The Variability Effect: An instructional approach to enhance mathematics learning

Author:
Likourezos, Vicki

Publication Date:
2020

DOI:
https://doi.org/10.26190/unsworks/2090

The Variability Effect: An instructional approach to enhance mathematics learning

Vicki Likourezos

A thesis in fulfilment of the requirements for the degree of Doctor of Philosophy

School of Education

Faculty of Arts and Social Sciences

2019

Surname/Family Name	**Likourezos**
Given Name/s	**Vicki**
Abbreviation for degree as given in the University calendar	**PhD**
Faculty	**Faculty of Arts and Social Sciences**
School	**School of Education**
Thesis Title	**The Variability Effect: An instructional approach to enhance mathematics learning**

Abstract

Based on cognitive load theory, the variability effect occurs when learners' exposure to highly variable tasks results in better test performance. Using four randomised controlled trials in the area of secondary and tertiary mathematics instruction, the present study investigated the effects of variability, with an emphasis on levels of instructional guidance and levels of learner expertise. Experiments 1, 2 and 4 hypothesised that learners who study fully-guided worked examples will yield higher post-test performance scores, compared to learners who attempt unguided problem-solving tasks (Hypothesis 1); and learners who study high-variability worked examples will yield higher post-test performance scores, compared to learners who study low-variability worked examples, with no difference being generated under problem-solving conditions (Hypothesis 2). Hypothesis 1 was not supported in Experiments 1, 2 and 4, while Hypothesis 2 was supported only in Experiment 2. The variability effect that was produced in Experiment 2 led to further investigation in Experiment 3, where it was hypothesised that more-experienced learners (experts) would demonstrate the variability effect, and less-experienced learners (novices) would demonstrate a reverse variability effect. This hypothesis was supported, producing a classic expertise reversal effect. In addition, in all four experiments, learners' cognitive load was evaluated by having each participant complete a subjective rating of difficulty scale upon completion of their learning tasks. The results supported the assumptions based on cognitive load theory: learners in the worked-examples groups experienced less cognitive load compared to the problem-solving groups (in Experiments 1, 2 and 4); novices experienced less cognitive load when solving low-variability problems compared to high-variability problems, and lower cognitive load was experienced by experts, compared to novices, for both high- and low-variability tasks (in Experiment 3); and cognitive load associated with the completion of high-variability tasks was higher compared to the completion of low-variability tasks (only in Experiment 4). Although it is well grounded in empirical evidence that learners should be provided with worked examples during the initial stages of learning, these results strongly suggest that learners should be initially presented with low-variability problems, and as their levels of knowledge advance, variability should increase.

Originality Statement

‘I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.’

Signed......................................

Date..

Inclusion of Publications Statement

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure. **Publications can be used in their thesis in lieu of a Chapter if:** • The student contributed greater than 50% of the content in the publication and is the "primary author", i.e. the student was responsible primarily for the planning, execution and preparation of the work for publication • The student has approval to include the publication in their thesis in lieu of a Chapter from their supervisor and Postgraduate Coordinator. • The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis

Please indicate whether this thesis contains published material or not.

☐ *This thesis contains no publications, either published or submitted for publication*

☐ *Some of the work described in this thesis has been published and it has been documented in the relevant Chapters with acknowledgement*

☒ *This thesis has publications (either published or submitted for publication) incorporated into it in lieu of a chapter and the details are presented below*

CANDIDATE'S DECLARATION I declare that: • I have complied with the Thesis Examination Procedure • where I have used a publication in lieu of a Chapter, the listed publication(s) below meet(s) the requirements to be included in the thesis.		
Name Vicki Likourezos	**Signature**	**Date (dd/mm/yy)**

<table>
<tr><td colspan="3">Postgraduate Coordinator's Declaration (to be filled in where publications are used in lieu of Chapters)
I declare that:
• the information below is accurate
• where listed publication(s) have been used in lieu of Chapter(s), their use complies with the Thesis Examination Procedure
• the minimum requirements for the format of the thesis have been met.</td></tr>
<tr><td>PGC's Name
A/Professor Scott Eacott</td><td>PGC's Signature</td><td>Date (dd/mm/yy)</td></tr>
</table>

For each publication incorporated into the thesis in lieu of a Chapter, provide all of the requested details and signatures required

