an individual student to master a learning concept. | Time as a variable for learning to be consolidated in long-term memory according to individual student needs. | Mastery Learning<br>Bloom (1968) |
| differentiated learning load | The quantity of flashcards being addressed at any one time, according to student aptitude. | The effects of variable quantities of flashcards being practised concurrently according to individual student needs. | Mastery Learning<br>Bloom (1968) |
| mixed content | The contiguous practice of mixed topic flashcards. | The effects of similar or different subject / topic flashcards being learned within the same session. | Unknown |

| Learning Strategy | Conceptual Definition | Operational Definition | Theoretical Assumption |
|---|---|---|---|
| explicit instruction | Clear guided instruction | The effects of explicit instruction on long-term memory. | Unknown |
| spaced learning | Practice distributed over time | The effects of the distribution of learning at intervals of varying duration on the transfer of learning to long-term memory. | Unknown |
| testing | Assessment of mastery criteria | Testing based on mastery criteria within a response time of ≤ 4 seconds. | Mastery Learning Bloom (1968) |
| feedback | An assessment of and response to student mastery of learning content. | Effective student feedback for learning. | Mastery Learning Bloom (1968) |
| Metacognitive Strategies: | | | |
| self-efficacy | A student's belief that they can succeed with learning tasks through feedback and perseverance. | The type and effect of instructional strategies that promote student self-efficacy and motivation, particularly with reference to learning mistakes. | Mastery Learning Bloom (1968) |
| goal setting & self-monitoring | Student understanding the immediate and long-term purpose of learning and knowing their current and potential achievement level. | The effects of learning goals and self-monitoring on learning attitudes and achievement. | Unknown |

### *Impacts and Outcomes*

A characterising and critical feature of interventions is that they are embedded and delivered in complex social settings (Pawson et al., 2004), in this case, the classroom. Pawson et al. (2004, p. 7), describe interventions as “fragile creatures… rarely if ever, is the ‘same’ programme equally effective in all circumstances because of the influence of contextual factors.” With this understanding, the researcher’s subjective anecdotal experiences of student results, are displayed within the final section of the logic model displayed in Figure 13.

*Figure 13.* Researcher’s anecdotal observations of teacher and student impacts.



The researcher's perceived management of differentiation and consolidation of classroom learning improved as the process was refined. Following lesson-based differentiation strategies, the DLCP provided a means of providing finer grained, self-paced student practice according to individual needs. Unexpectedly, over time the researcher also observed an increase in student motivation and self-efficacy.

**Stage 1 (b): Clarify the Review Purpose**

The overarching purpose of realist synthesis according to Pawson (2006, p. 3)

> "is to articulate underlying programme theories and then to interrogate the existing evidence to find out whether and where these theories are pertinent and productive. Primary research is examined for its contribution to the developing theory. The overall intention is to create an abstract model of how and why programmes work, which then can be used to provide advice on the implementation and targeting of any novel incarnation of the intervention".

Given the researcher's previous role of teaching in a gifted and talented programme, and general pedagogical knowledge, she had some understanding of differentiation and consolidation. The DLCP was thought to combine the principles and strategies of mastery learning and Bloom's revised taxonomy. Given that the intervention was assisting the researcher to better manage differentiation and consolidation, a greater theoretical understanding was sought. The purpose of this study was therefore to *review programme theory integrity* (Pawson et al., 2004). This approach provided the opportunity to identify and investigate the underlying theory of the DLCP to potentially facilitate evidence-informed maintenance or modification. The research purpose is further specified through the research questions.

### *Research Questions*

Research on complex interventions is unlikely to be comprehensive due to breadth (Pawson et al., 2004). Pawson et al. (2004) recommend the prioritisation of one aspect of the intervention, in this case the DLCP, and a key mid-range explanatory theory. The new theory of disuse (Bjork & Bjork, 1992) and the desirable difficulty framework (Bjork & Bjork, 2011) were selected and they are described further in the forthcoming chapter. Outlined below are the overarching and specific research questions.

#### Overarching Questions.

What is the DLCP programme theory? Can it be aligned with evidence-informed research, and if so, how?

#### *Learning Strategy Focus Questions.*

- What is the theoretical basis for the strategy, how is it expressed in practice and how does it compare to the expression of the strategy within the DLCP?
- Does research data inform DLCP maintenance or modification?

### *Significance*

Theorists raise concerns that research findings in cognitive psychology are not written directly for teachers, and therefore, have limited adoption in classroom teaching and learning (Dempster, 1988; Howard-Jones, 2014, Lovell, 2020, foreword by Sweller, p. 7). Therefore, the creation, annotation and evidence-informed revision of a teacher-developed intervention situated within this theoretical construct, may make an academic and practical contribution.

Teachers have limited time and support resources. Sharing the responsibility of practice supervision with a tutor, combined with a delivery method external to the regular lesson environment, may make individualised differentiation and consolidation possible within the constraints of the classroom.

Broader societal applications may be relevant for those with memory deficits or for research purposes. The DLCP may provide an instrument for testing various cognitive psychology learning strategy variables, as well as contextual implementation scenarios. The initial theoretical assumptions previously discussed, together with newly identified theories are explored and elaborated in Chapter 5: Theoretical Foundation.

## Chapter 5: Theoretical Foundation

The DLCP is a theory of theories consisting of macro to micro mechanisms of action embedded within the context of an educational social system (Pawson et al., 2004). Foundational to the DLCP is the belief that "if nothing has changed in long-term memory, nothing has been learned" (Kirschner et al., 2006, p.76). It is premised on two factors: the necessity of prior knowledge for new learning, and the importance of durable learning in long-term memory. Juxtaposing the researcher's DLCP experience (expert framing) with DLCP conceptual and operational definitions, the realist synthesis goal was to identify the underlying programme theories. Chapter 5 is the culmination of researcher's identification and assessment of potentially applicable theories.

The narrative review commences with foundational theories found within cognitive science and the implicit value they place on knowledge. Information processing theories and human cognitive architecture then define how, potentially through the DLCP, knowledge may be stored, expanded and utilised. Based upon cognitive load theory, instructional design theory is explored (Sweller et al., 2019). The new theory of disuse (Bjork & Bjork, 1992) is described and the learning strategies of relevance elaborated. The next section of the review introduces the metacognitive theories that may be applicable to the DLCP, related to self-regulation, self-efficacy and motivation. The final section compares the identified theories with prior assumptions, summarising their interactions and culminating in a DLCP theoretical model.

### Cognition

In an age of instant information access, the value of knowledge and how it is acquired is a topic of debate in education. At the turn of the twenty-first century, Lankshear et al. (2000, p. 20) stated that the desire for information could render knowledge "as either passe or in need of a serious reframing”. In recent years, the importance of subject area knowledge has been questioned.

Colloquially described as ‘twenty-first century learning’ some educators believe that a focus on higher order thinking skills will provide students with the best preparation for their future endeavours. Pedagogies that prioritise critical and problem-solving strategies frequently deem these skills as transferable across subject domains (Hirsch, 2010). Cognitive psychologists, however, propose knowledge-based reasoning claiming, “data from the last 30 years lead to a conclusion that is not scientifically challengeable; thinking well requires knowing facts” (Willingham, 2009, p. 28).

Knowledge can be defined philosophically as “the sum of what is known: the body of truth, information, and principles acquired by humankind” (Merriam-Webster, n.d.). In the context of cognitive psychology, Anderson and Schunn (2000) assign a dual focus for memory: the declarative knowledge of facts, and procedural knowledge—together with the ability to manipulate and apply them. Declarative knowledge is expressed through language, and procedural knowledge is demonstrated by action. The acquisition of knowledge has been classified by Geary (2008) from an evolutionary perspective as either biologically primary or biologically secondary, a division that has implications for learning.

Biologically primary knowledge is acquired implicitly through participation in society. It includes complex cognitive skills such as listening, speaking, facial recognition, social interactions, generic problem solving, planning and assessment of surroundings (Sweller et al., 2019). These skills are described as generic-cognitive rather than domain specific. Biologically secondary knowledge is culturally relevant and learned through instruction and learning effort (Sweller, 2016), for example, subject-based curriculum like the recognition of sound-symbol relationships in early literacy.

Biologically secondary knowledge builds upon primary knowledge, just as listening and speaking skills are foundational to reading and writing (Sweller, 2016). Through the experience of touching a hot stove, a child learns the biologically primary knowledge that doing so, can hurt. Biologically secondary knowledge follows when the child is taught about different types of danger that they may not have yet experienced. The focus of the DLCP is biologically secondary knowledge: domain-specific curriculum that requires the movement of knowledge from working memory to long-term memory.

Hattie and Donoghue (2016) describe secondary knowledge acquisition and consolidation in three phases: surface, deep and transfer, each with related consolidation. They recognise that depth of understanding begins with preliminary surface knowledge. Surface and deep knowledge are not mutually exclusive nor is one form more important than the other provided the goal is the acquisition of learning in long-term memory. Knowledge that begins as facts may, with ongoing instruction, evolve into related conceptual understandings (Hattie & Yates, 2014). In illustrating the differences between the learning outcomes of surface and deep objectives Hattie and Donoghue (2016) use the SOLO model (Structure of Observed Learning Outcomes). Another taxonomy that describes learning objectives is Bloom's revised taxonomy of the cognitive domain.

Bloom's revised taxonomy (Bloom et al., 1956; Krathwohl, 2002) was used by the researcher in the development of the DLCP and provided a descriptive vocabulary for learning objectives of increasing complexity. Surface knowledge may relate to *Remembering* and *Understanding.* Deep knowledge may be described by *Applying*, *Analysing* and *Evaluating* (previously termed *Synthesis* in the original taxonomy), and transfer may be described by the *Creating* learning objective. Both the SOLO and Bloom's taxonomies share the concept that the acquisition of surface knowledge is predominately cumulative (Young, 2008). Hattie and Donoghue (2016, p. 4) state that "one cannot move straight to higher order thinking (e.g., problem solving and creative thought) without [a] sufficient level of content knowledge”. This concurs with the developers of Bloom’s Taxonomy who assumed a “cumulative hierarchy; that is each simpler category was prerequisite to mastery of the next more complex one” (Krathwohl, 2002, p. 213; Young, 2008)—although the SOLO model does not extend the concept beyond surface knowledge (Hattie & Donoghue, 2016). The way knowledge is acquired, stored and accessed is reflected in current understandings of human cognitive architecture and is defined as the organisation of our cognitive structures (Kirschner et al., 2006).

Foundational to the DLCP, Atkinson and Shiffrin (1968) introduced the information processing theory, a multi-store model of memory that defined human cognitive architecture in terms of three memory stores: (a) the sensory register, (b) the short-term store and (c) the long-term store. As originally described, the sensory register is first to receive information from the environment which is only briefly accessible. With attention, the information is transferred to the short-term memory store, filtering out any irrelevant environmental input. Some information may be transferred to the long-term store. Relating human cognitive architecture to instruction and learning, Sweller et al. (1998) developed cognitive load theory, which may be applicable to the DLCP.

## Cognitive Load Theory

Cognitive load theory is an instructional design theory (Sweller, 1988, Sweller et al., 1998) within the field of cognitive psychology. According to the theory, the purpose of instruction is to successfully navigate the limitations of working memory to enable and assist students to secure new domain-specific knowledge in long-term memory (Chen et al., 2017). It was identified as highly relevant to the current study, with principles which may have implications for the DLCP. Cognitive load theory builds upon Atkinson and Shiffrin’s (1968) information processing theory and working memory theory (Lovell, 2020).

### *Working Memory*

In 1974, Baddeley and Hitch expanded the concept of short-term memory developing a working memory model that focussed on both storage and thinking. It contains an attention-controlling central executive, two passive stores: the phonological loop and visuospatial sketchpad, and a multimodal store termed the episodic buffer (Baddeley, 2017). The model explains how information is manipulated and retained during thinking and it relates to working memory performance in both adults and children (Swanson, 2015). Baddeley's model describes both domain specific storage and domain general capability for cognitive control and executive function (Conway et al., 2005). Whilst hypotheses, understandings of human cognitive architecture are foundational to cognitive science and are "widely accepted and quite noncontroversial" (Sweller et al., 1998, p. 289).

Working memory is the location of conscious processing and initial learning (Sweller, 2016). It has three limitations when novel biologically secondary knowledge is introduced. First, the ability to recall knowledge, if not rehearsed, may be forgotten within 30 seconds (Peterson & Peterson, 1959). Second, the capacity of working memory is restricted to storage of a small number of knowledge items, described as approximately seven by Millar (1956) and four by Cowan (2010). Third, the 'depletion effect', a recently proposed addition to cognitive load theory (Chen et al., 2017), suggests that working memory capacity may be depleted by extensive effort, and expanded after rest (due to the spacing effect).

The characteristics of working memory, the difficulty of the material and the design of learning tasks impose an information processing load which influences students' ability to manage new knowledge and transfer it to long-term memory (Sweller et al., 2019). The mental effort required to both understand and process information in working memory is defined as the cognitive load (Sweller et al., 2011).

A heavy cognitive load is deleterious to learning as working memory is limited not only by the number of elements it can attend to, but how many it can process at any one time (Millar, 1956). This is termed element interactivity, and such processing reduces the number of items that can be attended to simultaneously to three or less (Sweller et al., 2011). A learning task example is when students' conduct a novel science experiment.

As an instructional method, a science experiment may contain multiple and interacting elements; students must process what to do, how to do it, perform the experiment, and interpret the results. In the absence of any further supporting information, this instructional method applies a heavy cognitive load to the learning of scientific principles. Under suitable conditions, however, new knowledge received through working memory can be processed more effectively. Unnecessary demands on cognition can be managed through appropriately differentiating the learning material and optimising thinking processes that are related to learning (Sweller et al., 2019) through appropriate instructional design.

Cognitive load theory draws together evolutionary psychology (Geary, 2008) and the properties of human cognitive architecture in addressing instructional strategies for the learning of biologically secondary information. Five basic principles guide instructional procedures (Sweller et al., 2019):

1. Information store principle: Information, stored in unlimited long-term memory, is required for human cognition.
2. Borrowing and reorganising principle: Learning is predominately received from others which is integrated into individuals' schemata.
3. Randomness as genesis principle: Information not received through others can be gained only through problem solving 'generation and test' procedures.
4. Narrow limits of change principle: Working memory capacity is variable according to (a) knowledge in long-term memory, (b) on-going schema

development, (c) trial and error procedures and (d) working memory resource depletion through cognitive effort or expansion after rest (Chen et al., 2017).

5. Environmental organising and linking principle: Working memory limitations are reduced through the ability to transfer domain-specific secondary knowledge in chunks from long-term memory.

These properties of cognitive architecture are used to gain, arrange and store domain-specific knowledge in long-term memory for future use. Historically, Sweller et al. (1998) classified cognitive load into three categories: intrinsic, extraneous or germane.

Intrinsic cognitive load relates to the inherent complexity of the learning material. It is influenced by the processing demands of element interactivity in working memory, and the knowledge held by the learner in long-term memory. Novices in a subject-specific domain, as in primary school students in most domains, have fewer and less sophisticated schema resulting in a greater reliance on working memory and a higher risk of cognitive overload (Sweller et al., 1998; Sweller et al., 2019). Intrinsic cognitive load can only be reduced through increasing learners' expertise or modifying the learning content. For example, if instruction and practice is differentiated to relate to students' prior knowledge, it is better able to be integrated into existing schemata (Kirschner et al., 2020). Experts in a domain, however, do not require these accommodations and gain benefit from more sophisticated approaches such as problem solving (Sweller, 1988; 2019).

Extraneous cognitive load is determined by the presentation and procedures of instruction which increase or decrease element interactivity (Sweller et al., 2019). Strategies which have a low extraneous cognitive load include explicit instruction and worked examples. Intrinsic and extraneous loads are interconnected as element interactivity is based on the student's prior knowledge, the complexity of the learning content and the instructional design (Sweller et al., 2019).

If working memory's limited capacity is used to manage the requirements of the learning task (extraneous load), less will be available for managing the intrinsic load—the implicit difficulty of the material. In a departure from the past characterisation, the germane cognitive load (Sweller et al., 1998) is now thought to perform a redistributive role within working memory rather than an additional load (Sweller et al., 2019). The germane cognitive load manages the load imposed by extraneous activities. Any technique to reduce, manage or redistribute cognitive load supports the availability of working memory resources for the intrinsic cognitive load of learning (Sweller et al., 2019).

Cognitive load management was identified as relevant to many aspects of the DLCP which seeks the economy of mental effort in the following ways:

- It is based on classroom instruction which seeks to provide foundational supportive information.
- The teacher uses observation and assessments to identify students' developing knowledge to provide targeted remediation, consolidation or extension practice along learning progressions (intrinsic load management).
- The DLCP supports teacher management of the progression of learning from simple to complex through remembering, understanding and applying learning objectives (Bloom et al., 1956; Krathwohl, 2002) according to individual student needs (intrinsic load management).
- Content limited flashcards may assist novices to process knowledge through low element interactivity (intrinsic and extraneous load management).
- Tutors provide just-in-time corrective feedback and point-of-need explicit instruction whilst content is active in working memory. This may reinforce cognitive rules—a step towards automaticity (Sweller et al., 2019) (intrinsic and extraneous load management).

- The learning content changes but the DLCP remains consistent and predictable (extraneous load management).

The limitations of processing new information in working memory do not apply once knowledge been has transferred to long-term memory (Kirschner et al., 2006; Sweller et al., 1998).

***Long-Term Memory***

The difference between working memory and long-term memory is demonstrated through the following exercise;

> "If I show you 16 digits for five seconds and then ask you to reproduce them, you will probably fail: 4 8 7 1 9 4 7 5 0 3 8 5 8 6 0 4
>
> But if I show you the following 16 letters for five seconds, you will probably be able to reproduce them exactly: The cat sat on the mat" (Christodoulou 2014a, p. 31).

Remembering the digits relies completely on working memory. In contrast, the 16 letters are easily recalled because words, sound-symbol relationships, phonetic decoding and semantic knowledge can be easily retrieved from long-term memory by those who have learnt to read. This example illustrates the understanding that there are no known limits on the amount of information that can be stored and recalled into working memory over time (Tricot & Sweller, 2013; Sweller et al., 2019).

Long-term memory is the stored representation of learning: acquired knowledge and skills—the first principle of cognitive load theory (Sweller et al., 2019). Memories may be based on events (episodic), knowledge (semantic) or procedures (Duchesne & McMaugh, 2018). Memory processing is defined in three phases: encoding (the creation of schema or mental representations), retention (consolidation of encoded information) and retrieval (recall) (Brown et al., 2014).

**Encoding.**

Whilst not fully understood, encoding converts sensory input into chemical and electrical impulses that form mental representations of learning experiences (Brown et al., 2014). This may include concrete or abstract objects or information (Ericsson & Pool, 2016). The new representations are defined as memory traces (Brown et al., 2014). Thousands of units of information on any given topic are expanded, organised and categorised according to how they will be used (Sweller et al., 1998). This abstract knowledge structure was labelled a *schema* by Piaget in the 1920s (Piaget & Cook, 1952). During encoding, memory traces are easily altered and require stabilising for retention to occur (Brown et al., 2014).

**Retention.**

Retention, also described as consolidation, is a process that involves energy, time, biological resources, rest and sleep (Weinstein et al., 2019). The established schema of a topic facilitates the retention of new information through "combining lower level schemas into higher level schemas … by an active, constructive process" (Sweller et al., 1998, p. 255), thus why prior knowledge is acknowledged as essential for learning. Hattie and Yates (2014, p. xii) emphasise that "the more teachers understand the prior status of the student, and the more they are aware of the nature of success … the greater the probability of learning happening". The DLCP seeks to assist teachers to address students' prior knowledge through the differentiation of practice material.

Differentiation of instructional strategies, content and learning time are consistent with mastery learning principles (Bloom, 1968; Carroll, 1963). Carroll viewed aptitude as the amount of time required for a student to understand a concept or master a task. Therefore, if given the appropriate amount of time, all students should be capable of durable learning. In mastery learning, learning goals and success criteria determine progress along learning progressions (Block, 1971).

The importance of learning progressions is highlighted in the National STEM School Education Strategy (2015) with its recommendation to “extend national literacy and numeracy continuums to assist teachers to identify and address individual student needs according to the expected skills and growth in student learning at key points” (Education Council, 2015, p.9). Whilst the DLCP does not result in uniform achievement, in seeking to differentiate the practice of classroom instructional material, targeted learning along learning progressions may assist what Goss et al. (2015) consider learning’s most important objective: progress. Hattie (2015, p.3) defines this as “a year’s progress for a year’s input” for all students. Consolidation and retention are also enhanced through retrieval and rehearsal.

**Retrieval.**

Retrieval involves *reconstructing* the memories of past learning. When memories are activated after a delay, they may be subjective or altered (National Academies of Sciences, Engineering, and Medicine, 2018). Successfully retrieved memories may be enhanced, strengthened and connected to more recent experiences, a process termed reconsolidation (Brown et al., 2014). Once established in long-term memory, knowledge can re-enter working memory as a chunk as illustrated by the Christodoulou’s (2014a) *The Cat Sat on the Mat* example.

***Memory Chunking.***

A schema in long-term memory, broad in “size, complexity, and sophistication” (Sweller et al., 1998, p. 256) can be transferred to working memory as a single entity (De Groot, 2008) as described by the fifth principle of cognitive load theory (Sweller et al., 2019). This is termed memory chunking and it substantially reduces cognitive load. In the context of junior primary, students may be asked to solve an addition problem such as 7 + 4 + 5 + 6. If number bonds to ten, doubles knowledge and procedures such as partitioning are held within long-term memory, then the addition algorithm can mentally proceed as follows:

7 + 4 + 5 + 6

(7 + 3) + (6 + 6)

10 + 12

(10 + 10) + 2

With memory chunking, processing time is also reduced. Conversely, the student who does not have mastery of basic facts and strategies may have concrete materials and counting-on as their only option. Demonstrating this lack of mastery, Hopkins and Bayliss' 2017 investigation revealed that less than 50% of a cohort of 200 Australian Year 7 students had proficiency with simple mental addition. The flow on effects may include helplessness and a lack of confidence and enjoyment (Willingham, 2009). When automaticity is absent, deep learning is disrupted by the cognitive load of low-level processing (Hattie & Yates, 2014; Sweller et al., 1998). Sweller et al. (1998, p. 256) identify that it is "extensive practice" that leads to automation in long-term memory.

***Automaticity.***

Kirschner et al. (2006) define knowledge as automated: the ability to judge a given situation or problem and respond appropriately and efficiently. The DLCP seeks this automaticity for the content and skills delivered through classroom instruction. Automaticity is attained through training and practice (Ericsson & Pool, 2016).

Ericsson and colleagues developed a general theory of expertise based upon investigations into the habits and practices of exceptional performers from a wide range of fields (Ericsson & Pool, 2016). They identified a set of general principles collectively termed *deliberate practice* based upon the adaptability of the human brain and body. A student's potential is not something innate to be fulfilled, rather it is developed through dedication, practice and time. The general theory of expertise shares similarities with both mastery learning theory and the premises of Bloom's revised taxonomy, reflecting the DLCP philosophy that the practice of a skill improves performance (Rohrer et al., 2015).

Deliberate practice challenges homeostasis, forcing the brain to adapt to the demands required of it, ultimately producing increasingly efficient domain-specific mental representations (Ericsson & Pool, 2016). It is consistent with the first principle of cognitive load theory, that working memory is limited in capacity, but there are no known limits of long-term memory (Tricot & Sweller, 2013).

The defining characteristic of experts compared to novices is the quality and quantity of their mental representations—the knowledge in their long-term memory (Sweller et al., 2011). This knowledge is organised, accessible and proficient, which facilitates automaticity and higher order thinking skills such as analysis, evaluation and creativity (Bloom et al., 1956). The ease of the assimilation of new information within a domain is therefore greatly enhanced through established meaning and understandings (Ericsson & Pool, 2016).

Deliberate practice incorporates well-defined specific goals, focus, feedback and challenge, both requiring and developing self-efficacy and self-regulation (Ericsson & Pool, 2016). The characteristics of the general theory of expertise reflect the researcher's goals for the DLCP in the following ways:

- Practice is required for the automaticity of learning content and skills.
- A meaningful learning context is necessary, provided through the supporting foundation of classroom instruction.
- Prior knowledge is key. This is identified through teacher observation and assessment of developing learning to inform the differentiation of content based on learning progressions.
- Flashcards define specific student learning goals.

- The use of tutors assists with student focus, feedback, explicit instruction and metacognitive conversations pertaining to self-regulation and self-efficacy.

Knowledge automaticity is the foundation of reasoning (Krathwohl, 2002).

***Reasoning Skills.***

Cognitive psychologists state that the ability to manipulate information, or reason, is based upon domain-specific conceptual schema held in long term memory (Tricot & Sweller, 2013). The influence of domain-specific knowledge in primary-aged students was demonstrated by Schneider et al. (1989) in a study which compared the comprehension ability of soccer novices and experts. Both high and low literacy aptitude soccer experts of all ages, recalled more details and correctly identified inferences and contradictions in comprehension exercises than the soccer novices. The researchers concluded that the domain-specific prior knowledge of the soccer experts influenced their comprehension skills more than a greater aptitude in literacy.

The Schneider et al. (1989) study highlights that problem-solving and critical thinking skills, often implied to be independent in discussions of twenty-first century learning, retrieve domain-specific facts and procedures from long-term memory, the fifth principle of cognitive load theory (Sweller et al., 2019). If reasoning skills are based upon domain-specific knowledge, then they are highly limited in their transference from one subject to another. Whilst primary biological knowledge does confer some generic problem-solving ability (Sweller et al., 2019), Hirsch (2010, p. 218) explains that "an ability to think critically about chess does not translate into an ability to think critically about sailing". This perspective on reasoning compliments the cumulative nature (Krathwohl, 2002) of Bloom's learning objectives.

In a paper examining the revised Bloom's taxonomy, Anderson and Sosniak (1994) recount the developers' concern that teachers may overemphasise the knowledge (remembering) component of the taxonomy rather than the balance provided by the list. Christodoulou (2014a) suggests that, with the emphasis on twenty-first century learning and skills, the pendulum has now swung in the opposite direction. Duschesne and McMaugh (2018) caution against an overemphasis on higher order thinking skills due to the importance of foundational knowledge. The DLCP seeks the consolidation of the content of classroom instruction to facilitate reasoning. The memory processes of encoding, retention and retrieval are greatly influenced by forgetting.

Forgetting begins rapidly after encoding but eventually slows down (Weinstein et al., 2019). The concept of a timeline of forgetting was introduced by Ebbinghaus (1885 / 1913) who, through self-experimentation with nonsense syllables, created what became known as the forgetting curve. The new theory of disuse (Bjork & Bjork, 1992) reflects the classroom experience of the researcher in terms of the relationship observed between remembering, forgetting and learning.

**The New Theory of Disuse**

The original ‘law of disuse’ (Thorndike, 1913) asserts that memories are forgotten due to irregular access over time. Bjork and Bjork’s new theory of disuse (1992), however, presumes that the memory representations remain, but without regular access, retrieval strength is lost. Retrieval strength is a measure of the ease with which learning can be accessed in memory. A second characteristic, storage strength, is a measure of learning embedded in long-term memory. The relationship between these factors is the key concept of the theory (Bjork, 2011). When storage strength is high, retrieval strength is enhanced by restudying or retrieval (Bjork & Bjork, 2020). However, higher retrieval strength results in a smaller gains in storage strength as a result of restudy or retrieval. A loss of retrieval strength (forgetting) can therefore enhance durable learning (Bjork & Bjork, 2020).

### ***Performance Versus Learning***

Prior to the development of the DLCP, the researcher observed successful student performances in remembering the number bonds to ten during instruction, however, a delayed test revealed that the learning had not endured. This observation is consistent with the new theory of disuse which identifies that there is a dichotomy between the durability of knowledge and skills demonstrated *during* instruction (termed performance) and the relatively permanent knowledge and skills retained over time in long-term memory (termed learning). The researcher assumed "that performance during instruction provides a valid basis for judging whether the relatively permanent changes that will support long-term performance have or have not taken place" (Bjork & Bjork, 1992, p. 110).

Research in the mid-20th century demonstrated that learning may be occurring in the context of little change in performance during instruction, with the reverse also being true; demonstration during performance may not indicate durable learning in long-term memory (Bjork, 1999). If testing is conducted during a learning event, results may suggest that successful learning has taken place, however, only a *delayed* test will deliver an accurate assessment (Richland et al., 2005). Weinstein et al. (2018) provide an illustration of the durability of knowledge related to storage and retrieval strength. Should a student choose to cram study, the retrieval strength will be high, but the storage strength will be low. The performance in a test the next day may be satisfactory, however, the learning may not endure when reassessed after a delay. The fact that performance during instruction does not provide a "reliable index" of learning (Bjork & Bjork, 1992, p. 110) creates challenges for teachers and students. In the number bonds to ten example the researcher had presumed long-term stability of what, for most students, was working memory learning performances. Students themselves may also be misled by their performance.

#### Illusions of Knowledge.

Learning during instruction which is fluently retrieved (high retrieval strength) and perceived by students as easy, can be misunderstood as secure in long-term memory. Bjork (1999) describes this metacognitive state as illusionary. Illusions of competence based upon initial knowledge fluency, demonstrate limited storage strength over time. There are many reasons why it is important to have an accurate assessment of learning.

Overconfidence, based on performance during instruction, can result in the belief that no further attention to a topic is required. Erroneous assessments by teachers and students, may lead to insufficient consolidation and therefore on-going learning gaps. From a societal point of view, few would want to be served by individuals, such as pilots or surgeons, who overestimate their knowledge or capabilities (Brown et al., 2014). Immediate corrective feedback, such as found within the DLCP, may assist students to gain an accurate understanding of what they do and do not know. Storage strength can be enhanced by slowing down progress through increased difficulty in the form of the *thinking* required during retrieval (Bjork, 1999), a characteristic present in the DLCP, through testing students' recall over time.

### *Desirable Difficulties*

Certain mentally challenging learning conditions, termed desirable difficulties (Bjork, 1994), enhance long-term knowledge acquisition, retrieval and transfer (Bjork, 1999). "These difficulties are desirable because overcoming their challenges stimulates advantageous encoding and retrieval processes during learning, which results in durable learning" (Weissgerber et al., 2018, p. 177). As simply explained by Willingham (2009, p.124), "if you repeat the same thought-demanding task again and again, it will eventually become automatic"; the brain seeks to remember the information to avoid the thinking effort. The avoidance of disequilibrium relates well to the biological process of homeostasis (Ericsson &

Pool, 2016). Continual mental challenge over time may trigger a homeostatic brain response; learning in working memory moves into long-term memory.

Cognitive psychologists seek to identify learning strategies that create durable learning. Strategies of desirable difficulty are described as highly effective in stabilising learning in long-term memory (Bjork & Bjork, 2013; Brown, et al., 2014; De Bruyckere, 2018; Dunlosky et al., 2013; Weinstein et al., 2019; Weissgerber et al., 2018). They include spaced practice, retrieval practice (which includes elaboration) and interleaved practice. Weinstein et al. (2018) claim that, outside the field of cognitive science, few instructors are aware of the potential of these learning science strategies. These desirable difficulty learning strategies are potentially applicable to the DLCP.

**Spaced Practice.**

It has been known for over a century that an effective way to improve long term retention of learning is to provide learning opportunities which are distributed over time (Toppino & Gerbier, 2014). Distributed practice, also known as spaced practice, is described by Bjork as the "most robust effect from experimental psychology across the 130 years of research on human learning and memory" (ColumbiaLearn, 2018, 9:34). Its beneficial effect on learning has been described as "dependable and replicable" (Dempster, 1988, p. 627), "[large] and most robust" (Rohrer et al., 2015), "overwhelming" (Dunlosky, 2015, p 15), "tremendous" (Kang, 2016, p12), and "empirically well-established" (Chen et al., 2017. p. 498). As previously mentioned, formal research into the spacing effect began with Hermann Ebbinghaus (1885 / 1913), the creator of the forgetting curve.

Ebbinghaus (1885 / 1913) determined that when the repetition of syllables was spaced over three days instead of one, he could halve the number of repetitions required to learn them (Weinstein et al., 2018). Over many years, hundreds of studies followed to investigate various aspects of the interaction of memory, spaced practice and learning (Cepeda et al., 2009). Historically, much research has been based on empirical characteristics outside of ecologically relevant contexts (Dempster, 1988) however, there is increasing interest in educational research (Kapler et al., 2015). The spacing effect may have multiple and related causes under different conditions (Küpper-Tetzel, 2014) which will be addressed in Section 2: Theory Application. The DLCP spaces the testing of flashcards across intervals of one to seven days and, therefore, this mechanism may be implicated in the positive outcomes observed by the researcher in the classroom. A second potential DLCP learning strategy is retrieval practice.

**Retrieval Practice.**

Learning is often understood in terms of study and memory processes (Karpicke & Roediger, 2008). Testing was previously understood by the researcher as either formative, to inform student learning needs, or summative, to assess student learning. Research in cognitive psychology has identified a different purpose for testing: learning. Termed retrieval practice, investigations have determined that the retrieval of knowledge through testing leads to stronger long-term retention of learning (Brown et al., 2014). Student testing, however, is associated with some controversy within education.

Some educators express concerns that frequent testing—a feature of the DLCP—is unnecessary and irrelevant (Adesope et al., 2017; Brown et al., 2014). Retrieval testing however, recognises the nature of human cognitive architecture and has been identified as an effective learning strategy (Dunlosky et al., 2013; Pashler et al., 2007). Section 2 will further explore research on the evidence-informed use of retrieval testing and potential applications to the DLCP.

The learning strategy of elaboration is another form of retrieval practice (Brown et al., 2014). Elaboration asks students to explain or justify their understandings (Dunlosky et al., 2013). In doing so, they must search their memory to retrieve the relevant information which links to the learning at hand. This enhances schema development (van Merriënboer & Kirschner, 2018).

Spaced retrieval practice combines spaced practice (the spacing effect) and retrieval practice (the testing effect), and this strategy may represent key learning mechanisms within the DLCP. A third potential strategy of desirable difficulty within the DLCP is interleaved practice.

**Interleaved Practice.**

Students of all ages are ultimately required to integrate and apply learning of topics or concepts in exams or real-life contexts. They need to have the "specific knowledge to perform the familiar aspects of those problems, but, above all, have the necessary general and abstract knowledge to deal with the unfamiliar aspects of those problems" (van Merriënboer and Kirschner, 2017, p. 8). Van Merriënboer and Kirschner (2017) describe interleaved practice as a strategy that can facilitate this learning objective.

Inductive learning is the ability to generalise concepts and categories from exposure to multiple related exemplars (Kornell & Bjork, 2008). Interleaving involves interspersing (spacing) the practice of exemplars from different categories, rather than blocking them by category. Research identifies interleaved practice of low discriminatory learning material as an effective inductive learning strategy (Kornell & Bjork, 2008, Weinstein et al., 2019; Weissgerber et al., 2018; Kirschner & Neelan, 2018). Rohrer and Taylor (2007) demonstrated increased learning gains for geometric solid formulae when practised in the interleaved condition as compared to practice blocked by formula type. Learning gains have also been demonstrated based on the interleaving of high discriminatory mathematics practice examples. When high discriminatory mathematical practice examples are interleaved, the learner needs to retrieve both the type of strategy and the procedure required to solve it. This increases the desirable difficulty, and therefore the potential learning gains (Rohrer et al., 2014).

Cross-curricular flashcards within the DLCP are randomly interleaved when tested. Conditions facilitating the interleaving effect and potential implications for the DLCP are discussed in Section 2. Whilst cognitive strategies facilitate learning progress, metacognitive strategies assist students to monitor and control it (Flavell, 1979).

**Metacognition**

John Flavell established the term metacognition, which is defined as “awareness or analysis of one's own learning or thinking processes” (Merriam-Webster, n.d.). Metacognition in learning relates thinking to concepts of learning intentions, success criteria, feedback, progress, achievement (Hattie & Yates, 2014) self-regulation (Winne & Hadwin, 1998), self-efficacy, motivation (Bandura, 1986) and attributions for success or failure (Weiner, 1985).

The DLCP inherently provides students with explicit learning intentions, in flashcard format. Success criteria is applied to each learning item through the attempted retrieval of content during each practice session. Progress is demonstrated through the forward movement of learning items when successfully retrieved, or the return of flashcards to the Entry pocket when unsuccessful. This feedback provides students with realistic understandings of what they do and do not know (Bjork, 1999; Hattie & Yates, 2014). Achievement is highlighted when learning items are mastered, graduating from the process and tool.

Winne and Hadwin (1998) describe four stages of self-regulated learning: task definition, goal setting and planning, enactment and adaptation. Teachers and tutors manage these decisions on behalf of the students, however, the DLCP may act as a scaffold and model of self-regulated learning for independent study in the future.

Bandura’s (1986) social cognitive theory seeks to explain the influence of social context on student learning and behavior (Duchesne & McMaugh, 2018). Described as reciprocal determinism, Bandura divided the influence of social context into three elements: personal (cognitive) factors, behavioural responses and environmental considerations. Cognitive factors determine how external events are perceived based on knowledge, beliefs and emotions. These may elicit a behavioural response which in turn, may alter the environment. This he described as personal agency from which theories of self-efficacy and motivation emerge.

Bandura’s (1986) self-efficacy theory relates to the resilience and motivation that come from successful learning experiences. Within the DLCP, learning content is differentiated to address prior and developing student knowledge with the goal of achieving progress along learning continuums. Researcher observations of student responses to the DLCP over time, suggest that it may make a positive contribution to self-efficacy and motivation. Bandura’s (1986) self-efficacy theory may provide insights for DLCP revision. Teacher and tutor conversations during retrieval sessions may guide student thoughts on the reasons behind the success or failure they experience within the DLCP.

Attribution theory divides the perceived or actual causes of success or failure into internal or external loci, stability and controllability (Weiner, 1985), all of which may impact ego esteem needs and the expectations of future learning outcomes. In a school context, ego esteem needs relate to the desire to be positively recognised for learning performances and outcomes (Hattie & Yates, 2014). Internal stable factors, such as student aptitude, have a greater impact on ego esteem needs, than external factors such as a particularly difficult test (Weiner, 1985). Students who attribute failure to unstable factors, such as the application of effort, may suffer less impact to ego esteem. The application of effort may be controllable, in which case, a student may see potential in changing a future outcome. Metacognition and applications of attribution theory to self-regulation and the DLCP are explored in Chapter 8: Metacognitive Development Strategies.

## Revision of Theoretical Assumptions

The potential DLCP programme theories included in this chapter were derived as a result of the researcher's experience as a teacher and with the DLCP and they may help to explain some of the observed effects of DLCP in the classroom. The acquisition of automated knowledge is colloquially associated with the method of rote learning. The researcher wondered, and some teaching colleagues proposed, that the mechanism of action within the DLCP was rote learning. This strategy shares some similarities with the DLCP so, with greater theoretical understanding, it shall be reviewed.

### *Rote Learning*

The Merriam-Webster (n.d.) dictionary defines rote as "1: the use of memory usually with little intelligence [and] 2: mechanical or unthinking routine or repetition"—highly controversial concepts within education. In the early years of cognitive load research, John Sweller and colleagues surmised that, despite understandings of the nature of long-term memory and the value of automated knowledge, the implications of the theory had made little impact upon instructional design for fear of being associated with rote learning (Sweller et al., 2019). The erroneous association of rote learning with memorised information to the exclusion of higher order thinking is illuminated by cognitive load theory (Sweller et al., 2019).

A dichotomy exists for some educators, between knowledge (held in long-term memory) and creativity (or other higher order thinking skills) (Christodoulo, 2014a). However, according to the environmental organising and linking principle of cognitive load theory (Sweller et al., 2019), working memory is able to process knowledge from long-term memory as a chunk (Kalyuga, 2015). This informs and facilitates more advanced thinking, learning and creativity (Bloom et al., 1956; Willingham, 2009). The importance that the DLCP places on the movement of learning from working to long-term memory is consistent with this principle. Cognitive load theory also describes prerequisites for the acquisition of knowledge.

The second principle of cognitive load theory describes the importance of prior knowledge—an established schema—for the integration of new knowledge (Sweller et al., 2019). Rote learning is a technique which may, or may not, recognise prior knowledge or an appropriate learning context. The DLCP, however, seeks to support the practice of classroom instruction according to learning progressions and individual student needs. In addition, the new theory of disuse (Bjork & Bjork, 1992) and relevant desirable difficulty strategies, may provide an alternate explanation to rote learning as a mechanism of learning within the DLCP. In conclusion, the researcher suggests that mechanism of learning within the DLCP is not rote learning.

### *Potential DLCP Theories*

A general knowledge of spaced learning and two formal theoretical assumptions—Bloom's revised taxonomy (Anderson & Krathwohl, 2001; Bloom et al., 1956), and mastery learning (Bloom, 1968)—were the researcher's theoretical presumptions prior to the current study. This chapter has discussed potential additional theories which are displayed in Figure 14. Newly identified terminology replaces some of the previously used descriptions (Figure 15).

*Figure 14.* The baseline and revised DLCP programme theory.

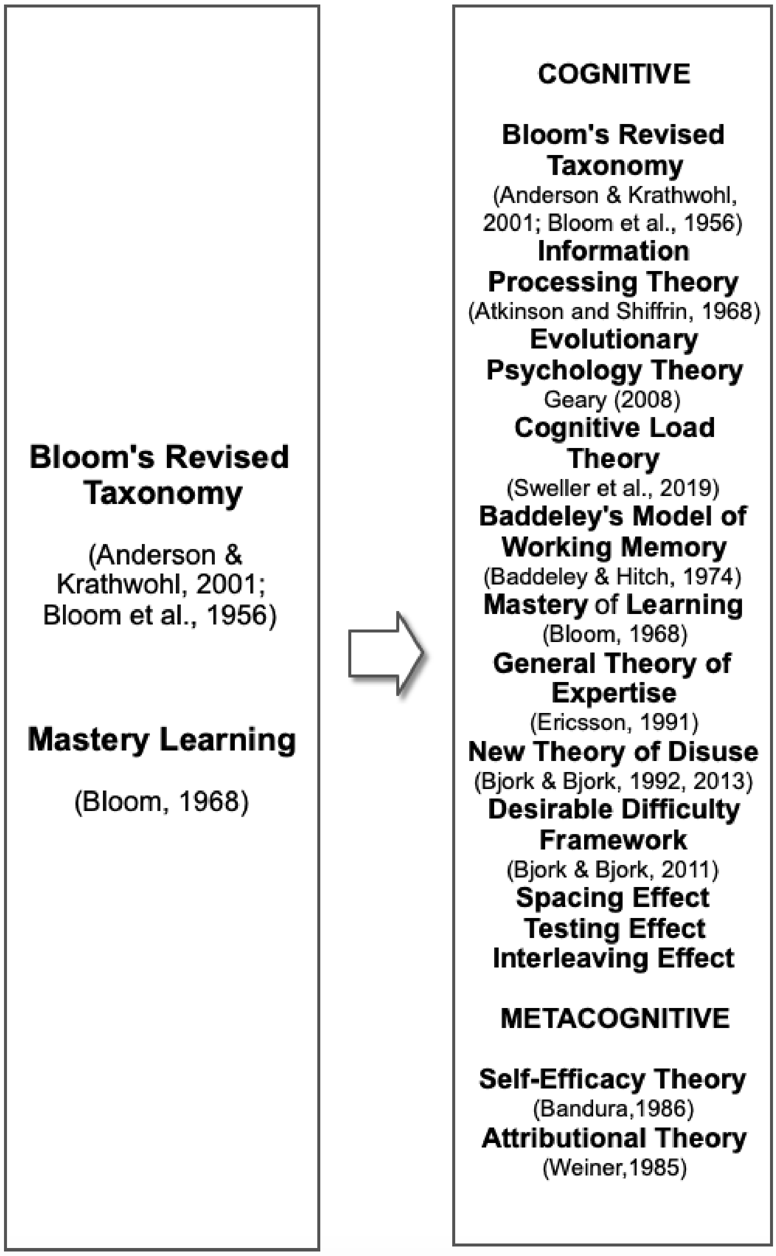
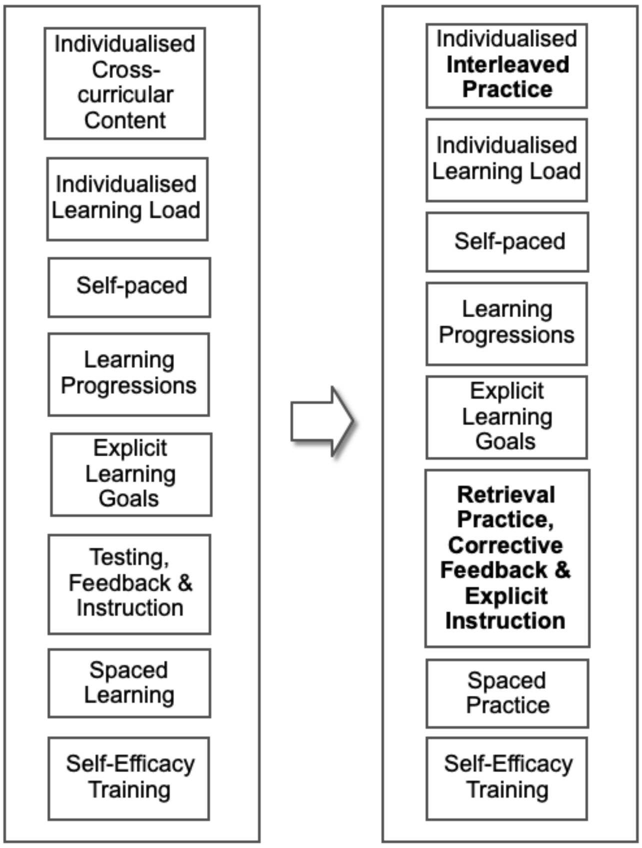


*Figure 15.* The baseline and revised DLCP component assumptions.