<table>
<tr><td colspan="7">Details of publication #1:
Full title: The variability effect: When instructional variability is advantageous
Authors: Vicki Likourezos, Slava Kalyuga and John Sweller
Journal or book name: Educational Psychology Review
Volume/page numbers: 31
Date accepted/ published: June 2019</td></tr>
<tr><td>Status</td><td>Published</td><td>X</td><td>Accepted and In press</td><td></td><td>In progress (submitted)</td><td></td></tr>
<tr><td colspan="7">The Candidate's Contribution to the Work
The candidate (primary author) was responsible for the planning and running of the experiments, and the collection and processing of data. The preparation of the work for publication was principally completed by the candidate (approximately 90%).</td></tr>
<tr><td colspan="7">Location of the work in the thesis and/or how the work is incorporated in the thesis:
The work described in sections 4.3 (Experiment 2) and 4.4 (Experiment 3) of Chapter 4, and in parts of sections 5.2 (Summary of Empirical Study) and 5.6 (General Conclusion) of Chapter 5, has been published.
The hypotheses, method, results and general discussion for Experiment 1 and Experiment 2 in the publication (see Appendix S) were restated and referred to as Experiment 2 and Experiment 3 respectively in this thesis.</td></tr>
<tr><td colspan="7">Primary Supervisor's Declaration
I declare that:
• the information above is accurate
• this has been discussed with the PGC and it is agreed that this publication can be included in this thesis in lieu of a Chapter
• All of the co-authors of the publication have reviewed the above information and have agreed to its veracity by signing a 'Co-Author Authorisation' form.</td></tr>
</table>

<table>
<tr><td>Supervisor's name
Professor Slava Kalyuga</td><td>Supervisor's signature</td><td>Date (dd/mm/yy)</td></tr>
</table>

Copyright Statement

Signed …………………………

Date ……………………………

Authenticity Statement

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis.'

Signed …………………………

Date ……………………………

Acknowledgements

To my caring and devoted daughter, Chloe, my eternal source of joy – you graciously encouraged me to gain self-direction so I can fulfil my passion and complete my doctoral thesis.

To my truehearted mother, Soula, who has always unfailingly supported me – you were the safe person I turned to when I changed my trajectory to embrace a vulnerable, purposeful life.

To my first-class supervisors, Prof. Slava Kalyuga and Emer. Prof. John Sweller – sharing your knowledge and giving constructive feedback allowed me to produce my best work.

To my exceptional fellows and staff members at the School of Education, UNSW Sydney – your collaborative support helped me sustain my vigour and become open-minded about research.

To Prof. Jonathan Kress, who unselfishly gave up his time so I could conduct my first experiments – your professionalism helped me achieve competence in conducting my research.

To the students who participated in my experiments – you played a pivotal role in my research. Even though I cannot disclose your names, I am very grateful for your invaluable contribution.

To my UNSW mathematics tutors – you fostered my deep love for mathematics. Because of your efforts, I have memories of fulfilling moments of being able to assimilate complex information.

To my compassionate father, Steve, who always said I was "strong" from when I was a child – I felt warmth and comfort when I spoke to your peaceful soul throughout my challenging times.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

cognitive load theory	CLT
long-term memory	LTM
long-term store	LTS
short-term memory	STM
short-term store	STS
working memory	WM

List of Appendices

Publication and Presentations from the Thesis

Publications and conference presentations associated with this thesis include:

Journal Article

Likourezos, V., Kalyuga, S., & Sweller, J. (2019). The variability effect: When instructional variability is advantageous, *Educational Psychology Review*, *31*(2), 479-497. doi: 10.1007/s10648-019-09462-8 (see Appendix S)

Conference Presentations: Papers and Poster

Likourezos, V., & Kalyuga, S. (2019). *The variability effect in high and low guidance instruction: A cognitive load perspective*. Paper presented at the 2019 Mathematics Education Research Group of Australasia Conference, Perth, Australia.

Likourezos, V. (2018). *The variability effect in high and low guidance mathematics learning groups: An experimental study*. Poster presented at the 2018 American Educational Research Association Annual Meeting, New York City, New York.

Likourezos, V. (2018). *A study in the effectiveness of different instructional approaches to enhance mathematics learning*. Paper presented at the 2018 UNSW Faculty of Arts & Social Sciences Postgraduate Student Conference, Sydney, Australia.

Abstract

Based on cognitive load theory, the variability effect occurs when learners' exposure to highly variable tasks results in better test performance. Using four randomised controlled trials in the area of secondary and tertiary mathematics instruction, the present study investigated the effects of variability, with an emphasis on levels of instructional guidance and levels of learner expertise. Experiments 1, 2 and 4 hypothesised that learners who study fully-guided worked examples will yield higher post-test performance scores, compared to learners who attempt unguided problem-solving tasks (Hypothesis 1); and learners who study high-variability worked examples will yield higher post-test performance scores, compared to learners who study low-variability worked examples, with no difference being generated under problem-solving conditions (Hypothesis 2). Hypothesis 1 was not supported in Experiments 1, 2 and 4, while Hypothesis 2 was supported only in Experiment 2. The variability effect that was produced in Experiment 2 led to further investigation in Experiment 3, where it was hypothesised that more-experienced learners (experts) would demonstrate the variability effect, and less-experienced learners (novices) would demonstrate a reverse variability effect. This hypothesis was supported, producing a classic expertise reversal effect. In addition, in all four experiments, learners' cognitive load was evaluated by having each participant complete a subjective rating of difficulty scale upon completion of their learning tasks. The results supported the assumptions based on cognitive load theory: learners in the worked-examples groups experienced less cognitive load compared to the problem-solving groups (in Experiments 1, 2 and 4); novices experienced less cognitive load when solving low-variability problems compared to high-variability problems, and lower