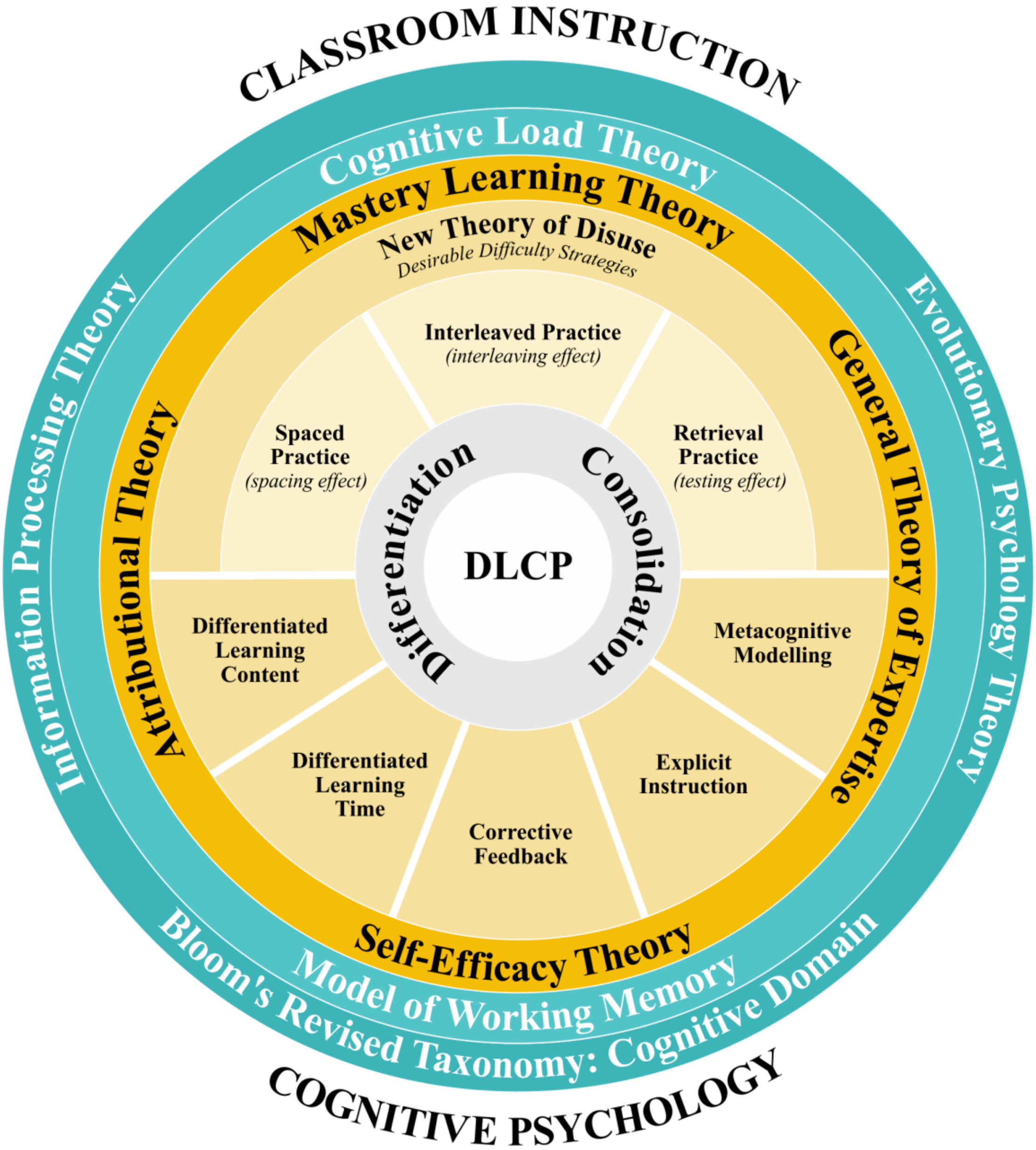

***Programme Theory Combinations***

The purpose of realist synthesis Stage 1(c) continues with the abstraction of identified theories into related theoretical categories, combinations and subsets. Appendix K displays a table that assisted the work of interpreting commonalities and relationships between the DLCP theoretical components which lead to the DLCP theoretical model.

**A DLCP Theoretical Model.**

A model of the DLCP theoretical assumptions is presented in Figure 16. Concentric rings identify the theoretical scope, moving from broad to specific in terms of direct impact on the DLCP. It is acknowledged that this model is introductory and may be modified as a result of future research.

*Figure 16.* A proposed theoretical model of the differentiated learning consolidation process.



The DLCP is based on the content of classroom instruction and a cognitive psychology theoretical foundation. Cognitive psychology examines multiple mental operations, such as perceiving, remembering and thinking, all of which have relevance to learning and the DLCP (Agarwal & Bain, 2019). Context and meaning are provided through (a) the classroom learning objectives, (b) classroom instruction, and (c) differentiation that relates individual student prior knowledge to learning progressions. Three theoretical constructs of knowledge form the broadest tier of the DLCP theoretical model.

Information processing theory defines human cognitive architecture in terms of three related memory stores: the sensory register, the short-term store and the long-term store (Atkinson & Shiffrin,1968). Evolutionary psychology theory partitions knowledge into biologically primary knowledge, and of relevance to the DLCP, biologically secondary knowledge gained through instruction and learning effort (Geary, 2008). Bloom's revised taxonomy (Anderson & Krathwohl, 2001; Bloom et al., 1956) communicates the importance of a knowledge foundation, across all learning objectives (Krathwohl, 2002). The DLCP goal of automated domain-specific knowledge may be used for the analysing and evaluating of information, as well as the creation of novel knowledge applications. The next tier includes Baddeley's model of working memory and cognitive load theory.

Baddeley's model of working memory expands the concept of short-term memory to include both storage and thinking (Baddeley & Hitch, 1974). Sweller (1988) introduced cognitive load theory to address the constraints of limited working memory through instructional design. Well-designed instruction will assist students to process new knowledge and move it into long-term memory through teacher management of the intrinsic load, the extraneous load and element interactivity. Relevant to the DLCP, it encompasses the need to address prior knowledge and to present instruction explicitly (Sweller et al., 2019). Moving inwards, the next tier addresses theories which are more specific to the DLCP.

The cognitive theories of expertise and mastery learning recognise the necessity of practice for knowledge automaticity (Ericsson & Pool, 2016). Mastery learning (Bloom, 1968) highlights the use of responsive instructional methods, success criteria and the provision of differentiated time to achieve learning goals according to individualised learning needs. Self-efficacy and attribution theories inform metacognitive development strategies.

Bandura's (2008) self-efficacy theory relates the achievement of mastery learning goals to the development of learning resilience and motivation. Additionally, it explains the potential metacognitive benefits of the effort and persistence that practice requires. The inner tiers describe the key components of the DLCP and the theories and learning strategies of direct relevance.

The new theory of disuse (Bjork & Bjork, 1992) introduces the strategies of desirable difficulty which, through effortful thinking, facilitate the movement of learning from working memory to long-term memory. The desirable difficulty strategies of spaced, retrieval and interleaved practice may form the chief mechanisms of learning through the DLCP.

**Theoretical Research Focus**

Stage 1(c) of a realist synthesis recommends the selection of a mid-range theory to narrow the research scope in addressing the 'why', 'when' and 'how' questions (Pawson, 2006). The selected theory is the new theory of disuse (Bjork & Bjork, 1992, 2013). This theory explains the movement of learning from working memory to long-term memory through the management of desirable difficulty. The desirable difficulty learning strategies of spaced practice, retrieval practice and interleaved practice, together with metacognitive development strategies, may be key mechanisms by which the goals of the DLCP are addressed and were therefore selected as the research focus of Section 2: Theory Application.

## Section 2: Theory Application

The DLCP was initially developed through informal action research in the absence of thorough theoretical understandings. Curiosity about the reasons behind the observed positive student results led to the current study to discover the DLCP theoretical foundation and potentially facilitate evidence-informed revision. The identified DLCP programme theory was presented in Chapter 5 and displayed as a theoretical model in Figure 16. Theoretical investigation suggested that the DLCP fit within the cognitivist theoretical framework and the field of cognitive psychology. Cognitive load theory was identified as an overarching construct of the DLCP, however, realist synthesis recommends the selection of a lower, mid-range theory of focus (Pawson et al., 2004). The new theory of disuse (Bjork & Bjork, 1992) and the theoretical framework of desirable difficulty (Bjork & Bjork, 2011) were selected.

The focus of Section 2: Theory Application are the DLCP mechanisms of action, the learning and metacognitive strategies which may have facilitated the outcomes observed in the researcher's classroom. Spaced practice, retrieval practice, interleaved practice and related metacognitive strategies were selected to test and refine the programme theory.

The synthesis of spaced retrieval (Chapter 6) and interleaved practice (Chapter 7) data used the methods of systematised literature review (Grant & Booth, 2009) and qualitative content analysis (Hsieh & Shannon, 2005) to investigate quantitative theoretical data, analyse and synthesise results (tabulated and narrative data) and apply findings as appropriate to the DLCP. The investigation of these strategies revealed the associated importance of metacognition. Using the metacognitive themes and results found within the previously addressed spaced retrieval and interleaving studies, research data on metacognitive development strategies (Chapter 8) were based on predominately qualitative studies accessed through secondary sources such books and synthesized research reports. These were explored through a narrative review. The goals of the systematised and narrative reviews of Section 2 were to:

- identify theory and studies of greatest relevance to the DLCP,
- assess prerequisite conditions,
- ascertain which results were theoretically applicable and practical within the process and broader intervention context,
- identify potential learning benefits and
- provide the reader with sufficient information to evaluate and adjudicate the conclusions (Pawson et al., 2004).

The format within each chapter of Section 2 was responsive the type of data and the nature of the application to the DLCP. In Chapter 6 Spaced Retrieval Practice, specific studies related to multiple individual or subsets of process components. Process maintenance and modifications were described accordingly. Chapter 7 Interleaved Practice, applied to only one component of the DLCP: the sequencing of flashcards. In this chapter, studies were divided according to the discriminability—low or high—of the learning content. In Chapter 8 Metacognitive Development Strategies, data was applied to the DLCP according to the three classifications of metacognition (Dunlosky & Metcalfe, 2009): metacognitive monitoring, knowledge and control.

The goal of the research synthesis was to derive an understanding of the mechanisms-of-action of the selected learning strategies and their application to the DLCP. The combination of cognitive load theory, as a ‘science of instruction’ and the new theory of disuse, a ‘science of learning’ (Desy et al., 2018), assisted the researcher to understand both instructional design factors and the facilitation of learning. The new theory of disuse and desirable difficulty framework is revisited as a foundation for the proceeding chapters.

**New Theory of Disuse: Desirable Difficulties**

Bjork (1994) introduced the term 'desirable difficulties' to describe the effortful thinking that his research identified led to durable learning. Desirable difficulties are thought to enhance encoding (storage strength) and recall (retrieval strength) as well as the potential to apply knowledge in novel contexts (transfer) "to yield knowledge and skills that are durable and flexible" (Bjork & Bjork, 1992, p. 109). Complementing cognitive load theory, desirable difficulty leads to the enhanced consolidation of schemas: the reorganising and stabilising of memory traces and connections with prior knowledge (van Merriënboer & Kirschner, 2018).

Desirable difficulties are thought to facilitate consolidation through a measure of the natural process of forgetting. The resultant mental effort to *reconstruct* knowledge from long-term memory strengthens the memory trace, modifying and connecting it with new learning. Termed reconsolidation, the struggle to reconstruct a memory may negatively impact performance in the present, but the thinking that is facilitated deepens learning and enhances the ability to retrieve knowledge in the future (Brown et al., 2014). Conversely, knowledge that is easily gained, understood and demonstrates successful retrieval in the short term, for example within the time frame of instruction, may not be durable in the long term. Illusions of knowledge may come undone when knowledge is assessed after an educationally relevant delay, such as an end-of-year exam (Bjork & Bjork, 2011).

It was suspected that the strategies of spaced practice, retrieval practice, interleaved practice and metacognitive development, had an integral involvement in the DLCP and that their mechanisms-of-action may be responsible for the student outcomes observed in the classroom. The theoretical investigation sought to identify potential revisions of the DLCP to create a more evidence-informed instructional design.

# Chapter 6: Spaced Retrieval Practice

Spaced retrieval practice combines spaced practice and retrieval practice. Initially these strategies will be discussed separately, however, the proceeding synthesis will focus on the combination.

## Spaced Practice

Some learning material is incidentally spaced and revisited within the classroom lesson context (Pashler et al., 2007). For example, a synthetic phonics programme usually introduces a series of phonics sounds which are first practised separately and then through the decoding of phonetic words. This is followed by reading practice using decodable books and ultimately, general reading. Similarly, mastered mental mathematics strategies may continue to be practised through their application in evolving mathematical contexts. Spaced practice is particularly beneficial when learning is not intrinsically reviewed (Pashler et al., 2007).

Researchers agree that spaced practice can enhance the durable learning of a wide variety of educational material (Küpper-Tetzel et al., 2014). Memory effects have been investigated and demonstrated for memorization tasks (Mettler et al., 2016), verbal learning tasks, mathematics and natural science learning materials (Küpper-Tetzel et al., 2014). Learning objectives include remembering, conceptual understanding and educationally relevant skills. Typically, however, learning in mainstream education rarely takes full advantage of the benefits identified by research in spaced practice (Dempster, 1988; Küpper-Tetzel et al., 2014; Mettler et al., 2016; Weinstein, 2019). Several reasons have been identified.

Instructional sequences invariably follow discrete curriculum objectives and resource materials (Taylor & Rohrer, 2010). Mathematics lessons and practice often follow the blocked sequence of textbooks and science curriculums move from one topic to the next.

Spiral curriculums aim to return to prior learning, however, the gap between presentations, potentially a year, may be broader than the capacity of students to remember the previous instruction (Mettler et al., 2016). Like the researcher, teachers may not recognise the difference between performance evidenced during instruction and long-term durable learning (Bjork, 1999) and may therefore overestimate the retention of knowledge. Multiple classroom factors may challenge the use of systematic spaced practice within lessons: group delivery (Kapler et al., 2015), crowded curriculums (Dinham, 2014), variations in prior knowledge and achievement and the logistics required to customize schedules of practice (Mettler et al., 2016). Yet, the learning benefits of spaced practice are undisputed (Chen et al., 2017).

Effect sizes resulting from spaced practice are large: Cohen's $d = 0.71$ (Hattie, 2009) to $d > 1$ (Cepeda et al., 2006), which makes incorporation of the strategy within educational contexts worth pursuing (Küpper-Tetzel et al., 2014). The scientific community notes, however, that there is a "missing bridge between research and practice" (Küpper-Tetzel et al., 2014, p. 72).

Within research investigations, spacing effects are compared to massed learning effects (Küpper-Tetzel et al., 2014). Massed instruction and practice occur within a single session or time frame without subsequent or on-going review. Investigations on spaced practice use similar terms including 'distributed practice' and 'inter-study interval effects' (also known as lag or gap effects) (Cepeda et al., 2006; Cepeda et al., 2009). Inter-study interval effects on learning have been identified for durations of minutes (Rowland, 2014), days, weeks, months and years (Dempster, 1988).

***Theories***

Spaced practice is yet to be explained by a unified theory that covers all aspects of empirical findings (Küpper-Tetzel et al., 2014; Smith & Scarf, 2017, Chen et al., 2017). The effects are linked to four non-mutually exclusive theories: contextual / encoding variability, the reconsolidation account, working memory resource depletion effect and study-phase retrieval. The final theory will be reviewed in the following section on retrieval.

**Contextual / Encoding Variability.**

Contextual variability proposes that factors present at the time of learning are associated with the memory (Glenberg, 1979). Such factors include aspects of the physical environment, like location, sounds or smells, the time of day, the mood of the learner and the way the material was studied. If practice is spaced, different contextual aspects will be associated with the memory trace over time. Multiple contextual factors may act as cue to retrieve the learning after a delay. This theory complements with the new theory of disuse which links the cramming of study with high performance (Bjork, 1999) through the likelihood of contextual overlap within a short time frame. Likewise, positive results at a delayed test may be explained by wider contextual variability over time when practice is spaced (Küpper-Tetzel et al., 2014).

**Reconsolidation Account.**

The reconsolidation account suggests that spacing builds the memory consolidation process across multiple presentations. Initial and subsequent repetitions of the same learning material integrate, resulting in an enhanced memory and the creation of more durable learning (Smith & Scarf, 2017). Sleep is thought to play an active role in consolidation: reactivating memories, reducing forgetting and developing generalizations. Like reconsolidation, sleep-consolidation benefits from spaced practice (Brown et al., 2014, Smith & Scarf, 2017).

**Deficient Processing Account.**

The deficient processing account suggests that spacing intervals, when very short (as in massed presentation), results in less focused attention (Mettler et al., 2016). Conversely, spaced presentation of greater durations, refreshes the attention and requires more effortful thinking. A recent theoretical explanation for the spacing effect is related to cognitive load theory (Chen et al., 2017).

#### Working Memory Resource Depletion.

According to cognitive load theory, working memory capacity is consistent, but varies between learners (Sweller et al., 1998). In Chen et al.'s (2017) investigation, the working memory capacity of primary school students was assessed after massed vs. spaced instruction. They identified that working memory capacity depletes after cognitive effort, particularly when the tasks have similar cognitive components and demands, as found when practice is massed. Conversely, their investigation revealed that working memory capacity recovered after the rest provided by spaced practice. Working memory resource depletion theory hypothesises that, with time held constant, massed instruction imposes heavier demands on working memory than learning material that is spaced. They recommend further research to identify if the rest provided by spacing is also applicable to a change in the type of cognitive activity, which may be applicable to the learning strategy of interleaving across domains. The second learning strategy, often associated with spaced learning, is retrieval practice.

### Retrieval Practice

Within the classroom, 'learning' is interpreted as the acquisition and encoding of new curricular content received through materials and activities (Karpicke, 2017). Testing is understood as either formative, to inform student learning needs or summative, to assess student learning. Research in cognitive psychology, however, has identified that retrieving learning from memory in activities such as quizzing, is a learning strategy and has a powerful effect on retention through the interruption of forgetting (Brown et al., 2014).

Testing as a concept, as previously mentioned, is frequently controversial with concerns raised about its place and importance in education (Brown et al., 2014). Helping to counter the more controversial aspects of testing, a key factor in retrieval practice is that it is most effective when presented as a low- or no-stakes event (Agarwal & Bain, 2019) and communicated explicitly to students as a strategy of learning rather than a form of assessment.

Theoretically described as the testing effect, retrieval includes the recall of facts, concepts or skills (Karpicke, 2017). Secondary in effectiveness, but also achieving learning gains, is the recognition of a correct answer, as found in multiple choice tests (Dunlosky, 2013). The synthesis will focus on recall as this is the usual requirement of the DLCP.

Pre-testing on new topics also demonstrates learning benefits through the creation of mental search sets which benefit subsequent study (Kornell et al., 2009). This aspect is relevant to the learning of new items when tested in the first pockets of the DLCP.

Since 2006, there has been a dramatic increase in research interest in retrieval practice, particularly concerning applications to education (Karpicke, 2017). Study designs often include initial instruction, followed by the comparison of two conditions: a retrieval task, such flashcard recall, and a study task such as the rereading of learning material (Brown et al., 2014). A final test is then administered at delays from immediate to several months or more (Karpicke, 2017).

Rowland's (2014) meta-analysis on laboratory studies identified the testing effect as robust. The benefits of retrieval practice have been observed through laboratory memory studies including the recall of word lists, word-pair learning, foreign language vocabulary and memorization of prose (Tran et al., 2015). Classroom studies include history facts (Carpenter et al., 2009), science facts and generalisations (Gluckman et al., 2014), vocabulary (Küpper -Tetzel et al., 2014; Sobel et al., 2011), maths facts (Schutte et al., 2015) and mathematics problems (Lyle et al., 2020).

### *Theories*

Retrieval practice enhances retention, with or without feedback, and stronger effects are created by free recall than recognition tasks (Carpenter & Delosh, 2006). Research points to the ‘act of retrieval’ as the mechanism of the testing effect. The effect is thought to be influenced by two general factors: desirable difficulty that results in retrieval effort (Bjork & Bjork, 2011) and the initial successful retrieval of the target information (Karpicke, 2017). Multiple, potentially non-mutually exclusive, theories are proposed for the effects of retrieval practice.

#### Transfer-Appropriate Processing.

Whilst not directly related to the mechanism of learning, transfer-appropriate processing theory highlights the advantage of practising the process required to recall a memory: the act of retrieval (Karpicke, 2017). Similarly, study-phase retrieval theory is also descriptive rather than explanatory, however, it is a prominent theory in spaced retrieval research literature.

#### Study-Phase Retrieval Theory.

Study-phase retrieval proposes that subsequent memory performances benefit from the ability to retrieve previous learning encounters with the same material (Kang et al., 2014). Complementing the new theory of disuse (Bjork & Bjork, 1992), spacing can be adjusted to manipulate the optimal effort (desirable difficulty) required (Mettler et al., 2016). When an item of low retrieval strength (low memory accessibility) is retrieved, the storage strength is enhanced benefiting future retrievals. Successful retrieval over time, reduces forgetting as compared to restudy conditions (Küpper-Tetzel et al., 2014). Also related to the desirable difficulty framework, Pyc and Rawson (2009) introduced the retrieval effort hypothesis which states that the optimal degree of difficulty will be achieved at the longest interval at which a successful retrieval can be achieved.

### Elaborative Retrieval Account.

When Carpenter and Delosh (2006) manipulated memory cue support, they identified that fewer cues enhanced retention. In the absence of cues (for example, non-multiple choice tests), the elaborative retrieval account suggests that less accessible memories result in greater processing (desirable difficulty), and therefore recall on a delayed test, as compared to easier restudy. This processing may enhance meaningful cues that link to the memory of successfully retrieved items which in turn, may assist future recall (Karpicke, 2017). The elaborative retrieval account view is supported by results demonstrating better retention at longer inter-study intervals and improved results for recall as compared to recognition retrieval tasks. The contextual / encoding variability theory is also related to memory cues.

### Contextual / Encoding Variability.

In addition to assisting to explain the spacing effect, the contextual / encoding variability theory is associated with retrieval practice. Like the spacing of to-be-learned material, it is assumed that contextual information is encoded at an initial retrieval practice event (Karpicke, 2017). During on-going retrieval sessions, memory searching includes seeking the previous contextual cues. When successfully retrieved, the initial context is upgraded with cues from the new context which continues through future retrievals.

Rowland's (2014) meta-analysis on retrieval practice focused on theoretical characteristics. He identified that whilst the exact mechanisms remain elusive, the results support the retrieval effort class of theories. Visible Learning (2019) identifies retrieval testing as having an effect size of 0.46.

## Synthesis

Spaced retrieval practice maximises the benefits of spacing and retrieval (Kang et al., 2014), interrupting forgetting and producing greater retention of learning at educationally relevant delays (Karpicke, 2017). Spacing and retrieval effects are frequently understood in terms of similar psychological mechanisms (Dempster, 1996).

Spaced retrieval studies often involve initial retrievals to establish a baseline criterion, an inter-study interval and second retrieval session, followed by a delay (the retention interval) prior to a final test. A delayed test avoids the confounding effect of working memory. These conditions may be compared with a single massed presentation that combines the total learning time, or the comparison of other variables such as duration of inter-study or retention intervals.

Two research areas address spacing effects: within-session and between-session inter-study intervals (Küpper-Tetzel et al., 2014). Within-session spacing involves a collection of to-be-learned items that are interspersed within a list. Spacing is achieved through one presentation in time followed by subsequent non-juxtaposed presentations over seconds or minutes. This form of spaced practice is termed interleaving. Between-session inter-study intervals involve durations ranging from day(s) to years (Küpper-Tetzel et al., 2014). The DLCP includes both types of spacing. Interleaving studies addressing spacing effects are included in this chapter. A more detailed discussion on interleaving and studies related to the nature of the learning material, will be discussed in the next chapter.

Research designs involving more than one session may use or compare different types of inter-study intervals of equal, expanding, contracting or adaptive durations. Optimising the space (lag effects) between retrieval sessions may have "substantial implications for real-world learning" (Mettler et al., 2016, p. 897).

The selected synthesis studies are focussed on relevance to the DLCP.
Specific goals are to:

- investigate initial mastery criterion parameters,
- identify the optimum type and duration of inter-study intervals and to

- determine retrieval testing best practice.

Two hundred and twenty-six citations were identified as relevant to spaced and or retrieval practice. SCOPUS was used to refine the search, identifying prominent authors and citation frequency. Subsequent snowballing and pearling techniques provided the most relevant citations. Purposive sampling identified thirty-four papers for abstract reading. Twenty-seven papers were read and eleven papers of relevance to the DLCP were selected for the data synthesis.

The DLCP intervention is displayed in Figure 4. Learning content is based on classroom instruction and the process is preceded by lesson-based activities and practice. The original process was divided into two phases, the Tutoring Phase and the Mastery Test Phase, delivered through a combination of tutor and teacher.

Each spaced retrieval session is both a learning event, through the desirable difficulty strategies, and an assessment of mastery through retrieval testing and corresponding flashcard movement through the tool. The tool consists of 12 pockets. Eight were originally used by the process and four were unallocated. Aspects of the process relevant to the spaced retrieval synthesis include:

- learning objective complexity,
- desirable difficulty management,
- learning load influence,
- response time criterion,
- inter-study interval optimisation,
- feedback and elaboration conditions and
- the nature of explicit instruction.

The synthesis begins with an analysis of the original Tutoring Phase.

***Tutoring Phase***

As displayed in Figure 17, the Tutoring Phase commenced with a teacher judgement on the appropriate quantity of learning items (the learning load). Learning load was thought to be a variable that could be manipulated according to student aptitude; it was presumed that fewer flashcards would make the mastery criterion easier to achieve. For any given learning item, the number of spaced retrieval sessions for each student was dependent on accuracy and response time (≤ 4 seconds). The mastery criterion was five consecutive successful recalls per learning item, which, at one retrieval test per session, required a minimum of five spaced retrieval sessions, after which, the item moved into the Mastery Test pocket to be assessed 1 – 7 days later. Unsuccessfully retrieved items, or those demonstrating a response time of over 4 seconds, resulted in those learning items returning to the Entry pocket.

Table 6 contains the research study summaries of relevance to the Tutoring Phase. These studies are interrogated to demonstrate their relevance to different aspects of the DLCP and to determine if they can inform evidence-informed process maintenance or modification. Study descriptions to enable stakeholders to evaluate and adjudicate the conclusions.

*Figure 17*. The original DLCP spaced retrieval sessions.

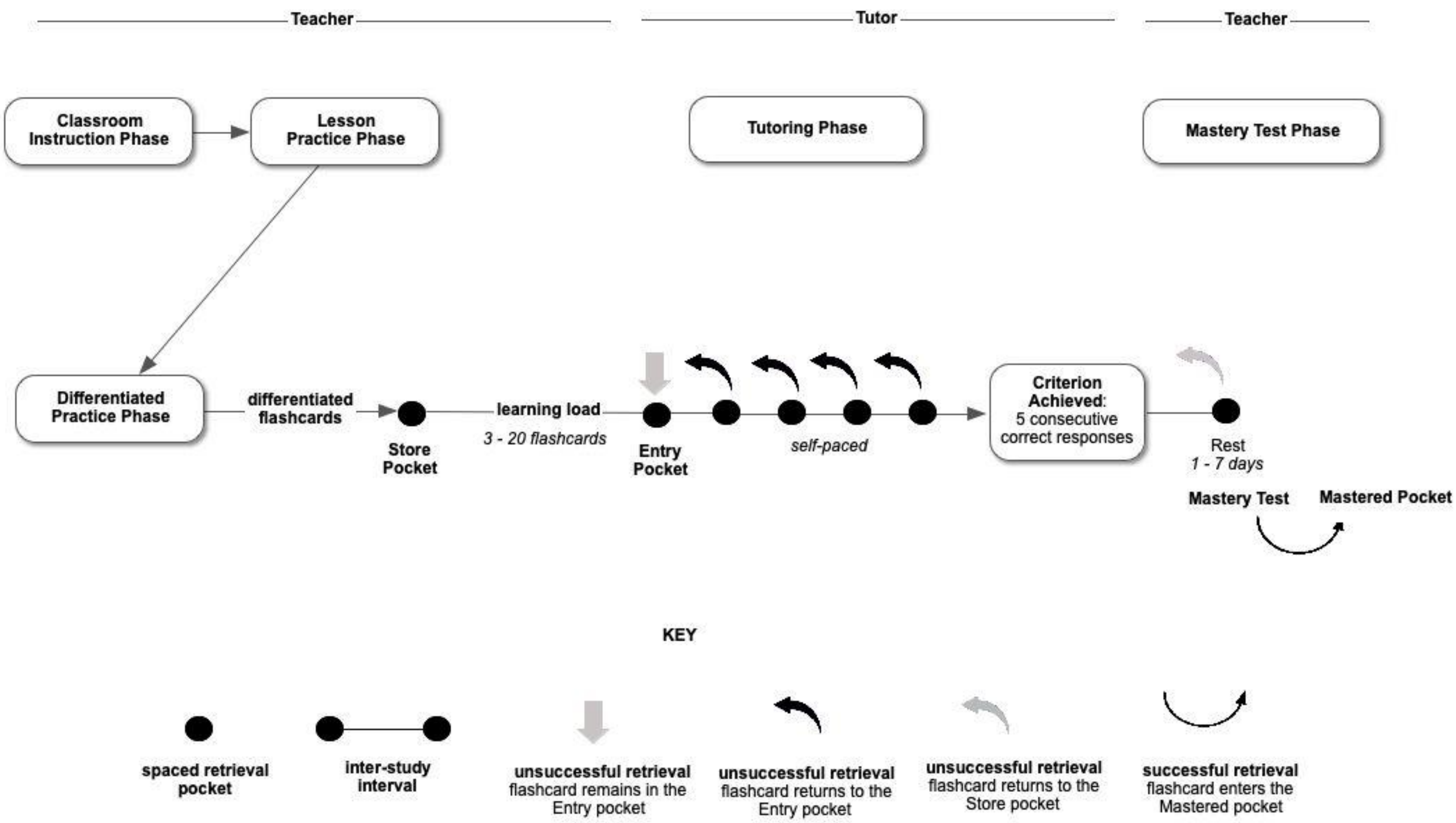


Table 6

*Synthesis Studies of Relevance to the Tutoring Phase*

| Study | Research goals and participants | Topic and approach | Findings |
| --- | --- | --- | --- |
| Learning Objectives:<br><br>Gluckman, M., Vlach, H. A., & Sandhofer, C. M. (2014). Spacing simultaneously promotes multiple forms of learning in children's science curriculum. *Applied Cognitive Psychology*, *28*(2), 266-273.<br><br>Lesson conditions:<br>massed<br>clumped<br>spaced | To examine the learning effects of spaced vs. massed lesson-based science instruction on memory (facts), simple generalisation (concepts) and complex generalisations (interrelated concepts).<br><br>*n* = 32<br>Classroom context<br>Year 1 and 2 students<br>Quantitative | Instruction on food chains within different biomes including information, concrete materials, demonstrations and explanations.<br><br>Pre-test<br><br>Lessons in three conditions<br><br>1-week delayed post-test (free recall, cued recall and forced choice questions) | Student results in the spaced condition outperformed the massed and clumped conditions on memory and generalisation questions.<br><br>Results in memory and generalisation were not significantly related, suggesting spacing supports different cognitive process for each form of learning. |

| Study | Research goals and participants | Topic and approach | Findings |
|---|---|---|---|
| Desirable Difficulty:<br><br>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. | Experiment 1. To test the retrieval effort hypothesis that successful but difficult retrievals enhance memory more than easier successful retrievals.<br>$n$ = 129<br><br>Experiment 2a. To repeat Exp. 1 using response time as a metric.<br>$n$ = 98<br><br>Undergraduates<br>Quantitative | The difficulty level of Swahili-English translation word pairs was manipulated by adjusting within-session (interleaved intervals) and initial criterion.<br><br>Learning event design: retrieval attempt + repeated retrievals of incorrect recalls to one successful retrieval within one session<br><br>7 initial study criterion levels<br>short vs. long interleaved intervals<br><br>short retention test: 25 mins<br>long retention test: 1 week | Longer interleaving intervals (increased difficulty) increased retention.<br><br>Increased initial criterion levels improved retention in the short term but the effect diminished with cumulative retrieval trials.<br><br>Conditions in which retrieval was more difficult produced greater memory benefits as compared to easier successful retrievals.<br><br>Response time was confirmed as an indicator of desirable difficulty. |

| Study | Research goals and participants | Topic and approach | Findings |
|---|---|---|---|
| Learning Load:<br><br>Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2017). Extending cognitive load theory to incorporate working memory resource depletion: Evidence from the spacing effect. *Educational Psychology Review, 1-19*.<br><br>Conditions:<br>massed practice (40 min session)<br>spaced practice (4 days) | To investigate the relationship between the spacing effect and working memory depletion.<br><br>primary school students<br><br>Experiment 1<br>*n*= 54<br>Year 4 students<br>Experiment 2<br>*n* = 61<br>Year 5 students<br>Quantitative | The addition of positive fractions with different denominators.<br><br>Schedule:<br>Worked example slides<br>Working memory test<br>Content post-test | Using a working memory capacity test, students obtained lower cognitive load ratings for spaced practice compared to massed practice.<br><br>Reduced working memory capacity may be attributed to resource depletion resulting from the greater mental effort associated with the massed condition. Conversely, working memory capacity may be reinstated through the rest provided by spacing. |

| Study | Research goals and participants | Topic and approach | Findings |
|---|---|---|---|
| Criterion Level:<br><br>Rawson, K., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General* 140, 283-302. https://doi.org/10.1037/a0023956 | To determine an optimum practice schedule that achieves durable and efficient learning for both short term and long-term retention.<br><br>Educationally relevant spaced retrieval across ≤ 10 days<br><br>Experiments 1, 2, 3<br>*n* = 130, 68, 335<br><br>Undergraduates<br>Quantitative | Three variations of a similar experiment:<br>Criterion levels - 1 to 4 within-session (interleaved) successful retrievals compared across 1 – 5 subsequent relearning sessions with different final test delays.<br><br>Learning event design:<br>retrieval attempt + repeated retrievals of incorrect recalls to one successful retrieval<br><br>Schedules were assessed for efficiency and learning durability. | Criterion Level Efficiency:<br>Initial learning (within session interleaving): practice to three successful recalls benefited early relearning trials with diminishing effects over cumulative trials.<br><br>Learning Durability:<br>At 1-month and 4-month delayed tests:<br>With the above criterion levels, three spaced retrieval sessions were optimum for efficiency and learning durability. |

| Study | Research goals and participants | Topic and approach | Findings |
|---|---|---|---|
| Adaptive Inter-Study Intervals:<br><br>Mettler, E., Massey, C. M., & Kellman, P. J. (2016). A comparison of adaptive and fixed schedules of practice. *Journal of Experimental Psychology: General*, *145*(7), 897-917. https://doi.org/10.1037/xge0000170 | Experiment 1. To determine if adaptive schedules of practice outperform fixed schedules (equal or expanding).<br>*n* = 72<br><br>Experiment 2. To determine if the benefits of adaptive schedules relate to individual students or learning items.<br>*n* = 48 | In a single session, learning items and filler examples required participants to match African country names with location using interleaving with an adaptive or fixed presentation schedule.<br><br>Pre-test<br>Learning phase with feedback<br>Immediate test<br>Delayed test (1 week) | Adaptive scheduling demonstrated greater learning gains at the delayed test, increased response times and reduced forgetting between tests:<br><br>Adaptive intervals M= 0.42<br>Fixed expanding intervals M = 0.31<br>Fixed equal intervals = M = 0.30<br><br>Greater learning gains are achieved when spacing is adapted according to on-going student performance, not learning item difficulty. |

| Elaboration Reviews |
|---|
| Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58. |
| Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction, 19*(5), 402-410. |
| Roediger III, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition, 1*(4), 242-248. |
| van Merriënboer, J. J., & Kirschner, P. A. (2017). *Ten steps to complex learning: A systematic approach to four-component instructional design*: Routledge. |

| Explicit Instruction Reviews |
|---|
| Chen, O., & Kalyuga, S. (2020). Exploring factors influencing the effectiveness of explicit instruction first and problem-solving first approaches. *European Journal of Psychology of Education, 35*(3), 607-624. https://doi.org/10.1007/s10212-019-00445-5 |
| Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction, 19*(5), 402-410. |

**Learning Objectives.**

Prior to the researcher's development of the DLCP, traditional mastery learning folders were frequently used as tools for fact learning. Likewise, research on spaced practice has traditionally focused on factual memory tasks, however in the last decade, studies have begun to investigate higher order learning objectives, such as generalisations. Generalisation tasks involve presenting learners with a range of information from which they are required to develop abstractions and applications to new contexts (Gluckman et al., 2014). A study by Gluckman et al. (2014) investigated if spaced practice could benefit both memory and generalisation learning objectives. The results of this study may inform the use of the DLCP beyond the learning objectives of remembering, understanding and applying (Anderson & Krathwohl, 2001; Bloom et al., 1956).

Set within the classroom, Gluckman et al.'s (2014) study used Year 1 and 2 elementary school participants studying food chains in the context of different biomes (grasslands, artic, ocean, swamp and desert). Lesson instruction included introductory science facts, concrete materials such as proportionally correct animal figurines, demonstrations and explanations. The topic facilitated the learning of information (vocabulary and facts), simple generalisations (large animals eat smaller ones) and more complex generalisations (food chain interdependency). Three spaced conditions were investigated:

- massed (no spacing) – four lessons in immediate succession,
- clumped – two lessons on one day and two lessons on the next and
- spaced – one lesson per day over four days.

One week after the final lesson for each condition, students received a post-test. Questions included free recall, cue-recall (memory tasks) and forced choice questions which assessed both memory and generalisation objectives.

The students in the spaced learning condition significantly outperformed the other conditions. Analysis revealed that memory and generalisation scores were not significantly related, indicating that the improvement in memory tasks was not responsible for the generalisation results. The authors suggest that different cognitive processes are involved for each type of learning with spaced practice supporting both lower and higher order learning objectives.

Within the Year 1 classroom, the foundational learning objectives of remembering, understanding and applying were the focus of the DLCP. However, given the results of Gluckman et al.'s (2014) study, practice material to facilitate higher order learning objectives may be possible if the tutor has the capacity to accurately assess the answers and provide relevant feedback. Student leader tutors, for example, may not have the appropriate level of knowledge. Future research may inform the potential of the DLCP to facilitate learning gains for objectives at higher levels of complexity.

**Desirable Difficulty.**

Desirable difficulty (Bjork & Bjork, 2011) is created through spacing the retrieval of learning items. Pyc and Rawson (2009) sought to further define the characteristics of desirable difficulty. They designed an investigation to test their retrieval effort hypothesis which states that "difficult but successful retrievals are better for memory than easier successful retrievals" (p. 437). Their study used interleaved learning items which were subject to different conditions.

Studies within the desirable difficulty framework frequently recommend an initial learning benchmark, known as a criterion level, as a foundation of further practice. Pyc and Rawson (2009) used the criteria of 1, 3, 5, 6, 7, 8 or 10 correct retrievals within the initial session to manipulate the difficulty of item retrieval in a subsequent retrieval session. The assumption was that a criterion of 10 correct retrievals would make subsequent recall easier than one successful retrieval. Secondly, they addressed the interleaving intervals between presentations of the same learning item. Half the participants had short intervals between these presentations, the remainder had long. The effects on retention were assessed at two intervals (25 minutes and 1 week).

Pyc and Rawson's (2009) results demonstrated that higher initial criterion levels improved retrieval at the short retention interval. This result is consistent with the desirable difficulty concept that performance during instruction produces short-term learning gains (Bjork, 1999). However, as the initial criterion level increased, the desirable difficulty decreased, as did retention at the 1-week delayed test, creating what is described as a curvilinear relationship. Further, longer intervals between presentations of the same item, increased difficulty and improved retention at the delayed test.

Pyc and Rawson's (2009) demonstration of the retrieval effort hypothesis highlights the relevance and importance of desirable difficulty management. This concept is of fundamental importance to the DLCP and will be addressed in different aspects of the process throughout the synthesis. Previously, the quantity of active learning items (learning load), was understood to be the major variable in the management of difficulty.

**Learning Load.**

Prior to first use of the DLCP, the researcher made a judgement on the appropriate number of active flashcards to be learnt by each student in any given session (the learning load) based on the learning aptitude of each student. The quantity ranged from a minimum of three flashcards, to over twenty. At the mastery test, the learning load could be adjusted.

In the researcher's experience, learning items that had arrived in the Mastery Test pocket usually had a high retrieval success rate. If, over the period of a few weeks, a student did not demonstrate this level of mastery, the first response was to ask the tutor how they were delivering the process. If the process was being followed faithfully, the researcher would respond by reducing the learning load, believing that less learning items would reduce the level of difficulty and so make recall easier. Two studies challenge this assumption.

First, Pyc and Rawson's (2009) investigation classifies desirable difficulty as effortful but successful retrieval. The quantity of learning items would therefore be irrelevant to the difficulty of each learning item. Second, the learning load may impact on each student's individual working memory capacity.

Chen et al.'s (2017) working memory resource depletion hypothesis suggests that working memory capacity is reinstated by the rest provided by spacing, and potentially (though not yet researched) by changes in topic. DLCP learning items are spaced; within and between pocket interleaving provides temporally short spacing, and retrieval sessions are spaced over days. If spacing refreshes working memory capacity, then the learning load may not negatively impact cognitive load. The results of Chen et al. (2017) and Pyc and Rawson (2009), suggest that learning load may be unrelated to the level of difficulty. The learning load does, however, relate to the practical aspect of time-on-task for students, tutors, and teachers during retrieval sessions.

If time was unlimited, students could learn many things. In the classroom context, however, the educator's goal of durable learning is highly restricted by this variable. Learning efficiency is, therefore, a very relevant component of instructional design. Rather than moderating difficulty, the learning load may be adjusted to optimise the overall average duration of the retrieval session.

Issues of time relate directly to the availability of tutors and the practicalities of a single classroom teacher conducting a weekly mastery test with all students. During the Tutoring Phase (students and tutors), the duration of the retrieval session varied between 3 and 8 minutes for up to 20 flashcards per student. In the Mastery Test Phase (students and teacher) the retrieval session was shorter, 2 to 4 minutes, due to only one pocket being tested.

In the classroom delivery mode, based on 5 minutes per student, one tutor could conduct retrieval sessions with approximately four or more students within the 20 minutes available before the school bell. In this case, approximately six tutors would be required to cover a class of 24 students. The researcher found this number to be an achievable goal, given the variety of tutor options described in Chapter 4. Concurrent Mastery Test Phase testing by the teacher may be staggered across the week at a rate of five students a day, or scheduled during other suitable timeslots, such as during independent student activities. If available, education assistants may also be involved. Based on the scenario described[2], reducing the learning load by five learning items, may result in an approximate 1-minute gain per student which would reduce the overall average duration of the retrieval session to around 16 minutes. Therefore, the learning load could be used to adjust the time-on-task according to time and tutor availability.

**Response Time.**

Pyc and Rawson's (2009) second experiment sought to replicate their Experiment 1 results, using response time rather than accuracy to determine difficulty. The items with lower criterion levels were more difficult and demonstrated longer response times which decreased with increasing criterion levels. Response time was also longer for items in the longer interval condition. These results confirmed that response time is an indicator of difficulty and the findings of Experiment 1.

In the original DLCP, each retrieval was interpreted only as a test of mastery. The researcher estimated that a response time of ≤ 4 seconds was considered to be a demonstration mastery for any given learning item. However, the new understanding that

---

[2] 24 students and 6 tutors = 4 students per tutor

20 flashcards ≈ 5 minutes per student ≈ 20 minute overall retrieval session duration

every spaced retrieval session within the DLCP is also a learning event, has implications for response time duration. Pyc and Rawson's (2009) study demonstrated that (a) successful but effortful retrievals had greater impact on retention than easy or unsuccessful retrievals and (b) both accuracy and response time were indicators of difficulty. The misplaced emphasis on mastery reduces the thinking time available for effortful retrieval, short-changing the potential benefits of desirable difficulty, therefore, the DLCP response time expectation requires modification.