cognitive load was experienced by experts, compared to novices, for both high- and low-variability tasks (in Experiment 3); and cognitive load associated with the completion of high-variability tasks was higher compared to the completion of low-variability tasks (only in Experiment 4). Although it is well grounded in empirical evidence that learners should be provided with worked examples during the initial stages of learning, these results strongly suggest that learners should be initially presented with low-variability problems, and as their levels of knowledge advance, variability should increase.

Introduction

Over the last 30 years, *cognitive load theory* (CLT hereafter) has developed into a robust instructional theory that has successfully identified, for novices, processes that foster learning and impediments that compromise learning. The theory is grounded in extensive research in cognitive processes arising from the interactions between working memory and long-term memory (see Sweller, Ayres, & Kalyuga, 2011, for a comprehensive review of the theory), and cognitive load (mental effort generated in working memory) that is experienced when processing instructional information or performing problem-solving tasks (Sweller, 1994; Sweller, van Merriënboer, & Paas, 1998).

In well-defined domains such as mathematics, the most common approach to learning is to initially provide students with worked examples that relate to new material. Substantial research from a cognitive load perspective confirms that initially providing appropriately designed worked examples is a more effective instructional technique than using problem-solving tasks for individuals who lack relevant knowledge structures for a particular task (Sweller et al., 2011). The variability technique is an enhancement and extension to the methodology of worked examples. Increasing problem variability exposes students to a greater range of tasks which facilitates the differentiation of relevant and irrelevant features of tasks (van Merriënboer & Sweller, 2005). Within the theoretical framework of CLT, the present study focuses on how variability fosters deeper understanding of problem-solving procedures by examining the effect of variability on worked-examples and problem-solving instructional formats, and the possibility of a reverse variability effect depending on learner prior knowledge, by comparing the performance of more- and less-experienced learners.

A worked example provides a detailed step-by-step solution to a problem so that learners do not need to try to work out the solution steps on their own. According to CLT, the *worked example effect* occurs when learning is enhanced for less-experienced learners if they are exposed to instruction that relies on studying worked examples, rather than instruction that directs them to attempt to problem solve without exposure to worked examples. However, with increasing expertise in a specific task domain, processing worked examples may inhibit learning because more-experienced learners will expend cognitive resources processing (integrating and cross-referencing) redundant information (information they already know). Correspondingly, the advantage of studying worked examples over problem solving ultimately reverses, so that solving problems becomes superior to studying worked examples for more-experienced learners (Kalyuga, Chandler, Tuovinen, & Sweller, 2001). Investigated within the cognitive load theoretical framework, this phenomenon is known as the *expertise reversal effect* (Kalyuga, 2007; Kalyuga, Ayres, Chandler, & Sweller, 2003).

Several studies have found a reversal in the effectiveness of instructional methods when there is a change in the level of learner knowledge in a particular domain (Kalyuga et al., 2003; Kalyuga, Chandler et al., 2001; Kalyuga & Renkl, 2010; Kalyuga, Rikers, & Paas, 2012; Tuovinen & Sweller, 1999). These studies showed problem solving to be a more effective strategy, compared to comprehensive guidance in the form of worked examples, for learners with high levels of prior knowledge in a domain, and an ineffective strategy for those learning new procedures and concepts for the first time. A contemporary issue that confronts CLT is whether providing novices with problem-solving tasks during the initial learning phase, can actually facilitate learning. CLT theorists argue that the only advantage in favour of problem solving has been limited to learners who have acquired high levels of prior knowledge and hence would benefit more from problem solving. Kalyuga et al. (2003) claim that if unnecessary instructional guidance is provided to expert learners, this imposes an additional cognitive load because experts will need to consume additional working memory resources to interpret the redundant information.

Learning occurs best when instructional designs are matched to the learner's level of expertise and consider fundamental characteristics of human cognitive architecture. CLT is linked to an established human cognitive architecture model which includes working memory (WM hereafter) and long-term memory (LTM hereafter). It is generally accepted that WM limitations – such as duration constraints, along with restricted storage and processing capacities for new information – require learners to avoid processing excessive amounts of interacting elements of information, otherwise cognitive overload may occur (Chandler & Sweller, 1996; Kester, Kirschner, & van Merriënboer, 2006; Mayer & Moreno, 2003; van Merriënboer & Kirschner, 2018).

Overloading WM inhibits learning, and consequently innovative designs have been developed by cognitive load theorists to reduce cognitive load to foster learning (Carroll, 1994; Cooper & Sweller, 1987; Gerjets, Scheiter, & Catrambone, 2004; Pollock, Chandler, & Sweller, 2002; Ward & Sweller, 1990). Accordingly, cognitive load optimisation can be achieved by providing instructional support to novice learners and removing unnecessary guidance as learners gain superior levels of proficiency in a specific domain. Despite WM's duration and capacity constraints, Sweller (2003, 2004) discusses no known limitations when information is retrieved from LTM, as information that is stored in LTM is extensive and comparatively permanent.

LTM contains elements of information that have been organised and stored in knowledge structures known as schemas (Paas, Renkl, & Sweller, 2004). According to van Merriënboer, Kirschner, and Kester (2003), schemas provide a bridge between a learner's prior knowledge and new information they need to know in order to perform a learning task. Schema construction incorporates chunking – a memory mechanism in which familiar units of information are grouped together to form a larger unit of information (Gobet et al, 2001; Miller, 1956). By chunking information into a larger unit of meaningful information, an individual can improve their WM capacity (Ericsson, Chase, & Faloon, 1980), and as a result, learners can hold large amounts of information in WM because a larger unit of information is easily dealt with as one element (van Merriënboer & Sweller, 2005).

Since a schema incorporates multiple elements of information into a single element (or chunk of information), Kalyuga and Sweller (2004) stipulate that it is possible for a single, high-level element to make WM more manageable, as a single element requires less WM capacity for processing, compared to the many, low-level elements it comprises. Learner expertise emerges from the construction of increasingly sophisticated schemas, by which easy concepts merge into more complicated ones. Sweller (2004) also maintains that if information that is processed in WM is not stored in LTM (in the form of schemas), no permanent learning has taken place.

In addition to human cognitive architecture, the type of material provided to learners is also critical to CLT. Researchers in CLT emphasise the importance of variability of tasks for learning and transfer. Research studies have shown that exposure to highly variable example-based instruction, compared to less variable, homogeneous examples, gives learners the opportunity to engage in deeper processing, enabling new knowledge to be adaptable to novel situations, resulting in enhanced transfer performance (Clark, Nguyen, & Sweller, 2006; Paas & van Merriënboer, 1994; Quilici & Mayer, 1996; van Merriënboer & Sweller, 2005).

Paas and van Merriënboer's (1994) study was the first study of variability from a cognitive load perspective to show that high-variability worked examples enhanced learning compared to low-variability worked examples. Despite high-variability examples requiring an increased use of WM resources (due to an increase in element interactivity), their experimental results showed that the *variability effect* was obtained using appropriate instructional designs. Notwithstanding the positive effects of high-variability tasks, surprisingly little is known about the advantages and disadvantages of providing high-variability tasks to learners with varying levels of prior knowledge in the domain.

Instructional procedures must be adapted for a learner's existing schematic knowledge base in LTM in a way that optimises cognitive load during learning (Kalyuga & Sweller, 2004). When schemas are acquired or automated, studies have demonstrated that more cognitive resources are freed up – allowing for greater capacity for creativity to enable transfer of learning (Cooper & Sweller, 1987; Kotovsky, Hayes, & Simon, 1985; Schneider & Shiffrin, 1977; van Merriënboer & Kirschner, 2018). Despite this, existing research has not explored whether learning and problem solving would improve with high- or low-variability tasks provided to learners with varying levels of prior knowledge. As such, this proposition provided the major impetus for the research undertaken for this thesis. By building on previous research that showed that the worked-example – high-variability combination yields superior transfer outcomes, the present study investigated the connections between low/high-variability worked examples and problem-solving tasks, and the connections between low/high-variability tasks provided to less-experienced and more-experienced learners in the domain.

This thesis is divided into two parts. Part I comprises the first two chapters: Chapter One reviews literature on human cognitive architecture underpinning CLT, and Chapter Two addresses the basic theoretical assumptions of CLT. Part II comprises the remaining three chapters: Chapter Three sets out the dependent and independent variables that embody the hypotheses which the present empirical study seeks to investigate; Chapter Four reports and discusses the empirical findings of four experiments comprising the present study, in light of previous findings that were generated by CLT; and Chapter Five concludes the thesis with a general discussion, including an overview of the results and the limitations of the present study, critical factors for educational practitioners and instructional designers to consider when tailoring tasks to learners with different levels of expertise, and recommendations for future research.