Changing the emphasis from mastery to effortful retrieval could be achieved by increasing the available response time (within practical limits). The researcher estimates that ≤ 10 seconds would provide sufficient time for both effort and criterion assessment, which may be assessed by future research. Response time could also be used to adjust the desirable difficulty: increasing the time available for struggling students or reducing it for students demonstrating quick correct answers. The concept of learning efficiency has been investigated through attempts to optimise schedules of practice according to criterion level.

**Criterion Level.**

Many spaced retrieval practice studies establish an initial criterion level for to-be-learned items, often for the methodological purpose of establishing a uniform baseline, however, a small number of studies have compared the effects of initial criterion levels with subsequent retention. Pyc and Rawson's (2009) results were based on one interval between the establishment of criterion levels and the delayed tests, however, a study by Rawson and Dunlosky (2011) compares criterion level with retention results over multiple spaced retrieval sessions. These results are of relevance to the DLCP Tutoring Phase.

Rawson and Dunlosky (2011) sought to assess the optimum amount of spaced retrieval practice for durable and efficient learning in three related experiments. Two student learning goals were considered: the need to both pass a course exam (short-term retention) and to maintain the necessary prior knowledge for a subsequent more advanced course (long-term retention: 1- 4 months).

The investigation assessed two variables: the effects of initial learning criterion and the effects of relearning after retrieval. Relearning sessions involved one successful retrieval for any item, or, for unsuccessfully retrieved items, repeated interleaved retrieval to the criterion of one successful recall. Step two of the original DLCP included repeated retrievals with the same goal. For the purposes of this discussion, ‘relearning sessions’ are described as spaced retrieval sessions as each session is spaced and commences with a retrieval attempt.

In the first session, Rawson and Dunlosky (2011) established criterion levels of 1, 2, 3 or 4 successful recalls. This was followed by one to five spaced retrieval sessions to determine the relationship between the initial criterion level and retention results. Learning efficiency was assessed by averaging the total number of retrieval trials per item at each criterion level for comparison with the learning gains at the five spaced retrieval sessions. The results demonstrated a relationship based on the desired durability of the knowledge.

Like Pyc and Rawson (2009), Experiment 1 identified that higher initial criterion levels improved learning over short durations. More specifically, Rawson and Dunlosky (2011) determined a criterion level of three successful recalls was optimum for students to remember learning for events such as a forthcoming test as relative retention gains diminished with additional spaced retrieval sessions (demonstrated in Experiments 2 and 3), a finding that will be further discussed later in the text. However, Rawson and Dunlosky (2011, p. 300) state, “we would still advocate that learners practice to a higher initial criterion” as research has identified that students [and perhaps teachers] are inclined to abandon independent practice after one successful recall, a view also supported by Kornell and Bjork (2007). Additionally, higher initial criterion levels also modestly assist subsequent relearning (Rawson & Dunlosky, 2011).

The concept of an initial criterion level had been previously interpreted by the researcher as a responsiveness to individual student needs; the DLCP content was based on prior knowledge, classroom instruction, ongoing assessment and learning progressions. However, the Tutoring Phase involves spaced retrieval intervals of short duration (day/s) analogous to the establishment of Rawson and Dunlosky's (2011) initial criterion level. The tutoring phase will henceforth be renamed the Criterion Phase, to reflect the initial learning goal of short-term retention over a period of days.

In Experiment 2, three spaced retrieval sessions resulted in a score of 56% on the delayed cued-recall test which was a 75% improvement on the score after one session, demonstrating the benefit of additional spaced retrieval sessions. The effect of higher criterion levels, however, attenuated with further spaced retrieval sessions.

In Experiment 3, Rawson and Dunlosky (2011) increased the number of spaced retrieval sessions, from one to five with a delayed cued-recall test conducted at one and four months. For the 1-month retention duration, an average of seven trials per item were required to meet the criterion in the initial session, though in subsequent sessions, only one or two trials were required to return items to criterion due to faster relearning, indicative of improved efficiency with an increasing number of relearning sessions. Using the time-on-task metric, in the first session, practice of 6.3 mins per item was required to establish the criterion level, but the second spaced retrieval session added only 0.7 minutes to practice time and the third, 2.7 mins. Retention scores improved at each spaced retrieval session, however, relative gains diminished in sessions four and five.

Similar results were achieved at the 4-month delayed cued-recall test[3]. The longer retention interval also identified that more spaced retrieval sessions equalled greater learning

[3] NB This test implicitly included an additional retrieval session due to the 1-month delayed test which was included in the data.

gains, however, as the number of sessions increased, the relative retention gains diminished. In terms of efficiency, smaller gains in retention (2% in Session 5) came at the cost of more time-on-task (additional practice trials): once again demonstrating a curvilinear relationship. This relationship, also demonstrated by Pyc and Rawson (2009), suggests learning efficiency may be optimised by a reduced number of spaced retrieval sessions, for a minimal sacrifice of learning gains.

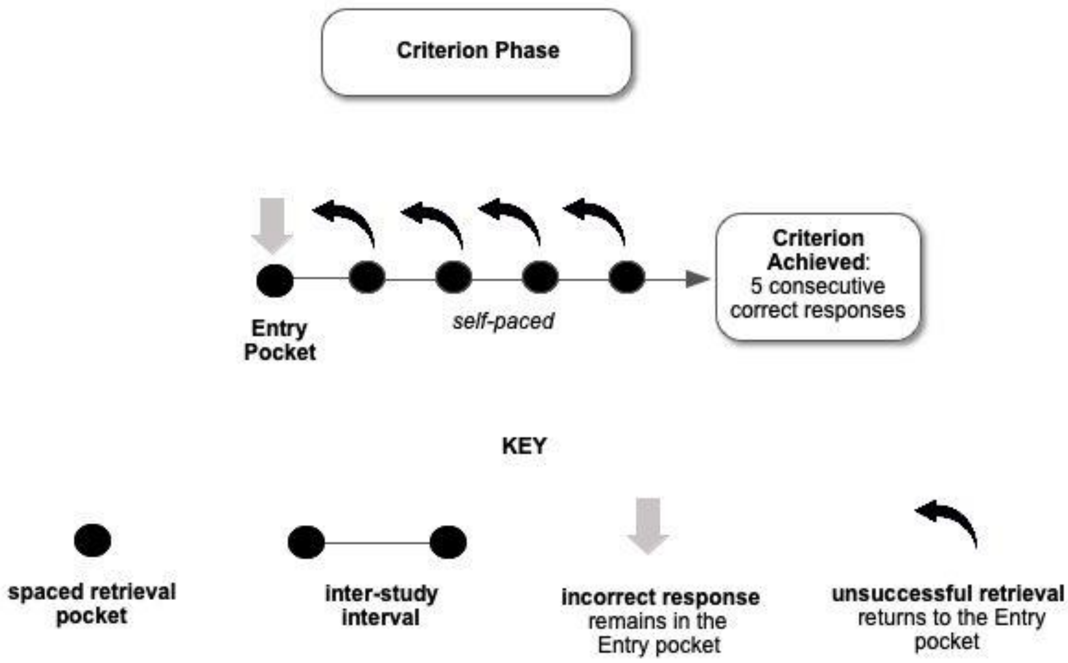
Rawson and Dunlosky (2011) concluded that practice schedules optimised for learning efficiency and durability are dependent on student or teacher learning retention goals. Students inevitably require both short and long-term retention. The authors selected a 3 + 3 schedule to address these goals: practice to the criterion of three correct retrievals during initial learning, followed by three subsequent spaced retrieval sessions. This combination may enhance retention goals and the increased efficiencies that come from a reduced number of retrievals. Whilst the research designs of Pyc and Rawson (2009) and Rawson and Dunlosky (2011) do not directly align with the instructional design of the DLCP, their results provide some principles for potential modifications. First, some differences need highlighting.

Dunlosky and Rawson (2011) and Pyc and Rawson (2009) both used repeated spaced retrieval practice (interleaved practice) within their initial sessions to achieve the required criterion levels. Interleaved practice uses temporally short intervals of seconds to minutes. This is different to the DLCP which involves only one retrieval attempt per session during the Criterion Phase, resulting in longer spaced intervals of hours and days. Whilst the Rawson and Dunlosky (2011) study did not address the variable of spacing, they note that the retrieval test at 1-month provided an expanded retention interval which may have positively influenced the results at the 4-month retrieval test, as greater spacing is associated with improved memory effects (Cepeda et al., 2008; Dunlosky et al., 2013; Kang, 2016). This

DLCP condition may, therefore, not be disadvantageous to comparisons. The effects of spacing will be addressed in a forthcoming section.

Pyc and Rawson (2009) using temporally short interleaved retrieval, and Rawson and Dunlosky (2011) using multiple retrieval sessions, both demonstrated a curvilinear relationship between the number of retrieval sessions and retention, a factor of relevance to the original criterion level of the DLCP. In the original DLCP, if recall was not successful in any spaced retrieval session, the item was returned to the Entry pocket which resulted in additional retrieval sessions. The criterion level was set at five consecutive successful recalls for any given learning item, as illustrated in Figure 18.

*Figure 18.* The criterion level process for the original DLCP.




This high criterion level held the potential to produce multiple spaced retrieval sessions. As an example, an unsuccessful retrieval at the third pocket results in the learning item returning to the Entry pocket. If followed by continuous successful retrievals, a minimum of eight spaced retrieval sessions will be required to reach the criterion level of five consecutive retrievals. Table 7 illustrates the number of spaced retrieval sessions resulting from one unsuccessful retrieval in the third pocket, one to five times.

Table 7

*Number of Unsuccessful Retrievals at Pocket 3 Versus Total Number of Retrieval Sessions with a Criterion Level of Five Consecutive Retrievals*

| Unsuccessful retrievals | Number of Spaced Retrieval Sessions |
|---|---|
| 1 | 8 |
| 2 | 11 |
| 3 | 14 |
| 4 | 17 |
| 5 | 20 |

The range of eight to twenty spaced retrieval attempts is excessive for two reasons. First, Rawson and Dunlosky (2011) demonstrated that criterion levels higher than three successful initial recalls had diminishing returns on retention results. Second, Pyc and Rawson, (2009) demonstrated that as criterion levels increase, desirable difficulties and therefore long-term retention, also decrease. The original criterion level of five consecutive retrievals within the DLCP could, therefore, be counterproductive to each item's desirable difficulty. The principles demonstrated in both studies suggest that reducing the criterion level of the DLCP may improve both learning efficiency and durability.

The criterion level could be reduced in two ways. First, the number of pockets within the Criterion Phase could be reduced from five to three. Second, rather than a requirement for consecutive successful retrievals, successful retrievals could be cumulative. This could be achieved if unsuccessfully retrieved items are returned to their pocket of origin, rather than to the Entry pocket (Figure 19). Using the same scenario as previously described, an unsuccessful retrieval at the third pocket (followed by successful retrievals) would result in a cumulative total of four retrieval sessions rather than eight. Table 8 compares the two criterion level conditions.

*Figure 19.* The Criterion Phase process for the modified DLCP.

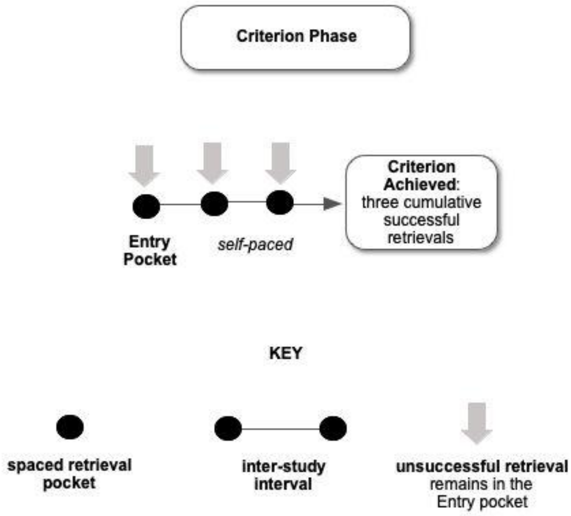


Table 8

*Number of Unsuccessful Retrievals at Pocket 3 Versus Total Number of Retrieval Sessions Comparing Three Cumulative Retrievals with Five Consecutive Retrievals*

| Unsuccessful retrievals | Number of Spaced Retrieval Sessions | |
|---|---|---|
| | Three Cumulative Successful Retrievals | Five Consecutive Successful Retrievals |
| 1 | 4 | 8 |
| 2 | 5 | 11 |
| 3 | 6 | 14 |
| 4 | 7 | 17 |
| 5 | 8 | 20 |

Although this modification may still result in more than three initial successful retrievals (Rawson & Dunlosky, 2011), the increased spacing afforded by only one retrieval per session may increase the desirable difficulty for potential learning gains. Whilst the studied articles suggest a substantial reduction criterion levels and number of spaced retrieval sessions, future research on the Criterion Phase, could investigate optimum learning and efficiency conditions.

**Inter-Study Intervals.**

Inter-study intervals within the original DLCP were set according to practical classroom considerations. For example, during the first year of implementation, the DLCP was a homework strategy and scheduled Monday to Thursday creating four inter-study intervals of 1 day, and an interval of four days over the weekend. In the following year, practice sessions were scheduled prior to the morning school bell. To suit the classroom timetable, these sessions were scheduled on a Monday, Tuesday, Thursday and Friday resulting in two inter-study intervals of 1 day, one of 2 days and another of 3 days over the weekend. Other factors influencing sessions and intervals included school events such as excursions and student absences.

The difficulty of a retrieval is based on the storage strength of a learning item for an individual student at any given point in time (Bjork & Bjork, 1992). The optimum inter-study interval for each student will be a successful retrieval of their learning items at the longest duration (Pyc and Rawson, 2009). Mettler et al., 2016 (p. 898) acknowledges that “it is logistically difficult for educators to customize schedules of practice for individual students and topics”.

Prior knowledge, performance accuracy and response time are indicators of storage strength (Mettler et al., 2011) and can be used for manipulations of desirable difficulty. As previously described, the DLCP seeks to address prior knowledge through teacher assessments and observations. Performance accuracy for any given learning item is addressed through mastery criteria and the movement of flashcards. Flashcards move forward to the next pocket when successfully retrieved or, if retrieval is unsuccessful, are returned to the pocket of origin. To facilitate effortful thinking, a response time of $\leq$ 10 seconds is the revised measure of storage strength for a successful retrieval.

To address storage strength and provide desirable difficulty, Mettler et al. (2011) developed a software-based learning system called Adaptive Response-Time-based Sequencing (ARTS) system. The adaptive system uses an algorithm that assigns a priority score to each item which facilitates dynamically scheduled spacing intervals. Mettler et al., (2016) designed two experiments comparing the ARTS system to fixed scheduling.

In Experiment 1, twenty-four African countries and their map locations were learned (with related unassessed filler items) through spaced retrieval (with feedback) in the interleaved condition. There were two between-subject conditions based on schedule presentation: adaptive spacing (ARTS) or fixed spacing. The fixed space items were further divided into equally spaced or expanding and all items were presented four times. Two primary dependent measures were used across items: accuracy and response time. A post-test was administered immediately after the learning session, followed by an identical test one week later.

The adaptive schedule used real-time performance data. Relevant factors included: accuracy, response time, number of trials since the last presentation and enforced desirable difficulty delay. Additionally, a priority score based on the ARTS algorithm defined storage strength with items competing for presentation at the next trial. The fixed equal condition, spaced items at equidistant interleaved intervals, with the distance between items expanding in the other fixed condition.

At the delayed post-test, the adaptive condition outperformed the fixed schedules for item accuracy. Forgetting was measured in all conditions by subtracting the delayed test results from the immediate post-test. The adaptive condition resulted in less forgetting than the fixed conditions. Additionally, the adaptive condition demonstrated a trend towards faster response times. In summary, "adaptive spacing based on ongoing assessments of learning [storage] strength yields greater learning gains than fixed schedules" (Mettler et al., 2016, p 897). Mettler et al.'s (2016) second experiment identified that learning gains improve when spacing is adapted according to on-going performance rather than item difficulty.

Mettler et al. (2016) used adaptive inter-study intervals for learning items in the interleaved condition and obtained results consistent with the desirable difficulty framework (Bjork & Bjork, 2011). Whilst interleaving is a potential DLCP learning strategy (which is addressed in the forthcoming chapter), the results of their investigation are not directly applicable to the inter-study intervals *between* spaced retrieval sessions. However, the desirable difficulty principles applied to the ARTS algorithm and the learning gains achieved through individual adaptation may be theoretically applied and tested empirically in future research. The relevant principles include real-time responsiveness, accuracy (mastery criteria), response time and the potential expansion or contraction of spacing intervals according to storage strength. These factors together with Pyc and Rawson’s (2009) principle that successful but effortful retrievals are the most advantageous for durable learning, may inform between-session inter-study interval modifications during the Criterion Phase.

The original DLCP used real-time performance data by adjusting the number of retrieval sessions based on item performance accuracy, and the use of response time data has now been included. The original between-session inter-study intervals, however, related only to classroom practicalities; classroom and homework schedules were fixed based on factors such as tutor availability, before school commitments or parent participation during homework sessions. The researcher’s experience suggests that the adjustment of inter-study intervals may be of most relevance to either low or high aptitude students.

The researcher has observed that most students experience successful retrieval of new learning items within the first few spaced retrieval sessions, however, a small number of students continue to struggle. Adaptive intervals may hold great potential for these students to achieve effortful, but successful retrievals (Pyc & Rawson, 2009).

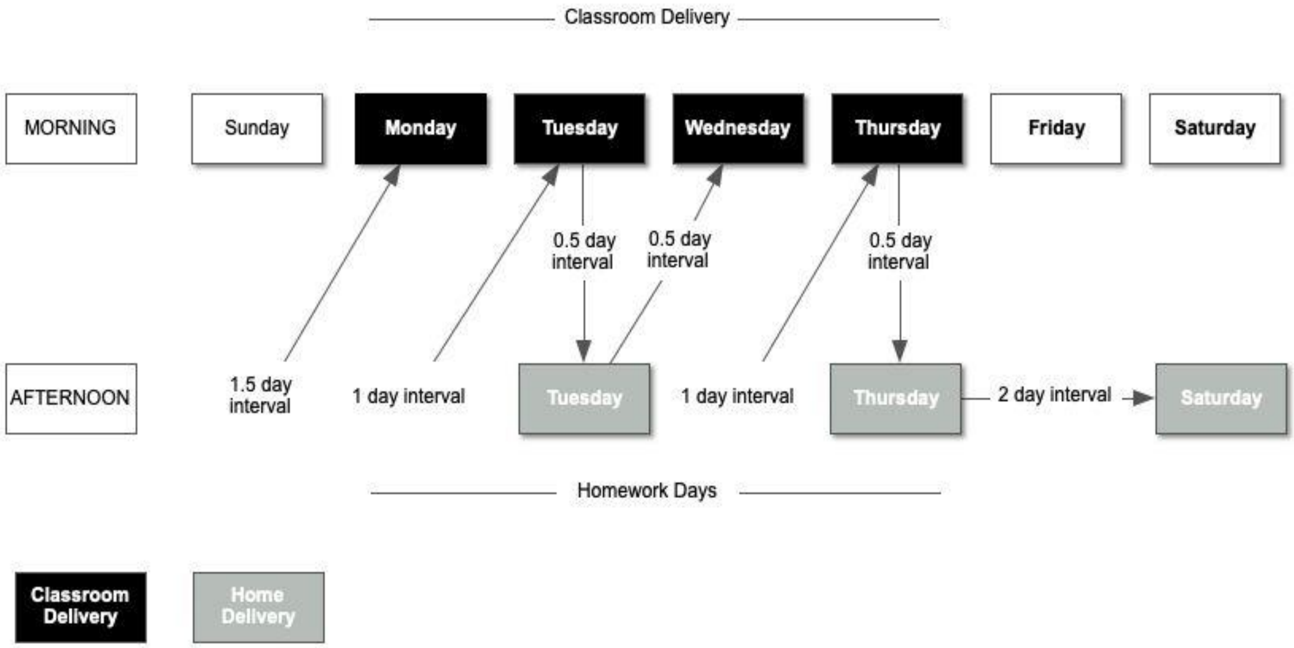
During the Criterion Phase, tutors could be requested to advise the teacher of any students demonstrating continued retrieval difficulties. Poor results during the Mastery Test Phase would also raise teacher awareness. Based on the principles of desirable difficulties (Bjork & Bjork, 2011), and the ARTS system research (Mettler et al., 2016), shorter inter-study intervals could be incorporated for students failing to achieve successful retrievals at the standard intervals. Any adjustments to the spacing between intervals must, however, be practical and achievable in the classroom context.

One option could be to schedule additional spaced retrieval sessions in between the established classroom intervals through a combination of classroom and home delivery. For example, parents of struggling students could be asked if they would be willing to conduct a small number of spaced retrieval sessions at home. Figure 20 provides an example of how the inter-study interval could be reduced to address low storage strength and potentially provide more tailored desirable difficulty for these students.

Within this sample schedule, the longest inter-study interval is 2 days. Monitoring student progress at the longer intervals may assist in evaluating the suitability of the selected schedule. Inter-study intervals could be contracted (by adding sessions) or expanded (by reducing the number of sessions) per week according to the student's progress. Whilst the researcher observed a small group for whom the established intervals created too much difficulty, they also noted a small number of students for whom the desirable difficulty was too low. The adaptation of between-session inter-study intervals may also be responsive to their learning needs.

Durable learning is not associated with fast, easy acquisition (Bjork & Bjork, 1992; Pyc & Rawson, 2009). If high aptitude students demonstrate quick response times and successful retrieval early in the Criterion Phase, the level of desirable difficulty may be insufficient. Initially, the classroom teacher would need to ensure that the learning content was appropriate to the student's prior knowledge and followed advanced learning progressions. With these conditions met, desirable difficulty could be increased by the tutor disregarding easier learning items during one or more spaced retrieval sessions, thereby creating a longer inter-study interval for those items. Alternatively, the student may have a reduced total number of spaced retrieval sessions per week.

*Figure 20.* Criterion Phase sample schedules for a student with low storage strength with delivery through homework and classroom delivery.




From the middle primary years, high aptitude peers may be included in the tutor team for one session per week to increase the interval to their next session. This would increase the desirable difficulty they experience and may enhance durable learning. Assessments of response time (≤ 10 seconds) and accuracy may therefore facilitate the dynamic adaptation of inter-study intervals according to desirable difficulty, for students finding retrieval at the standard intervals too difficult or not difficult enough for optimal learning.

**Elaboration.**

Elaboration is a form of retrieval practice that asks *how* and *why* questions of learners. When asked for explanations of newly learned content or strategies students search their memory for the cognitive schemas related to the subject at hand (Dunlosky et al., 2013; van Merriënboer & Kirschner, 2018). When found, the new learning links to the discovered associations and analogies (prior knowledge), to create meaningful contexts (Dunlosky et al., 2013; Kalyuga, 2009; van Merriënboer & Kirschner, 2018). The established cognitive schema is enhanced (elaborated) and due to additional retrieval pathways, becomes more accessible in the future. The term ‘knowledge elaboration’ describes this process.

The benefits of elaboration have been demonstrated for learners of different ages (from upper primary school), stages and abilities (Roediger & Pyc, 2012). As a retrieval strategy, asking questions to facilitate elaboration is described as elaborative questioning or interrogation (Dunlosky et al., 2013). Self-interrogation or explanation are terms used for independent learners.

The benefits of elaborative interrogation are most pronounced when questioning is tailored to a student’s prior knowledge (Dunlosky et al., 2013). Under these conditions, the effects of elaborative questioning have been shown as robust for different types of factual (Dunlosky et al., 2013) and complex (Sumeracki et al., 2019) information.

The original DLCP followed retrieval testing with checking for understanding and repeated retrievals. Rawson and Dunlosky (2011) demonstrated that repeated retrievals within the criterion phase, accrued minimal learning gains. Facilitated by the tutor, elaborative interrogation could replace the repeated retrievals of the original DLCP after the completion of testing at each pocket. Elaborative questions may benefit memory for both successful and unsuccessful retrievals. Examples of elaboration questions may include:

- Why did you choose this strategy?
- How did you remember this …?
- Can you explain your thinking?
- How does this relate to …?
- Could this apply to …?
- Which part did you find tricky?
- Can you think of a way to remember it next time?

Rowland's (2014) meta-analysis noted that increased semantic processing and elaboration may lead to greater testing effect benefits. Future research on the use of elaborative interrogation within the DLCP, may shed light on the effectiveness of its inclusion for young students across different subject areas and levels of learning. In addition, to stimulating thinking and learning, student responses provide feedback to tutors to guide their explicit instruction and provision of supportive information (van Merriënboer & Kirschner, 2018; Sweller et al., 2019). The relationship between elaborative interrogation and metacognition is discussed in Chapter 8, Metacognitive Development Strategies.

**Explicit Instruction.**

Cognitive load theory informs instructional design through explanations of cognitive architecture (Sweller, 1988). Working memory capacity limits the rate and scope of knowledge acquisition. The expertise reversal effect identifies that, whilst advanced learners benefit from independent problem-solving practice, novice learners in a knowledge domain require explicit instruction and supportive information to manage cognitive load (Kalyuga, 2009). Through one-on-one elaborative interrogation, trained tutors may have the potential to assess student knowledge, control cognitive load and tailor instructional guidance according to changing student learning needs (Kalyuga, 2009).

Nickrow et al. (2020, p. 1) describe tutoring "as one of the most versatile and potentially transformative educational tools in use today". Their meta-analysis of interventions identified an overall pooled effect size estimate of 0.37 with impacts on learning strongest for teacher / paraprofessional directed programmes, earlier grades and those conducted during school. Through tutoring, the DLCP may facilitate reduced extraneous cognitive load through the demonstration of worked examples in mathematics, followed by a related problem-solving example for the novice student (Chen & Kalyuga, 2020). Additional point-of-need examples could be added to the student's folder for further spaced retrieval practice. Worked examples and further resources may be provided to the tutor by the teacher in print form or displayed during the classroom retrieval session. Following the Criterion Phase, learning items entered the Mastery Test Phase.

***Mastery Test Phase***

The Mastery Test Phase was facilitated by the classroom teacher and included an inter-study interval and a final spaced retrieval session. In the original DLCP, learning items achieving the criterion level moved from the last active pocket, into the Mastery Test pocket to await the mastery test (Figure 21). This movement occurred during any session between the previous mastery test and the next. As mastery testing usually occurred weekly, the inter-study interval was between 1 and 7 days. At the mastery test, successfully retrieved items moved to the Mastered pocket. In the original DLCP, unsuccessfully retrieved items were returned to the Store pocket to recommence the process from the start.

Principles from studies on the optimum number of retrieval sessions (Rawson & Dunlosky, 2011) and adaptive intervals (Mettler et al., 2016) combine with studies on inter-study and retention intervals (Table 9) to provide insights into potential modifications of the Mastery Test Phase.

*Figure 21.* The original DLCP Mastery Test Phase.

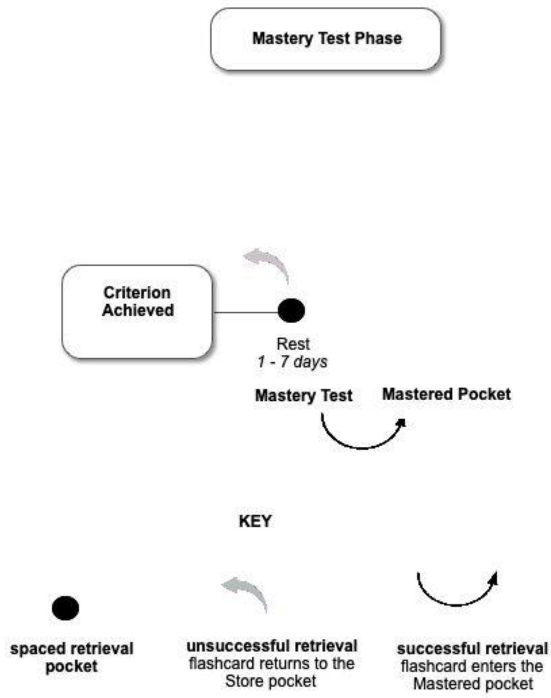


Table 9

*Synthesis Studies of Relevance to the Mastery Test Phase*

| Study | Research goals and participants | Topic and approach | Findings |
|---|---|---|---|
| Inter-study Intervals for Retention: | | | |
| Kang, S., Lindsey, R., Mozer, M., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review*, *21*(6), 1544-1550. https://doi.org/10.3758/s13423-014-0636-z | To investigate the relative efficacy of equal and expanding spaced retrieval practice schedules across educationally relevant training durations.<br><br>Participants: adults<br>*n* = 37<br><br>Training period: 28 days | Learning of 60 Japanese-English word pair translations were compared in two scheduling conditions: equal and expanding.<br><br>Initial study period<br>Session 1: Three cycles of retrieval practice for all items<br>Sessions 2-6: Three cycles of retrieval practice according to allocated schedule<br><br>Corrective feedback<br>Delayed test: 56 days | When retrieval practice occurs over days or weeks expanding scheduling produces a better average performance, faster acquisition and slight retardation of forgetting during training. |

| Study | Research goals and participants | Topic and approach | Findings |
|---|---|---|---|
| Küpper-Tetzel, C. E., Kapler, I. V., & Wiseheart, M. (2014). Contracting, equal, and expanding learning schedules: The optimal distribution of learning sessions depends on retention interval. *Memory & Cognition*, 42(5), 729-741. https://doi.org/10.3758/s13421-014-0394-1 | To investigate the optimal distribution of three learning sessions for the retention of paired associates with retention intervals from 0 to 35 days.<br><br>Participants:<br>university students<br>*n* = 210<br><br>Study duration: 7 days | Arbitrary word pairs were learnt over three sessions within one week. Analysis of results compared equal, contracting or expanding intervals with retention intervals of 0, 1, 7 or 35 days.<br><br>Tests: free and cued recall<br><br>Session 1: Pairs retrieved to a criterion of two successful cued-recalls with feedback<br>Sessions 2 & 3: Retrieval practice with feedback | The optimal schedule varied with the required retention interval. A contracting schedule was advantageous for 1 and 7-day retention intervals. Equal and expanding schedules produced better performance at the 35-day retention interval, demonstrating a 43% advantage compared with the contracting schedule.<br><br>Expanding schedules maintained higher performance during learning. |

| Study | Research goals and participants | Topic and approach | Findings |
|---|---|---|---|
| Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science in the Public Interest, 19*(11), 1095-1102. https://doi.org/10.1111/j.1467-9280.2008.02209.x | To assess final test performance as a function of lag (inter-study interval) and retention interval to determine optimal lag.<br><br>Participants:<br>mixed ages and nationalities<br>*n* = 1354<br><br>26 lag / retention intervals<br><br>Equated study times<br><br>Context: laboratory | Learning session 1: 32 facts learned to a criterion of one correct recall<br><br>Lag period variable<br><br>Review session: two retrievals for each fact with feedback<br><br>Retention interval variable<br><br>Two tests without feedback:<br>1. Recall<br>2. Recognition (multiple choice) | Optimal lag intervals are a function of the required retention interval.<br><br>The optimal gap compared to no gap resulted in an overall increase in recall of 64%.<br><br>As a proportion of test delay, the optimal lag for 1 year retention is 5 – 10% (18 – 36 days). |

| Study | Research goals and participants | Topic and approach | Findings |
|---|---|---|---|
| Lyle, K. B., Hopkins, R. F., Hieb, J. L., & Ralston, P. A. (2020). How the Amount and Spacing of Retrieval Practice Affect the Short- and Long-Term Retention of Mathematics Knowledge. Educational Psychology Review, 32(1), 277-295. https://doi.org/10.1007/s10648-019-09489-x<br><br>Conditions:<br>baseline<br>increased spacing<br>increased retrieval<br>increased spacing and retrieval | To investigate the increase in the amount and spacing of retrieval practice for short and long delays in mathematics.<br><br>Quantitative<br><br>Participants:<br>Undergraduates<br>*n* = 62<br>within-semester short day<br>*n* = 51<br>between-semester long day<br><br>Context:<br>authentic undergraduate course | 32 targeted precalculus learning objectives<br><br>Class tutorial for each objective<br><br>Quiz questions requiring problem solving with feedback<br><br>Short delay test:<br>within-semester exam<br><br>Long delay test:<br>between-semester diagnostic readiness exam for proceeding calculus course | At short delay: increased retrieval and spacing practice boosted retention compared to other conditions.<br><br>At long delay: additional spacing boosted retention at a similar magnitude observed at the short delay test but was not influenced by increased retrieval practice. |

**Number of Sessions.**

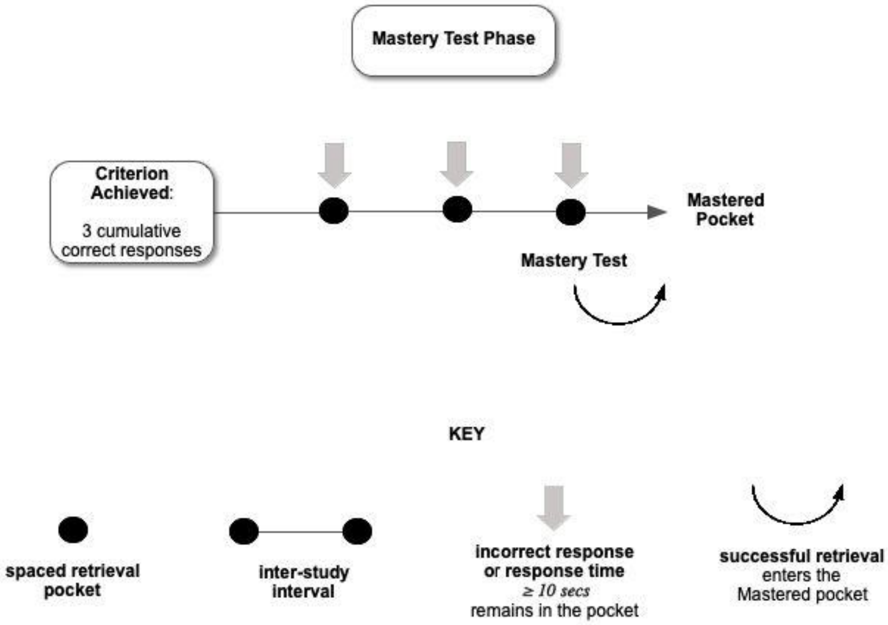
Rawson and Dunlosky (2011) identified a 3 + 3 combination of spaced retrieval sessions: three during initial learning to establish a criterion level, followed by three further spaced retrieval sessions. Two Mastery Test Phase modifications may be incorporated in relation to their findings.

First, as displayed in Figure 21, the original Mastery Test Phase included only one spaced retrieval session, the final mastery test. The addition of two more spaced retrieval sessions would bring the process into alignment with Rawson and Dunlosky's (2011) study. Following the reallocation of pockets during the Criterion Phase, the remaining pockets may be repurposed to facilitate these additional spaced retrieval sessions.

The original mastery learning folder tool contained four unassigned pockets. The process modification which reduced the criterion level from five consecutive successful retrievals to three cumulative retrievals created another two spare pockets. Three of these six pockets may be utilised to provide additional spaced retrieval sessions.

Second, sending unsuccessfully retrieved items from the Mastery Test pocket back to the Store pocket returns them to the start of the process and therefore subjects them to potentially excessive retrievals (Rawson & Dunlosky, 2011). As revised in the Criterion Phase, unsuccessfully retrieved items could be returned to the Mastery Test pocket for another retrieval attempt following the intervening interval (usually 1 week). Figure 22 displays these modifications. Optimum inter-study intervals for durable learning have been a topic of research debate.

*Figure 22.* The revised Mastery Test Phase spaced retrievals.

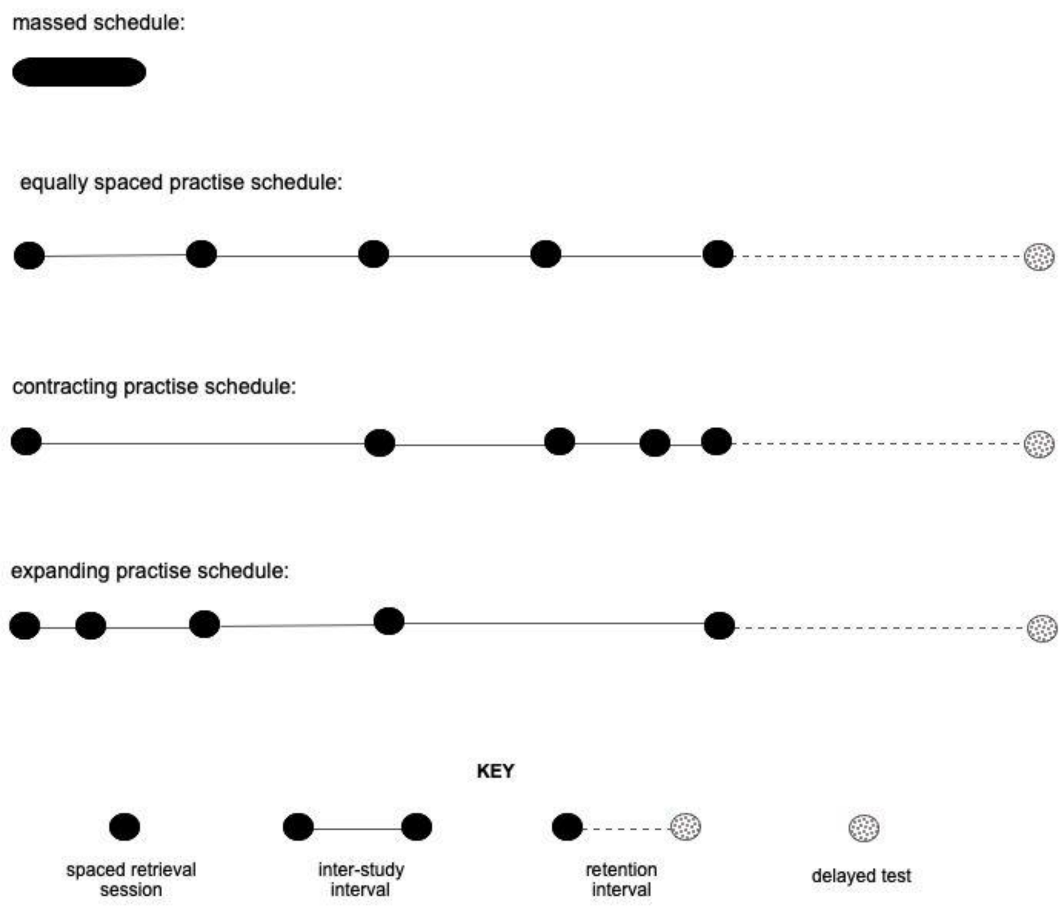



**Inter-Study Intervals.**

Consistent with the desirable difficulty framework (Bjork & Bjork, 2011), Mettler et al.'s (2016) study identified that the adaptation of inter-study intervals based on ongoing student performance achieved learning gains beyond fixed (equal or expanding) schedules of practice. The Criterion Phase was revised by the option to adapt the number of spaced retrieval sessions to moderate difficulty for selected students. The modified Mastery Test Phase now contains additional retrieval sessions and inter-study intervals which may be scheduled to optimise learning duration effects. The Mastery Test Phase is therefore better described as the Retention Phase.

Researchers have attempted to identify optimal inter-study intervals based on equally spaced, contracting and expanding practice schedules (Figure 23). In a detailed review of retrieval practice, Karpicke (2017) stated that equal, contracting and expanding intervals had not demonstrated statistically significant differences, however, Kang et al. (2014) claim that this is because the research designs have consisted of predominately single learning sessions with effects assessed by a single final test, rather than on performance throughout a training period. Additionally, interval testing has been based on spacing of short duration in the interleaved condition. Whilst Rowland's (2014) meta-analysis confirms learning gains at intervals in the order of minutes, Kang et al. (2014) suggest that a single learning session is unlikely to demonstrate results relevant to the retention of learning at educationally meaningful delays. In response, they designed an investigation to focus on the more educationally relevant schedule of multiple spaced retrieval opportunities and an assessment of recall throughout training as well as at a delayed test.

*Figure 23.* Potential Retention Phase inter-study interval schedules.




*Note.* Adapted from "Student Instruction Should Be Distributed Over Long Time Periods", by D. Rohrer, 2015, *Educational Psychology Review*, 27(4), p. 636 (https://doi.org/10.1007/s10648-015-9332-4).

Kang et al. (2014) used Japanese-English word pair translations to investigate equal and expanding spaced retrieval practice schedules across educationally relevant durations, followed by a 56-day delayed test. Their results indicated that when retrieval practice occurs over days or weeks, an expanding schedule produces a better average performance, faster acquisition and a slight retardation of forgetting during training when compared with equal spacing. In the same year, Küpper-Tetzel et al. (2014) conducted a similar experiment to determine the optimal scheduling of three learning sessions comparing equal, contracting or expanding intervals at different retention durations.

In Küpper-Tetzel et al.'s (2014) study, university student participants learnt arbitrary noun pairs to the criterion of two successful retrievals. Memory performance was assessed immediately, and at retention delays of 1, 7 or 35-days by free and cued-recall tests for each practice schedule. A contracting schedule optimised results at the 7-day retention interval, however, at the most educational relevant interval of 35-days, the equal and expanding schedules outperformed the contracting schedule by 43%. Additionally, analysis revealed that expanding practice produced better performance during learning. They concluded that the results demonstrated an association between schedule condition and the required retention interval.

The laboratory studies of Kang et al. (2014) and Küpper-Tetzel et al. (2014) suggest that expanding intervals may provide optimal learning gains at educationally relevant delays. Rawson and Dunlosky (2011, p. 298) confirm that "the diminishing returns of increasing relearning [spaced retrieval] sessions may be overcome by expanding the interval between later sessions" a finding consistent with increasing the desirable difficulty. However, the absence of contextually relevant studies with younger students must be noted. Until empirical DLCP research provides greater clarity, expanding intervals reflect accessible results. A well-known study by Cepeda et al. (2008) may provide insight into the optimal duration of these expanding inter-study intervals.

**Interval Durations.**

The goal of education and the focus of the DLCP, is long-term learning retention. Minimally, learning from one year level should endure to provide a baseline of prior knowledge relevant to the curriculum in the following year. Cepeda et al.'s (2008) study sought to investigate the duration of inter-study intervals for the retention of learning over substantial periods of time, as required within educational contexts. Their comprehensive investigation used a large adult population (*n* = 1354), multiple delayed tests and retention intervals of up to one year. With study times equated, this systematic study used 26 inter-study vs. retention interval conditions to identify educationally relevant durations for durable learning.

Participants within Cepeda et al.'s (2008) study learnt 32 facts to the criterion of one successful retrieval. This was followed by the assigned inter-study interval after which a review session was administered with a criterion level of two successful retrievals (with feedback). The prescribed retention interval was then applied, followed by a recognition and recall test.

The conclusion of their analysis was that inter-study intervals are contingent on the desired or required retention interval, a result also reported by Küpper-Tetzel et al. (2014), Pyc and Rawson (2009) and Rawson and Dunlosky (2011). The results are consistent with encoding variability and study-phase retrieval theories (Cepeda et al., 2008). The recall test was of greatest relevance to the DLCP and these results are displayed in Table 10.

Table 10

*Inter-study Intervals (Optimal Lag) for Different Retention Intervals and Improvement Compared with No Lag for Recall Testing*

| | Days | | | |
|---|---|---|---|---|
| Retention Intervals | 7 | 35 | 70 | 350 |
| Optimal Lag | 1 | 11 | 21 | 21 |
| Improvement | 10% | 59% | 111% | 77% |

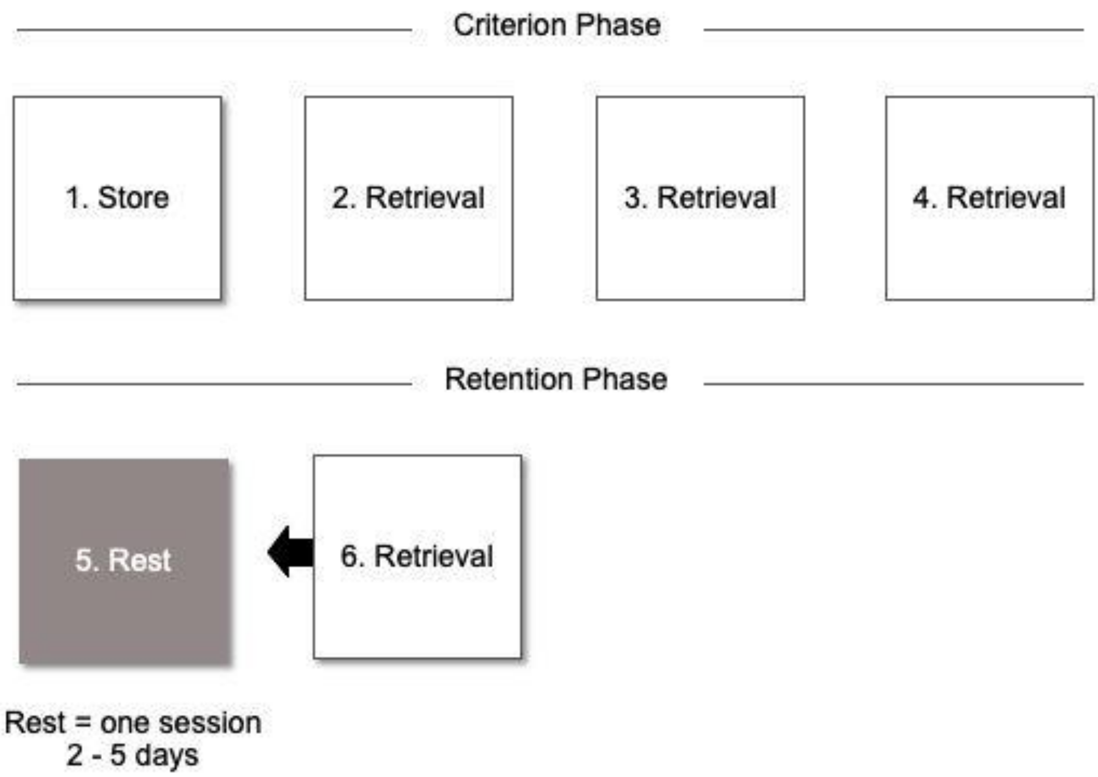
With the criterion established, an inter-study interval of 1 day resulted in a retention interval of 7 days. A 35-day retention interval was achieved by inter-study interval of 11 days. A retention interval of 70 and 350-days was the closest to the ideal retention interval of 1 year. This was achieved by an inter-study interval of 21 days. These inter-study intervals are potentially applicable to the DLCP, moderated by the understanding that Cepeda et al.'s (2008) results were based on adult populations. Future research using the DLCP may provide an opportunity to refine the application of Cepeda et al.'s (2008) results through investigations with younger students.