PART I: LITERATURE REVIEW

Chapter 1: Human Cognition

1.1 Introduction to Chapter 1

The goal of this chapter is to provide a historical review of the assumptions relating to the characteristics of *human cognitive architecture*, upon which CLT is based on. Human cognitive architecture refers to the components and properties of the human cognitive system whereby humans are able to think, learn, and problem solve (Sweller et al., 2011). This chapter will focus on the key memory structures of this architecture: WM (previously known as short-term memory) and LTM. Additionally, the conceptualisation of CLT and its link with biological evolution will be discussed with reference to a comparative framework comprising five principles. Analogies between human cognition and biological evolution will reveal how thinking, learning and problem solving are all natural occurrences (Sweller & Sweller, 2006).

1.2 Human Cognitive Architecture

The concept of human cognitive architecture is derived from examining the components of the human memory system (Sweller, 2003). The framework of human cognitive architecture is based on WM, which deals with the conscious activity of processing new elements of information (that need to be learned and constructed into meaningful knowledge), and LTM, where information is permanently stored in the form of schemas of varying size, complexity and degree of automaticity (Sweller et al., 1998). CLT is based on the way these two major components of human cognitive architecture are arranged and interconnected (Sweller et al., 2011).

Baddeley's (1992) widely used WM model assumes a limit on complex cognitive activities (such as learning, reasoning, and language comprehension), in terms of temporary storage and manipulation of information. Sweller (1994) claimed that this limitation could be overcome through schema acquisition and automation. Constructed schemas only develop into automated schemas if they are continuously applied across consistent problem situations (van Merriënboer & Sweller, 2005). Because schema automation eliminates or reduces the need for conscious processing in WM, automation makes it possible for familiar tasks to be performed smoothly and accurately – a mechanism by which maximum WM capacity is made available for unfamiliar tasks to be learned with maximum efficiency.

Without the dual functions of schema construction and automation, van Merriënboer and Sweller (2005) claimed that WM is incapable of dealing with complex, unfamiliar tasks, for which no schemas are available. However, when mandatory skills are automated, this can increase the availability of WM resources that can become available for managing complex interactions between unfamiliar elements (van Merriënboer & Sweller, 2005).

Human cognitive architecture can be used to explain how different skills and experiences account for the ways novices and experts reason and plan solutions. For example, when a novice in a mathematics domain attempts to solve a problem, they do so by backward reasoning. They begin with the known end, or by envisioning the desired end (the goal state), and then attempt to solve the problem by using a general means-ends search (Sweller, 1999). Because a novice has not yet had extensive experience solving similar problems, they will reason between the superficial features of the question by recalling and applying a sequence of individual formulae that relate to the specific parts in the question (initial problem state). This imposes considerable cognitive load and diverts attention from critical features of the problem that are crucial for learning (Ayres & Sweller, 1990;

Cooper & Sweller, 1987; Owen & Sweller, 1985). In contrast, an expert in a mathematics domain is able to solve a problem by forward reasoning. They are able to generate an integrated representation of the problem description based on principles. Because an expert possesses domain specific knowledge, they are able to analyse and sub-group the problem within a generalised solution (schema) from their LTM. This enables them to solve the problem in a single mental step which demands minimal WM load because they are able to see the entire solution as one unit. This difference between novices and experts was described in Larkin, McDermott, Simon, and Simon's (1980) research which used two simulation models that solved elementary physics problems in ways that were analogous to novice and skilled human problem solvers when they produced solutions on paper.

Adding to the above, when a task requires several elements to be simultaneously manipulated in WM, cognitive load levels become "naturally high" (Sweller, 1994, p. 295). The concept of a high-element interactivity task explains why material may be difficult to learn and understand, because the elements need to be mentally organised into a coherent knowledge structure and integrated with relevant existing knowledge (Mayer & Moreno, 2003). In this sense, expertise develops as learners develop cognitive schemas that incorporate interacting elements.

Sweller et al. (1998) claimed that material that is low in element interactivity entails sequential learning and requires just a few elements to be held in WM at a given time. A low-element interactivity task is very simple since no schemas are acquired or integrated with other schemas. Furthermore, the level of expertise of a learner will determine the level of element interactivity. This is because a single element for a higher-ability learner might be equivalent to a large number of interacting elements for a lower-ability learner (van Merriënboer & Sweller, 2005).

1.2.1 The modal model of human cognition.

The theoretical framework adopted by this thesis, which is underpinned by human cognitive architecture, is rooted in the *modal model* of human cognition. In the late 1950s, information processing models were created by cognitive scientists (Shiffrin & Atkinson, 1969; Waugh & Norman, 1965) who built on the memory processes which had traditionally been broken into three stages by memory researchers: the acquisition of new information into a system, the storage of this information within the system, and the retrieval of information from the system when it was required. The main features of the framework that guided these information processing models came to be known as the modal model of memory. The modal model assumed three separate memory stores (sensory, primary and secondary), which were connected; that is, information could be transferred between them (Healy & McNamara, 1996).