The goal of inter-study intervals is to provide sufficient spacing to create effortful but successful retrievals (Pyc & Rawson, 2009). Within the Criterion Phase, the first goal is to create sufficient storage strength for an effortful but successful retrieval at short durations (days). During the first two spaced retrieval sessions of the Retention Phase, the aim is for effortful but successful retrievals as learning items move forward *within* the process. The final inter-study interval seeks to facilitate a long retention interval, approximately one year, during which time a student may be exposed to related lesson-based content.

As previously discussed, the modified pocket arrangement includes three Criterion Phase pockets and three Retention Phase pockets, leaving four assigned pockets. These pockets may be repurposed in the Retention Phase to create longer inter-study intervals through the inclusion of rest pockets that simply hold the learning items.

Expanding intervals may be facilitated through a delineation between the tutor sessions that occur over a period of days, and the sessions that are conducted weekly, by the classroom teacher. Figure 24 displays the pockets addressed during the tutor sessions. As indicated, the tutor could facilitate the first short rest interval of the Retention Phase by

*Figure 24.* Inter-study intervals facilitated by the tutor.



skipping the retrieval testing of items in the Rest pocket for one session as further explained below.

As previously described, retrieval testing by tutors commences from Pocket 6 and works backwards. Successfully retrieved items move forward, and unsuccessfully retrieved items return to their pocket of origin. After testing the Pocket 6 items, the tutor moves the flashcards from Rest Pocket 5 into Pocket 6 without testing them. There they remain until the next session. This facilitates a short one session rest interval, which, depending on the practice schedule of three to five sessions per week, will be between two and four days. A supplementary video demonstration of the tutor session procedure can be accessed via the link below[4]. The remainder of the Retention Phase is conducted by the classroom teacher once a week.

Figure 25 displays the six remaining pockets of the Retention Phase. Retrieval testing by the classroom teacher commences from Pocket 11, Mastery Test, and works backwards to Pocket 7. As there are three rest pockets and two retrieval pockets, the process is quick. Items successfully retrieved during the testing of Pocket 11 move to the Mastered pocket, where they may be added to a cumulative total or graphed (Appendix J). Periodically, mastered flashcards may be removed from the tool and stored in a zip-lock bag. The teacher checks student understanding of the unsuccessfully retrieved items and returns them to the same pocket for another attempt, one week later. The Mastery Test pocket now receives the bundle of resting learning items from Pocket 10 and Pocket 10 receives the resting items from Pocket 9. The teacher then conducts a spaced retrieval session on the learning items within Pocket 8, with successfully retrieved items moving forward, or returning to the same pocket if not retrieved, for another retrieval attempt at the next teacher session. Pocket 8 then

[4] Tutor Session Animation: https://youtu.be/fLncM4W842s

receives the resting learning items from Pocket 7. During the teacher directed session, additional learning items may be added to the Store pocket or the learning load may occasionally be adjusted to manipulate total classroom time-on-task if required. A supplementary demonstration can be accessed via the link below[5].

Figure 25 also highlights the inter-study intervals. The second expanding rest interval spans pockets seven and eight where learning items rest for a total of 7 – 14 days. A successful retrieval after this interval, may facilitate retention of sufficient duration for a challenging but successful retrieval at the final mastery test, 21 days later[6]. The resting of flashcards in pockets nine to 11 results in the largest inter-study interval of 21 days which is associated with a retention interval of 70 – 350 days (Cepeda et al., 2008). Of practical relevance to the DLCP, Cepeda et al., (2008) found that whilst a 21-day inter-study interval was optimal for the longest retention, the rate of subsequent forgetting after this interval was very slow. This provides a measure of flexibility in the timing of the mastery test should contextual factors cause a delay. A supplementary animation of the inter-study intervals can be accessed via the link below[7].

---

[5] Weekly Teacher Session Demonstration: https://youtu.be/SPitkw6knLU

[6] Cepeda et al.'s (2008) results demonstrated a retention interval of 35 days for an inter-study interval of 11 days.

[7] Inter-Study Intervals Animation: https://youtu.be/rKSlqdeDcKo

*Figure 25.* The second and third expanding inter-study intervals within the Retention Phase administered by the teacher.

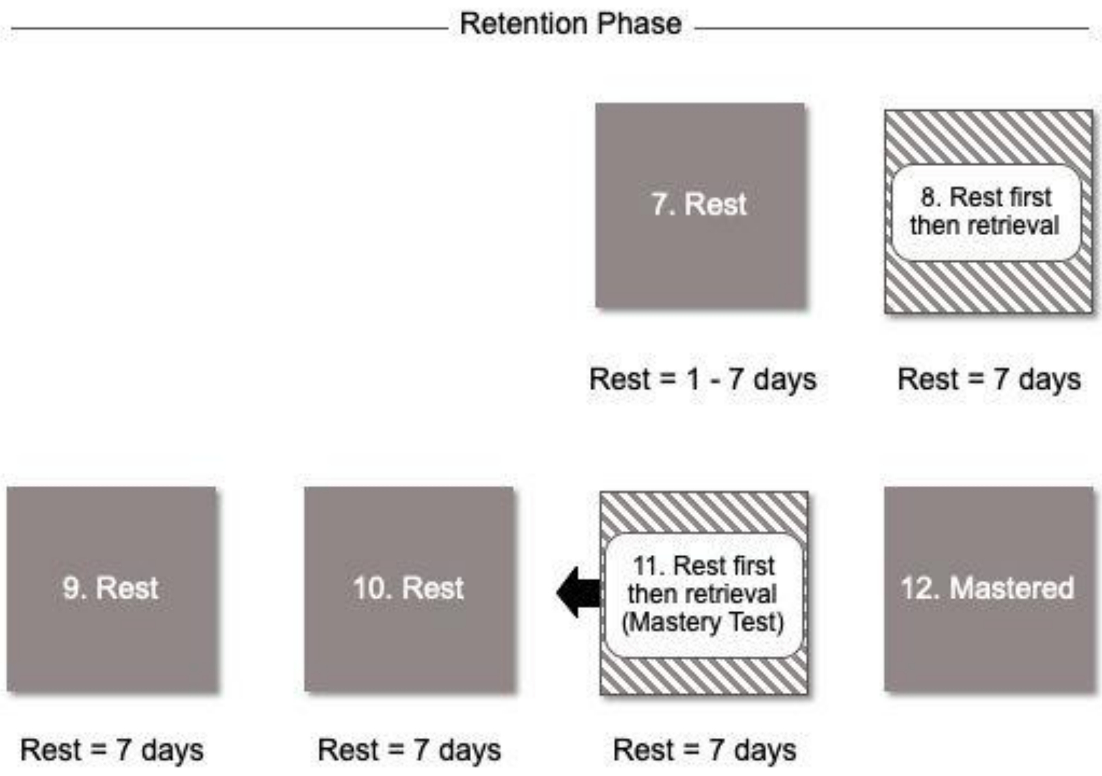
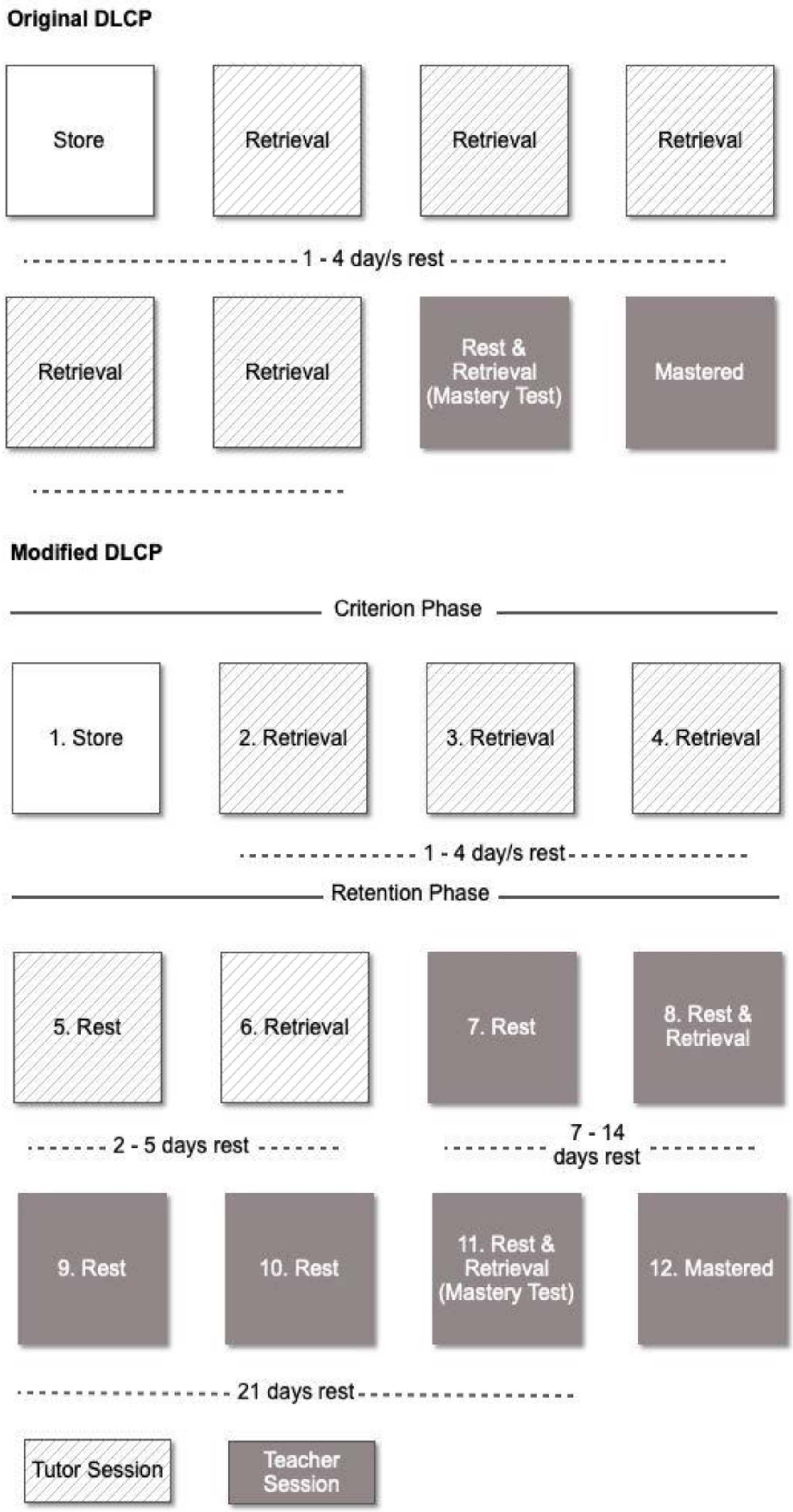


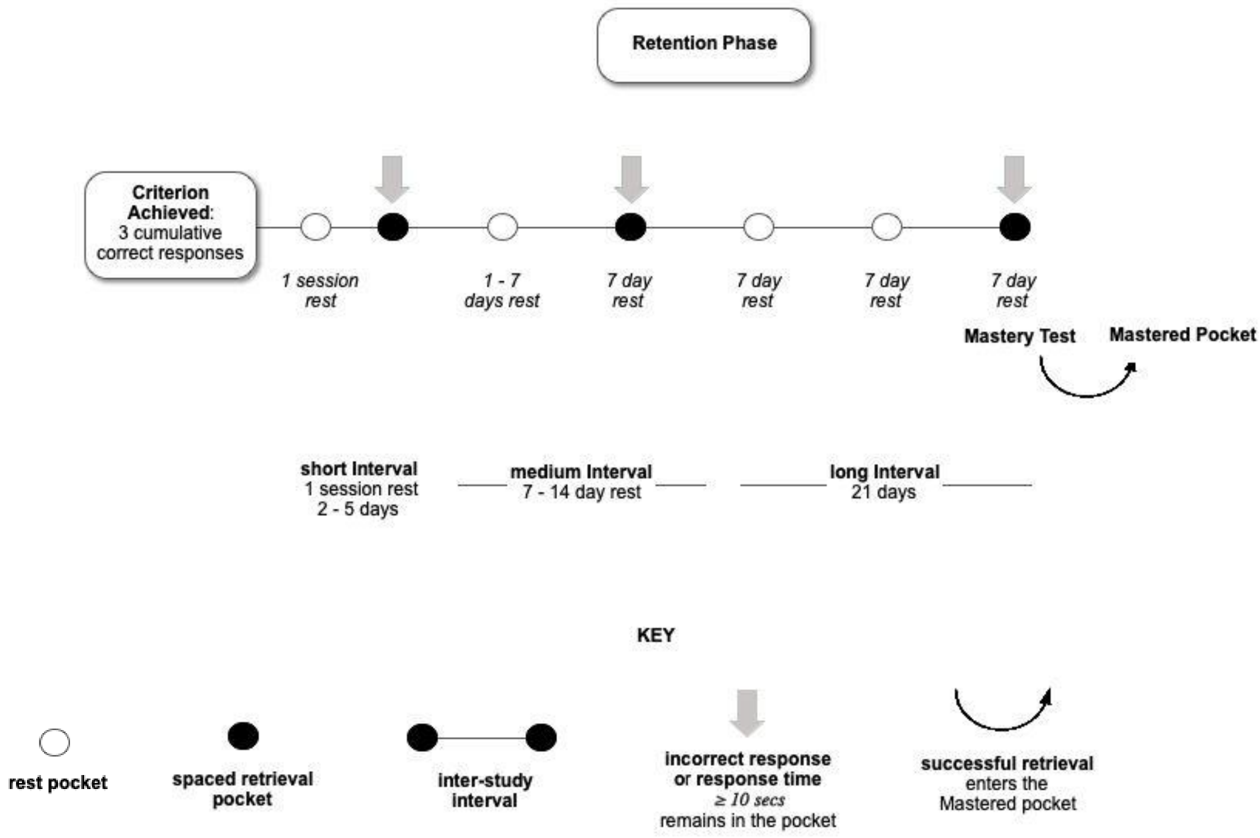

Figure 26 compares the original DLCP pocket allocation with the modified version and includes the usual roles of tutor and classroom teacher. It should be noted that schools using the DLCP assign roles according to their available resources. For example, in some schools, teachers manage the allocation of learning items but use education assistants to conduct the entire process.

*Figure 26.* Original and modified DLCP pocket allocation.



The Cepeda et al. (2008) study provides valuable insights into inter-study intervals for learning retention at educationally meaningful durations. Modifications to the DLCP will require further research to determine if the inter-study and retention intervals identified by Cepeda et al.'s (2008) study translate to school-aged populations. Figure 27 summarises the Retention Phase spaced retrieval modifications. To conclude the synthesis of spaced retrieval research applications, a study by Lyle et al. (2019) set in an educational context, ties together several findings.

*Figure 27.* The revised Retention Phase.




Lyle et al. (2020) investigated three spaced retrieval conditions to determine how the retention of classroom learning is affected by the following spaced retrieval conditions: increased spacing, increased retrieval and increased spacing and retrieval. The participants were undergraduate students participating in a regular within-semester pre-calculus course.

Retrieval practice took the form of quiz questions. A baseline condition was established through the presentation of three quiz questions for each target learning objective at the first quiz session. In the spaced condition, the inter-study intervals were manipulated; quiz questions (on the same learning objective) were distributed across the first four weeks of the course in an expanding schedule (days 1, 7 and 14). The amount of retrieval was varied by the provision of one or two questions at these sessions. Increasing only the amount of retrieval was achieved by presenting all six quiz questions on each learning objective at the first session.

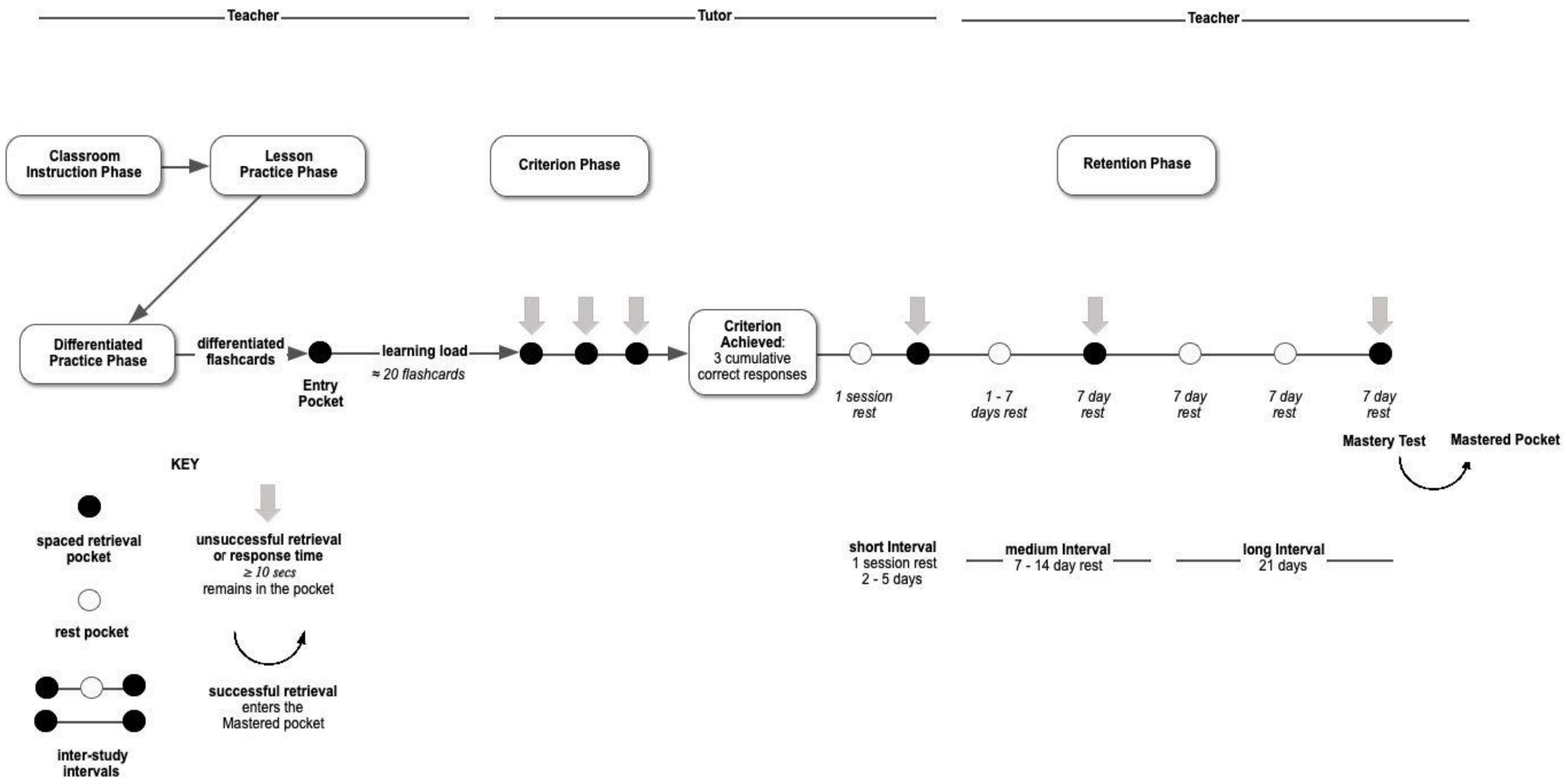
Two retention intervals were assessed. First, a within-semester condition in the form of a pre-calculus course exam was administered at a short delay of 4 weeks after the final spaced retrieval session. Second, a between-semester condition was presented in the form of a diagnostic readiness exam for the forthcoming calculus course. This involved a longer delay of 12 weeks after the final spaced retrieval session.

Consistent with desirable difficulties, their study revealed that increased spacing reduced quiz performance during the instructional phase but led to retention gains at both the pre-calculus exam and diagnostic readiness exam. Due to the nature of the forgetting curve, the proportion correct at the longer retention interval was less than at the shorter duration, however, the spacing effect was similar for both when compared to the baseline condition.

Consistent with the previous studies, increased retrieval practice improved retention at the short delay, though the effect size was smaller than that achieved by spacing the practice. Similarly, Lyle et al. (2020) identified that increased retrieval practice did not influence retention at the longer delay, rather, long term retention was best served by spaced practice, which in their study, was presented in an expanding interval schedule.

Lyle et al.'s (2020) classroom study demonstrated similar results to previous studies that were conducted in less educationally relevant contexts. They confirm that spaced practice is the major factor influencing learning retention at educationally meaningful delays which may support the inclusion of expanding DLCP spaced intervals. Figure 28 displays the features of the DLCP that have been maintained or modified through the spaced retrieval synthesis.

*Figure 28.* Summary of the maintenance and modifications of spaced retrieval sessions within the DLCP.



## Future Research

In addition to the investigation of variable conditions within the DLCP, an evidence-informed DLCP may provide an instrument for testing a variety of variables using relevant curriculum materials and naturalistic classroom conditions. Table 11 provides examples of potential research.

Table 11

*The DLCP as a Potential Instrument for Contextually Relevant Investigations*

| Investigation | Related Studies |
|---|---|
| Comparison of retention results using learning objectives at different levels Bloom's taxonomy. | Gluckman et al. (2014) |
| Contextually relevant applications of the DLCP for students of different ages and abilities. | Lyle et al. (2020) |
| Manipulation of variables to determine working memory effects. | Chen et al. (2017)<br>Kalyuga (2009)<br>Chen & Kalyuga (2020) |
| Criterion Phase:<br>Manipulations of desirable difficulty using the variables of response time, criterion level, adaptive inter-study intervals, number of spaced retrieval sessions and inter-study interval durations.<br>Elaboration and explicit instruction variables. | Mettler et al. (2016)<br>Pyc & Rawson (2009)<br>Rawson & Dunlosky (2011)<br>Dunlosky et al. (2013)<br>van Merriënboer & Kirschner (2018)<br>Kalyuga (2009) |
| Retention Phase:<br>Manipulations of desirable difficulty using the variables to determine optimal inter-study intervals for retention at educationally relevant delays, interval scheduling and optimal number of retrieval sessions.<br>Elaboration and explicit instruction variables. | Cepeda, et al. (2008)<br>Kang, et al. (2014)<br>Küpper-Tetzel et al. (2014)<br>Lyle, et al. (2020)<br>Dunlosky et al. (2013)<br>van Merriënboer & Kirschner (2018)<br>Kalyuga (2009) |

Spaced practice and retrieval practice feature in several lists describing effective cognitive psychology learning strategies. The report “Organising Instruction and Study to Improve Student Learning” commissioned by the Institute of Education Sciences, U.S Department of Education, includes both strategies in their seven recommendations for schools (Pashler et al., 2007). More recently, the US National Council on Teacher Quality report “Learning About Learning” (2016) confirms their selection of spaced and retrieval practice (Pomerance et al., 2016). In their detailed monograph, “Improving Students' Learning with Effective Learning Techniques: Promising Directions from Cognitive Psychology”, Dunlosky et al. (2013) attributed a high utility factor to spaced and retrieval practice, ranking them as the most effective strategies from a review of ten instructional strategies. In a related article, Dunlosky (2013, p. 16) states that, in combining spaced and retrieval practice, “many students will begin to master material they never thought they could learn”. Visible Learning (2019) research on student achievement influences assigns an effect size of 0.79 for deliberate practice, 0.46 for retrieval practice and 0.65 for spaced practice compared to massed. It is hoped that future empirical research will assess the application of spaced retrieval research to the DLCP and create a process that is evidence-informed, contextually relevant and manageable for classroom teachers.

# Chapter 7: Interleaved Practice

Learning is influenced by the way it is presented and sequenced (Carvalho & Goldstone, 2019). Learning ideally moves from simple to complex (Kirschner & Neelan, 2018) through remembering, understanding and applying objectives (Anderson & Krathwohl, 2001; Bloom et al., 1956). The DLCP supports classroom instruction guided by macro sequences such as learning progressions and micro sequencing involving the juxtaposition of learning content in flashcard format.

Investigations within cognitive psychology have identified micro sequencing effects which may have applications to classroom instruction and interventions. The desirable difficulty strategy of interleaved practice is a micro sequencing condition appropriate to certain learning conditions (Birnbaum et al., 2013; Kornell & Bjork, 2008). Chapter 5: Theoretical Foundation, raised the hypothesis that the mixing of content within the DLCP may be identified as interleaving.

Interleaved practice involves the shuffling or interspersing of learning examples from different domains or categories, rather than grouping by topic (which is termed blocking). A representation of blocked versus interleaved practice is displayed in Figure 29. This chapter includes interleaving definitions, key elements of theory, tabulated and narrative descriptions of synthesis studies and potential applications to the DLCP.

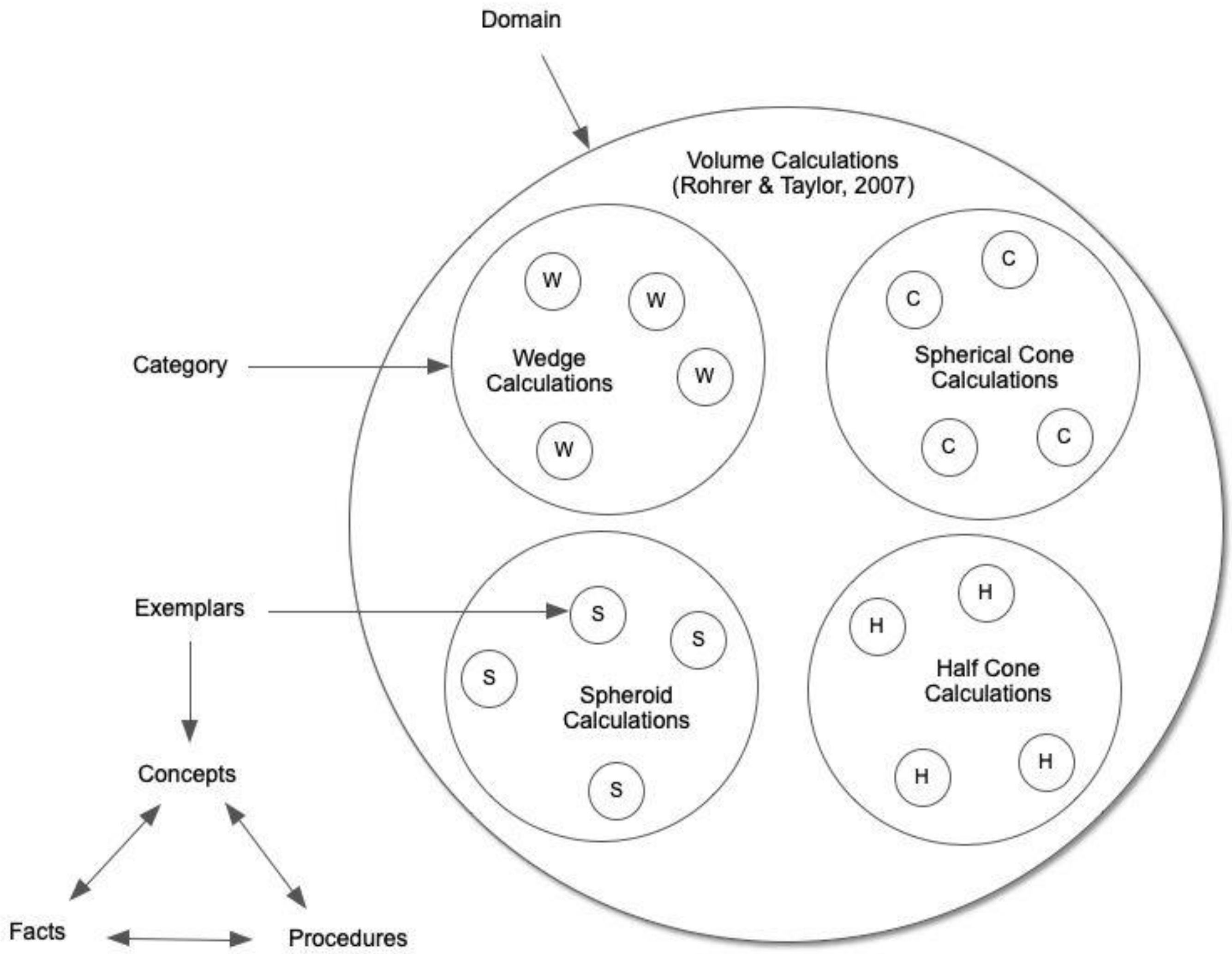
*Figure 29*. Blocked versus interleaved practice.


Interleaved practice involves sequencing within and between different hierarchies of to-be-learned information.  Hierarchical definitions used within the thesis discussion, are described as follows:

- A domain is defined "as a sphere of knowledge" (Merriam-Webster, n.d.). It may be broad, such as mathematics or more specific, for example, an understanding of chess.
- A category is a subset of a domain. It is defined as "any of several fundamental and distinct classes to which entities or concepts belong" (Merriam-Webster, n.d.). Categories may contain similar examples (low discriminatory) or dissimilar examples (high discriminatory).
- Examples within a category are known as exemplars, may be a concept (including facts) or a procedure. Mathematics examples include long division and formula calculations.

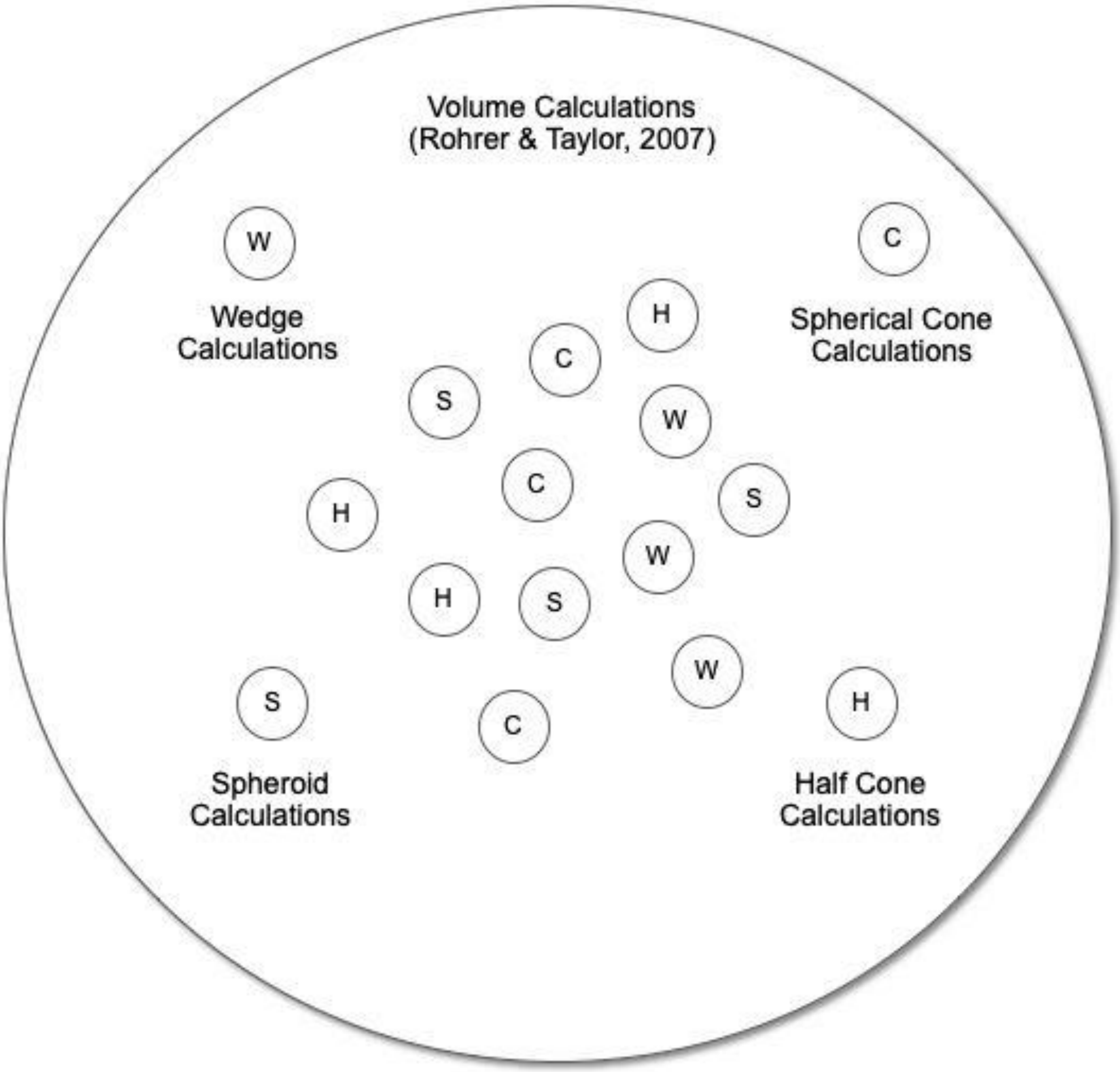
Figure 30 illustrates these classifications in the blocked condition for an investigation conducted by Rohrer and Taylor (2007).

*Figure 30.* Organisational hierarchies displayed in the blocked sequencing condition.



Rohrer and Taylor's (2007) study provides a general example of an interleaving investigation. College students were presented with related exemplars (formulae) to calculate the volume of four different geometric solids. Half of the participants practised the volume calculations blocked by the type of solid, practising the formula for the volume of a spheroid, followed by the formula for a half-cone, then the other two geometric solids sequentially (Figure 30). The remaining participants received the same overall number of practice items presented in the interleaved condition (Figure 31). By nature, blocked practice states or provides the relevant formula. Conversely, the interleaved condition facilitates the practice of both formula selection and the solution procedure (Taylor & Rohrer, 2010). In this investigation, greater positive learning gains were achieved through interleaving the practice questions than by blocking them. Category structure is one classification used to distinguish interleaving studies.

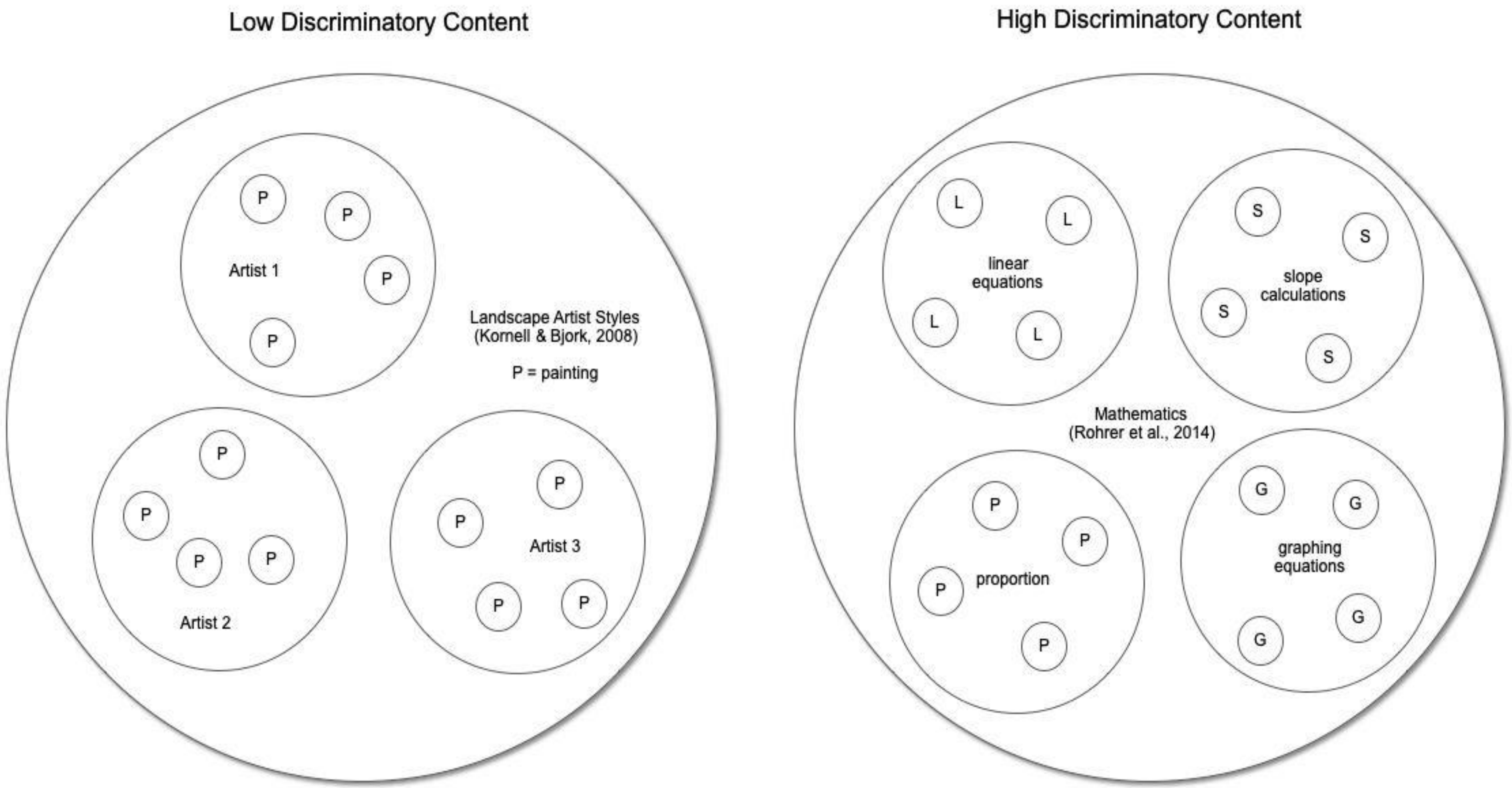
*Figure 31.* Rohrer and Taylor's (2007) interleaved sequencing condition.



Within interleaving investigations, category structure is based on the discriminability of exemplars. Discriminability may be described as low (having highly similar exemplars) or high (low similarity exemplars). Like Rohrer and Taylor's (2007) investigation, most interleaving studies are based on low discriminatory perceptual features (Brunmair & Richter, 2019). Other examples include the identification of artists' work by style (Kornell & Bjork, 2008), the recognition of butterflies (Birnbaum et al., 2013), and the identification of different bird species from the same taxonomic order (Wahlheim et al., 2011). Second in prevalence (Brunmair & Richter, 2019), are studies within the mathematics domain such as the selection of appropriate fraction arithmetic rules (Patel et al., 2016) and formula calculations of of faces, corners, edges and sides for different geometric solids (Taylor & Rohrer, 2010). Different populations have been studied, from three-year-olds (Vlach et al., 2008) to older learners (Kornell et al., 2010). Ostrow et al. (2015) explored interleaved practice with middle school students in an information technology learning context, using an adaptive tutoring platform. Each of these studies identified that the interleaving of practice exemplars achieved greater learning gains than the blocking of exemplars.

A smaller sample of studies has investigated the interleaving of high discriminatory learning content such as word pairs from different subject areas (Hausman & Kornell, 2014), superficially dissimilar mathematics problems (Foster et al., 2019; Rohrer et al., 2014; Rohrer et al., 2015) and mixed domain concepts (Yan & Sana, 2020). Figure 32 provides an illustration of low and high discriminatory learning content displayed in the blocked condition.

*Figure 32.* Diagrammatic representation of low or high discriminatory learning content.



Different category structures and schedules of practice may support different types of learning. Ultimately however, students in all year levels are required to integrate and apply the learning objectives of topics or concepts. They need to have the "specific knowledge to perform the familiar aspects of those problems, but, above all, have the necessary general and abstract knowledge to deal with the unfamiliar aspects" (van Merriënboer & Kirschner, 2017, p. 8). The ability to generalise conceptual understandings (category learning) through exposure to multiple related exemplars is described as inductive learning (van Merriënboer & Kirschner, 2018).

**Inductive Learning Effects**

Traditionally, blocking learning by topic was thought to facilitate inductive learning through highlighting the similarities between exemplars in a category (Kornell & Bjork, 2008). Logically, the spacing of exemplars through interleaving should make the associations more difficult to discern (Kornell & Bjork, 2008). However, research suggests that when within- and between-category exemplars are low discriminatory, interleaved practice does facilitate learning (Kirschner & Neelan, 2018; Kornell & Bjork, 2008; Weissgerber et al., 2018). Research on interleaving is dominated by studies focussed on inductive learning using low discriminatory material, for example, the previously described practice of formulae algorithms (exemplars) for different geometrical solids (categories) (Rohrer & Taylor, 2007). Under certain conditions, inductive learning may also be facilitated by the blocking of practice (Carvalho & Goldstone, 2015; Carvalho & Goldstone, 2020). Two theories specifically address inductive learning through interleaving and/or blocked conditions: the discriminatory contrast hypothesis (Kang & Pashler, 2012) and sequential attention theory (Carvalho & Goldstone, 2015).

### *Theories*

#### Discriminative Contrast Theory.

Kurtz and Hovland (1956) first proposed a discrimination hypothesis for the effects of interleaving which has since been replicated and refined by other studies (for example, Carvalho & Goldstone, 2014; Guzman-Munoz, 2017; Kang & Pashler, 2012; Kornell & Bjork, 2008; Wahlheim et al., 2011 and Zulkiply & Burt, 2013). Discriminative contrast theory proposes that categories and exemplars which are very similar, for example perceptual features of bird species, helps the learner to better identify and remember the differences between the related categories (Carvalho & Goldstone, 2014). This result is thought to be achieved by juxtaposing different exemplars over time (Foster et al., 2019) leading to the ability to classify novel examples. In addition to perceptual features, the need for discrimination is also applicable to many types of mathematical problems (Rohrer & Taylor, 2007) at nearly every level (Rohrer et al., 2015). The sequential attention theory (Carvalho & Goldstone, 2014) seeks to explain how the discriminative contrast theory is applicable in both interleaved and blocked practice schedules under certain conditions.

#### Sequential Attention Theory.

Sequential attention theory (Carvalho & Goldstone, 2015) proposes that interleaved and blocked schedules highlight different aspects of the to-be-learned material. Enhanced learning will be achieved using the sequencing schedule that highlights the most challenging feature of the material. “Learners focus their attention on and encode mostly differences between objects of different categories and similarities among objects of the same category” (Carvalho & Goldstone, 2015, p. 7). When exemplars of different categories are similar (low-discriminatory features), interleaving will benefit category learning though discriminative contrast. Conversely, when within-category similarity is low (high-discriminatory features), the challenge is to determine how the exemplars are similar; blocking exemplars may facilitate the identification of the feature that characterises the category. According to sequential attention theory, "any situation that changes the relative importance of differences between categories versus similarities within categories should show similar results" (Carvalho & Goldstone, 2015, p. 8).

### *Meta-Analysis*

Brunmair and Richter (2019) conducted a meta-analysis on interleaved studies designed for inductive learning effects. They used Hedge's g for effect sizes for studies with sample sizes less than 20[8]. A moderate overall interleaving effect (Hedges' g = 0.42) was determined, however there was variety in effect based on setting and the type of learning material. The studies based upon paintings had the largest effect (g = 0.67), followed by mathematics (g = 0.34). Low or negative effects were identified for some language-based content. The results supported the sequential attention (Carvalho & Goldstone, 2015) and discriminative contrast theories (Birnbaum et al., 2013; Kang & Pashler, 2012). Theories related to the interleaving of both low and high discriminatory material are thought to be non-mutually exclusive (Foster et al., 2019). Many researchers propose that the study-phase retrieval theory may also have an involvement in inductive learning (Foster et al., 2019; Guzman-Munoz, 2017; Rohrer et al., 2015).

## Spaced Learning Effects

Whilst most interleaving studies relate to the investigation of low discriminatory learning material (Brunmair & Richter, 2019), learning gains have also been observed through the interleaving of high discriminatory content (Rohrer et al., 2014). The theory of relevance for this category structure is study-phase retrieval.