The main feature of the modal model was the short-term store (STS hereafter) of information because long-term learning was assumed to be dependent on the information held in this STS, until it transferred to the long-term store (LTS hereafter). Baddeley (1986) clarified this by claiming that the probability of learning was “a direct function of the amount of time an item resides in STS” and that the STS was “responsible for encoding the incoming material in a range of different ways” (p. 16). This suggests that the STS was considered to be a temporary storage system where information was manipulated and learning was limited.

Certain aspects of the formulation of the modal model's multistore system were objected to. For example, Tulving and Patterson (1968) opposed the notion that information was transferred from one store to another. Similarly, Shallice and Warrington's (1970) findings opposed the idea that information needed to enter STS before entering LTS.

By the early 1970s, many problems began to beset the general modal model. An increasing number of new techniques and ongoing changes to the model resulted in more complex components (see Sweller et al., 2011, for a review). For example, it was proposed that there was a loop connecting LTM to sensory memory which allowed permanently stored information to influence briefly held stimuli for initial perceptual processing. This proposition suggests that information does not flow through memory in a linear fashion; that is, by entering sensory memory and proceeding to WM for further processing, prior to entering LTM. Instead, information is processed simultaneously between WM and LTM which are interconnected components within the memory system; with LTM influencing initial processing.

1.2.2 Short-term memory.

Research on memory performance which supports the modal model of memory can be traced back to James (1890), who first distinguished "primary" memory, which he described as immediate concerns held momentarily in consciousness, from "secondary" memory, which he assigned to unconscious, lasting memories. It was not until the late 1950s that the distinction between short-term memory (STM hereafter) and LTM was more fully developed (Broadbent, 1957; Brown, 1958). STM was used with reference to tasks that contained small amounts of information that needed to be retained for brief moments, and LTM was used with reference to information that was stored for more than a few seconds.

An early argument that information processing was constrained by STM was discussed in Miller's (1956) classic article, where he claimed that a normal person's memory span could handle approximately seven, plus or minus two, chunks of information at any point in time. The general limitation on human information processing, as represented by Miller's (1956) model, suggests that to process information more efficiently, the size of the chunks needs to be expanded so that several elements of information are organised into a single, meaningful unit of information. Although many cognitive psychologists have accepted this narrow memory range of around seven items (across many domains), Cowan (2001) argued, drawing on a wide variety of data on capacity limits, that the common capacity of STM was around three to five chunks. This implies that the number of chunks that can be reliably stored is dependent on the type of processing required; that is, capacity is limited by the number of chunks and the level of processing of these chunks (for example, a complex cognitive activity such as problem solving, which requires more processing, would result in the storage of fewer chunks).

As well as being limited in capacity, studies have suggested that STM is limited in duration. For example, Peterson and Peterson's (1959) study confirmed that memory in STM faded away as a function of time (time-related decay) if information was not refreshed by rehearsal. By measuring how well university students could recall a trigram (a three-consonant syllable, such as 'ABC') after undertaking a continuous verbal activity (counting backwards by three from a given number), Peterson and Peterson (1959) found that after three seconds, students had forgotten half of the information, and after eighteen seconds, students had forgotten almost everything. However, Cowan (2001) discussed several studies that revealed forgetting in STM was more a result of interference (where old information was replaced by new information) than time-related decay. For example, Waugh and Norman's (1965) experiment showed that unrehearsed verbal information that was interfered with intervening information resulted in forgetting, irrespective of time-related decay.

1.2.3 Working memory.

Traditionally, the prominent framework for memory research postulated that information entered STM after being initially processed by sensory memory – the component which perceived, recognised, and assigned meaning to incoming stimuli. STM has since been replaced with WM because of the distinction between the subprocesses in STM that passively maintained information and the active processing of information in WM (Baddeley, 2001). In his book on WM, Baddeley (1986) referred to the common perception that "memory might not be a single monolithic system but might have two or more components" (p. 3).

According to Baddeley (1992), the concept of WM evolved from the idea that there was a unitary STM system. The concept postulated that there was a more complex framework, where the temporary storage of information was used to process complex cognitive tasks such as "language comprehension, learning, and reasoning" (p. 255). Baddeley and Hitch (1974) proposed that the simultaneous storage and manipulation of information in WM required three components: the central executive (which controlled the overall system), and two subsystems (which were responsible for the maintenance of spatial and verbal information): the visuo-spatial sketchpad (the visual channel) and the phonological loop (the auditory channel). Despite some criticism which indicated that the WM model placed too much emphasis on the interpretation of a tripartite structure rather than the processes underlying the subsystems, Baddeley and Hitch's WM model has proven useful in explaining the way in which information is temporarily stored as part of its central role in complex cognitive processing (Tulving & Craik, 2000).

To understand the usefulness of theorising the existence of a tripartite system, Miyake and Shah (1999) surveyed common and diverse conceptualisations of ten comprehensive WM theories. They discussed the combination of an executive control with specialised storage systems which displayed key differences between verbal and visual material. Additionally, Baddeley's (2001) further analysis of the controlling central executive led to an update to his multicomponent WM model with the incorporation of a fourth system, the "episodic buffer", which formed an interface between the visual and auditory subsystems. Irrespective of ongoing debates concerning the validity of the multicomponent WM model, cognitive scientists have continued to use this model in many CLT studies, and it has led to an increased level of knowledge about human cognitive architecture.

1.2.4 Long-term memory.

The distinction between temporary storage of information (in WM) and permanent storage of information (in LTM) continues to be a central characteristic of all prominent information processing theories. LTM refers to the unconscious component of the memory system where unlimited amounts of information, in the form of schemas, are stored. According to Tulving and Craik (2000), the transfer of information from WM to LTM is the most vital part of information processing, which has been identified to be a specific function of WM. Unlike the capacity and duration constraints when novel information is dealt with in WM, there are no known limitations when WM deals with information retrieved from LTM (Ericsson & Kintsch, 1995; Sweller, 2003, 2004).

Information in LTM is coded in terms of its meaning. For example, in relation to verbal material, the real distinction between STM and LTM, as shown by Baddeley and Dale (1966), was their different coding characteristics. STM appeared to be predominantly acoustic, while LTM appeared to be mainly semantic (meaningful encoding). Moreover, Shiffrin and Atkinson (1969) argued that forgetting from STM occurred after less than 30 seconds, while material was forgotten from LTM either very slowly or not at all. Shiffrin (1975) stated that once sensory information entered STM, it was initially encoded automatically. At a later stage, additional rehearsal, such as maintenance or coding, occurred to facilitate the complete transfer from the STS to the LTS.

According to Sweller (1999), LTM is not just a passive repository of memorised facts; rather, it contains an unlimited storage of "sophisticated structures that permit us to perceive, think and solve problems" (p. 10). Since WM can only deal with a limited number (usually no more than two or three) of novel interacting elements, the expansion of learners' processing ability is possible because of the schemas brought from LTM to WM (Paas, Renkl, & Sweller, 2003). Schemas retrieved from LTM organise and store a vast number of low-interacting elements, that would normally have exceeded the processing capacity of WM if each interacting element needed to be processed individually (Paas, Renkl, & Sweller, 2003).

As mentioned, schemas allow problem solvers to group elements of information into appropriate categories according to the manner in which the information will be used to solve a similar category of problems (Chi, Feltovich, &

Glaser, 1981). Knowledge, according to cognitive science, can be distinguished between *declarative* knowledge and *procedural* knowledge (Chi & Ohlsson, 2005; Schraw, 2006). Declarative (or descriptive) knowledge refers to static information such as facts (for example, ‘A’ is the first letter of the alphabet), concepts (for example, the abstract phenomena of happiness), and the relationship between concepts which form integrated conceptual knowledge in a particular domain. Procedural (or implicit) knowledge refers to knowing how to process or manipulate information (for example, driving a car). This distinction indicates that even if the declarative knowledge of two problems may be the same (i.e., common elements are contained in each problem), the procedural knowledge involved in solving the problems may differ (i.e., different processes are required to solve each problem).

1.2.4.1 Novice-expert differences.

The role of LTM in human cognitive performance was accounted for in de Groot’s (1965), and Chase and Simon’s (1973a, 1973b) pioneering research on chess players’ expertise. They theorised that after many years of practice, chess experts’ superiority in memorising chessboard configurations was obtained from encoding a large number of specific arrangements of chess pieces in terms of familiar, well-organised, integrated memory patterns of information. The ability to chunk and organise new information into schemas enabled chess experts to quickly recognise meaningful patterns of information, such as successful chess moves and the implications of such moves (de Groot, 1965).

Chase and Simon (1973a, 1973b) found that expert and novice chess players relied on a similar number of chunks in STM with reference to the patterns of the chess pieces on the chessboard. On the one hand, the chess expert’s performance was superior, compared to the novice’s performance, because they possessed chunks of information that were substantially more complex. Experts automated parts of the problem-solving process contained in these more complex chunks which enabled them to quickly recognise specific patterns of information (chunks) presented on the chess board so that they could encode and recall chess board configurations. On the other hand, novices recognised fewer chess board configurations which required them to encode arrangements in terms of individual chess pieces. The inability of novices to recognise meaningful patterns of information increased demand on their conscious attention to search for better moves. De Groot (1965) explained that the superior recall of expert chess players when briefly presented with chessboard configurations was based on their ability to chunk information and organise it into schemas.