---

[8] Hedge's g (sample size $< 20$) and Cohen's d (sample size $\geq 20$) are comparable.

### *Theories*

#### Study-Phase Retrieval.

Interleaving spaces the within- and between-category retrieval of learning material (Rohrer et al., 2019). Birnbaum et al. (2013) observed that greater spacing between exemplars within the same category had positive memory effects. As previously discussed, study-phase retrieval is a theory based on spacing and testing (retrieval) effects.

When practice is blocked, the category defines the concept or procedure, so this information does not need to be retrieved (Rohrer et al., 2015). However, when practice is interleaved (spaced), the retrieval of concept and/or procedure information requires the learner to revisit previous thinking, increasing the desirable difficulty (Bjork & Bjork, 2011), which reinforces the memory pathway for future retrievals (Foster et al., 2019; Guzman-Munoz, 2017). Recall is required for two steps: the selection of a strategy and the execution of it (Rohrer et al., 2015) for both low discriminatory and high discriminatory content (Rohrer et al., 2014; Rohrer et al., 2015, Foster et al., 2019).

In conclusion, there are three main non-mutually exclusive hypotheses which seek to explain the learning benefits of different practice sequences:

- the discriminatory-contrast hypothesis (interleaving for inductive category learning),
- sequential attention theory (interleaved or blocked sequences for inductive category learning) and
- study-phase retrieval hypothesis (inductive or non-inductive learning).

To explore interleaved practice within the DLCP, literature was examined to identify the alignment and relationship between sequencing theories and potential DLCP maintenance or modification.

## Synthesis

Classroom instruction is frequently modelled on the sequencing format of textbooks which inherently present learning content in the blocked condition (Rohrer et al., 2014; Rohrer & Taylor, 2007). Blocked practice within lessons naturally follows. The DLCP aims to differentiate and consolidate the content of this instruction, however, the self-paced mastery process results in the overlapping of cumulative cross-curricular learning material. Therefore, by default, flashcards within the DLCP may be interleaved by both domain and category.

Three hundred and four citations were identified as potentially relevant to interleaving within the DLCP through initial database searching. SCOPUS was used to refine the search using more specific criteria. Forty-seven articles were identified. Snowballing and pearling techniques revealed further relevant citations with forty-one papers selected for abstract reading. Thirty-one papers were read in part (11) or whole (20). Sixteen papers were identified as having potential to shed light on the interleaving of low and high discriminatory learning material. Purposive sampling was employed for the selection of ten studies for data synthesis.

The classifications of high and low discriminability are used to divide studies within the synthesis to provide insight into evidence-informed application of interleaved and blocked conditions to the DLCP. Each classification includes study descriptions to enable stakeholders to evaluate and adjudicate the conclusions.

### *Low Discriminatory Material*

Low discriminatory learning material features in studies related to inductive learning. Learning gains are attributed to the identification of differences between features of sequentially presented exemplars. The studies of most relevance to this synthesis are (a) related to school curriculums, (b) focussed on remembering, understanding and applying

(Anderson & Krathwohl, 2001; Bloom et al., 1956) and (c) those potentially applicable to the DLCP. Consequently, most studies selected for this section are related to mathematics learning.

Table 12 summarises the data from five studies, comparing the interleaving of low discriminatory material to the blocked learning condition. The first study by Kornell and Bjork (2008) introduces the finding that interleaving supports inductive learning. The next three studies focus on classroom mathematics learning using student populations from Year 4 to Year 7. The final study by Carvalho and Goldstone (2014), compares interleaved and blocked sequencing conditions for the inductive learning of low or high discriminatory material.

Table 12

*Synthesis Studies of Low Discriminatory Interleaved Learning Material*

| Study | Research goals and participants | Topic and approach | Findings |
|---|---|---|---|
| Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological science, 19*(6), 585-592.<br><br>Domain:<br>fine arts - paintings<br>Categories: artists<br>Exemplars: paintings by artist | To investigate blocked versus interleaved practice of low-discriminatory material for inductive learning.<br><br>Participants: undergraduates<br>Experiment 1a. *n* = 120<br>Experiment 1b. *n* = 72<br>Experiment 2. *n* = 80<br><br>Context: laboratory | Experiment 1a. Multiple paintings by different named artists presented in the blocked or interleaved condition.<br><br>Experiment 1b. As above with participants grouped by condition.<br><br>Experiment 2. Novel paintings including studied artists, to test recognition by style as familiar or non-familiar.<br><br>Quantitative<br>Post-test | Interleaved study led to more effective inductive learning than blocked study for recall of artist name and recognition of painting style.<br><br>Exp 1a.<br>Interleaved Mean = 61%<br>Blocked Mean = 35%<br>Exp 1b.<br>Interleaved Mean = 59%<br>Blocked Mean = 36%<br>Exp 2.<br>Interleaved Mean = 77%<br>Blocked Mean = 67% |

| Study | Research goals and participants | Topic and approach | Findings |
|---|---|---|---|
| Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24*(6), 837-848.<br><br>Domain: geometry<br>Categories: prisms<br>Exemplars: algorithms | To compare the effects of interleaved versus blocked practice of low-discriminatory formulae calculations.<br><br>Participants: Year 4 students (*n* = 24)<br>Context: classroom | Formula selection and calculation of the number of faces, corners, edges and sides of different geometric solids.<br><br>Tutorial provided<br>Quantitative<br>Practice session and test, post-test<br><br>Group 1. Interleaved practice<br>Group 2. Blocked practice | Interleaving impaired results during the practice session, however, it improved scores on a delayed test due to the increased discrimination in pairing formulae with the appropriate solution procedure.<br><br>Post-test scores:<br>Interleaved 78%<br>Blocked 38% |

| | | | |
|---|---|---|---|
| Patel, R., Liu, R., & Koedinger, K. R. (2016, August). *When to block versus interleave practice? Evidence against teaching fraction addition before fraction multiplication* [Paper presentation]. Cognitive Science Society 38th Annual Meeting, Philadelphia, PA, United States.<br><br>Domain: fractions<br>Categories: addition, multiplication and denominator conversion<br>Exemplars: algorithms | Experiment 1.<br>To compare the effects of interleaved versus blocked practice of low-discriminatory fraction addition and multiplication calculations and to assess the transfer of knowledge when applied to novel fraction division calculations.<br><br>Participants: Year 6 students (*n* = 70)<br>Context: classroom | Addition and multiplication of fractions with same and different denominators, followed by novel fraction division calculations.<br><br>Instruction provided<br><br>Quantitative<br>Pre-test, mid-test and post-test with corrective feedback<br><br>Group 1 Participants:<br>Blocked practice<br>Period 1: 24 fraction additions<br>Period 2: 24 fraction multiplications<br><br>Group 2 Participants:<br>Interleaved practice<br>Period 1: 24 randomised questions<br>Period 2: 24 randomised questions | Interleaving facilitated practice of the decision to convert fraction denominators, and improved accuracy results in the post-test.<br><br>Interleaving improved the transfer of knowledge to the division of fractions.<br><br>Post-test scores:<br>Interleaved 79%<br>Non-interleaved 68%<br><br>Transfer to fraction division post-test:<br>Interleaved 70%<br>Non-interleaved 57% |

| | | | |
|---|---|---|---|
| Rohrer, D., Dedrick, R. F., Hartwig, M. K., & Cheung, C. N. (2019). A randomized controlled trial of interleaved mathematics practice. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000367<br><br>Domain: algebra<br>Categories: graph, inequalities, expressions and circles calculations<br>Exemplars: algorithms | To assess the efficacy of interleaved practice and evaluate the feasibility of implementation in the classroom.<br><br>Participants: Year 7 students<br>Fifty-four classes ($n$ = 787)<br>Study duration: 5 months<br>Caveats:<br>Students took more time to complete questions in the interleaved condition.<br>Test benefits may be smaller with shorter test delays.<br>Initial blocked practice may be desirable.<br>Practice included corrective feedback.<br>Interleaved practice within this regime included three strategies: interleaving, spacing and retrieval. | Four types of algebra problems (graph, inequalities, expressions and circles) were interleaved with unrelated filler problems and compared with the blocked condition.<br><br>Teacher assistance was provided during assignment completion.<br><br>Quantitative<br>initial classroom blocked practice<br>cluster randomized controlled trial<br><br>Practice phase (8 assignments),<br>Review worksheet and test<br><br>Group 1. Interleaved practice<br>Group 2. Blocked practice | Interleaving produced large learning gains in a delayed test, a result which may have been enhanced by<br>• increased discrimination ability to pair problem type with solution procedure<br>• incorporation of between-session spaced practice due to study duration and<br>• retrieval practice.<br><br>Test Scores<br>Interleaved 60.7%<br>Non-interleaved 37.6%<br>Effect size $d$ = 0.83 |

| Study | Research goals and participants | Topic and approach | Findings |
|---|---|---|---|
| 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. | Experiment 1.<br>Compared high within- and between-category similarity with low within- and between-category similarity in interleaved versus blocked conditions. | Study of blob figures ("alien cells") to identify species by perceptual features.<br><br>Quantitative<br>Study task with corrective feedback<br>Generalisation task (no feedback) | The interleaved condition achieved greater generalisation for novel items of high-similarity categories.<br><br>(Blocked condition results will be discussed within the high discriminatory section.) |
| Domain: visual discrimination<br>Categories:<br>high within-category similarity<br>low within-category similarity<br>high between-category similarity<br>low between-category similarity<br>Exemplars: blob figures | Participants: undergraduates<br>high similarity/low discriminatory<br>(*n* = 29)<br>low similarity/high discriminatory<br>(*n* = 32) | Group 1. High similarity condition<br>- blocked & interleaved exemplars<br><br>Group 2. Low similarity condition<br>- blocked & interleaved exemplars | |

Prior to Kornell and Bjork's (2008) investigation, the blocked presentation of related low discriminatory examples was thought to better facilitate the abstraction of principles and concepts (inductive learning) through highlighting the similarities of exemplars (Kornell & Bjork, 2008). In what became a seminal study, Kornell and Bjork (2008) sought to determine the size of the blocking effect and the presumed reduced effectiveness of the interleaved condition by assessing students' ability to identify the work of different artists. The results "caused a small stir in the field of applied cognitive psychology" (Guzman-Munoz, 2017, p. 421). When participants were tested with a novel painting by a studied artist, results indicated that artists' whose work had been interleaved were more successfully identified than those that had been blocked. Overall, 78% of participants better recognised the new works when the artists had been presented in the interleaved condition. This study generated interest in the application of interleaving to educationally relevant materials and contexts. Taylor and Rohrer (2010) followed their 2007 study with another investigation related to geometric solids, with Year 4 students.

With spacing held constant, participants in the blocked or interleaved condition were assessed on their ability to select the appropriate formula and compute the number of faces, corners, edges or sides of different geometric solids (Taylor & Rohrer, 2010). Results indicated that the interleaved practice condition impaired performance during practice, however, it doubled the subsequent scores on a delayed test, 77% versus 38%. The authors claim that the interleaved condition required the participants to remember both the formula as well as the solution procedure, so when presented with a delayed test, they were better prepared to discern the formula and complete the related procedure. Both groups produced a similar amount of fabrication errors (where a formula not encountered in the investigation was used). The discrimination errors, however, which involved selecting the wrong formula from the instructional set, were 10% for the interleaved group and 46% for the students who used blocked practice. The desirable difficulty "incurred during the practice session proves to be ultimately worthwhile" for the low discriminatory learning material (Taylor & Rohrer, 2010, p. 844).

The effectiveness of interleaving for reducing discrimination errors was also tested by Patel et al. (2016) using fraction addition and multiplication calculations with Year 6 student participants. The tasks, which only differed in appearance by the type of operator (+ or x), had very different solution procedures. Errors in these types of fraction problems are most frequently due to incorrect strategy selection. Students were assessed on the key requirement to first recognise if the conversion of denominators was appropriate before completing the calculation and the remaining procedure. The blocked group received fraction addition questions in one lesson period, followed by fraction multiplication questions in the next, all with corrective feedback. The interleaved group received randomised questions in both periods. The mean accuracy of the interleaved group on the post-test was 79% compared with 68% for the blocked condition. Progress graphs on the decision to convert the fraction before the procedure provided insight into the post-test scores. The interleaving group were slower to master this skill than the blocked group, however, their slower progress ultimately resulted in higher scores. Patel et al. (2016) concluded that interleaving learning gains were achieved through the more extensive practice in discerning the necessity of denominator conversion. An additional post-test was given on fraction division to assess the transfer of knowledge. The interleaved group scored 70% compared to the blocked group score of 57%, suggesting that the interleaved condition facilitated better learning transfer with related novel material.

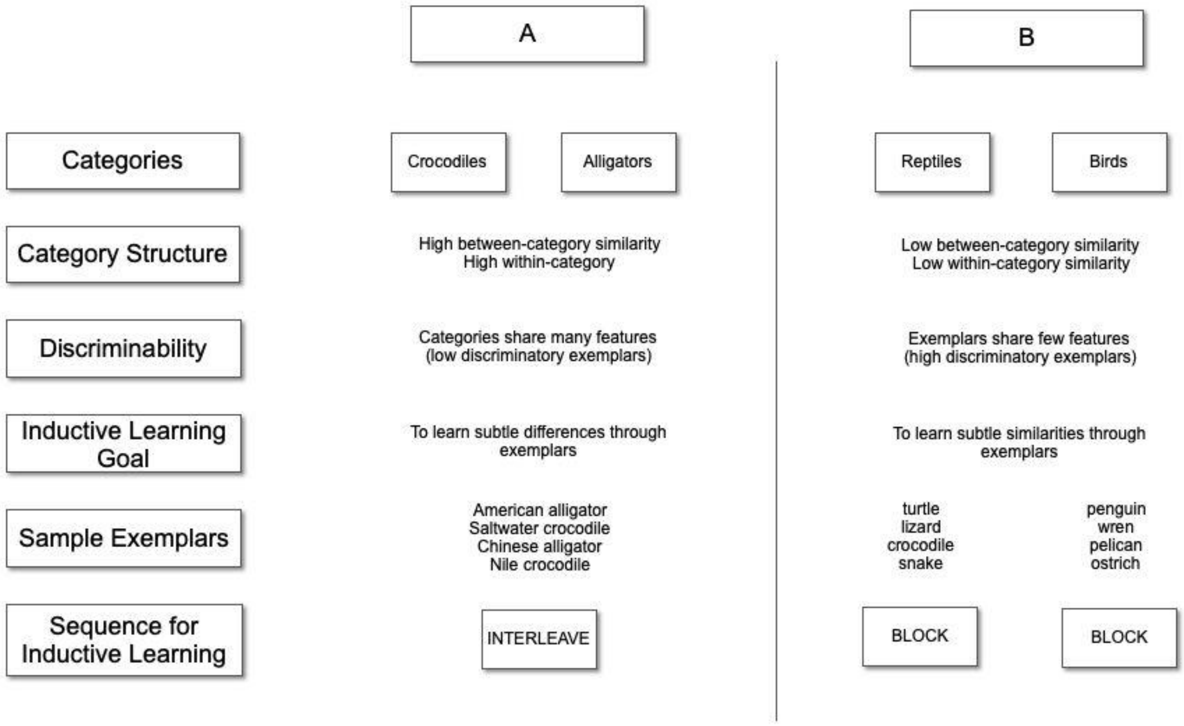
In 2019, Rohrer et al. conducted an extensive study with ecologically relevant materials (assignments), classroom procedures (including a pre-exam review) and a large sample (54 Year 7 classes). With spacing held constant, low discriminatory algebra problems in the interleaved condition were compared with the same questions in the blocked condition.

Practice questions were provided in eight assignments over 4 months, followed by a review of learning and a delayed test 33 days later. The interleaved practice condition produced a mean 60.7% accuracy compared to the blocked condition (37.6%). The effect size was large ($d$ = 0.83). In discussing the magnitude of the effect, the authors highlight that the interleaved condition may have enhanced spacing and retrieval effects due, in part, to the duration of the regime (5 months). Additionally, learning may have benefited from the research design which included initial blocked practice.

The interleaving studies conducted by Taylor and Rohrer (2010), Patel et al. (2016) and Rohrer et al. (2019) were based on low discriminatory material: similar exemplars from related mathematical categories. In each study, interleaving resulted in inductive learning gains relative to blocking. However, Carvalho and Goldstone (2014) proposed that similarities or differences within- or between-categories may modulate sequencing effects in inductive learning. Dissimilar exemplars within categories may challenge the identification of what characterises each category, in which case, students would need to need to discern what is the *same* between exemplars. Sequential attention theory proposes that, under these conditions, inductive learning may be better facilitated by blocking exemplars. Carvalho and Goldstone (2014) created an investigation to test this theory.

Undergraduate students studied one category structure type in both the interleaved and blocked condition, followed by a generalisation (transfer) task in which they had to classify novel items. Concurring with the results of previously presented studies, the interleaved condition achieved greater generalisation for novel items *between* high-similarity (low discriminatory) categories. The blocked presentation, however, improved performance for novel items *within* a low-similarity category. Carvalho and Goldstone's (2014) results suggest that category structure will influence the selection of interleaving or blocking for inductive learning gains. Figure 33 describes these conditions using arbitrary content. These conditions will be further discussed in the next section on high discriminative material.

*Figure 33.* Interleaving versus blocking for inductive learning based on category structure.




In summary, this section of the synthesis has focussed on investigations of low discriminatory material for inductive learning. Some studies, Carvalho and Goldstone (2014) and Kornell and Bjork, (2008), guided theoretical parameters, the remaining studies were selected according to DLCP relevance and criteria stated in the methodology. Table 13 collates the key findings. Given the volume of studies and to assist with cumulative concluding discussions, study numbers have been allocated to all research papers.

Table 13

*Data Extraction of Key Findings for Inductive Learning of Low Discriminatory Learning Content Within the DLCP*

| Study | Authors | Key Findings | Relevance | Limitations |
|---|---|---|---|---|
| **1** | Kornell, N., & Bjork, R. A. (2008) | Interleaved practice leads to more effective learning than non-interleaved study. | Theoretical<br>Learning objective: understanding nuances of artist style | Context: laboratory<br>Content: perceptual-visual-paintings<br>Participants: undergraduates<br>Test delay: 15 secs |
| **2** | Taylor, K., & Rohrer, D. (2010) | Interleaved practice increased discrimination ability to pair problems and solution procedure. | Context: classroom<br>Participants: Year 4 students<br>Content: mathematics<br>Instruction: provided<br>Learning objective: application | Test delay: 1 day<br>Sample: (*n* = 24) |
| **3** | Patel, R., Liu, R., & Koedinger, K. R. (2016) | Interleaving facilitated practice of the decision to convert fraction denominators and improved accuracy and transfer results. | Context: classroom<br>Participants: Year 6 students<br>Content: mathematics<br>Instruction: provided<br>Learning objective: application | Test delay: 80 minutes |

| | | | | |
|---|---|---|---|---|
| **4** | Rohrer, D., Dedrick, R. F., Hartwig, M. K., & Cheung, C. N. (2019) | Interleaving produced large learning gains after a study duration of 5 months and a delayed test.<br><br>Learning gains may have benefited from initial blocked learning. | Context: classroom<br>Participants: Year 7 students<br>Content: mathematics<br>Instruction: provided<br>Filler problems: mathematical<br>Sample: large (*n* =787)<br>Test delay: 1 month<br>Learning objective: application | |
| **5** | Carvalho, P. F., & Goldstone, R. L. (2014) | On the generalisation (transfer) task, the interleaved condition achieved greater generalisation for novel items of high-similarity (low discriminatory) categories. | Theoretical<br>Learning objective: remembering / recognition | Context: laboratory<br>Content: perceptual-visual-blob figures<br>Participants: undergraduates<br>Test delay: immediate |

**DLCP Features to Maintain.**

The DLCP may be consistent with the interleaving studies presented within the low discriminatory section in the following ways:

- the presentation of initial blocked instruction in the lesson context,
- the use of interleaved mathematics material,
- practice of both strategy selection and solution procedure (Taylor & Rohrer, 2010),
- relevance to the remembering, understanding and applying learning objectives (Anderson & Krathwohl, 2001; Bloom et al., 1956) and
- the potential to use category structure to determine sequencing conditions.

The foundation of the DLCP is prior knowledge and classroom instruction which is usually followed by blocked in-class practice activities. Initial blocked practice may highlight shared theoretical constructs of potential learning benefit to subsequent low discriminatory interleaving (Carvalho & Goldstone, 2014; Rohrer et al., 2019). Studies 2, 3 and 4 and many other interleaving studies reviewed, provided instruction prior to interleaved practice. Initial instruction and blocked practice are recommended, particularly for young or less skilled learners or more complex learning (Carvalho & Goldstone, 2019; Dunlosky et al., 2013; Rohrer et al., 2014; Rohrer et al., 2015). Interleaving within the DLCP, is consistent with these principles.

Mathematics is a focus area within the DLCP. Brunmair and Richter's (2019) meta-analysis identified this domain as the second most prevalent within the interleaving studies reviewed. Results from Studies 2, 3 and 4, suggest inductive learning gains for the interleaving of low discriminatory mathematics material. These findings may support the continued use of mathematics content within the DLCP. Two examples of low discriminatory learning content used within the DLCP are the practice of mental maths strategies and word problems.

Study 2 demonstrates the learning benefits of practising strategy selection and solution procedures. Figure 11 (Chapter 4 Baseline Data) illustrates how the DLCP facilitates Year 1 mental maths strategy and solution practice. Additionally, students frequently find word problems difficult as highly similar questions may require different solution procedures (Taylor & Rohrer, 2010). Interleaved word problems are practised within the DLCP and may assist students to discern the appropriate strategy.

The DLCP recognises the importance of moving students through the learning objectives of remembering, understanding and applying (Anderson & Krathwohl, 2001; Bloom et al., 1956). Interleaving Study 5 involved recognising and remembering key visual characteristics to facilitate categorisation of novel exemplars (remembering). Study 1 involved participants being able to discern nuances between artists' painting styles (understanding). Studies 2, 3 and 4 involved practise of mathematics strategy selection and/or solution procedure (application objectives). The use of different learning objectives within the synthesis studies resonates with the purpose of the DLCP.

Based on the sequential attention theory, Study 5 highlights a consideration of category structure in sequencing decisions. In Study 3, when applying the discriminative-contrast theory to learning, Patel et al. (2016, p. 2074) recommends "careful cognitive task analysis to support the decision of when to block or interleave". These factors will be addressed in the section on high discriminatory material.

**DLCP Features to Modify.**

Inductive learning achieves results through the presentation of examples. Discriminative contrast and attention are enhanced when low discriminatory exemplars from related categories are presented contiguously; the juxtaposition highlights the differences between categories, making them easier to discern and remember (Carvalho & Goldstone, 2014; Kornell & Bjork, 2008). Studies 2, 3 and 4 demonstrated learning gains through the interleaving of low discriminatory contextually relevant mathematics calculations. Whilst the DLCP includes mathematics content, the interleaved condition is often cross curricular and randomly sequenced which is inconsistent with these studies. As Brunmair and Richter's (2019) meta-analysis revealed an effect size of g = 0.34, close to a moderate[9] educational benefit, it would be worthwhile to consider if and how the DLCP could fulfil the required conditions. Two options exist to facilitate advantageous micro sequencing of low discriminatory mathematics content for inductive learning gains within the DLCP.

First, the process could be used exclusively for the interleaved practice of low discriminatory within- and between-category mathematics exemplars. Following instruction, teachers could use online flashcards or create their own, to span the variety of practice tasks required, for example, related mental mathematics strategies. This arrangement would preserve the self-paced mastery process and may facilitate inductive learning according to the discriminative contrast and sequential attention theories. A limiting factor is that the precise mixing of exemplars, as found within studies, would be interrupted by the self-paced mastery process. For example, an incorrectly identified exemplar returning to the pocket of origin, may, at the next retrieval test, be situated next to a related exemplar. To rectify this condition, tutors could be requested to modify the sequence when they remove flashcard bundles from each pocket rearranging them so that different exemplars are juxtaposed prior to retrieval testing, which is discussed later in the text. Further research within the DLCP could ascertain if the random mixing of low discriminatory mathematics material produced a statistical difference when compared to the precise juxtaposition of related but different categories

[9] Hattie (2009) describes Cohen's d = 0.4 as an educationally relevant effect size of moderate efficacy.

(Table 14, A vs. B). This approach would prevent the inclusion of unrelated mathematics practice tasks and other subject area content. Alternatively, depending on tool type and cost effectiveness, subject specific folders could be used.

Second, the interleaved related mathematics content could be bundled within the DLCP. Study 4 provided assignments that contained related interleaved practice problems presented contiguously, followed or preceded by unrelated maths filler problems. Learning gains were still achieved. The flashcards for inductive learning could be fastened in the required interleaved sequence (no identical categories presented sequentially). A mastery criterion could be applied to the bundle, for example a minimum score, which would determine if the bundle moved forward or returned to the pocket. Before moving on to the remaining unrelated content, the tutor could check for understanding of the bundle concepts and instruct as required. Future DLCP research could compare this condition with the sorted interleaved condition (Table 14, B vs. C). Further research is required for the use of low discriminatory within- and between-category exemplars in the DLCP to identify potential process and tool accommodations that are both effective and practical within the classroom context.

Table 14

*Future Research Conditions to Compare Inductive Learning Gains using Low Discriminatory Within- and Between-Category Exemplars in the DLCP*

| | Condition | Process/Tool Accommodations |
|---|---|---|
| A | random interleaved content | • dedicated mathematics mastery tool<br>• tutor conducts retrieval testing of flashcards in random condition |
| B | sorted interleaved content | • dedicated mathematics mastery tool<br>• tutor sorts flashcards into category juxtaposed sequence prior to retrieval testing each pocket |
| C | fastened interleaved content in the juxtaposed condition | • other content is within the tool<br>• mastery criterion is applied to the low discriminatory bundle to guide movement forward or back to the pocket of origin<br>• tutor checks for understanding, provides corrective feedback and instruction after retrieval testing the bundle prior to retrieval testing the remaining unrelated content |

The initial analysis of low discriminatory content for inductive learning suggests that applications within the primary school context may be limited. The priority of one-on-one mastery assessment necessitates a broad range of tutors, from older student leaders through to education assistants who may or may not have the capacity to make category structure decisions for the precise juxtaposition of low discriminatory material exemplars. However, this requirement may be manageable for older independent learners or more relevant to the subject specific nature of high school classes.

According to Kirschner and Neelen (2018), studies of interleaving have most frequently been applied to low-discriminatory content to enable students to generalise categories and apply that knowledge to novel examples. Rohrer et al. (2014) note that this restrictive boundary condition would limit the applicability of interleaving within the classroom context, where practice often involves mathematical problems that are easily distinguishable. Some studies have, therefore, sought to explore the interleaving of high discriminatory learning material.

***High Discriminatory Material***

Research examples of high discriminatory content include (a) different categories within the mathematics domain, for example, a combination of interleaved algorithms on linear equations, word problems on proportion, graphing equations and slope calculations (Rohrer et al., 2015), and (b) unrelated domains, for example, statistics and physics, with categories in a variety of interleaved or blocked conditions (Yan & Sana, 2020). These studies more closely align with the original interleaved DLCP condition and are, therefore, investigated in greater depth. Six studies were selected to assess learning in high discriminatory sequencing conditions.

The first three studies were conducted in the classroom and therefore, contextually relevant. The study by Rohrer et al. (2014) introduces the interleaving high discriminatory mathematical categories. Rohrer et al. (2015) adjust the same experimental design to include both a student review and a more relevant test delay to better reflect teaching practice. Foster et al. (2019) build on the theoretical foundation established by Taylor and Rohrer (2010) discussed in the low discriminatory section, designing an investigation to evaluate the relative contributions of discriminative contrast versus spaced practice. The next study by Hausman and Kornell (2014), was selected as it investigated content in the language domain: the interleaving of high discriminatory word-pair associations. An investigation by Yan and Sana

(2020) is then reviewed as they present a unique study that investigates relative learning gains of both within and between domain conditions. Finally, Carvalho and Goldstone's (2014) study is revisited to review their results on the high discriminatory condition within their investigation. Based on relevance to the DLCP, this collection of studies presents confirmatory and contradictory results of the interleaving effect and contributes perspectives on potential evidence-informed revision of the process. Table 15 summarises each study's design and results.

Table 15

*Synthesis Studies of High Discriminatory Interleaved Learning Material*

| Study | Research goals and participants | Research topic and approach | Findings |
|---|---|---|---|
| Rohrer, D., Dedrick, R. F., & Burgess, K. (2014).<br>The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review*, *21*(5), 1323-1330.<br><br>Domain: mathematics<br>Categories: linear equations, proportion word problems, graphing equations and slope calculations<br>Exemplars: algorithms | To investigate the effects of interleaving on the learning of high discriminatory mathematical problems.<br><br>Participants: Year 7 students ($n$ = 140) | Computation of four types of mathematical problems: linear equations, word problems on proportion, graphing equations and slope calculations.<br><br>Quantitative<br>Pre-test<br>Ten assignments over 9 weeks<br>2-week delayed post-test<br><br>Group 1. Interleaved practice<br>Group 2. Blocked practice | Interleaving strengthened the association between problem type and solution strategy.<br><br>Post-test scores:<br>Interleaved 72%<br>Blocked 38% |

| | | | |
|---|---|---|---|
| Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology, 107*(3), 900.<br><br>Domain: mathematics<br><br>Categories: slope and graph problems<br><br>Exemplars: algorithms | 1. To improve ecological validity in the assessment of interleaving by providing an end of trial learning review.<br>2. To determine if the learning benefits of interleaving decrease over time.<br><br>Participants: Year 7 students (*n* = 126) | The content under investigation was slope and graph problems, interleaved with the following unrelated topics: fractions, proportions, percentages, statistics and probability.<br><br>Quantitative<br>Ten practice assignments<br>Learning review session<br>Post-test: 1- or 30-day delay<br><br>Group 1. 1-day delay post-test,<br>Graph problems: interleaved<br>Slope Problems: blocked<br>Group 2. 30-day delay post-test<br>Graph problems: blocked<br>Slope Problems: interleaved | The study determined that the interleaved condition conferred learning benefit after a short and long delay.<br><br>Post-test scores (1-day delay):<br>Interleaved 80%<br>Non-interleaved 64%<br>Post-test scores (30-day delay):<br>Interleaved 74%<br>Non-interleaved 42% |

| | | | |
|---|---|---|---|
| Foster, N. L., Mueller, M. L., Was, C., Rawson, K. A., & Dunlosky, J. (2019). Why does interleaving improve math learning? The contributions of discriminative contrast and distributed practice. *Memory & Cognition*. https://doi.org/10.3758/s13421-019-00918-4<br><br>Domain: mathematics<br>Categories: wedge, spheroid, spherical cone, half cone<br>Exemplars: algorithms | Experiment 2. To assess the relative contribution of the effects of discriminative contrast and distributed practice (spacing effect) or evidence for combined effects.<br><br>Conditions –<br>(a) blocked practice<br>(b) interleaved practice<br>(c) remote-interleaved (volume calculations spaced by non-volume mathematical calculations)<br>(d) remote-blocked (blocked presentation of volume and non-volume calculations)<br><br>Participants: university students | Calculation of the volume of four three-dimensional geometric shapes and non-volume problems (fraction addition/division and permutations)<br><br>Quantitative<br><br>Pre-test<br><br>Tutorial & Practice<br><br>1 week delay final test | Experiment 2<br>Interleaving and remote-interleaving results were similar and superior to blocked practice, supporting the distributed-practice hypothesis (spacing effect). |

| | | | |
|---|---|---|---|
| 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.<br><br>Domain: English<br>Categories: translation word pairs, definition word pairs<br>Exemplars: word pairs | Experiment 4. To investigate the interleaving of high discriminatory material and determine the effects of interleaving mixed subject word pairs.<br>Within session:<br>(a) interleaved flashcards,<br>(b) semi-blocked flashcards (half and half)<br>Between session:<br>(c) Indonesian translation pairs or anatomy term / definitions<br><br>Participants: adults | Recall of Indonesian / English translation word pairs (I), and anatomy term / definition word pairs (T).<br><br>Quantitative<br><br>Practice session<br>1 week post-test delay<br><br>Two interleaved conditions and one non-interleaved. | Within session spacing produced the same results for mixed (TATA …) and unmixed (AAA …TTT) conditions. These spaced conditions were superior to the unmixed massed condition of the study of (I) and (T) in separate sessions. |

| | | | |
|---|---|---|---|
| Yan, V. X., & Sana, F. (2020). Does the interleaving effect extend to unrelated concepts? Learners' beliefs versus empirical evidence. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000470<br><br>Domains: statistics and physics<br>Categories:<br>Statistical tests –<br>chi-square, Wilcoxon signed-ranks and Kruskal-Wallis<br>Physics –<br>inertia, acceleration, and action/reaction<br>Exemplars: word problems linked to textbook titles | Experiment 3<br>To examine study schedules of conceptual concepts within and between different domains.<br><br>Conditions –<br>(a) domain and concepts blocked<br>(b) domains blocked, concepts interleaved<br>(c) domains interleaved, concepts blocked<br>(d) domains interleaved, concepts interleaved<br><br>Participants: undergraduate students ($n$ = 157) | To adjust the scheduling condition of concepts within two unrelated domains (statistics and physics) to compare blocked versus interleaved conditions with within domain, and between domain conditions.<br><br>Quantitative<br><br>Three concepts from each domain<br><br>Study phase: six-word problems per concept<br>Final test: three-word problems per concept | All conditions showed learning gains.<br><br>Interleaving at the domain (c) OR the concept level (b) was similar (Mean = 0.59, Mean = 0.58) and achieved higher results than integrated interleaving (d) (Mean = 0.40) or blocking (a) (Mean = 0.46). |

| | | | |
|---|---|---|---|
| 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.<br><br>Domain: visual discrimination<br>Categories:<br>high within-category similarity<br>low within-category similarity<br>high between-category similarity<br>low between-category similarity<br>Exemplars: blob figures | Experiment 1.<br>Compared high within- and between-category similarity with low within- and between-category similarity in interleaved versus blocked conditions.<br><br>Participants: undergraduates<br>high similarity/low discriminatory<br>(*n* = 29)<br>low similarity/high discriminatory<br>(*n* = 32) | Study of blob figures ("alien cells") to identify species by perceptual features.<br><br>Quantitative<br>Study task with corrective feedback<br>Generalisation task (no feedback)<br><br>Group 1. High similarity condition<br>- blocked & interleaved exemplars<br><br>Group 2. Low similarity condition<br>- blocked & interleaved exemplars | On the generalisation task, blocked presentation improved performance for novel items of low-similarity (high discriminatory) categories. |

Rohrer et al. (2014) investigated the utility of interleaving using mathematical problems which were superficially dissimilar: linear equations, word problems on proportion, graphing equations and slope calculations. Their investigation was classroom-based and used an ecologically relevant assignment format. Year 7 students participated in blocked or interleaved practice of each problem type over nine weekly assignments with a delayed test after two weeks. The mean test scores were 72% (interleaved practice) and 38% (blocked practice). As anticipated with dissimilar learning material, and consistent with Study 2, student work displayed very few discrimination errors (5% interleaved, 4% blocked), suggesting that discriminative contrast was not a major factor in results. The authors concluded that, in the context of dissimilar problems, interleaving "strengthen[ed] the association between each kind of problem and its corresponding strategy" (Rohrer et al., 2014, p. 1323). Conversely, blocking inherently removed the opportunity to practice strategy selection and therefore, failed to reinforce the related association. Rohrer et al. (2014) determined that the spacing achieved through interleaving was a key explanation for the large positive learning gains demonstrated for the high discriminatory material used in their investigation.

In the following year, Rohrer et al. (2015) adjusted the experimental design to include a student review and two post-tests at 1 and 30 days. The review session reduced the variable associated with particular question types occurring closer to the post-test (in the interleaved condition) than blocked questions, although they note that this is an inherent characteristic of interleaving. Additionally, classroom instruction, which is frequently blocked, is often supplemented with a cumulative review prior to important assessments like exams, and so this inclusion better reflected teaching practice. The 30-day post-test was used to determine if interleaving benefits decreased over time.

The content under investigation was slope and graph problems, interleaved with the following unrelated algorithms: fractions, proportions, percentages, statistics and probability. The mean test scores after a delay of one day were: 80% interleaved condition, 64% blocked; and after a 30-day delay, 74% interleaved and 42% blocked. The study confirmed the benefit of the high discriminatory interleaved condition over blocked practice after a short and long delay, also inspiring confidence in the results of their previous investigation (Rohrer et al., 2014). The results of Rohrer et al. (2014) and Rohrer et al. (2015) suggest that spacing alone may produce an interleaving effect, irrespective of discriminative contrast.

Taylor and Rohrer's (2010) findings, discussed in the low discriminatory section, had firmly supported positive learning benefits for the interleaving of similar formulae, attributing the results to the association between strategy selection and solution procedure facilitated by the spacing effect. However, their study design did not rule out the theoretical contribution of discriminative contrast (Foster et al, 2019). To avoid the conflation of discriminative contrast and distributed practice[10], Foster et al. (2019) designed an investigation to determine the theoretical contribution of each. Their interleaving study was modelled on the previously discussed investigation by Rohrer and Taylor (2007), who used interleaved versus blocked volume calculations based on geometric solids. In addition to volume calculations, they included unrelated mathematical content such as fraction addition, fraction division and permutation problems.

Foster et al. (2019) designed Experiment 2 to determine if the learning gains of interleaved practice could be attributed to distributed practice alone (the spacing effect).

[10] The mechanism of the distributed-practice hypothesis is study-phase retrieval or contextual/encoding variability, both based on the spacing effect.

Their study design included four practice conditions: blocked, standard-interleaved (as per Rohrer and Taylor, 2007) remote-blocked and remote-interleaved:

- Condition 1. In the blocked condition, participants received a tutorial on calculating the volume of a specific solid followed by a practice set of problems on that solid. This sequence repeated for each solid.
- Condition 2. In the standard-interleaved condition (testing discriminative contrast), all four tutorials on the volume of geometric solids were delivered consecutively and the subsequent practice of formulae retrieval and calculations were interleaved. A higher result for this condition compared to Condition 4 would suggest a discriminative contrast benefit.
- Condition 3. The remote-blocked group received the wedge volume tutorial, followed by wedge volume practice questions. Subsequent tutorial and practice sessions followed in a fixed block sequence for the remaining topics.
- Condition 4. In the remote-interleaved condition (testing the distributed practice theory), the session commenced with the tutorials on wedge volume, fraction addition, fraction division and permutations. This was followed by participants practising wedge volume formula retrieval and calculations interleaved with the unrelated content. In this condition, discriminative contrast is not applicable so if wedge volume results are comparable with the standard-interleaved condition, then learning benefit from distributed practice would be confirmed.

One week after completion of the practice phase, participants were tested on four novel questions for each problem type, being requested to respond with both the formula (for volume problems) and the answer. Consistent with similar studies (Carvalho & Goldstone, 2014; Rohrer & Taylor, 2007; Taylor & Rohrer, 2010), both interleaved conditions outperformed the blocked condition for formula retrieval and final test performance.

The authors had predicted that if discriminatory contrast contributed to the interleaving effect, then the size of the effect would be larger for standard-interleaved versus the standard-blocked group than the remote-interleaving versus the remote blocked group. This did not occur; the results for formula retrieval were statistically comparable in the remote-interleaved and standard-interleaved conditions[11]. This outcome indicates that distributed practice alone was responsible for more accurate formulae retrieval. Like Rohrer et al. (2014) and Rohrer et al. (2015), the authors attribute learning gains to the spaced practice of both formula selection and solution procedure. The authors conclude that discriminative contrast may not always be required for learning gains through interleaving in mathematics, a result of relevance to the DLCP which shall be discussed in the forthcoming section.

The DLCP frequently includes language domain content such as sight words, phonetic decoding and spelling. There are substantially fewer interleaving studies within the language domain compared to those based on perceptual features and mathematics. One example is a study by Hausman and Kornell (2014) who investigated the interleaving of high discriminatory unrelated interleaved word pairs: Indonesian / English translations and anatomy term / definitions. In Experiment 4, they compared three conditions across two learning sessions separated by 48 hours. The conditions were (a) interleaved flashcards, (b) semi-blocked flashcards (Indonesian word pairs together and anatomy word pairs together) in both sessions and (c), Indonesian pairs practised in one session, and anatomy pairs in the next. On the delayed test, one week later, the recall rates on condition (a) interleaved and (b),

---

[11] Solution procedures were inconclusive as results were also influenced by unrelated computation errors.

semi-blocked were similar to the fully blocked condition. Whilst the interleaving of unrelated topics did not demonstrate a negative effect, the anticipated positive within-session spacing effect due to the interleaved condition was not demonstrated. The authors recommended further research to determine if high discriminative nature of the interleaved flashcard content was detrimental to the usual learning advantage of spacing. Before exploring the application of this investigation to the DLCP, the study design requires consideration.

First, the nature of the material used in Hausman and Kornell's (2014) investigation may be subject to the implications of other research on sequencing and memory. The selected categories are described as highly discriminatory. Under these conditions, sequential attention theory, recommends that categories are blocked for practice to facilitate category generalisations (Carvalho & Goldstone, 2014), however, in their investigation blocking results did not demonstrate a learning advantage.

Second, whilst the within- and between-category similarity may have been low, language translations and science definitions are both word-based. Brunmair and Richter's (2019) meta-analysis identified negative effects for inductive learning of words through interleaving such as "names that belonged to different conceptual categories, pronunciation rules, or translations in different languages" (p. 1035). Inductive learning for concepts within expository text also had low positive utility. Concept characteristics, such as those described within the language domain, may limit the applicability of interleaving, however, the principles of cognitive load theory may also contribute a theoretical perspective.

Hausman and Kornell's (2014) study design specifically excluded participants with prior knowledge of either topic and did not provide any initial instruction, therefore, a semantic context (Bjork, 1994), highly relevant for the language domain (Schneider et al., 1989), was not created. As Yan and Sana (2020) highlight, Hausman and Kornell's investigation tested factual memory, rather than conceptual learning. Facts are valuable to education, however, the borrowing and reorganising principle of cognitive load theory states that information (received through others, for example, instruction) is integrated into a student's schemata based on prior knowledge structures (Sweller et al., 2019). Without prior knowledge or instruction, these learning pre-requisites were not addressed.

Third, the cognitive load conditions may have been high. The cognitive load theory narrow limits of change principle states that working memory capacity is variable according to prior knowledge, on-going schema development, trial-and-error procedures and depletion due to cognitive effort (or expansion after rest) (Sweller et al., 2019). In the search for meaning, the participants may have sought a semantic association between the word pairs, however, without prior knowledge or instruction, the association between word pairs would have been arbitrary. The cognitive load of trial-and-error matching procedures and resultant working memory resource depletion may have contributed to the ambivalent results. The presentation of four language domain exemplars (Indonesian/English translations/anatomy term/definitions), the lack of instruction and semantic understanding from prior knowledge, plus a heavy cognitive load may have confounded the comparison of the interleaved and blocked conditions. Consequently, the Hausman and Kornell study may not provide insight into potential DLCP modifications for learning material within the language domain.

The previously discussed high discriminatory studies have addressed the interleaving of different categories within the same domain in mathematics and language. The three mathematical studies attributed learning gains to the spacing effect, whilst the results were inconclusive for Hausman and Kornell (2014). The next study investigates interleaved learning of high discriminatory material, both within and between domains and categories which reflects the original DLCP sequencing condition.

Yan and Sana (2020) recognise that within education, students are required to learn multiple concepts in a variety of domains, concurrently. It was this reality that resulted in the

DLCP moving from subject specific, to unrelated cross-curricular learning content. The authors reflect on the need for undergraduates to create appropriate practice schedules for concepts across multiple domains. Two domains were selected for their investigation, statistics and physics. The purpose of their third experiment was to determine the effectiveness of different schedules of interleaving versus blocking for domain versus concept level practice material. The following four practice schedules were investigated (Figure 34):

(a) domain and concepts blocked

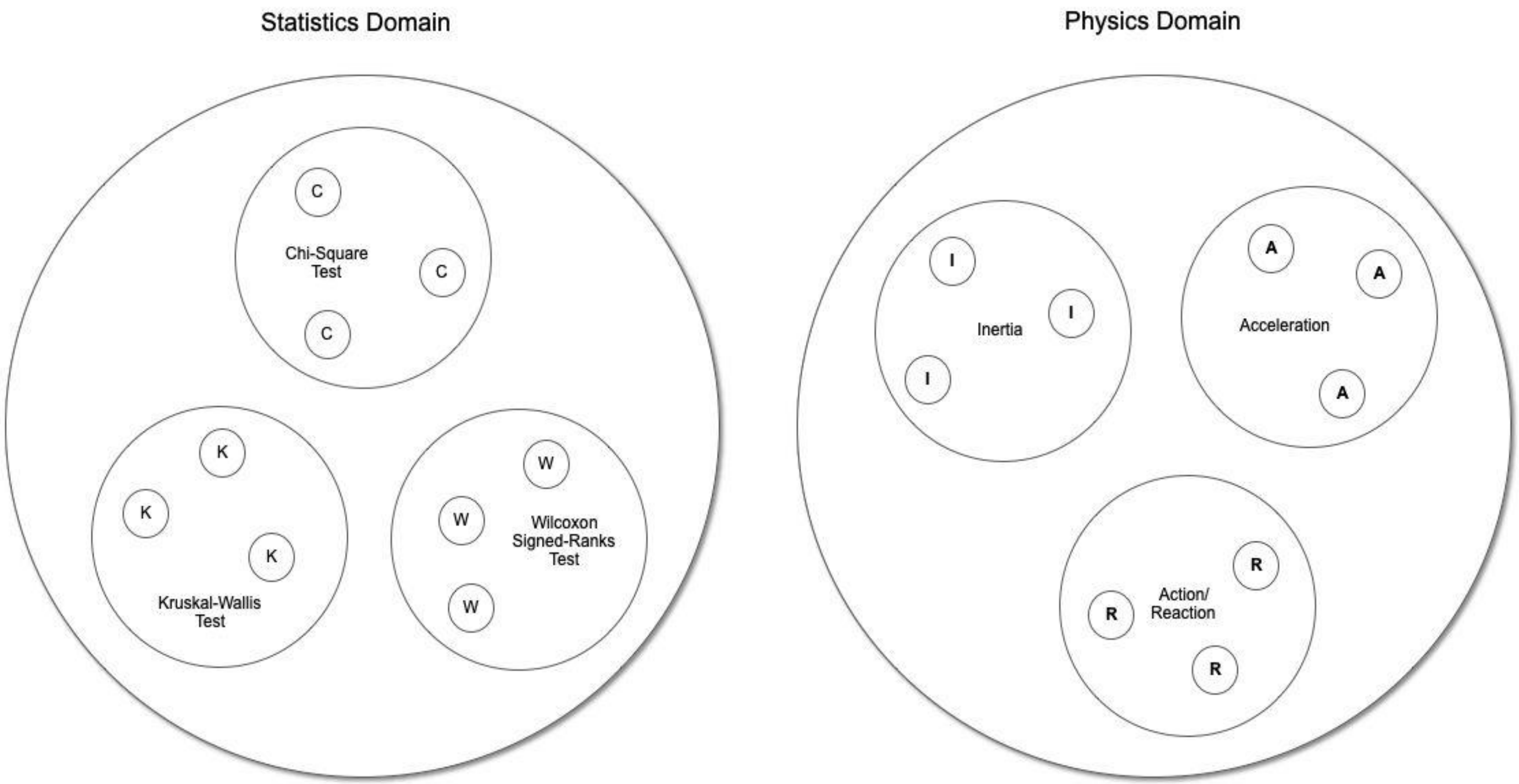
(b) domains blocked and concepts interleaved

(c) domains interleaved and concepts blocked

(d) domains interleaved and concepts interleaved

Yan and Sana’s (2020) participants were required to identify three concepts in each domain through the study of word problem examples. They were told that they would see multiple word problems from six textbooks based on the six conceptual themes. The concept in each word problem was coded (not named) by display with a specific textbook title. The memory test required participants to match novel word problems to the appropriate textbook title. These results may inform DLCP revision.

*Figure 34.* Yan and Sana's (2020) Experiment 3 domain versus concept conditions.




Condition A (domain and concepts blocked): C C C W W W K K K **I I I A A A R R R** (displayed above)
Remaining conditions:
Condition B (domains blocked-concepts interleaved): C K W C K W C K W **I A R I A R I A R**
Condition C (domains interleaved-concepts blocked): C C C **I I I** K K K **A A A** W W W **R R R**
Condition D (domains interleaved-concepts interleaved): C **I** K **A** W **R** C **I** K **A** W **R** C **I** K **A** W **R**

Consistent with previous studies, interleaving was shown to be superior to blocking. Conditions (a) and (d) were similar (M = 46%, M = 40%). Conditions (b) and (c) demonstrated the optimal memory sequence for mixing concept and domain material (M = 59%, M = 58%). The results of conditions (b) and (c) suggest, that when mixing practice material, interleaving should occur at either the concept or domain level, but not both (condition d) or neither (condition a). A randomised version of condition (d) is the original arrangement of flashcards within the DLCP. Carvalho and Goldstone's (2014) study results for high discriminability material in the within-category interleaved condition also contributes a relevant perspective.

Carvalho and Goldstone (2014) investigated if category structures (high or low similarity) influenced memory when presented in the interleaved or blocked condition. The high discriminatory material (low within-and between-category similarity) determined that the blocked condition improved inductive learning for novel items (Figure 33, B) where the challenge was to determine how the exemplars were similar.

In summary, this section of the synthesis has focussed on sequencing condition research using high discriminatory learning material. The results of these studies will be discussed according to relevance and potential application to the DLCP. The key findings of interest are annotated in Table 16. Study numbers continue from Table 13 to assist the cumulative discussion.

Table 16

*Key Findings for the Learning of High Discriminatory Learning Content Within the DLCP*

| Study | Authors | Key Findings | Relevance | Limitations |
|---|---|---|---|---|
| **6** | Rohrer, D., Dedrick, R. F., & Burgess, K. (2014) | When compared to the blocked condition and in the absence of discriminative contrast conditions, interleaving strengthened the association between high discriminatory problem type and solution strategies, suggestive of the spacing effect. | Context: classroom, assignment-based<br>Content: mathematics<br>Participants: Year 7 students<br>Sample: $n$ = 140<br>Instruction: provided<br>Duration: 9 weeks<br>Test delay: 2 weeks<br>Test type: calculations<br>Learning objective: application | |
| **7** | Rohrer, D., Dedrick, R. F., & Stershic, S. (2015) | When compared to the blocked condition and in the absence of discriminative contrast conditions, interleaving strengthened the association between high discriminatory problem type and solution strategies, suggestive of the spacing effect. Interleaving benefits diminished by less than 10% after an educationally relevant test delay of 30 days. | Context: classroom, assignment-based, review<br>Content: mathematics<br>Participants: Year 7 students<br>Sample: $n$ = 126<br>Instruction: provided<br>Duration: 3 months<br>Test delay: 1 day versus 30 days<br>Test type: calculations<br>Learning objective: application | |

| Study | Authors | Key Findings | Relevance | Limitations |
|---|---|---|---|---|
| **8** | Foster, N. L., Mueller, M. L., Was, C., Rawson, K. A., & Dunlosky, J. (2019) | When compared to the block conditions, interleaving strengthened the association between problem type and solution strategy without the influence of discriminatory contrast. Learning gains were attributed to distributive practice (the spacing effect). | Content: mathematics<br>Filler problems: mathematical<br>Sample: *n* =126, divided between three conditions<br>Instruction: provided<br>Test delay: 1 week<br>Test type: calculations<br>Learning objective: application | Context: laboratory<br>Participants: university students |
| **9** | Hausman, H., & Kornell, N. (2014) | Recall rates of word pair associations on the interleaved, semi-blocked and fully blocked conditions demonstrated similar results. | Content: Indonesian / English translation pairs and anatomy term / definition word pairs<br>Sample: *n* = 77 between two conditions<br>Test delay: 1 week<br>Learning objective: remembering (recall) | Context: laboratory<br>Participants: adults<br>Duration: 11 days<br>Instruction: none<br>Test type: cue recall test<br>Lack of semantic cues |

| Study | Authors | Key Findings | Relevance | Limitations |
|---|---|---|---|---|
| **10** | Yan, V. X., & Sana, F. (2020) | Interleaving at either the domain or concept level, demonstrated greater classification accuracy than interleaving both levels or blocking at either level. | Content: statistics and physics concept recognition<br>Sample: *n* =157 divided between four conditions<br>Learning objective: understanding (classification) | Context: laboratory<br>Test delay: 10 min distractor task (puzzles)<br>Participants: undergraduates<br>Test type: classification |
| **11** | Carvalho, P. F., & Goldstone, R. L. (2014) | On the generalisation (transfer) task, blocked presentation improved performance for novel items within a category that contained high discriminatory exemplars. | Theoretical<br>Learning objective: remembering (recall and recognition) | Context: laboratory<br>Content: perceptual features<br>Participants: undergraduates<br>Test delay: immediate |

#### DLCP Features to Maintain.

Sequencing research with educationally relevant, high discriminatory learning content is most prevalent within the domain of mathematics. Interleaved content has demonstrated learning gains attributed to study-phase retrieval theory through spaced practice. Mathematical flashcard content within the DLCP is usually high discriminatory.

Most mathematics textbooks present learning content in the blocked condition and practice activities within lessons often reflect this condition (Rohrer et al., 2014; Rohrer & Taylor, 2007). The self-paced mastery process and resultant cumulative mathematics practice may be a benefit of the original DLCP. Additionally, teachers may also choose to add past content for revision purposes. Several modifications to the DLCP may be possible in response to other findings of high discriminatory interleaving studies.

#### DLCP Features to Modify.

The findings of Studies 6 and 7 confirm spacing and retrieval effects for high discriminatory mathematics material. Study 6 demonstrated an almost doubling of results when unrelated mathematics content was interleaved as compared to the blocked condition. Similar results were achieved in Study 7, which also identified that the interleaved condition conferred learning benefit after both a short and long delay, consistent with Rawson and Dunlosky's (2011) investigation, described in Chapter 6. The more extensive research design of Study 8 sought to evaluate the relative contribution of discriminative contrast theory versus distributed practice. Concurring with Studies 6 and 7, the spacing of practice was thought to reinforce the association between formula selection and solution procedure, regardless of discriminability. Currently, mathematics content within the DLCP is randomly interleaved with content from other domains and categories and, therefore, does not conform to the domain specific condition upon which these studies are based.

The previous section described two options for the interleaving of low discriminatory mathematics content: (a) the use of a dedicated mathematics folder with secondary sorting prior to the retrieval testing of flashcards, or (b) mathematics flashcards appropriately juxtaposed and fastened to divide them from unrelated content. The results of Studies 6, 7 and 8, however, demonstrate that learning gains are achievable for unrelated (high discriminatory) mathematics content. This suggests the possibility that all mathematics concepts within the DLCP may be interleaved. This condition could be applied to the DLCP by a small change in tutor procedure.

When tutors remove cross-curricular flashcards from each pocket prior to retrieval testing, they could simply group the mathematics flashcards randomly and test them contiguously, before other subject area content. This simple procedural change would not add significant time or difficulty to the process and should be within the capability of all tutors. Future research would be required to determine if random interleaving made a statistical difference to the learning gains associated with spacing and retrieval effects (Table 17, A). Additionally, these research variables could identify the presence of a statistical difference between the low versus high versus mixed discriminatory material within the DLCP.

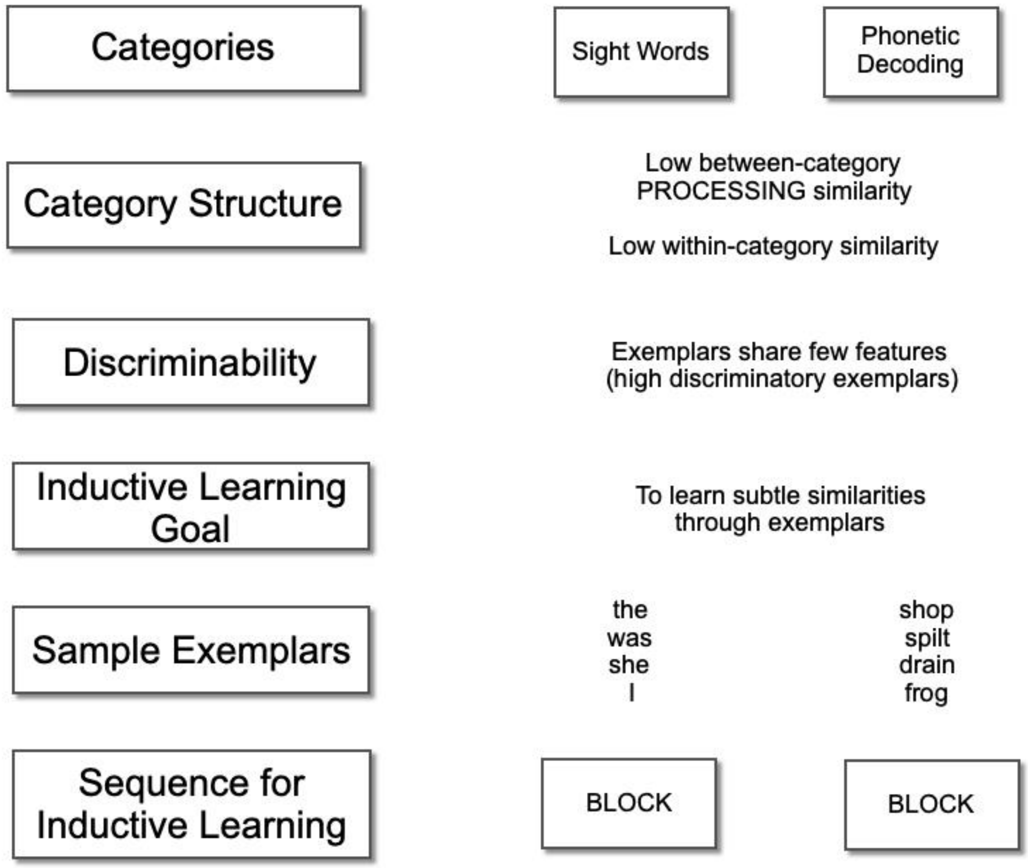
The DLCP currently includes the interleaving of language-based content including sight words, phonetic decoding and spelling. Anecdotally, the researcher has observed learning gains within this domain, however, Brunmair and Richter's (2019) meta-analysis (25 studies on expository text concepts and 13 studies related to words) identified little evidence for inductive learning through interleaving on the language-based topics investigated. One of the studies by Hausman and Kornell (2014), is unique in comparing high discriminatory language concepts, however, the nature of Study 9 does not supply sufficient evidence to theoretically discount potential interleaving learning benefits within the language domain. Rohrer et al. (2014) note that broader applications of the results of interleaving mathematics content remain unknown. Contextually relevant DLCP research is needed to understand the factors influencing the results of interleaving within the language domain, particularly the potential for non-inductive learning gains through spacing and testing effects (Table 17, C). Carvalho and Goldstone's (2014) sequential attention theory raises the possibility of an alternative sequencing condition for language-based content.

Sequential attention theory supports the discriminative contrast theory when applied to high discriminatory material within- and between-categories through studying exemplars in the blocked condition (Carvalho & Goldstone, 2014). This condition may facilitate learner ability to discern the characteristics of a category. Study 11 was based on perceptual features of blob figures, so direct applications to classroom learning are not possible without further research. However, phonetic decoding, sight words and spelling, are highly discriminatory in terms of the procedures used to identify their exemplars; phonetic words are decoded, sight words use a combination of phonetic components and memory, and spelling is based on combinations of segmenting, memory, semantics and rules. The within-category discriminability of exemplars of these topics is also high, as most words look or sound different.

According to sequential attention theory, the relative importance of differences between categories (highlighted by interleaving) or similarities within categories (highlighted by blocking) determines the best sequencing condition (Carvalho & Goldstone, 2015). Language-based topics such as phonetic decoding, sight words and spelling, may achieve greater learning gains through the blocked rather than interleaved condition (Figure 35). In decoding phonetic words contiguously, students will practice the same sound in the context of different words and, as a result may determine the similarity and predictability of the sound-symbol relationships. Likewise, sight word learning may benefit from students observing some similarities in their component sounds and rules. Future DLCP research

(Table 17, C) could compare blocked and interleaving conditions. Study 10 may provide insights into sequencing options that combine both the mathematics domain and language-based categories within the DLCP.

*Figure 35*. Potential blocking by category for language-based topics.




Yan and Sana (2020) identify that there is a general research gap in interleaving studies based on unrelated learning concepts. Their study compared interleaved versus blocked conditions to domain versus conceptual (category) learning content (Figure 34). They identified that learning gains were achieved through the interleaving of either domains or categories, but not both which was the condition of the original DLCP. When the results of Study 10 are compared with the previously discussed studies, complementary parallels emerge.

Interleaving studies of high (and low) discriminatory mathematics content, have focussed on mathematics as a separate domain. Results have been achieved through both discriminative contrast theories (including sequential attention theory) and study-phase retrieval. Within the original DLCP, categories were randomly interleaved across domains. The results of synthesis studies may be applied to the DLCP if mathematics flashcards are first separated from the other domains. Study 10 adds weight to the previously described process modification of grouping mathematics flashcards as they are removed from each pocket prior to retrieval testing. Pending future research, exemplars from different mathematics categories would continue to be randomised. The following DLCP applications based on Study 10 are related to language-based topics.

Using two domains, Study 10 identified that "learning was best supported when one level was interleaved and the other level was blocked": conditions (b) and (c) (Yan & Sana, 2020, p. 9). Flashcard sequencing within the DLCP may be modified to apply their results whilst also being practically feasible.

First, in keeping with the results of Studies 6 – 8 and 10, the mathematics domain flashcards could be blocked when removed from each pocket, prior to retrieval testing. Categories within the mathematics domain, however, would be practised in the randomly interleaved condition. Theoretically, learning gains may be achieved through study-phase retrieval. This condition corresponds to Yan and Sana's (2020) condition (b).

Second, the remaining categories, for example sight words, phonetic decoding and spelling could each be blocked for contiguous retrieval practice of exemplars. This condition also corresponds to Yan and Sana's (2020) condition (b), however, at the level of category and exemplar. To achieve these conditions, tutors could be requested to separate flashcards into 'topics' described as spelling, sight words, phonetic decoding, or other topic areas, as they are removed from each pocket prior to retrieval testing. For simplicity, the mathematics domain, could also be described as a topic. Figure 36 illustrates these conditions. In summary, these conditions may facilitate:

- Blocked language *categories*, with exemplars practised within-category. Based on the sequential attention theory (for high discriminatory learning), learning gains may be achieved through students becoming aware of the similarities between exemplars, together with study-phase retrieval (Studies 10 and 11) and
- Blocked mathematics *domain* for the interleaved practice of exemplars from high discriminatory categories with learning gains based on study-phase retrieval (Studies 6 – 8 and 10).

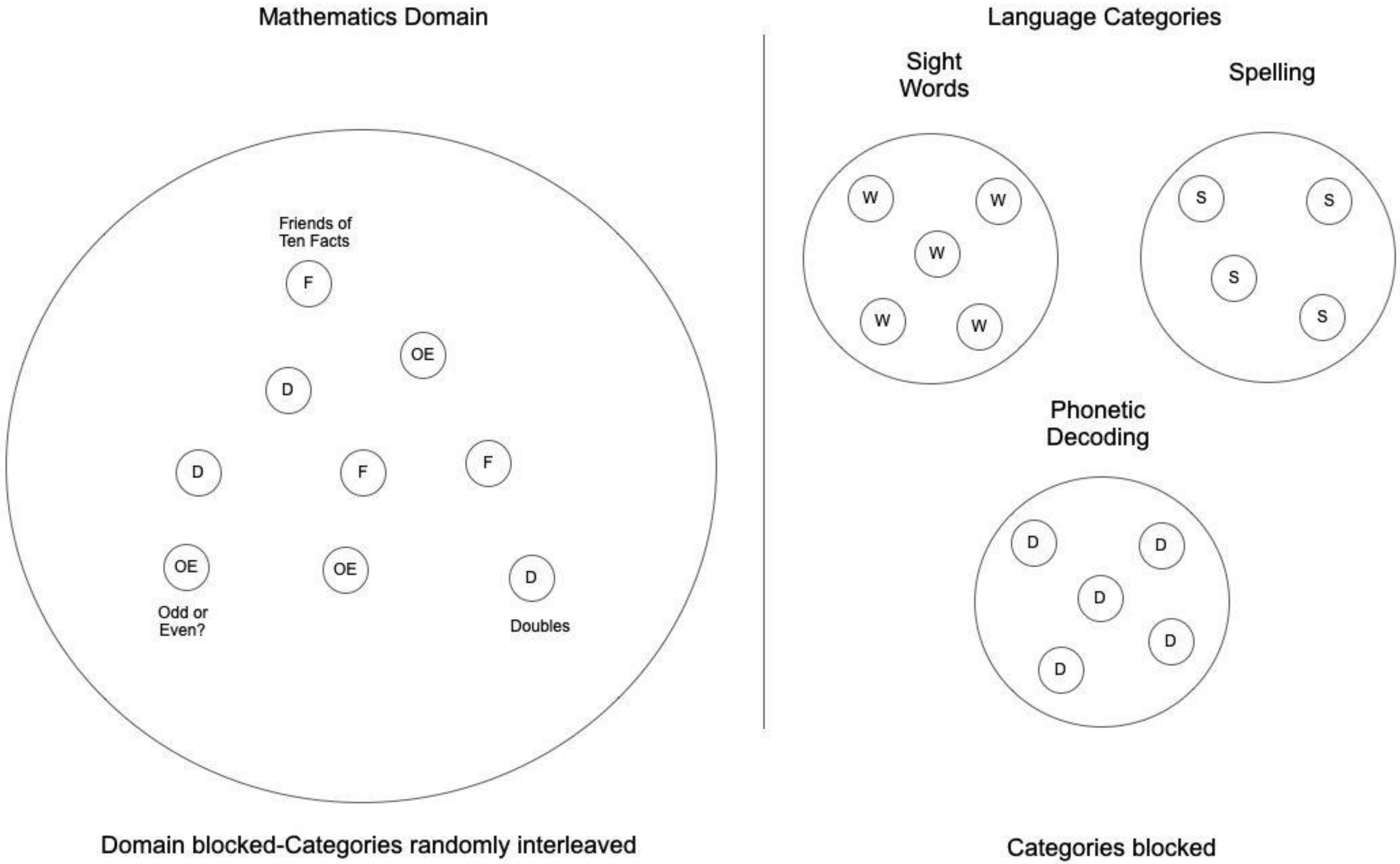
Future DLCP research could compare various combinations of these conditions (Table 17, B, C and D).

Table 17

*Future Research Conditions to Compare Learning Gains using High and Low Discriminatory Material, Within- and Between-Category Exemplars in the DLCP*

| | Research Conditions |
|---|---|
| | Mathematics |
| A | juxtaposed low discriminatory within-category content<br>versus<br>juxtaposed high discriminatory between-category content<br>versus<br>randomly mixed high discriminatory between-category content<br>versus<br>randomly mixed high and low discriminatory content |
| B | mathematics domain blocked, categories interleaved versus categories blocked |
| | Language |
| C | Categories: sight words, phonetic decoding and spelling<br>interleaved categories versus blocked categories |
| | Additional Cross-curricular Topics |
| D | other categories blocked versus categories interleaved |

*Figure 36.* Sequencing modifications within the DLCP prior to further research.



The application of sequencing conditions from the synthesis studies, relate to the interleaving of flashcards within each pocket. The DLCP, however, has a second level of interleaving: that which occurs between pockets. Each pocket will repeat the blocked topics it contains, for example, mathematics, phonetic decoding, sight words and spelling. Research was not identified to address this condition and empirical research is required.

In a review of cognitive science learning strategies, Dunlosky et al., (2013) describe interleaving as having a moderate utility. They suggest inconsistent results prior to 2013 may relate to a failure to fully understand the theoretical mechanisms involved, for example, how the amount of initial instruction and practice may influence the benefits of interleaving. Visible Learning (2019) identified that interleaved practice sits in the zone of desired effects with an effect size of 0.47. However, further research on interleaving within the DLCP is required to determine the validity of the described synthesis applications.

In general research terms, the amount of initial instruction, the relationship of the strategy to student ability and the complexity of the learning content would benefit from further research (Dunlosky et al., 2013). Rohrer (2012) lists the following research priorities:

- the effects in the classroom using meaningful procedures and time frames,
- combinations of blocking and interleaving,
- analysis of cost, ease of use and efficacy and
- student reactions to the additional learning effort required.

The application of theory to the DLCP has thus far addressed spaced retrieval and interleaved practice. To conclude these chapters, the revised process will be compared with the two approaches with which it shares similarities: the traditional mastery learning folder strategy (Appendix F) and the Leitner box (“Leitner system”, 2020).

**Related Interventions**

Early in the research journey, a description of the traditional mastery learning folder approach was discovered (Appendix F). Unfortunately, no research was identified for this tool or process. Due to their shared characteristics and era, the traditional process may have been based on the Leitner box ("Leitner system", 2020). These approaches share similarities with the DLCP, therefore some key differences will be clarified.

The Leitner system uses the spaced retrieval of flashcards and their return to the first compartment of a box when incorrect ("Leitner system", 2020), as in the traditional folder process and the original DLCP. Based on the investigation by Rawson and Dunlosky (2011), the revised DLCP requires three cumulative retrievals to reach the criterion identified as sufficient for short term retention, and three for longer term retention.

Unlike the traditional folder process, the revised DLCP and the Leitner system share the characteristic of expanding intervals between retrievals. Rather than arbitrary intervals based on session scheduling, the revised DLCP links the duration of intervals to retention goals (Cepeda et al., 2008).

Flashcards in the traditional folder process have the potential to advance through the pockets within a single session. Whilst recall performance within the session may demonstrate accuracy, according to the new theory of disuse (Bjork, 1999) learning gains may not endure in the long term. The DLCP mastery process, however, permits the movement of flashcards forward by only one pocket per session when successfully retrieved, resulting in spaced practice over days and weeks rather than minutes and days. A more detailed comparison of the similarities and differences between the three approaches is presented in a Venn diagram in Appendix L. Additional effort is a feature of desirable difficulty strategies, therefore, an understanding of metacognition and methods to enhance self-regulation and self-efficacy, may be of great benefit to teachers, tutors and students.

# Chapter 8: Metacognitive Development Strategies

The multifaceted concept of metacognition was first developed by Flavell (1979) and is defined as an “awareness or analysis of one's own learning or thinking processes” (Merriam-Webster, n.d.). Dunlosky and Metcalfe (2009) describe three facets: metacognitive monitoring, metacognitive knowledge and metacognitive control. Monitoring is the ability to reflect on the current state of one’s thinking. Knowledge consists of the understandings of metacognition, and metacognitive control is the ability to regulate a cognitive activity. Developing theories suggest that metacognition may be “a ‘higher-order’ cognitive process closely linked and implicated in our executive function (Roebers & Feurer, 2016)” (Duchesne & McMaugh, 2016, p. 271). An investigation of metacognition is highly relevant to the desirable difficulty strategies of spaced retrieval practice and interleaved practice and, therefore, to the DLCP.

Substantial investments of time and personal effort are required by students to achieve learning goals (Hattie & Yates, 2014), particularly when using strategies of desirable difficulty. As students receive feedback they are confronted by exactly what they do and do not know. With support from their teacher, tutor and the nature of the process, they may learn to monitor their progress. Metacognitive control is needed to persist with DLCP learning tasks and delay gratification with the ultimate objective to “welcome errors as opportunities to learn” (Hattie & Yates, 2014, p. xi). Under optimum conditions, this self-regulation can be ‘taught and caught’ potentially leading to self-efficacy and motivation (Hattie & Yates, 2014). A narrative review was used to study the metacognitive theory potentially applicable to the DLCP.

This chapter reviews relevant literature on metacognition to determine the alignment of the DLCP with current research and identify potential evidence-informed maintenance or modification. Research findings are classified and discussed according to the three components of metacognition: monitoring, knowledge and control. Incorporated within these topics are aspects of relevance to the DLCP and the selected desirable difficulty strategies including judgements of learning, feedback, self-regulation, self-efficacy and motivation. Metacognition develops over time with increasing understanding of the skills that enhance learning, including the ability to "plan, monitor, evaluate and self-regulate" learning (Duchesne & McMaugh, 2016, p. 142).

## Metacognitive Monitoring

Dunlosky and Metcalfe (2009) describe metacognitive monitoring as a learning 'power strategy'. The DLCP may provide a context to develop student metacognitive monitoring, initially through the direction of teachers and tutors and, in later educational pursuits, through self-regulation.

### *Judgements of Learning*

Whilst young children have been shown to have an awareness of thinking, their metacognitive skills are limited (Duchesne & McMaugh, 2016). Students of all ages, however, may demonstrate inaccurate metacognitive monitoring (Dunlosky & Metcalfe, 2009). When studying, independent students judge their revision to determine if they have learnt content with sufficient automaticity to recall in an exam. When using flashcards to self-test, however, Kornell and Bjork (2008) identified that students frequently drop correct flashcards too early in the mistaken belief that the content is secure in long-term memory. Such misjudgements of learning, called illusions of knowledge, are common and have been extensively studied (Birnbaum et al., 2013; Dunlosky & Metcalfe, 2009; Yan et al., 2016). Metacognitive monitoring is calibrated and enhanced through feedback (Brown et al., 2014).

#### Feedback.

Research identifies that overestimates of learning are reinforced by certain study methods (Dunlosky et al., 2013). Strategies such as blocked practice and massed learning focussed on encoding (storage), deliver a sense of fluency with learning material and may foster misplaced student confidence in their understandings (Brown et al., 2014). The potential misrepresentations of learning that these strategies generate are prevalent in the classroom (Dunlosky et al., 2013). The influence of this feedback on illusions of knowledge was reviewed to determine if and how the DLCP could facilitate more accurate judgements of learning.

#### ***Blocked Practice.***

Within the classroom, traditional textbooks often introduce new material and related practice examples in succession, blocked by topic or type (Taylor & Rohrer, 2010). Content in lessons and homework practice (Brown et al. 2014) may follow suit. This pedagogy arose from the belief that learning is better retained if practised immediately after it is understood (Rohrer & Taylor, 2006).

Termed the overlearning strategy (as distinct from overlearning as a description of the degree of mastery), it involves immediate practice of multiple examples of the same type. In striving for efficiency, this strategy may enable teachers and students to economise learning: minimizing the number of practice items, the time involved and the effort and motivation required (van Merriënboer & Kirschner, 2017). As identified in Chapter 7: Interleaved Practice, these short-term learning gains may be advantageous to reach a baseline criterion when a new topic is introduced, however, without further intervention, long term retention may fail, particularly in mathematics when discrimination between solution options is required (ColumbiaLearn, 2018; Rohrer et al., 2005). As a result, illusions of knowledge may distort perceptions of progress for both students and teachers, creating the impression that no further practice is required. Yan et al., (2016) identify three reasons for the illusions of knowledge metacognitive phenomenon.

First, as previously identified by Kornell and Bjork (2008), learning in blocked format feels fluent. The authors informed study participants that 90% of students achieved better results through interleaved practice, however, the majority of students associated their learning gains with blocked practice. The authors responded, “We know of no experiment that can match the current findings in terms of sheer inaccuracy of judgements” (Kornell & Bjork, 2008, p. 591). Second, test results during practice in blocked format suggest the successful acquisition of learning. As previously described, Rohrer and Taylor (2007) compared interleaving versus blocked practice for student consolidation of the formulae used to calculate the volume of geometric solids. Results during practice were consistent with studies that show that performance during instruction is impaired due to practice being more difficult, slow and less accurate (Rohrer & Taylor, 2010). In a delayed test, however, the results were reversed with interleaving returning a three-fold advantage, supporting the finding that performance during instruction is not a reliable index of learning (Bjork, 1999). Third, rather than a focus on the mechanics of memory, students may believe that learning style is individualistic (Yan et al., 2016), a theory that is not supported in research literature (Brown et al., 2014). The counter-intuitive nature of the benefits of interleaving have been replicated in many studies (Birnbaum et al., 2013).

Based on the DLCP revisions described in Chapter 7, the use of the desirable difficulty strategy of interleaving—after initial blocked practice in class—may increase student thinking effort and slow down learning, to decrease erroneous perceptions of fluency (Rohrer & Taylor, 2010) and facilitate more accurate judgements of learning. When solely relied upon, massed encoding practice also gives students a sense of fluency with learning material (Brown et al., 2014).

***Massed Encoding Practice.***

In the primary school context, massed practice on a particular topic is that which occurs within defined time frames such as consecutive lessons or weekly themes. It focusses on *storing* knowledge rather than its retrieval. Examples of massed practice include instruction, reading and viewing topic-based content. Conversely, when learning is spaced, students are required to think again and again about the topic as the revisit previous learning. As students progress through school, more study self-regulation is required. Popular encoding strategies include rereading notes and highlighting text (Dunlosky et al., 2013).

Rather than the establishment of durable knowledge, ease in following an argument or understanding a concept when rereading, may only create the illusion of it. Though retrieval strength is high under these conditions, storage strength is low and, therefore, knowledge may not be recalled at a delay (Roediger & Pyc, 2012). This inaccurate metacognitive monitoring may persist until the material is later tested, at which time the student could discover that the knowledge, presumed secure, is unable to be retrieved.

A teaching or self-study focus on massed encoding techniques creates a problem for students: a failure to know what they do and do not know and therefore a failure to use metacognitive control to address learning more effectively (Brown et al., 2014). For younger students, this may mean a lack of opportunity—or failure to communicate—what they do not understand. Changes to the DLCP introduced in Chapter 6: Spaced Retrieval Practice, may assist teachers and students to counteract the illusions created by this form of instruction and practice.

Due to the specificity of DLCP learning goals, spaced retrieval practice inherently provides fine-grained accurate feedback through the movement of learning items forward to the next pocket when correct, or their return to the pocket of origin when incorrect. Students therefore understand exactly what they do and do not know, and tutors and teachers can respond immediately with further feedback and instruction.

Through the DLCP, students who experience an appropriate degree of forgetting will require increased thinking to reacquaint with the learning material. Adaptive spacing of intervals within the process may provide a means to moderate desirable difficulty. When combined with retrieval practice, rather than encoding, both storage strength and retrieval strength are assessed which may improve both teacher and student judgements of learning (Dunlosky & Metcalfe, 2009). The research of Chapter 6: Spaced Retrieval Practice introduced an additional form of retrieval to the DLCP: elaborative interrogation or, for more independent students, self-explanation.

Elaborative questioning of students by tutors during DLCP requires students to reflect and explain their understandings. It was identified in Chapter 6, that this strategy may enhance learning more than the repeated retrievals used by the original process. As a learning strategy, elaborative interrogation is thought to link new learning to established understandings (van Merriënboer & Kirschner, 2018). As a metacognitive strategy, the depth of students' declarative knowledge may become self-evident, potentially preventing or recalibrating illusions of knowledge and demonstrating the need for further instruction. The monitoring of learning may also be enhanced by the tracking of progress by teachers and students (Brown et al., 2014) as developed within the original DLCP.

When learning items enter the DLCP, they remain within the folder until mastered, therefore, teachers may use checklists to record new topic bundles as they enter the Store pocket, a more efficient routine than recording individual mastered flashcards as they leave the process. These checklists may assist teachers to monitor student progress along learning progressions. An on-going tally of mastered flashcards (Appendix J) provides a visual record to students of their progress which may support metacognitive attributes such as self-efficacy; 'I persevered with this learning content and I have mastered it.' These forms of teacher and student feedback may improve metacognitive monitoring and therefore, should be maintained within the DLCP.

Based on moderate evidence, the Education Endowment Foundation identifies feedback as having a high impact strategy for very low cost (EEF, n.d.). Within the DLCP, feedback and metacognitive monitoring are facilitated through learning goals, strategies of desirable difficulty, mastery criterion, assessment, immediate and corrective feedback and teacher/student progress tracking. Interleaved and spaced retrieval practice may provide teachers with an option to provide cumulative practice following blocked and massed instruction to prevent or recalibrate student illusions of knowledge.

"Good judgement is a skill one must acquire, becoming an astute observer of one's own thinking and performance" (Brown et al., 2014, p. 105). Whilst advantageous for learning, accurate feedback may challenge student ego esteem needs (Hattie & Yates, 2014). Metacognitive knowledge may assist students' on-going development in this area.

**Metacognitive Knowledge**

Metacognitive knowledge is understandings of metacognition that can be 'declared', that is, explained with words (Dunlosky & Metcalfe, 2009). Some common classroom learning strategies like massed learning, create inaccurate metacognitive knowledge (Logan et al., 2012). Similarly, the perceived learning outcomes of the desirable difficulty strategies used within the DLCP, may be counterintuitive. Fluency with to-be-learned materials feels like learning: difficulties and challenge do not (Bjork, 1999).

The foundational principle of the desirable difficulty framework is that thinking effort facilitates the movement of learning from working memory to long-term memory (Bjork & Bjork, 2011). These strategies are therefore designed to create learning challenge, a condition which may result in students underestimating both their progress and achievement. Logan et al. (2012) investigated if providing college students with contrasting experiences (massed vs. spaced practice), learning feedback and instruction on the benefits of spaced learning, would improve their judgements of learning.

Experiments 1 – 3 of Logan et al.'s (2012) investigation demonstrated small gains in judgement of learning accuracy through experience and feedback when spacing effects were large. Experiment 4, however, included the provision of explicit instruction on the spacing effect. Whilst the magnitude of learning due to the spacing effect was still underestimated, this condition produced a significant improvement in student judgements of learning in the spaced practice condition. The development of teacher, tutor and student metacognitive knowledge may hold the potential to circumvent student ego esteem challenges and positively impact student progress.

The Education Endowment Foundation Metacognition and Self-Regulation Guidance Report contains a series of recommendations for applying metacognitive research evidence in the classroom (EEF, 2019). These understandings may be applied to the DLCP to improve metacognitive knowledge outcomes. The first recommendation is for teacher professional development.

Recommendation 1 highlights the importance of explicit teacher *training* in metacognition and how students learn (EEF, 2019), also an Australian Professional Standard for teachers (AITSL, 2011). Prior to the current study, the researcher provided professional development in the instructional design and implementation of the DLCP, however, metacognitive factors were not addressed. Future teacher and tutor training could include instruction in desirable difficulty theory and associated metacognitive effects to highlight that all learning is not necessarily as it seems; student performance during instruction does not guarantee durable knowledge (Bjork, 1999). These understandings may enhance teacher and tutor responsiveness to student ego esteem needs, as well as improve the communication of these theories to students.

The second recommendation is for teachers to explicitly *teach students* cognitive and metacognitive strategies (EEF, 2019). In providing tutors and students with explicit instruction on how learning occurs and the factors that influence success within the DLCP, metacognitive understandings may grow and facilitate further cognitive gains. Metacognitive knowledge is also enhanced when modelled by teachers and tutors.

Recommendation 3 suggests teachers use ‘think alouds’ to *model* metacognitive monitoring and control skills across all subject areas in classroom learning (EEF, 2019). Within the DLCP this modelling could be incorporated into teacher and tutor feedback to students. For example, a tutor could say, “When I look at 4 + 5 + 6 + 5, I think to myself, ‘I know that four and six are friends of ten, and that five and five are doubles. I’d start with those numbers first.’” Students who observe teacher and tutor thinking are guided in their own which may enhance the outcomes of the DLCP. Think alouds also reduce the cognitive load of learning tasks (Duchesne & McMaugh, 2016). They may assist students to understand the calculated use of challenge within desirable difficulty strategies and why they have been included in the classroom learning programme (Brown et al., 2014). For example, a tutor could say, “I’m so glad that this question is making you think! Keep thinking … you know that it will help you to remember the answer next time”.

Recommendation 4 recognises the importance of an appropriate level of student cognitive *challenge* for the development of metacognition and self-regulation (EEF, 2019). Willingham (2009) however, cautions that avoiding failure is more motivating than the equivalent return on a successful experience. He notes that curiosity is sparked by knowledge gaps, not chasms; if a knowledge gap is perceived to be manageable—and there is a known process by which it can be closed—curiosity may motivate the effort required. Already an objective within the DLCP, teachers and tutors who understand the reasons and ways of manipulating desirable difficulty within the process, may be able to create the appropriate challenge to success ratio for the learning content contained within it. Cognitive load is also pertinent to an appropriate level of challenge.

The DLCP seeks to manage cognitive load for the development of automated knowledge through:

- a consistent procedure,
- appropriately levelled learning goals identified through classroom instruction,
- incremental learning progressions,
- low element interactivity,
- moderated desirable difficulty learning strategies,
- corrective feedback and
- point-of-need explicit instruction.

An understanding of potential ego esteem challenges may assist teacher and tutor to anticipate discouragement and be responsive to student needs by communicating their knowledge of the nature of human cognitive architecture, the way we learn and emotional expectations in the context of desirable difficulty. This may encourage students to persevere. Students with learning science theoretical understandings may experience the thrill of progress and come to recognise that errors and difficulty are a means to an end: learning (Hattie & Yates, 2014). Recommendation 5 encourages educators to promote and develop *metacognitive talk.*

Chapter 6: Spaced Retrieval Practice, introduced the incorporation of elaborative questioning following retrieval attempts. The DLCP may provide a scaffold for the development of these ‘how did you remember’ and ‘why is this so’ questions to encourage student to think about thinking and participate in metacognitive talk. Teachers and tutors may be encouraged to model the language of metacognition within lessons and the DLCP. Through modelling, guidance and feedback from teachers and tutors, increased student understanding of metacognitive monitoring and knowledge, may provide a foundation for the beginnings of self-regulation, also known as metacognitive control (Dunlosky & Metcalfe, 2009).

**Metacognitive Control**

Flavell (1979) identified a continuum of student metacognitive understandings from pre-school to late primary school with associated effects on learning. Young students may only distinguish between understanding and not understanding. Flavell suggests that such early competencies may be used as building blocks for advancing metacognition.

***Self-Regulation***

Independent students who self-select efficient learning strategies, such spaced retrieval, are likely to monitor their learning more accurately. These judgements in turn, may inform further ‘what, when, why and how’ decisions about on-going study. Younger school students are dependent on their teachers to manage these decisions, however, the DLCP may provide both a means of learning and a model for potential future self-regulation.

Winne and Hadwin (1998) describe self-regulated learning as consisting of four stages: task definition, goal setting and planning, enactment and adaptation. Table 18 displays the expression of these stages within the DLCP and highlights that the process is recursive (Dunlosky & Metcalfe, 2009).

Table 18

*The DLCP as a Teacher Tool for the Development of Student Metacognitive Control*

| Self-regulation Stages | DLCP Expression |
|---|---|
| task definition | • student prior knowledge<br>• differentiated learning intentions |
| goal setting and planning | • learning items in flashcard format<br>• mastery criteria<br>• learning progressions |
| enactment | • spaced retrieval practice<br>• elaborative interrogation<br>• interleaved practice<br>• teacher and tutor feedback, modelling and instruction of cognitive and metacognitive knowledge<br>• teacher, tutor and student metacognitive monitoring |
| adaptation | • self-paced mastery process<br>• differentiated number of retrievals<br>• differentiated desirable difficulty through inter-study intervals manipulation |

Metacognitive knowledge and monitoring, and the DLCP functioning as a tool and model of metacognitive control, may contribute to the development of student self-efficacy and motivation.

**Reciprocal Determinism.**

Bandura's (1986) social cognitive theory research increasingly focussed on self-regulation, self-efficacy and motivation, emphasising learners' personal agency (cited in Duchesne & McMaugh, 2016). His reciprocal determinism model describes interactions between learners' personal cognition, behavioural responses and the external environment. Personal cognition and behavioural responses may find expression in the 'external environment' of the DLCP.

Personal cognitive factors such as a learner's knowledge, beliefs, emotions, and self-efficacy (Duchesne & McMaugh, 2016) may influence how the DLCP is perceived and how an individual will respond to it. As previously discussed, metacognitive knowledge on the nature and experience of desirable difficulty learning strategies, may influence students' behavioural responses when they succeed or fail to accurately retrieve a learning item. For example, if a student knows that the struggle to remember learning assists in the formation of long-term memories, their beliefs about failure may change. Negative emotional reactions may be reinterpreted (Bandura, 1974). Incidentally, the DLCP will provide practise in making errors and students may begin to see them as a part of the learning process (Hattie & Yates, 2014). If perseverance ultimately leads to successful outcomes, the ability to delay the gratification of a successful retrieval may grow. To summarise, metacognitive knowledge may lead to more accurate beliefs about learning, altered beliefs about failure may lead to emotional regulation and, mastery experiences, to growing self-efficacy (Bandura, 2010; Bloom, 1968).

**Self-Efficacy.**

Bandura (1974, p. 77) defines self-efficacy as "the belief in one's ability to influence events that effect one's life and control over the way these events are experienced". When learners experience success, their expectations of future success are enhanced and they are more likely to apply the effort and persistence required (National Academy of Science, Education and Medicine, 2018). The DLCP may structure the environment, to raise student belief in their own capabilities, through challenging but successful learning experiences (Bandura, 1974). Success is measured by progress (Bandura, 1974; Goss et al., 2015), not performance.

Self-efficacy has a major influence on student learning and is "as influential on their performance as their actual abilities" (Dunlosky & Metcalfe, 2009, p. 205). Kirschner and

Hendrick (2020) clarify, however, that self-efficacy is domain specific. A strong sense of self-efficacy in reading, may not translate to confidence in mathematics.

As discussed in Chapter 4: Baseline Data, the researcher anecdotally observed improvement in self-efficacy as students came to predict that learning items would move forward through the DLCP over time. Research identifies that when students sense that learning is within their control, it has a powerful effect on their motivation (Duchesne & McMaugh, 2016). According to Bandura (2008, p. 1)

> Among the mechanisms of agency none is more central or pervasive than beliefs of personal efficacy. This core belief is the foundation of human motivation, well-being, and accomplishments. Unless people believe they can produce desired effects by their actions they have little incentive to act or to persevere in the face of difficulties. Whatever other factors serve as guides and motivators, they are rooted in the core belief, that one has the power to effect changes by one's actions.

The DLCP relates to motivation as an emergent phenomenon; it can be developed over time and through student experiences with learning (NASEM, 2018).

**Motivation.**

Motivation is described as a "condition that activates and sustains behavior toward a goal" (NASEM, 2018, p. 109). NASEM (2018) identify that cognitive theories focus on intrinsic motivation. Three psychological principles are foundational to this form of motivation.

The first principle is an incremental view of intelligence (malleable versus fixed learning capacity). Foundational to the DLCP is the importance of knowledge. The more a student knows on a topic the more readily they can assimilate new learning and therefore experience success (Brown et al., 2014). Mastery learning theory (Bloom, 1968) supports an incremental view of intelligence with mastery learning goals subject to the variables of time and effort.

The second principle of intrinsic motivation is that learning material should reflect a student's prior knowledge. Appropriately differentiated learning content within the DLCP based on foundational knowledge in long-term memory, may assist students to manage intrinsic cognitive load and persevere with the mental effort of learning (Sweller, 2016). Achievable learning goals are most effective when specific and proximally attainable (Duchesne & McMaugh, 2016).

The third principle of intrinsic motivation is a learning orientation towards mastery rather than performance. The DLCP theories of mastery learning (Bloom, 1968) and self-efficacy (Bandura, 2008) reflect this learning goal. According to Bloom (1968), a student who masters appropriately differentiated learning content can be profoundly affected. Affective changes include:

- subjects once disliked are positively reframed,
- feeling of control over learning are developed,
- motivation for learning is increased and
- self-concept is enhanced.

The researcher's anecdotal observations support the claims resulting from mastered knowledge. According to Bandura (2008, p. 2), "the most effective way of building a strong sense of efficacy is through mastery experiences". Self-efficacy beliefs are influenced by the factors to which students attribute their success or failure which in turn influence motivation, performance and affective reactions (Bandura, 1974).

***Attribution Theory.***

To seek a cause for success or failure is to believe that the same success, or the avoidance of failure, is possible in the future (Weiner, 1985). Causal analysis of performance may be defined by locus (internal or external factors), stability factors and their controllability. In an educational setting, attribution theory is the perceived causes to which students attribute their academic success or failure. Kirschner and Hendrick (2020, p. 103) note that "the *perceived* cause of academic performance is as significant as the *actual* cause" (emphasis in original). In education settings, the dominant causes relate to ability and effort. Expectations and emotions arising from causal attributions can have a major influence on learning motivation (Weiner, 1985).

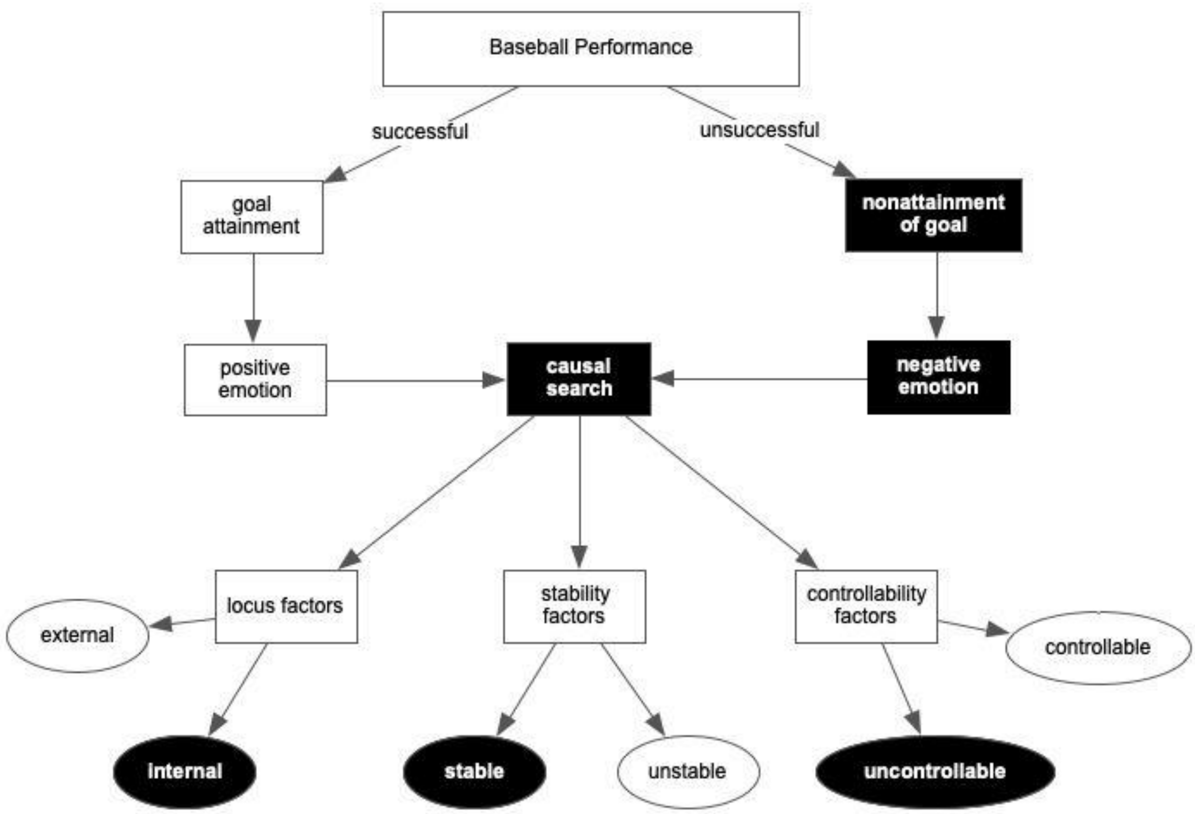
Internal factors have a greater impact on positive self-esteem for successful experiences or lower self-esteem for unsuccessful ones (Weiner, 1985). An example of an internal locus of causality is ability. If a student feels that they inherently lack the ability to succeed, motivation may be challenged. If, however, they feel that failure was due to an external factor, such as the unusual difficulty of an assessment, their self-esteem and motivation may not be as severely impacted (Weiner, 1985). The relative stability of causal factors influences student expectations of success (Weiner, 1985).

Stable causal factors are those that are constant or enduring, ability in a subject area is an example (Weiner, 1985). Unstable factors are those that are temporary or fluctuating, for example the application of effort. Kirschner and Hendrick (2020) explain that students who attribute failure to internal, stable causes like low ability, may suffer from poor ego esteem and disengage in learning while those who attribute failure to external unstable causes, such as lack of help or bad luck, may miss the opportunity to improve. Students who recognise unstable causal factors such as effort and perseverance, however, may have higher expectations of potential success. A third causal property is controllability.

Locus of causality and stability factors may both be influenced by factors of volitional control (Weiner, 1985). It may be within students' power to adjust the application of effort, however, mood or fatigue may be less controllable. Weiner (1985) provides an example the three dimensions of locus, stability and controllability using the scenario of a poorly performing Little League baseball player.

Figure 37 displays the achievement-related motivational sequence. The boy attributes his poor play to low ability: causal factors which are internal, stable and uncontrollable. He feels he is not a good player, therefore does not expect to be able to play well in the future. Given a situation that he believes he cannot control, he decides to stay home from the next game. The result is hopelessness. A different scenario may see a student achieving a performance goal and identifying that a decision to change (internal/controllability factors) to a different study strategy (external factor) led to improved learning (unstable factor). A positive affective response may follow.

*Figure 37.* An achievement-related motivational episode based on Weiner's (1985) scenario.



The previously discussed reciprocal determinism model (Bandura, 1986) describes interactions between learners' personal cognition, behavioural responses and the external environmental context, in this case, the DLCP. Bandura (2010) describes the creation of an environment, based on mastery experiences, that builds coping skills to potentially manage threats to self-esteem. The result being that students may be able to "perform successfully despite themselves" (Bandura, 2010, p. 6). An examination of the underlying causal properties of the DLCP in relation to Weiner's (1985) locus, stability and controllability may identify revisions to facilitate expectations of success and the management of affective outcomes.

In primary school settings, the DLCP is teacher directed and tutor delivered which, as displayed in Table 19, provides a high level of controllability. On behalf of students, the teacher seeks to manage the control of multiple causal factors in the pursuit of successful but effortful retrieval experiences:

- Being outside of the self, external factors have little impact on self-esteem (Weiner, 1985). Metacognitive discussions, instruction, monitoring and modelling may be used to attribute unsuccessful retrievals to the nature of the desirable difficulty strategy which is external.
- Success ascribed to internal factors positively influence self-esteem and feelings of pride (Weiner, 1985). For successful retrievals, teachers and tutors can attribute the result to internal factors; the student's thinking effort that moves learning into long-term memory over time.
- The stability of a cause relates to the relative expectancy of future success (Weiner, 1985). Whilst the learning items change, the DLCP is stable. The process is consistent and predictable.

Over time, students may experience and predict successful outcomes through the DLCP.

Table 19

*Suggested DLCP Internal and External Factors, Stability Factors and their Controllability*

| | Internal (student) | | External (DLCP) | |
|---|---|---|---|---|
| | Controllable | Uncontrollable | Controllable (teacher) | Uncontrollable |
| Stable<br>constant<br>enduring | | ability | DLCP strategy:<br>- session scheduling<br>- task characteristics and difficulty<br>- learning goals<br>- learning progressions<br>- participation<br>- teacher/tutor feedback, cognitive/metacognitive instruction, modelling. | interruptions<br>support |
| Unstable<br>temporary<br>fluctuating | engagement<br>thinking effort<br>persistence | mood<br>fatigue | | teacher / tutor expertise |

*Note.* Adapted from "How Learning Happens: Seminal Works in Educational Psychology and What They Mean in Practice," by P.A. Kirschner and C. Hendrick, 2020, Routledge.

Fulfilment of the following DLCP conditions may influence learning achievement and the development of self-regulation, self-efficacy and motivation.

Teachers

- ensure that students' prior knowledge is accurately addressed
- provide appropriate incremental learning goals based on learning progressions
- highlight learning goals and define success criteria,
- respond to student results by adjusting the level of desirably difficulty, if required, through manipulating interval duration to facilitate effortful but successful retrievals,

- know and explicitly teach cognitive and metacognitive knowledge on desirable difficulty theory to tutors and students,
- model metacognitive monitoring through feedback and elaborative interrogation and
- provide a supportive learning environment attributing unsuccessful retrievals to (external) desirable difficulties, successful retrievals to (internal) perseverance and an expectation of progress to the (stable and predictable) DLCP.

Well managed desirable difficulty with metacognitive understandings is key to student progress, the expectancy of future success, and potentially, learning motivation.

**Future Research**

The DLCP utilises and models metacognitive strategies. The effects of the DLCP on metacognitive development may be researched through comparing results, student attitudes, self-efficacy and motivation with variables that assess teacher and tutor metacognitive instruction and modelling.

The EEF (2019) identifies that, based on extensive evidence, the development of metacognition and self-regulation skills have a high impact on learning for a very low cost and should therefore, be seriously considered by schools seeking to improve student achievement. They also note that the development of these skills can be difficult to achieve in practice, particularly within the preferred context of content knowledge. With further research, the DLCP may provide a framework to contribute to the development of these skills for improved student outcomes and future independent study endeavours. Duchesne and McMaugh (2016, p. 113) conclude … "numerous studies have demonstrated that metacognition can be taught (Pennequin, Sorel & Mainguy, 2010) and that metacognitive knowledge and skill is related to academic achievement, to reasoning (Kuhn, 2000) and to intelligence (Sterberg, 1985)".

# Chapter 9: Conclusion

The initial goal of the thesis was to determine the theoretical foundations of a practice strategy—the Differentiated Learning Consolidation Process—which was developed in the classroom. The subsequent aim was to ascertain if, and how, these theories could refine the DLCP within the constraints of the classroom context.

## Background

After several years teaching in an upper primary gifted and talented programme, the researcher returned to mainstream classroom teaching. A learning focus on foundational knowledge and skills in the early primary years highlighted the difficulties of managing effective differentiation and consolidation of learning according to individual student needs. In response, informal classroom level action research was undertaken to develop a practice process that was both effective and suited to the limited resources of the classroom. A traditional tool, known anecdotally as a mastery learning folder, facilitated the delivery of the process. With ongoing modifications to the process, tool and implementation, the DLCP provided the researcher with the ability to better manage the consolidation of differentiated student learning. The experience led to the development of a hardcover mastery learning tool and the provision of professional development for interested teachers. Beyond some elementary understandings of mastery learning and the revised Bloom's taxonomy, the theoretical foundation of the process was unknown. It was hoped that a more thorough understanding of evidence-informed theory may lead to further process revision and improved student outcomes.

This chapter will summarise and reflect upon the research. The effectiveness of the methodology will be considered, and the major findings highlighted, followed by an assessment of the limitations of research in terms of methodology, validity and DLCP implementation. Potential future research will then be discussed, and the thesis will conclude with the significance of the research.

**Methodology Reflections**

The extensive search for an appropriate research design for the current study was well rewarded with the discovery of the realist synthesis logic of enquiry. This methodology facilitated the identification and incorporation of diverse and relevant data including the researcher's experience, baseline features of the process, potentially related programme theory and applications. Together, this data informed the revision of the DLCP.

The realist synthesis approach facilitated the fulfilment of the overarching and specific research goals, delivering applications and insights beyond the researcher's expectations, as well as ideas for future research. Additionally, the methodology's focus on theory may have assisted to address the potential bias of investigating a self-developed intervention.

**Major Findings**

The first overarching research question was 'What is the DLCP theory?' Section 1: Theory Identification, utilised the researcher's experience (Chapter 2) and the baseline annotation of the DLCP (Chapter 4) to contextualise the research. These chapters provided a rationale for the researcher's expert framing role: to identify and adjudicate between potential theories and their application (Pawson et al., 2004). A narrative literature review (Chapter 5) sourced data to identify potential DLCP programme theory. These chapters informed the second overarching research question; 'Can the DLCP programme theory be aligned with evidence-informed research, and if so, how?' which was addressed in Section 2: Theory Application. A summary of the major findings is provided in two sections: theory identification and theory application.

### *Theory Identification*

The search for relevant and applicable theory identified that the DLCP sits within the field of cognitive psychology. Two broad theories of relevance were identified: cognitive load theory (Sweller et al., 2019) and the new theory of disuse (Bjork & Bjork, 1992).

#### Cognitive Load Theory.

The focus of cognitive load theory on the aetiology of knowledge acquisition through effective instructional design (Sweller et al., 2019), provided the researcher with contextual understandings for the research that followed. Cognitive load theory directly addresses the purpose and goal of the DLCP: to secure the learning of classroom instruction within students' long-term memory. The DLCP may:

- reflect the understandings of human cognitive architecture and accommodate the limitations of working memory,
- address the importance of prior knowledge through the presentation of appropriately differentiated learning content, and
- have the potential to adjust element interactivity, through the complexity of learning item content designed for novice or expert learners.

These features may assist in the development of schema in students' long-term memory. The DLCP may also have an association with working memory expansion after the rest provided by spaced learning (Chen et al., 2017). Whilst not the application focus of this study, cognitive load theory will continue to influence the development and implementation of the process.

#### The New Theory of Disuse.

The new theory of disuse (Bjork & Bjork, 1992) provided many theoretical insights applicable to the DLCP. It was the researcher's observation that performance during instruction on the number bonds to ten was an unreliable index of durable learning that motivated the initial search for more effective learning strategies. Therefore, a theory that addressed the enhancement of memory storage and retrieval strength was a welcome discovery.

The review of research literature suggested that the DLCP may be consistent with the premises of the new theory of disuse and utilise the desirable difficulties strategies (Bjork & Bjork, 2011) of spaced, retrieval, and interleaved practice. New DLCP insights based on these theories were highlights of the research.

First, what was previously considered to be merely mastery testing, was identified as retrieval practice, a strategy of learning. Second, under certain conditions, the use of mixed flashcard content may align with the learning strategy of interleaving. Third, the inter-study intervals of practice—not the number of retrievals—were identified as the major influence on the duration of learning. Cognitive load theory and the new theory of disuse assisted the identification and refinement of the remaining, more specific DLCP theories.

#### DLCP Programme Theory.

Pawson et al. (2004) describe an intervention's programme theory as a theory of theories. Section 1: Theory Identification culminated in the proposed combination of DLCP programme theory as displayed in Figure 16. Section 2: Theory Application, presented the investigation of the desirable difficulty learning strategies of spaced retrieval and interleaved practice, together with metacognitive development strategies, to provide key data for the revision of the DLCP.

### *Theory Application*

The application of theory is described within three DLCP phases: the selection of learning content, the Criterion Phase and the Retention Phase. This is followed by the metacognitive applications which are more general in nature.

**Learning Content.**

Prior to the current study, the DLCP focussed on the learning objectives of remembering, understanding and applying. Using an ecologically relevant context and educational materials, Gluckman et al. (2014) identified that spacing practice achieved learning gains for generalisations. Potential may exist for the DLCP to incorporate higher order learning objectives when tutors have the capacity to assess student responses and provide appropriate feedback.

**Criterion Phase.**

The Criterion Phase involves establishing the retention of learning content over short durations prior to the longer inter-study intervals of the Retention Phase. The following key theory applications may be applied to the DLCP within this phase:

- Maintain the use of blocked practice during lessons to provide a foundation for interleaved mathematics practice (Rohrer et al., 2019).
- Request that tutors separate flashcards into the mathematics domain and language categories prior to the retrieval testing of each pocket to facilitate the most suitable sequencing arrangement for mixed curricular content (Carvalho & Goldstone, 2015; Yan & Sana, 2020).
- Maintain the focus on student prior knowledge and learning progressions to create effortful but successful retrievals. Introduce the adaption of interval durations to manage desirable difficulty based on the needs of selected students (Mettler et al., 2016; Pyc & Rawson, 2009).
- Provide a longer retrieval response time to enhance effortful thinking (Pyc & Rawson, 2009).
- Aim for successful mastery experiences through differentiated learning content, self-paced learning and the management of desirable difficulty to

foster the development of student self-efficacy and motivation (Bandura, 2010; Bloom 1968).

- Reduce the number of retrieval sessions to improve process efficiency (Pyc & Rawson, 2009; Rawson & Dunlosky, 2011).
- Maintain the use of feedback and replace repetitive retrievals with elaborative questioning and metacognitive talk to provide students with accurate judgements of learning and enhanced metacognition (EEF, 2019; Karpicke, 2017; Logan et al., 2012; Roediger & Pyc, 2012; van Merriënboer & Kirschner, 2018).

Many of the Criterion Phase findings also apply to the spaced retrieval sessions within the Retention Phase. The distinguishing feature of the Retention Phase is the introduction of longer inter-study intervals.

**Retention Phase.**

The Retention Phase seeks to establish durable knowledge over longer intervals. The key theory application was the introduction of rest pockets to incorporate three expanding inter-study intervals for effortful but successful retrievals and recall at educationally relevant delays (Cepeda et al., 2008; Kang et al., 2014; Küpper-Tetzel et al., 2014; Rawson & Dunlosky, 2011). The desirable difficulty goal of effortful retrieval has metacognitive implications. Key metacognitive theory applications relate to the process as a whole.

**Metacognitive Applications.**

Extensive evidence supports the association between metacognitive skills and high impact learning outcomes (EEF, 2019). The DLCP may provide a framework to systematically foster the development of student metacognitive skills within the authentic context of content knowledge (Hattie & Donoghue, 2016). Research data identified that metacognitive knowledge assists in the development of metacognitive monitoring and control which may benefit teachers, tutors and students.

Implementation of the DLCP has always included the training of teachers, tutors and students in the process and their roles within it, but has not addressed theoretical understandings. Future training could incorporate cognitive load theory, the new theory of disuse, explanations of desirable difficulty learning strategies and metacognitive development, in formats appropriate to the participants for metacognitive gains in monitoring and control. Teacher and tutor metacognitive talk and modelling hold the potential to enhance student learning.

The development of student metacognitive understandings may assist students to develop more accurate judgements of learning (Logan et al., 2012) and circumvent potential ego esteem challenges when illusions of knowledge are recalibrated through the use of desirable difficulty. This knowledge may also create more accurate beliefs about the nature of learning including the expectation of more errors, increased thinking demands and the possibility of discouragement during practice (Rohrer & Taylor, 2010). Desensitisation to these challenges may assist students to develop the perseverance required to see results improve (Hattie & Yates, 2014; Logan et al., 2012). Furthermore, the DLCP may function as a model of self-regulated learning for additional metacognitive gains.

The DLCP incorporates task definition, goal setting, enactment and adaptation of tasks according to results (Winne & Hadwin, 1998). If tutors—particularly of older students—make these associations explicit, it may encourage similar student organisation when their time comes for independent study, as required for school exams. Similarly, attributions of success or failure hold potential to positively impact student metacognition (Weiner, 1985).

Tutor metacognitive talk and modelling could include attributions of unsuccessful retrievals to external factors such as the nature of desirable difficulty. Successful retrievals may be attributed to the internal factor of perseverance and the predictability of the process could facilitate an expectancy of future success. These attributions may protect ego esteem (Hattie & Yates, 2014) and encourage self-efficacy and motivation (Weiner, 1985). The limitations of the research address the methodology, internal and external validity, and potential implementation constraints.

**Limitations**

The limitations of a realist synthesis approach are, perhaps, its defining feature. Pawson et al. (2004) describe the approach as involving "so many grey zones … so much off-piste work, so much wallowing in the subtle and contextual [and] so much negotiation of meaning with real-world practitioners"—such is research in the context of complex interventions within complex social systems. Specifically, the study findings are subject to limitations based on internal validity, external validity and issues of classroom implementation.

***Internal Validity***

Stakeholder participation and expert framing are essential components of realist synthesis and have a role to play in assessing internal validity. Realist synthesis requires expert framing in making judgements of data relevance and applicability, for example, in the purposive sampling and application of research studies. This was the researcher's responsibility. The researcher is an expert in the original DLCP, however, not in the identified cognitive psychology theories or their application. It is acknowledged, therefore, that the findings of this study are fallible.

In this context, internal validity is promoted through clarity on the part of the researcher (Pawson et al., 2004). This includes explaining selections, providing the reader with sufficient study information and presenting reasoning to enable stakeholders—particularly experts in the field—to evaluate and adjudicate the conclusions (Pawson et al., 2004). Judgement may follow in the form of refutation. Such feedback is encouraged as "exposure to scrutiny and critique is thus the engine for the revision and refinement of programme theories" (Pawson et al., 2004, p. 38) and potentially, in this context, a reworking of applications within the DLCP.

***External Validity***

The study involved the application of pertinent research to components of the DLCP, and, as such, does not claim a wider generalisation. Many teachers, however, have suggested that the DLCP could be developed into an online or software application. It is the researcher's belief that it would be difficult to replicate several features in a virtual context: the responsiveness to specific classroom learning objectives, the degree of individualised differentiation and the nuanced face-to-face interactions between teacher, tutor and student. Additionally, applications of the current research are subject to the limitations of the school and classroom context.

***DLCP Implementation***

Whilst the revision of the DLCP sought to recognise and address classroom context issues, it is acknowledged that it is the availability of resources and subjects (for example, teachers and tutors) that make interventions work and one size does not fit all (Pawson & Tilly, 1997; Pawson et al., 2004). Pertinent factors which may limit the successful implementation of the DLCP include the level of stakeholder commitment to the implementation process, the provision of adequate training, financial considerations and the availability of tutors. The range of implementation scenarios is broad.

Previous discussion sought to provide examples of DLCP use in various school scenarios. Implementation may involve a single student addressing the practise of learning content in one subject area or topic. Conversely, a primary school may implement the intervention across all year levels to support the differentiation and consolidation of cross-curricular learning objectives. Effective implementation across all scenarios requires consideration of several factors by administrative decision makers and teachers, including:

- an understanding of the DLCP programme theory to facilitate confidence in contextual application decisions,
- a realistic assessment of the available school resources, such as tutor options or the suitability of different delivery modes, and
- stakeholder time, effort and commitment.

The implementation of any intervention is an iterative process (Pawson et al., 2004). Hattie and Yates (2014, p. viii) note that students require "substantial investments of time, energy, structured tuition, and personal effort" to develop knowledge mastery regardless of inherent ability: the same could be said of the stakeholders responsible for the DLCP. A key factor to implementation success is initial and on-going professional development.

Whilst local embedding and adaption is presumed, researcher experience has determined that training, feedback and discussion are essential for intervention fidelity and the achievement of DLCP objectives. Ideally, tailored training would be provided for administrative decision makers, teachers, tutors and parents. This could include traditional face-to-face provision, coursework, videos, guidelines, or checklists, delivered at staff meetings, online, at parent gatherings or less formally. Through this guided consultation, stakeholders could ascertain what works "for whom, in what circumstances, in what respects, and how" (Pawson et al., 2004, p. v). After designated periods of use, it would be beneficial to review the implementation chain and any points of contention (Pawson et al., 2004). A commitment by stakeholders to adequate training and communication is key. Financial factors are also an important consideration in school contexts.

The Education Endowment Foundation Teaching and Learning Toolkit (Higgins et al., 2016, p. 4) describes any learning intervention costing less than $140 Australian dollars per student per year as “very low” and the DLCP intervention may fall within this parameter. Training in the process is currently provided online without charge. Volunteer parent tutors and the use of available education assistants who may already be assigned time to the class, may not incur additional costs. The cost of handmade or hardcover mastery learning folders or alternative tools are low. The appointment or extension of education assistant time would aid implementation, however, this would require more substantial funding. Whilst subject to school implementation decisions, the DLCP may be defined as a very low-cost strategy, and therefore potentially accessible to a wide range of schools.

Tutor support has been described throughout the thesis as one of the most important and potentially limiting factors of DLCP implementation. According to individual school contexts, a variety of tutor and scheduling options were described to maximise the accessibility of students to one-on-one DLCP delivery. Tutors can include parent helpers, education assistants and older student leaders. The training, knowledge and experience of tutors may impact student results, particularly in metacognition, however, the strategies of desirable difficulty remain the key mechanisms of learning. Teacher management options for the one-on-one delivery of Retention Phase sessions were also described, for example, staggering student testing—and the time involved—across the week.

Pawson et al. (2004, p. 38) conclude, that despite the limitations of realist synthesis—and the researcher would add, the limitations of the context—“much can be achieved through the drip, drip, drip of enlightenment”. As stakeholders gain experience with the intervention and observe the results, new possibilities for effective implementation may emerge. In the researcher’s experience, teachers are innovative in response to the limitations of their classroom context when they observe a learning return on investment.

**Research Recommendations**

With reference to studies of spacing, Kapler et al. (2015, p. 39) describe the necessity of “ecologically valid studies that use educationally relevant materials, timescales, and methods”. Similarly, De Bruyckere (2018) emphasises that research must move from the laboratory into classrooms. Based on the researcher’s observations of DLCP outcomes, the current study began with the question, ‘It works in practice, but does it work in theory?’ With a rudimentary baseline of DLCP programme theory established, the research focus needs to return to practice.

First, the most pertinent research will be to quantify the effectiveness of the revised DLCP to facilitate student progress along the learning progressions of classroom instruction. Subject to future supervisor feedback and methodology investigation, an empirical study is proposed with a focus on assessing individual student progress using authentic classroom learning materials. Second, research questions arising from section two of the current study could be addressed.

Future studies could investigate the parameters of various conditions within the DLCP in the classroom context with student populations. For spaced retrieval practice, desirable difficulty could be manipulated using the variables of response time, criterion level, adaptive inter-study intervals, number of spaced retrieval sessions and inter-study interval durations for retention at educationally relevant delays. For interleaved practice, learning gains could be compared in the blocked versus interleaved condition and the use of low or high discriminatory content. The use of domain and category sequencing could also be explored.

Third, DLCP metacognitive development strategies could be investigated through qualitative studies related to student attitudes, self-efficacy and motivation. Finally, the

DLCP may also be used as a research instrument for topics and applications related to its theoretical constructs in a variety of contexts beyond the field of education.

**Significance**

The motivation for the classroom development of the DLCP was two-fold. First was the researcher's experience that performance during instruction proves to be an unreliable index of durable learning and, second, was the need to address the diversity of individual mastery of learning content, an ever-present reality within most classes. The current study has developed a new theory, the DLCP, which may assist teachers to address these differentiation and consolidation concerns. A defining feature of the DLCP is the attempt to be responsive to the complexity of the classroom.

Whilst *based* on the learning content of classroom instruction, the implementation of the practice process is situated *outside* the restrictive lesson context and shared between teacher and tutors. Tutors facilitate individual feedback and instruction, and, through the mastery process, students can practise individualised learning objectives at their own pace across days, weeks or months. These factors combine to assist teacher management of learning differentiation and consolidation.

Consolidation was defined in the thesis as durable knowledge in long-term memory. Desirable difficulty learning strategies such as spaced retrieval and interleaved practice, have been identified as highly effective in creating durable learning (Dunlosky et al., 2013; Pashler et al., 2007), however, according to theorists, they have yet to reach their potential in terms of classroom application (Bjork & Bjork 2011; Dempster, 1988, Howard-Jones, 2014; Lovell, 2020, foreword by Sweller, p.7). Kapler et al. (2015) describe multiple implementation challenges including group delivery, potential peer distraction, variations in prior knowledge, and time constraints. Spacing learning across lessons over days, weeks and months, is described as logistically very difficult within the lesson context, as are "customize[d] schedules of practice for individual students and topics" (Mettler et al., 2016, p. 898). The DLCP incorporates the use of these effective consolidation strategies through avoiding the difficulties associated with the lesson context.

A requirement of the Australian Professional Standards for Teachers (Australian Institute for Teaching and School Leadership, 2011) is that teachers know how students learn and how they can teach effectively. Küpper-Tetzel (2014, p. 72) describes the transition of desirable difficulty theory to implementation as a "missing bridge between research and practice". With reference to spaced practice, Weinstein et al. (2018) note that teachers need more "evidence-based tools and guidelines for direct implementation in the classroom". The DLCP may provide the means for teachers to learn and practice cognitive load theory and the strategies of desirable difficulty which may then be incorporated into other aspects of their classroom learning programmes.

Teachers face widening learning gaps of up to six years within Australian classrooms (Goss et al., 2015) and Australian educational standards as reflected by PISA are cause for concern (Donnelly, 2019). If the prior knowledge of students is not recognised or addressed, then learning may continue to be compromised (Tomlinson, 2015). Differentiation provides equitable access to learning progress, yet the means of provision is challenging for many teachers (Goss et al., 2015). The DLCP may provide one option to address these difficulties guided by (a) teacher assessment and observation of student learning needs, (b) learning progressions, (c) mastery learning strategies and (d) explicit instruction and practice of the content of classroom instruction.

Beyond the dependence of younger students on teacher management of learning, classroom use of the DLCP may introduce students to potential future study routines through:

- the modelling and practice of desirable difficulty learning strategies,
- familiarity with illusions of knowledge,

- the development of student beliefs that effortful thinking can lead to mastered learning, and subsequently
- the development self-efficacy, motivation and self-regulation in learning.

"The end of knowledge is power" (Hobbes, 1839, p. 7) because knowledge in long-term memory is the foundation of reasoning (Tricot & Sweller, 2013). Our education systems hold an enormous responsibility to provide students of all backgrounds with equity and agency through knowledge acquisition. The DLCP does not seek or expect uniform student achievement, however, targeted practice of the content of classroom instructional guided by learning progressions may facilitate progress.

The overarching goal of the thesis was to present a practice process that had been adjudicated and revised by theory to potentially assist the researcher and other teachers with managing learning differentiation and consolidation. Foundational to all conclusions, was the understanding of the complexity of the classroom which requires any intervention to be effective and practical. Well managed desirable difficulty with metacognitive understandings is key to student progress and the expectancy of future success. Over time, students may experience and attribute successful outcomes to perseverance with the desirable difficulty strategies of the DLCP. For the researcher, this goal was epitomised by a Year 1 student with special needs, who, after an inability to recall a phonics sound during a mastery test, said with a big smile, "I'm still working on that one". In many ways, the current study has taken DLCP research to the starting line. It is hoped, however, that the theoretical focus has progressed the process towards a more evidence-informed foundation.

# References


Agarwal, P. K., & Bain, P. M. (2019). *Powerful teaching: Unleash the science of learning*. Jossey-Bass.

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659-701. https://doi.org/10.3102/0034654316689306

Airasian, P. W. (1971). *Mastery learning: Theory and practice* (J. H. Block, Ed.) (p. 93). Holt, Rinehart and Winston.

Anderson, J. R., & Schunn, C. (2000). Implications of the ACT-R learning theory: No magic bullets. In R. Glaser (Ed.), *Advances in Instructional Psychology, 5*, (pp. 1-33). Erlbaum. https://bit.ly/374AfIw

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Longman.

Anderson, L. W., & Sosniak, L. A. (1994). *Bloom's taxonomy: A forty-year retrospective*. NSSE.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation, 2*, 89-195.

Australian Bureau of Statistics. (2017). *Student (FTE) to teaching staff (FTE) ratios 2001 – 2016*. https://bit.ly/2LmD3IF

Australian Institute for Teaching and School Leadership. (2011). *Australian professional standards for teachers*. https://www.aitsl.edu.au/teach/standards

Ausubel, D. P., Novak, J. D., & Hanesian, H. (1968). *Educational psychology: A cognitive view*. Holt, Rinehart and Winston.

Baddeley, A. (2017). *Exploring working memory: Selected works of Alan Baddeley.* Routledge.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8*, 47-89. Elsevier.

Bandura, A. (1974). Behavior theory and the models of man. *American Psychologist*, *29*(12), 859-869. https://doi.org/10.1037/h0037514

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall, Inc.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman.

Bandura, A. (2008). An agentic perspective on positive psychology. In S. J. Lopez (Ed.), *Positive Psychology* (Vol. 1, pp. 167-196). Praeger.

Bandura, A. (2010). Self-efficacy. *The Corsini encyclopedia of psychology*, 1-3.

Bartholomew, L. K., Parcel, G. S., & Kok, G. (1998). Intervention mapping: a process for developing theory and evidence-based health education programs. *Health Education & Behavior*, *25*(5), 545-563.

Bell, J. (2005). *Doing your research project. A guide for first-time researchers in education, health and social science* (4th ed.). Open University Press.

Biesta, G. (2007). Why “what works” won’t work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory*, *57*(1), 1-22.

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.

Bjork, R. (n.d.) How we learn versus how we think we learn: Desirable difficulties in theory and Practice. UCLA Bjork Learning and Forgetting Lab. https://bjorklab.psych.ucla.edu/research/

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185 – 205). MIT Press.

Bjork, R. [Go Cognitive] (2011, November). T*he theory of disuse and the role of forgetting in human memory* [Video]. You Tube. https://www.youtube.com/watch?v=Hv6Vye1JCjo&feature=emb_logo

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance. Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (p. 435–459). The MIT Press.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes, Vol. 1. From learning theory to connectionist theory; Vol. 2. From learning processes to cognitive processes* (pp. 35-67). Lawrence Erlbaum Associates.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, and J. R. Pomerantz (Eds.) & FABBS Foundation, *Psychology and the real world: Essay illustrating fundamental contributions to society* (pp. 56-64). Worth Publishers.

Bjork, R. A., & Bjork, E. L. (2013). Optimizing treatment and instruction: Implications of a new theory of disuse. In L. G. Nilsson & N. Ohta (Eds.), *Memory and society: Psychological perspectives* (pp. 109-133). Psychology 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.

Block, J. H. (Ed.) (1971). *Mastery Learning Theory and Practice*. Holt, Rinehart and Winston Inc.

Bloom, B. S. (1968). Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation Comment*, 1(2), n2. https://files.eric.ed.gov/fulltext/ED053419.pdf

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Handbook I. Cognitive domain.* David McKay.

Booth, A., & Carroll, C. (2015). How to build up the actionable knowledge base: The role of 'best fit' framework synthesis for studies of improvement in healthcare. *BMJ Qual Saf*, *24*(11), 700-708.

Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful literature review.* Sage.

Brown, P. C., Roediger, H. L., III, & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. The Belknap Press.

Bruning, R., Schraw, G., & Ronning, R. (1999). *Cognitive psychology and instruction* (3rd ed.). Merrill.

Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, *145*(11), 1029. https://doi.org/10.1037/bul0000209

Carey, B. (2014). *How we learn*. Macmillan.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*(2), 268-276.

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of US history facts. *Applied Cognitive Psychology: The Official*

*Journal of the Society for Applied Research in Memory and Cognition*, *23*(6), 760-771.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, *64*(8), 723-733.

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). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*, *6(*505), 1-12.

Carvalho, P. F., & Goldstone, R. L. (2019). When does interleaving practice improve learning? In J. Dunlosky & K. A. Rawson (Eds.), *Cambridge Handbook of Cognition and Education* (pp. 411–436). Cambridge University Press.

Carvalho, P. F., & Goldstone, R. L. (2020). The most efficient sequence of study depends on the type of test. *Applied Cognitive Psychology*. Advance online publication. https://doi.org/10.1002/acp.3740

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, *56*(4), 236-246.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354-380.

Chen, H. T. (1990). *Theory-driven evaluations*. Sage.

Chen, H. T. (2012). Theory-driven evaluation: Conceptual framework, application and advancement. In R. Strobl, O. Lobermeier, & W. Heitmeyer (Eds.) Evaluation von

Programmen und Projekten für eine demokratische Kultur (pp.17-40). Springer. https://doi.org/10.1007/978-3-531-19009-9_2

Chen, O., & Kalyuga, S. (2020). Exploring factors influencing the effectiveness of explicit instruction first and problem-solving first approaches. *European Journal of Psychology of Education*, *35*(3), 607-624. https://doi.org/10.1007/s10212-019-00445-5

Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2017). Extending cognitive load theory to incorporate working memory resource depletion: Evidence from the spacing effect. *Educational Psychology Review, 2*(30), 483-501.

Christodoulou, D. (2014a). Minding the knowledge gap: The importance of content in student learning. *American Educator, 38*(1), 27-33.

Christodoulou, D. (2014b). *Seven Myths About Education*. Routledge.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). Routledge.

ColumbiaLearn. (2018, November 12). SOLER Symposium | Robert A. Bjork | Low Memorial Library, October 11th, 2018 [Video]. YouTube. https://www.youtube.com/watch?v=ouocL4ssJLY

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Cooper, H. (1984). *The integrative research review: A systematic approach*. Sage Publications.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (Vol. 2). Sage Publications.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51-57. https://doi.org/10.1177/0963721409359277

Creswell, J. (1998). *Qualitative inquiry and research design: Choosing among five traditions.* Sage Publications.

De Bruyckere, P. (2018). *The Ingredients for great teaching.* SAGE Publications.

De Bruyckere, P., Kirschner, P. A., & Hulshof, C. (2020). *More urban myths about learning and education: Challenging eduquacks, extraordinary claims, and alternative facts*. Routledge.

De Bruyckere, P., Kirschner, P. A., & Hulshof, C. D. (2015). *Urban myths about learning and education*. Academic Press.

De Groot, A. (2008). Thought and choice in chess. Amsterdam University Press. https://ebookcentral.proquest.com/lib/ECU/detail.action?docID=437584

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, *43*(8), 627-634.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 317-344). Academic Press

Denzin, N. K., & Lincoln, Y. S. (2011). *The Sage handbook of qualitative research.* Sage.

Denzin, N., & Lincoln, Y. (1994). *Handbook of qualitative research*. Sage Publications.

Desy, J., Busche, K., Cusano, R., Veale, P., Coderre, S., & McLaughlin, K. (2018). How teachers can help learners build storage and retrieval strength. *Medical Teacher*, *40*(4), 407-413.

Didau, D. (2019). Learning Myths. In C. Hendrick & R. Macpherson (Eds.), *What does this look like in the classroom? Bridging the gap between research and practice* (pp. 163 - 186). Learning Sciences International.

Dinham, S. (2014, September 11). Primary schooling in Australia: Pseudo-science plus extras times growing inequality equals decline [Keynote]. In *What Counts as Quality in Education?* (pp. 7 – 15). Australian College of Educators National Conference, Adelaide, SA, Australia. https://files.eric.ed.gov/fulltext/ED571027.pdf

Donnelly, K. (2019, December 6). Students betrayed by faddish schools of thought. The Australian. https://bit.ly/374NsRB

Duchesne, S., & McMaugh, A. (2016). *Educational psychology for learning and teaching* (5th ed.). Cengage Learning Australia.

Duchesne, S., & McMaugh, A. (2018). *Educational Psychology for Learning and Teaching*. http://public.eblib.com/choice/PublicFullRecord.aspx?p=6335538

Dunlosky, J. (2013). Strengthening the student toolbox: Study strategies to boost learning. *American Educator, 37*(3), 12-21.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Sage.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58.

Ebbinghaus, H., (1913). Memory: A contribution to experimental psychology. *Educational reprints, no. 3*. Teachers College, Columbia University.

Education Council. (2015). *National STEM School Education Strategy: A comprehensive plan for science, technology, engineering and mathematics education in Australia.* https://bit.ly/376N45j

Education Endowment Foundation (n.d.) *Developing effective learners*. https://bit.ly/3gWF1LO

Education Endowment Foundation (n.d.) *Feedback and monitoring pupil progress*. https://bit.ly/2Ki1J4o

Education Endowment Foundation. (2019). *Metacognition and self-regulated learning guidance report.* https://bit.ly/3a0hBDN

Ericsson, A., & Pool, R. (2016). *Peak: How all of us can achieve extraordinary things.* Random House.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906.

Foster, N. L., Mueller, M. L., Was, C., Rawson, K. A., & Dunlosky, J. (2019). Why does interleaving improve math learning? The contributions of discriminative contrast and distributed practice. *Memory & Cognition, 47*(6), 1088-1101. https://doi.org/10.3758/s13421-019-00918-4

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

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*(2), 95-112.

Gluckman, M., Vlach, H. A., & Sandhofer, C. M. (2014). Spacing simultaneously promotes multiple forms of learning in children's science curriculum. *Applied Cognitive Psychology*, *28*(2), 266-273.

Goss, P., Hunter, J., Romanes, D., & Parsonage, H. (2015). *Targeted teaching: how better use of data can improve student learning*. Grattan Institute. https://bit.ly/3gIuV0U

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, *26*(2), 91-108.

Guskey, T. R. (2007). Closing achievement gaps: Revisiting Benjamin S. Bloom's "Learning for Mastery". *Journal of Advanced Academics*, *19*(1), 8-31.

Guzman-Munoz, F. J. (2017). The advantage of mixing examples in inductive learning: a comparison of three hypotheses. *Educational Psychology, 37*(4), 421-437. https://doi.org/10.1080/01443410.2015.1127331

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hattie, J. (2015). *What doesn't work in education: The politics of distraction*. Pearson. https://bit.ly/2KhUpWs

Hattie, J. A., & Donoghue, G. M. (2016). Learning strategies: A synthesis and conceptual model. *npj Science of Learning*, *1*(1), 1-13. https://doi.org/10.1038/jpjscilearn.2016.13

Hattie, J., & Yates, G. C. (2014). *Visible learning and the science of how we learn*. Routledge.

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.

Hendrick, C. & Macpherson, R. (2019). *What does this look like in the classroom?* Learning Science International.

Higgins, S., Katsipataki, M., Villanueva-Aguilera, A.B.V., Coleman, R., Henderson, P., Major, L.E., Coe, R. & Mason, D. (2016). *The Sutton Trust-Education Endowment Foundation teaching and learning toolkit*. Education Endowment Foundation. https://bit.ly/2Ll14zL

Hirsch Jr, E. D. (2003). Reading comprehension requires knowledge—of words and the world. *American Educator, 27*(1), 10-13.

Hirsch Jr, E. D. (2010). *The schools we need and why we don't have them*. Anchor. http://estrada.cune.edu/facweb/brent.royuk/portfolio/Hirsch_Excerpts.pdf

Hirsch Jr, E. D. (2016). *Why knowledge matters: Rescuing our children from failed educational theories*. Harvard Education Press.

Hobbes, T. (1839). Computation or logic of philosophy. In W. Molesworth (Ed.), *The English Works of Thomas Hobbes of Malmesbury* (Vol. 1, pp. 7). John Bohn.

Hopkins, S., & Bayliss, D. (2017). The prevalence and disadvantage of min-counting in seventh grade: Problems with confidence as well as accuracy? *Mathematical Thinking and Learning, 19*(1), 19-32. https://doi.org/10.1080/10986065.2017.1258613

Howard-Jones, P. (2014). *Neuroscience and education: A review of educational interventions and approaches informed by neuroscience*. Education Endowment Foundation. https://bit.ly/33ZzC0U

Hsieh, H., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277-1288. https://doi.org/10.1177/1049732305276687

Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction, 19*(5), 402-410.

Kang, S. H. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 12-19.

Kang, S. H., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26*(1), 97-103.

Kang, S. H., Lindsey, R., Mozer, M., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review, 21*(6), 1544-1550. https://doi.org/10.3758/s13423-014-0636-z

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

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. T. Wixted (Ed.), *Learning and memory: A comprehensive reference* (pp. 487-514). Academic press.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966-968.

Kirschner P. A., & Neelen, M., (2019, May 21) What we already know determines what, how and how well we learn. *3starlearningexperiences*. https://bit.ly/2JTJbI3

Kirschner, P. A., & Hendrick, C. (2020). *How learning happens: seminal works in educational psychology and what they mean in practice*. Routledge.

Kirschner, P. A., & Neelan, M. (2018). Interleaving: Variety is the spice of learning. *3starlearningexperiences*. https://bit.ly/2W92rE6

Kirschner, P. A., & van Merriënboer, J. J. G. (2013). Do learners really know best? Urban legends in education. *Educational Psychologist*, *48*(3), 169-183. https://doi.org/10.1080/00461520.2013.804395

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. https://doi.org/10.1207/s15326985ep4102_1

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*(2), 219-224.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*(6), 585-592.

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging, 25*(2), 498.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 989.

Krathwohl, D. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 42*(4), 212-218.

Kuhlthau, C. C., Maniotes, L. K., & Caspari, A. K. (2015). *Guided inquiry: Learning in the 21st century*. ABC-CLIO.

Küpper-Tetzel, C. E. (2014). Understanding the distributed practice effect: Strong effects on weak theoretical grounds. *Zeitschrift für Psychologie, 222*(2), 71.

Küpper-Tetzel, C. E., Kapler, I. V., & Wiseheart, M. (2014). Contracting, equal, and expanding learning schedules: The optimal distribution of learning sessions depends on retention interval. *Memory & Cognition, 42*(5), 729-741. https://doi.org/10.3758/s13421-014-0394-1

Kurtz, K. H., & Hovland, C. I. (1956). Concept learning with differing sequences of instances. *Journal of Experimental Psychology, 51*(4), 239.

Lau, C., Kitsantas, A., Miller, A. D., & Rodgers, E. B. D. (2018). Perceived responsibility for learning, self-efficacy, and sources of self-efficacy in mathematics: A study of international baccalaureate primary years programme students. *Social Psychology of Education, 21*(3), 603-620.

Lehtiranta, L., Junnonen, J., Kärnä, S., & Pekuri, L. (2016). The constructive research approach: Problem solving for complex projects. In B. Pasian (Ed.), *Designs, methods and practices for research of project management*, (pp. 95-106). Gower.

Leitner system. (2020, January 26) In *Wikipedia*. https://en.wikipedia.org/wiki/Leitner_system

Logan, J. M., Castel, A. D., Haber, S., & Viehman, E. J. (2012). Metacognition and the spacing effect: the role of repetition, feedback, and instruction on judgments of learning for massed and spaced rehearsal. *Metacognition and Learning, 7*(3), 175-195.

Lovell, O. (2020). Sweller's cognitive load theory in action (T. Sherrington, Series Ed.). John Catt Educational.

Lyle, K. B., Hopkins, R. F., Hieb, J. L., & Ralston, P. A. (2020). How the amount and spacing of retrieval practice affect the short- and long-term retention of mathematics knowledge. *Educational Psychology Review, 32*(1), 277-295. https://doi.org/10.1007/s10648-019-09489-x

McMullen, F., & Madelaine, A. (2014). Why is there so much resistance to Direct Instruction? *Australian Journal of Learning Difficulties, 19*(2), 137-151.

McNamara, S., & Moreton, G. (1997). *Understanding differentiation: A teacher's guide*. Routledge. https://www.taylorfrancis.com/books/9780203390108

Merriam-Webster. (n.d.). Category. In *Merriam-Webster.com dictionary.* Retrieved June 6, 2020, from https://www.merriam-webster.com/dictionary/category

Merriam-Webster. (n.d.). Domain. In *Merriam-Webster.com dictionary.* Retrieved June 6, 2020, from https://www.merriam-webster.com/dictionary/domain

Merriam-Webster. (n.d.). Metacognition. In *Merriam-Webster.com dictionary.* Retrieved October 19, 2020, from https://www.merriam-webster.com/dictionary/metacognition

Mettler, E., Massey, C. M., & Kellman, P. J. (2016). A comparison of adaptive and fixed schedules of practice. *Journal of Experimental Psychology: General, 145*(7), 897-917. https://doi.org/10.1037/xge0000170

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.

National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press.

Nickow, A. J., Oreopoulos, P., & Quan, V. (2020). *The impressive effects of tutoring on preK-12 learning: A systematic review and meta-analysis of the experimental evidence.* [EdWorkingPaper: 20-267]. Annenberg Institute at Brown University. https://doi.org/10.26300/eh0c-pc52

Ostrow, K., Heffernan, N., Heffernan, C., & Peterson, Z. (2015). Blocking vs. interleaving: Examining single-session effects within middle school math homework. In *International conference on artificial intelligence in education* (pp. 338-347). Springer.

Palmberger, M., & Gingrich, A. (2013). Qualitative comparative practices: Dimensions, cases, and strategies. In U. Flick (Ed.), *The SAGE Handbook of analyzing qualitative data* (pp. 94-108). Sage.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning. IES Practice Guide.* National Center for Education Research. https://files.eric.ed.gov/fulltext/ED498555.pdf

Patel, R., Liu, R., & Koedinger, K. R. (2016). When to block versus interleave practice? Evidence against teaching fraction addition before fraction multiplication [Paper]. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, Austin, TX, USA (pp. 2069-2074). Cognitive Science Society. https://cogsci.mindmodeling.org/2016/papers/0360/index.html

Pawson, R. (2006). Realist synthesis: New protocols for systematic review. *Evidence-based policy*. SAGE Research Methods. http://dx.doi.org/10.4135/9781849209120

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage Publications.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2004). Realist synthesis: An introduction. *RMP Methods Paper 2/2004. ESRC Research Methods Programme.* University of Manchester.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*(3), 193-198.

Piaget, J., & Cook, M. (1952). When thinking begins. In *The origins of intelligence in children* (Vol. 8, No. 5, pp. 25-36) International Universities Press. https://bit.ly/3gLL05G

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.

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140*(3), 283-302.

Richland, L. E., Bjork, R. A., Finley, J. R., & Linn, M. C. (2005). Linking cognitive science to education: Generation and interleaving effects. In *Proceedings of the twenty-seventh annual conference of the Cognitive Science Society* (Vol. 27, pp. 1850-55). https://bit.ly/3qWpSOT

Roebers, C. M., & Feurer, E. (2016). Linking executive functions and procedural metacognition. *Child Development Perspectives, 10*(1), 39-44.

Roediger III, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition, 1*(4), 242-248.

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20*(9), 1209-1224.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science, 35*(6), 481-498.

Rohrer, D., Dedrick, R. F., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review, 21*(5), 1323-1330.

Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology*, *107*(3), 900-908.

Rohrer, D., Dedrick, R. F., Hartwig, M. K., & Cheung, C. N. (2019). A randomized controlled trial of interleaved mathematics practice. *Journal of Educational Psychology, 112(1), 40–52.* https://doi.org/10.1037/edu0000367

Rohrer, D., Taylor, K., Pashler, H., Wixted, J. T., & Cepeda, N. J. (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology*, *19*(3), 361-374.

Rycroft-Malone, J., McCormack, B., Hutchinson, A. M., DeCorby, K., Bucknall, T. K., Kent, B., Schultz, A., Snelgrove-Clarke, E., Stetler. C.B. Titler, M., Wallin, L., & Wilson, V. (2012). Realist synthesis: Illustrating the method for implementation research. *Implementation Science, 7*(1), 1–10. https://doi.org/10.1186/1748-5908-7-33

Schneider, W., Korkel, J., & Weinert, F. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology, 81*(3), 306-312. https://doi.org/10.1037/0022-0663.81.3.306

Schutte, G. M., Duhon, G. J., Solomon, B. G., Poncy, B. C., Moore, K., & Story, B. (2015). A comparative analysis of massed vs. distributed practice on basic math fact fluency growth rates. *Journal of School Psychology, 53*(2), 149-159. https://doi.org/10.1016/j.jsp.2014.12.003

Scott, C., Stone, B., & Dinham, S. (2001). I love teaching but … international patterns of teacher discontent. *Education Policy Analysis Archives,* 9(*28), 1-7.* Arizona State University.

Smith, C. D., & Scarf, D. (2017). Spacing repetitions over long timescales: A review and a reconsolidation explanation. *Frontiers in Psychology, 8*(962). https://doi.org/10.3389/fpsyg.2017.00962

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology, 25*(5), 763-767.

Swanson, H. L. (2015). Intelligence, working memory, and learning disabilities. In T. C. Papadopoulos, R. K. Parrila, & J. R. Kirby (Eds.), *Cognition, intelligence, and achievement* (pp. 175-196). Academic Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Sweller, J. (2016). Working memory, long-term memory, and instructional design. *Journal of Applied Research in Memory and Cognition, 5*(4), 360-367. https://doi.org/10.1016/j.jarmac.2015.12.002

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

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *1-32*. https://doi.org/10.1007/s10648-019-09465-5

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*(6), 837-848.

Thorndike, E. L. (1913). *The psychology of learning* (Vol. 2). Teachers College, Columbia University.

Tomlinson, C. A. (2000). Reconcilable differences: Standards-based teaching and differentiation. *Educational Leadership*, *58*(1), 6-13.

Tomlinson, C. A. (2015). Teaching for excellence in academically diverse classrooms. *Society*, *52*(3), 203-209. https://doi.org/10.1007/s12115-015-9888-0

Tomlinson, C. A., Brimijoin, K., & Narvaez, L. (2008). *The differentiated school: Making revolutionary changes in teaching and learning.* ASCD. https://bit.ly/3oLKU0Q

Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. *Psychology of Learning and Motivation,* (Vol. 60, pp. 113-189). Elsevier.

Tran, R., Rohrer, D., & Pashler, H. (2015). Retrieval practice: The lack of transfer to deductive inferences. *Psychonomic Bulletin & Review*, *22*(1), 135-140.

Tricot, A., & Sweller, J. (2013). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review*, *26*(2), 265-283. https://doi.org/10.1007/s10648-013-9243-1

van Merriënboer, J. J., & Kirschner, P. A. (2018). *Ten steps to complex learning: A systematic approach to four-component instructional design.* Routledge.

Visible Learning (2019). Visible learning 250+ influences on student achievement [Infographic]. *Corwin SAGE*. https://us.corwin.com/sites/default/files/250_influences_chart_june_2019.pdf

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition, 109*(1), 163-167.

W. K. Kellogg Foundation. (1998). Logic model guide. https://bit.ly/3nfMOXn

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. https://doi.org/10.3758/s13421-010-0063-y

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548-573.

Weinstein, Y., Madan, C. R., & Sumeracki, M. A. (2018). Teaching the science of learning. *Cognitive research: Principles and implications, 3*(1), 1-17. https://bit.ly/37ZWrD9

Weinstein, Y., Sumeracki, M., & Caviglioli, O. (2019). *Understanding how we learn*. Routledge.

Weissgerber, S. C., Reinhard, M. A., & Schindler, S. (2018). Learning the hard way: Need for cognition influences attitudes toward and self-reported use of desirable difficulties. *Educational Psychology, 38*(2), 176-202. https://doi.org/10.1080/01443410.2017.1387644

Willingham, D. T. (2009). *Why don't students like school? A cognitive scientist answers questions about how the mind works and what it means for the classroom*. John Wiley & Sons.

Wilson, J., & Murdoch, K. (2012). *Learning for themselves: Pathways for thinking and independent learning in the primary classroom*. Taylor and Francis.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304): Lawrence Erlbaum Associates Publishers.

Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J., & Pawson, R. (2013). RAMESES publication standards: Realist syntheses. *BMC Medicine, 11*(21). https://doi.org/10.1186/1741-7015-11-21

Yan, V. X., & Sana, F. (2020). Does the interleaving effect extend to unrelated concepts? Learners' beliefs versus empirical evidence. Advance online publication. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000470

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General, 145*(7), 918-933.

Young, S. F. (2008). Theoretical frameworks and models of learning: tools for developing conceptions of teaching and learning. *International Journal for Academic Development, 13*(1), 41-49. https://doi.org/10.1080/13601440701860243

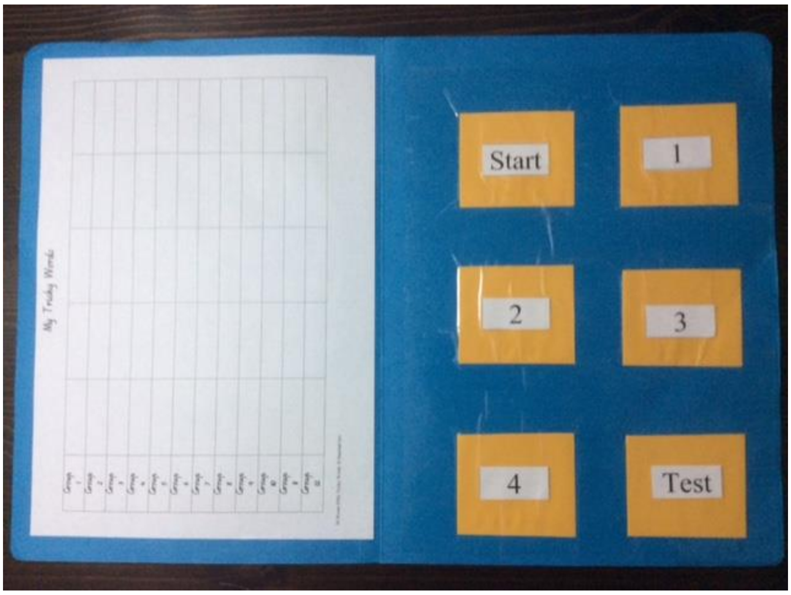
Zhang, L. (2016). Is inquiry-based science teaching worth the effort? *Science and Education, 25*(7-8), 897-915. https://doi.org/10.1007/s11191-016-9856-0

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition, 41*(1), 16-27.

## Appendices

## Appendix A

## The Typical Format of a Traditional Mastery Learning Folder



## Appendix B

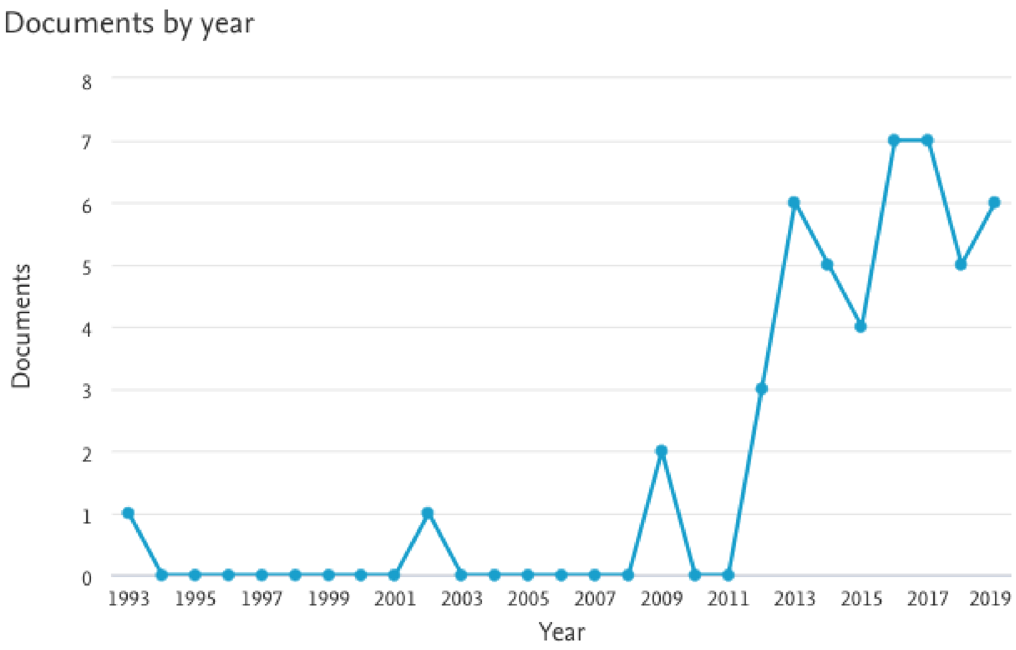
## SCOPUS Search Results Analysis

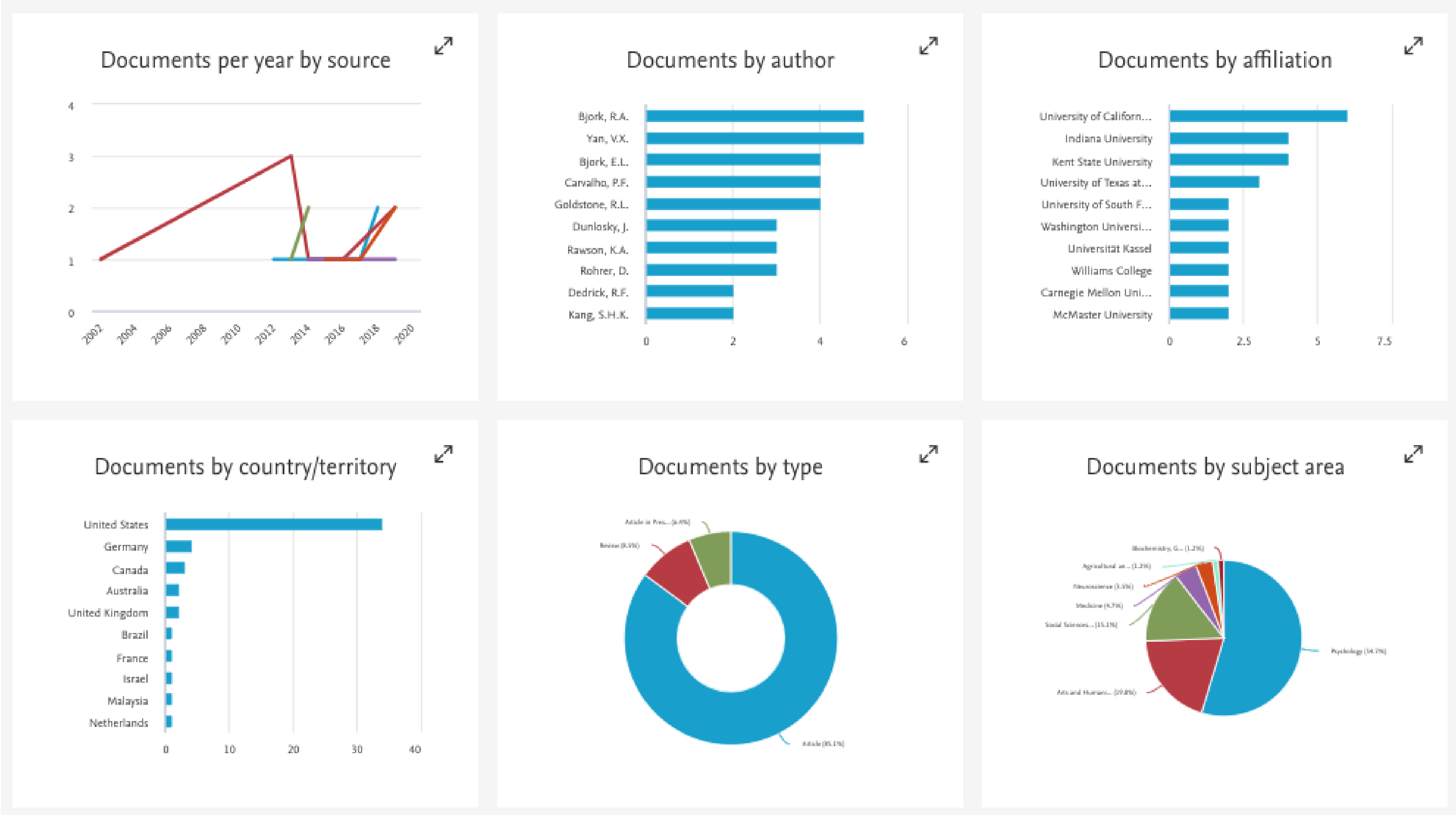
TITLE-ABS-KEY ( "interleaved practice" OR "interleaving" AND learning ) AND ( LIMIT-TO ( SUBJAREA , "PSYC" ) )

47 document results

Select year range to analyze: 1993 to 2019 Analyze

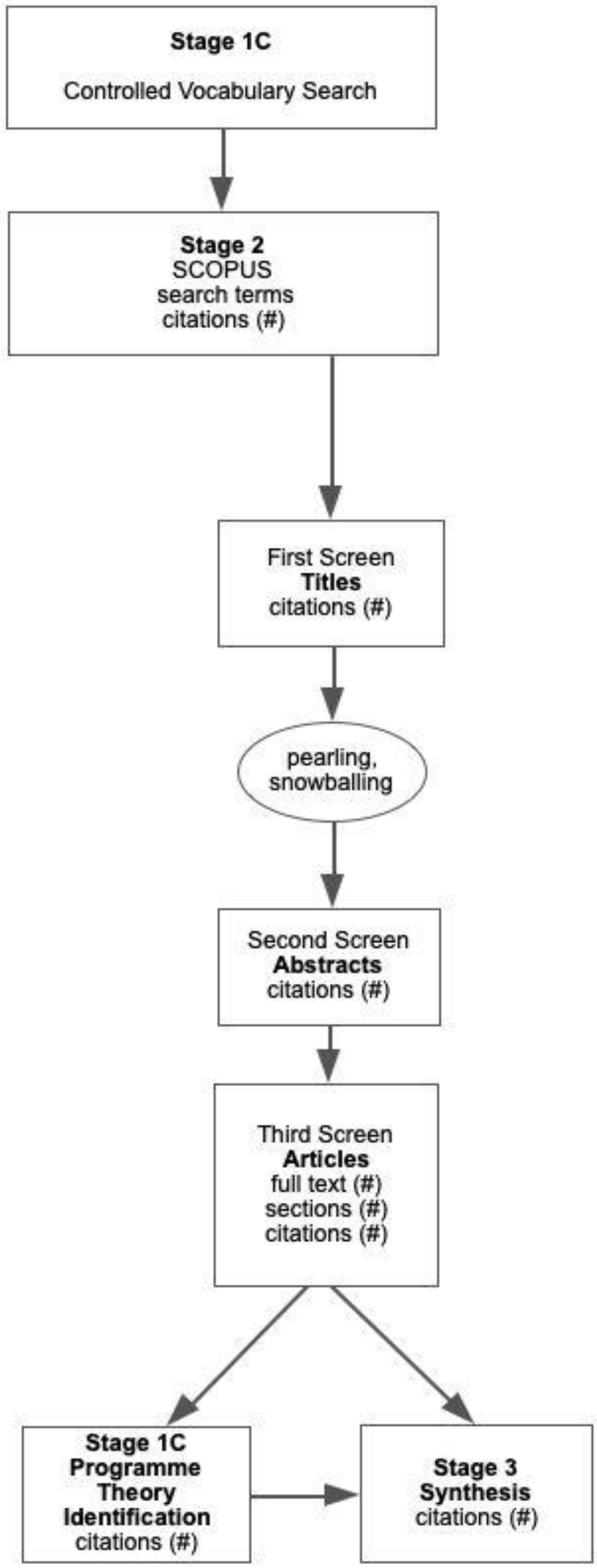
| Year | Documents |
|---|---|
| 2019 | 6 |
| 2018 | 5 |
| 2017 | 7 |
| 2016 | 7 |
| 2015 | 4 |
| 2014 | 5 |
| 2013 | 6 |
| 2012 | 3 |
| 2011 | 0 |
| 2010 | 0 |




Documents per year by source
4
3
2
1
0
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
Documents by author
Bjork, R.A.
Yan, V.X.
Bjork, E.L.
Carvalho, P.F.
Goldstone, R.L.
Dunlosky, J.
Rawson, K.A.
Rohrer, D.
Dedrick, R.F.
Kang, S.H.K.
0
2
4
6
Documents by affiliation
University of Californ...
Indiana University
Kent State University
University of Texas at...
University of South F...
Washington Universi...
Universität Kassel
Williams College
Carnegie Mellon Uni...
McMaster University
0
2.5
5
7.5
Documents by country/territory
United States
Germany
Canada
Australia
United Kingdom
Brazil
France
Israel
Malaysia
Netherlands
0
10
20
30
40
Documents by type
Documents by subject area

**Appendix C**

**Learning Strategy Evidence Data Flowchart**



## Appendix D

### Third Screening Sample Table Layout

| Reference | Format / Research Type | Population | Read | Notes |
|---|---|---|---|---|
| Weinstein, Y., Sumeracki, M., & Caviglиolo, O. (2019). *Understanding how we learn*. New York, NY: Routledge. | book - section | n/a | section | |
| De Bruyckere, P. (2018). *The Ingredients for Great Teaching*: SAGE. | book - section | n/a | section | |
| Del Missier, F., Sassano, A., Coni, V., Salomonsson, M., & Mäntylä, T. (2018). Blocked vs. interleaved presentation and proactive interference in episodic memory. *Memory, 26*(5), 697-711. | journal primary | undergrads | | |
| 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*. doi:10.1016/j.jarmac.2018.05.005 | journal primary | undergrads | full text | |
| Weissgerber, S. C., Reinhard, M., & Schindler, S. . (2018). Learning the hard way: Need for cognition influences attitudes toward and self-reported use of desirable difficulties. *Educational Psychology Review, 38*(2), 176-202. doi:http://dx.doi.org.ezproxy.ecu.edu.au/10.1080/01443410.2017.1387644 | journal primary | undergrads | discussion | |
| Select **Abel, M., & Roediger, H. L. (2017). Comparing the testing effect under blocked and mixed practice: The mnemonic benefits of retrieval practice are not affected by practice format. *Memory & Cognition, 45*(1), 81-92.** | **journal primary** | **undergrads** | **full text** | |
| Rau, M. A. (2017). Conditions for the Effectiveness of Multiple Visual Representations in Enhancing STEM Learning. *Educational Psychology Review, 29*(4), 717-761. doi:10.1007/s10648-016-9365-3 | journal review | n/a | | |
| Select Sana, F., Yan, V. X., & Kim, J. A. (2017). Study sequence matters for the inductive learning of cognitive concepts. *Journal of Educational Psychology, 109*(1), 84. | journal primary | undergrads | full text | |

**Appendix E**

**Sample Data Extraction Table**

Table X

*Interleaving Data Extraction Summary*

| Non-mutually Exclusive Theories: Discriminative Contrast Theory, Study-phase Retrieval (spacing effect) Theory, Sequential Attention Theory | | | |
|---|---|---|---|
| Author (date) | Research goals and participants | Research topic and approach | Findings |
| Kornell and Bjork (2008) | 1. To investigate the advantage of non-interleaved over interleaved practise on inductive learning of low-discriminatory material.<br><br>2. To investigate participants' subjective assessment of non-interleaved versus interleaved practise in the context of induction.<br><br>Participants: undergraduates | The study of multiple paintings by different artists with artwork presented consecutively (non-interleaved) or non-contiguously (interleaved).<br><br>Quantitative<br>Post-test with control group<br><br>Group 1. Non-interleaved practise<br>Group 2. Interleaved practise | 1. Interleaved study lead to more effective inductive learning than non-interleaved study.<br><br>2. Participants' metacognitive judgements of learning were the opposite of their performance. They believed that their inductive learning was enhanced by non-interleaved study, whereas results indicated that interleaved study was more effective. |
| Taylor and Rohrer (2010) | To compare the effects of interleaved and non-interleaved practise of low-discriminatory formulae calculations.<br><br>Participants: Year 4 students | Computation of the number of faces, corners, edges and sides of different geometric solids.<br><br>Quantitative<br>Practise session and test, Post-test<br><br>Group 1. Interleaved practise<br>Group 2. Non-interleaved (blocked) | Interleaving material impaired results during the practise session, however, it improved scores on a delayed test due to the increased discrimination ability to pair problems with their appropriate solution procedure (formulae).<br><br>Post-test scores:<br>Interleaved 78%<br>Non-interleaved 38% |

**Appendix F**

**Traditional Tool and Instructions**

## MASTERY LEARNING

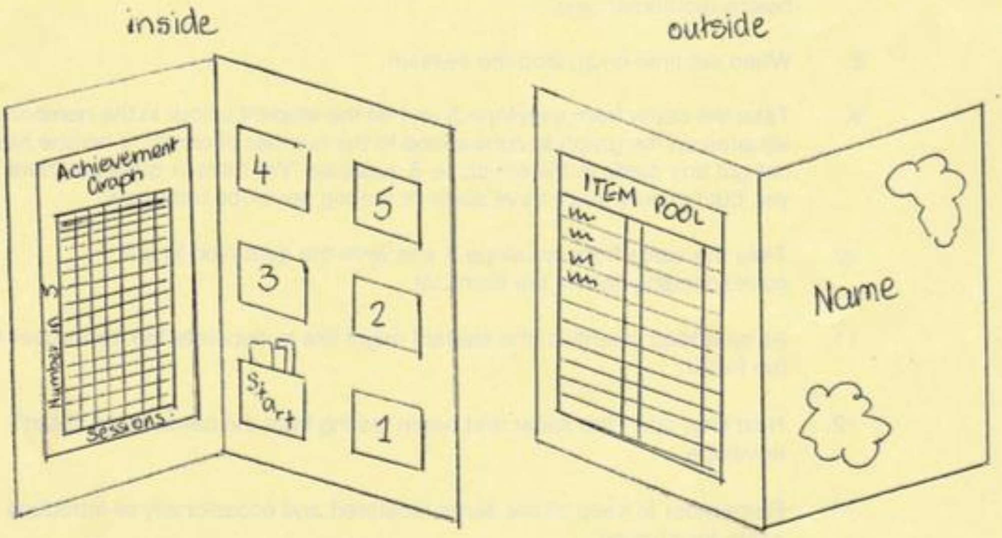
Mastery Learning means that the student has internalised a skill until the correct response becomes automatic. This technique requires the student to give the correct response five times consecutively and thus suggests that he/she has mastered the skill. It can successfully be used in the classroom with children paired off according to level of the skill they are learning (eg from basic sight words to tables). One of the advantages of having students paired and taking turns, is that all students get the chance to be the "teacher" as well as the learner. It is also a very effective way of learning skills at home.

**Materials Required:**

1 Manila folder.
6 Library pockets or envelopes.
1 Reproducible sheet for item pool (see sample).
1 Reproducible sheet for Achievement Graph.
Glue, felt pen and scraps of card suitable for presenting items to be learned.
Bank of skills to be mastered.

**Making The Package:**

1. On the inside right section glue the pockets in 3 rows.
2. Write "Start" on the bottom pocket and number 1-5 upward from the "Start" on the rest.
3. Write the list of skills to be mastered on the Item Pool sheet and fasten to the back of the folder.
4. Fasten the Achievement Graph sheet to the folder opposite the envelopes.
5. Write the skills to be mastered on cards (including some items the student has already mastered) and place in "Start" pocket. Don't use too many cards at each turn. Young students should start with 5 or 6.




F:\Psychologist\MasteryLearning030924.doc Page 1

**Using the Package:**

It is important that the package be used every day, preferable at the same time and for not more than 10 – 15 minutes.

1. Take the cards from the "Start" envelope.
   Show (if reading) the student each card in turn. Those that are read correctly are placed on the right hand pile. Don't forget to praise him/her each time the student gives a correct answer.

   Those that are incorrect go on the left hand pile. Encourage the child if he/she makes a mistake – eg if he/she says "top" for "tap" say "that was a good try – nearly right. The word is 'tap'. What is the word? – Student: 'tap'"

   Remember – whenever the student makes an error, correct immediately by:
   Modelling – the teacher says the item.
   Test – the child says the item.

2. When all cards have been shown, take the group on the left (incorrect ones) and place them in the "Start" envelope. The group on the right (correct ones) goes into envelope 1.

3. Take the cards from envelope 1 and retest. Those that are correct go into envelope 2 and those that are incorrect go back to the "Start" envelope.

4. Then take the cards from the "Start" envelope and re-test – the correct ones go into envelope 1 and the incorrect go back to the start.

5. Re-test the cards from envelope 1 and then those from envelope 2.

6. Return to "Start" – re-test envelopes 1, 2 and 3.

7. Return to "Start" – re-test envelopes 1, 2, 3 and 4. NB always shuffle cards before recommencing.

8. When set time is up, stop the session.

9. Take the cards from envelope 5 and let the student colour in the number of squares on the graph to correspond to the number of cards. If he/she has not got any cards in the envelope 5, just say, "We haven't quite got there yet, but I'm sure you'll have some in the top envelope tomorrow.

10. Take the cards from envelope 5 and write the date next to the corresponding item in the Item List.

11. As an added incentive, the student might like to decorate the front cover of the folder.

12. Next day, open the folder and begin testing from the cards in the "Start" envelope.

13. Remember to keep all the items mastered and occasionally re-introduce some for revision.

## Appendix G

## Development of the DLCP from 2010 – 2016

| Tool or Process Modification | Learning or Management Objective (in italics) |
|---|---|
| 2010<br>Each student had a mastery learning folder containing four pockets labelled 1, 2, 3 and Test (Appendix I).<br>Students received subject specific remediation, consolidation or extension flashcard content based on assessments of their learning. | • Content was based on *classroom learning* programmes.<br>• The mastery learning folder learning content was *individualised*.<br>• Use of targeted content *condensed learning time*. |
| The tutor conducted the DLCP as a part of student homework in a five to ten-minute session, four days per week.<br>Procedure:<br>• Step 1. Test the content in every pocket, move it forward to the next pocket if correct.<br>• Step 2. Check for understanding of the content and assist the student to practice through rehearsal.<br>A weekly mastery test was conducted by the teacher on the content in Pocket 4. Correct flashcards were removed from the folder. Incorrect flashcards were returned to the first pocket. At this time, new content was added to the Start pocket by the teacher. | • The *tutor* was a parent or sibling. Older student leaders or an education assistant tutored students who did not participate in homework.<br>• The process was *quick*.<br>• Step 1, *testing first* introduced a 24-hour *spaced interval* for daily sessions Monday to Thursday and a 72-hour interval from Friday to Monday.<br>• Testing provided immediate student *feedback* on progress.<br>• Step 2 attempted to ensure student *understanding* of content and recall through rehearsal.<br>• Student progress was *recorded*.<br>• It was thought that incorrect content returning to the start would provide *more remediation time* and facilitate *self-paced* progress. |
| Initial positive results led to trialling the incorporation of learning material from multiple subject areas. | • The folder content became *cross-curricular*. |

| Tool or Process Modification | Learning or Management Objective (in italics) |
| --- | --- |
| 2011<br>The DLCP changed from a homework strategy to a classroom strategy with delivery before the start of formal lessons in the morning.<br>Three additional pockets were added, and the pockets renamed Start, 1, 2, 3, 4, and Test (Appendix I).<br>Upon reaching the Test pocket, flashcards were not revised.<br>A weekly mastery test was conducted by the teacher on the flashcards that had reached the Test pocket. Correct flashcards were removed from the folder and incorrect flashcards were returned to the Start pocket. | • *Tutors* were parent helpers, student leaders or education assistants, when available.<br>• The rest *interval* between a flashcard arriving in the Test pocket and being tested by the teacher was between *one and six days* which was thought to give some indication of learning in long-term memory.<br>• Student learning was recorded within standard teacher subject area records.<br>• |
| 2012<br>A "Store" pocket was added before the Start pocket to contain inactive future learning content.<br>Based upon individual student, an appropriate number of flashcards (active learning load) was transferred from the Store into the Start pocket, prior to the first use of the tool. This was adjusted according to results.<br>As each flashcard arrived at the inactive Test pocket, the tutor or student transferred a new flashcard from the Store into the Start pocket. | • *Up to 20 flashcards* of new material could be added to the Store pocket at any time without affecting the number of active flashcards being learnt, increasing teacher *flexibility*.<br>• The active *learning load* became *differentiated*.<br>• The mastery test day became *flexible* as new content was transferred from the Store pocket into the Start pocket by the tutor/student instead of the teacher.<br>• With a one-to-one correspondence between flashcards leaving and entering the active pockets the learning load was automatically *maintained*.<br>• |

| Tool or Process Modification | Learning or Management Objective (in italics) |
|---|---|
| 2013<br>A hardcover, commercial version of the folder was manufactured to overcome the time inefficiencies of the handmade version. A bee theme was introduced, and the pockets were renamed Store, Hive (replacing Start), Daisies (1 – 4) and Test. A Mastered pocket was added (Appendix I).<br><br>Students graphed the cumulative total of mastered flashcards each week. | • The hardcover folder *reduced teacher preparation time*.<br>• Clear pockets enabled students to see their learning content move through the folder.<br>• Tutor instructions were printed on the folder.<br>• Through recording the quantity of mastered flashcards, students saw a *visual representation* of their progress over time (Appendix J).<br>• Flashcards were discussed as *learning goals*.<br>• The bee theme *enhanced conversations* about the nature of the folder process with the goal of developing student *self-efficacy*. |
| An online bank of free printable flashcard sheets was created and arranged according to the Australian Curriculum descriptors (K - 3). Flashcard content included knowledge, skills, routines and behavioural goals. | • Flashcards were designed to *support* curriculum, classroom routines and behavioural goals.<br>• The flashcard bank reduced teacher *preparation time* and was available to all teachers using handmade or commercial folders. |
| 2014 – 2016<br>On-going use | |
| 2017 Commenced research MEd | |

## Appendix H

## Original DLCP Instructions

The Mastery Learning Folder is an organizational tool designed to enable teachers and parents to support individual student progress through targeted remediation, consolidation or extension of the learning occurring in the classroom. It is complemented by free printable flashcards from the website.

Step 1. To determine the learning content of the Mastery Learning Folder, teachers may use their classroom assessments, parents may use flashcards provided by their school or students can be pre-tested using flashcards from the website.

Step 2. The relevant flashcards are placed into the Store pocket. Based on the ability of the learner, four to twelve flashcards are moved into the Hive pocket.

Step 3. Each day, the flashcards in the active pockets are first tested. Beginning at the Hive, correct flashcards are moved forward to the next pocket whilst incorrect flashcards return to the Hive. The flashcards are then learnt. This short, daily test and learn process repeats until each flashcard arrives at the Test pocket. At this time, a new flashcard is transferred from the Store pocket into the Hive.

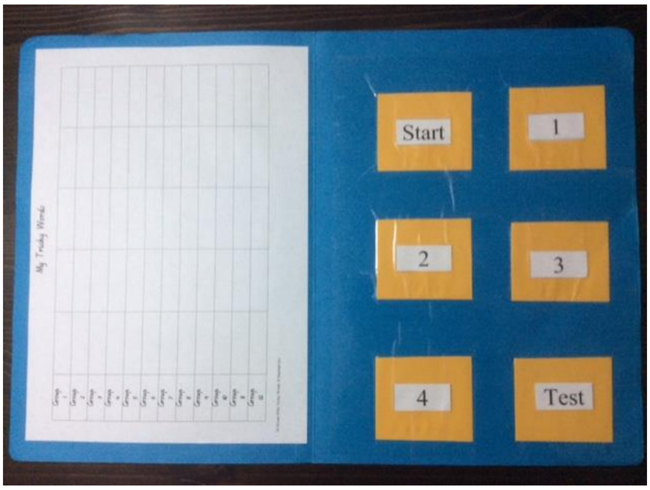
Step 4. With a collection of flashcards in the Test pocket, the learner is ready to be formally assessed. Correct flashcards move to the Mastered pocket whilst incorrect flashcards go back to the Hive. Students may keep a visual record of their mastered flashcards by recording the total number on a graph. When the number of mastered flashcards reaches twenty, they should be transferred to a storage bag or box.

## Appendix I

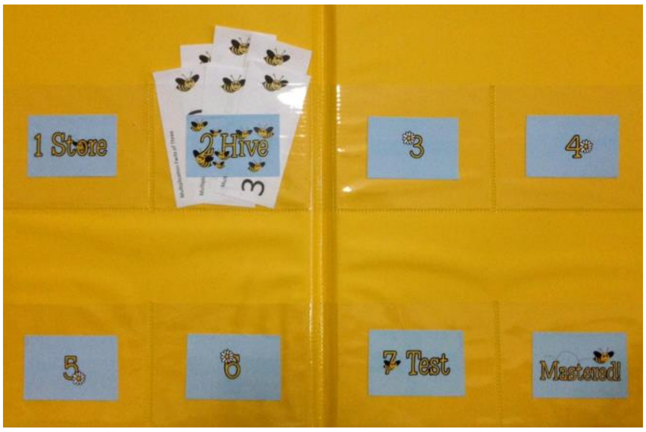
## Evolution of the Mastery Learning Folder Tool

2010 - 2011


2011 - 2012



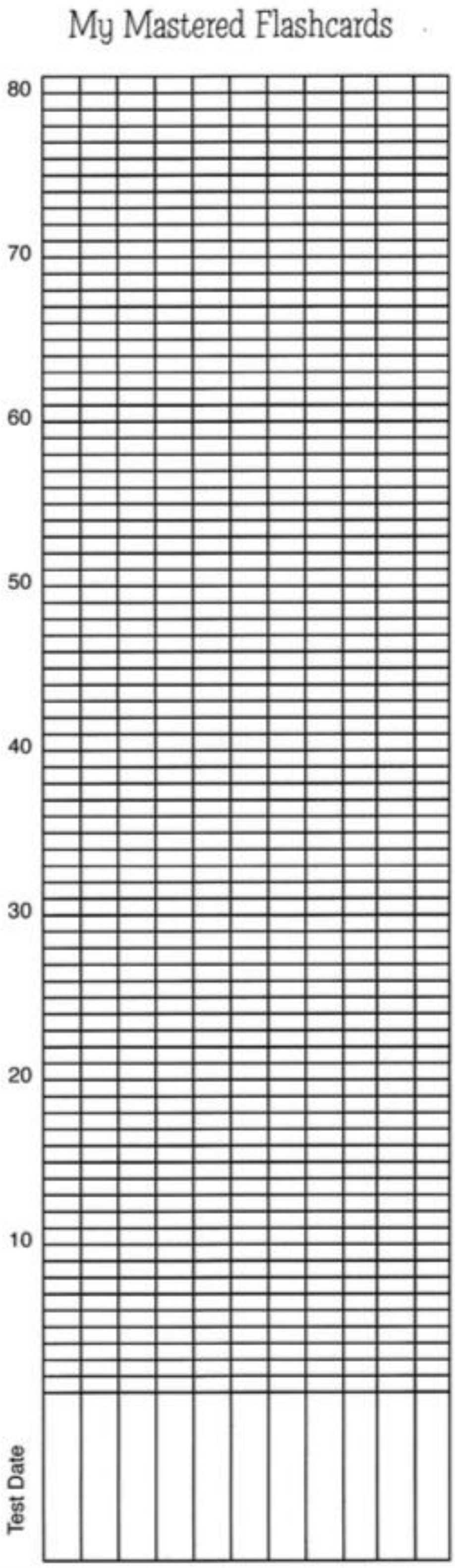

2013 - 2014




2015 - 2016

## Appendix J

## Sample Self-Monitoring Format




Name:

### My Completion Log

| Week | | | | |
|---|---|---|---|---|
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |

## Appendix K

## Understanding of Commonalities and Relationships Between the DLCP Theoretical Components Prior to Section Two Investigation

| | Information Processing Theory | Bloom's Revised Taxonomy: Cognitive | Evolutionary Psychology Theory | Baddeley's Model of Working Memory | Cognitive Load Theory | Mastery Learning | General Theory of Expertise | New Theory of Disuse | Spacing Effect | Testing Effect | Interleaving Effect | Bandura's Self-Efficacy Theory | Attributional Theory |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Learning Objectives:* | | | | | | | | | | | | | |
| Remembering<br>Understanding<br>Applying | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | |
| *Differentiation:* | | | | | | | | | | | | | |
| Learning Progressions | | ■ | ■ | | ■ | ■ | ■ | | | | | ■ | |
| Learning Content | | ■ | ■ | | ■ | ■ | ■ | | | | | ■ | |
| Learning Time | | | | | | ■ | ■ | | | | | ■ | |
| Learning Load | | | | ■ | ■ | ■ | | | | | | ■ | |
| Tutor Instruction | | | | | ■ | ■ | ■ | | ■ | ■ | ■ | ■ | ■ |
| Tutor Feedback | | | | | ■ | ■ | ■ | | ■ | ■ | ■ | ■ | ■ |
| *Consolidation:* | | | | | | | | | | | | | |
| Learning Goals | | ■ | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | |
| Deliberate Practice | | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | |
| Spaced Practice | | | | | ■ | ■ | ■ | ■ | ■ | | | ■ | |
| Retrieval Practice | | | | | ■ | ■ | ■ | ■ | | ■ | | ■ | |
| Mixed Topic Practice | | | | | ■ | ■ | ■ | ■ | | | ■ | ■ | |
| Success Criteria | | ■ | | | | ■ | ■ | ■ | | | | ■ | |
| Corrective Feedback | | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | |
| Explicit Instruction | | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | |
| *Metacognition:* | | | | | | | | | | | | | |
| Metacognitive Goals | | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| Self-Efficacy Training | | | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| Motivation | | | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |

## Appendix L

## Comparison Between the Revised DLCP, the Leitner System and the Traditional Mastery Learning Folder Approach