CLT is concerned with the way learners develop expertise in a domain during learning and problem solving – a process that requires learners to circumvent WM capacity limitations. By mindfully combining simple elements into more complex elements, to facilitate continuous construction of more complex schemas, skilled performance development is made possible. Sweller (1989) clarified this by arguing that schema acquisition (construction and automation) are the constituents of skilled problem-solving performance. When learners practise a task extensively, schemas can become automated – resulting in schemas being processed unconsciously, further reducing the cognitive load on WM (Paas, Renkl, & Sweller, 2003). Hence, freeing up WM capacity, by schema construction and automation, enables learners to process and integrate new information with prior knowledge in WM, before new information is encoded in LTM (Sweller et al., 2011).

There is ample evidence which shows that the expansion of WM capacity and the enhancement of memory performance occur when higher levels of memory skill are developed and relevant prior knowledge is acquired within a task domain (Kalyuga et al., 2003). Ericsson and Kintsch (1995) showed that "subjects must acquire encoding methods and retrieval structures that allow efficient storage and retrieval from LTM", in order to meet the specific demands of WM for a certain activity (p. 239). Within a complex domain, the acquisition of a greater number and sophistication of schemas account for the difference between novices and experts (Chase & Simon, 1973a, 1973b). Cooper and Sweller (1987) suggested that schema acquisition, more than schema automation, plays a major role in skilled problem-solving performance.

As discussed above, organising and storing information in LTM is not the only function served by schemas. Since there are no limitations on the magnitude, complexity and refinement of schemas, these vast arrays of interrelated elements can also effectively boost WM capacity. That is why Sweller et al. (1998) claimed that complex schemas, held in WM as a single entity, enable experts to encode and understand the elements of an intellectual task into an entity with one or few elements, yet novices, on the other hand, who do not possess relevant schemas, must remember and process each element individually.

Research conducted by Chi et al. (1981) revealed that the process of categorisation and representation of physics problems differed between experts and novices. The expert-novice difference was related to the content of the problem schemas. Even though novice schemas contained elaborate declarative knowledge, these schemas were "poorly formed" (p. 122), or contained "fewer explicit procedures" (Chi et al., 1981, p. 140). On the other hand, expert schemas contained a considerable amount of procedural knowledge which, at the very least, contained possible solution methods. The results from Chi et al.'s (1981) experiments revealed that the use of a cueing strategy, employed by the expert, involved analysing the problem, categorising the problem, and then selecting and applying the associated principles in their knowledge base to solve the problem representation. By contrast, the novice searched for a particular solution based on the literal surface features explicitly stated in the problem statement. Unlike the expert, the novice was unable to activate an internal problem-solving schema (category knowledge), as a response to some cue in the externally presented problem, that could provide the general form necessary to solve the specific category of problems. This indicates that an expert can call upon their rich schemas (which contain strong skills and extensive problem-solving processes) to guide them in interpreting and solving problems, which in turn results in a more efficient learning processes.

1.3 Evolutionary Perspective on Human Cognition

A framework for understanding natural information processing systems was developed by Sweller and Sweller (2006). This framework conceptualised CLT in evolutionary terms by linking human cognitive processes with biological evolutionary processes, and Geary's (2008) work on biologically primary and secondary knowledge.

Biologically primary knowledge is knowledge that has evolved over thousands of generations. It is acquired unconsciously and effortlessly without instruction, during immersion in a human society, rather than by explicit instruction. Learning to speak and listen in a native language are some examples. Because we have evolved to acquire such knowledge automatically, it does not need to be taught. On the other hand, biologically secondary knowledge needs to be explicitly taught and requires conscious effort, with most subject areas taught in educational institutions belonging to this category. Learning how to read and write are some examples (Geary, 2008). The modal model described in section 1.2.1, applies to biologically secondary rather than biologically primary information. CLT is strongly associated with the acquisition of biologically secondary knowledge.

According to Sweller (2008), human cognition and biological evolution have important similarities in how they generate new information, accumulate and reserve that information, and apply and reuse that information indefinitely. It is argued that the manner in which all natural information processing systems, including human cognition, operate, is based on five fundamental principles (Sweller, 2004; Sweller & Sweller, 2006). These principles reflect the deep connection between the functions and processes of the human cognitive system and biological evolution: *information store principle* and LTM, *borrowing and reorganising principle* and knowledge acquisition, *randomness as genesis principle* and search-based problem solving, *narrow limits of change principle* and WM of less-knowledgeable learners, and *environmental organising and linking principle* and WM of more-knowledgeable learners.

1.3.1 Information store principle.

The information store principle relates to the very large store of schematically organised information in human LTM (required for the adaption to complex human cognitive activity), and similarly, to biological information held in a genome, that is encoded in its DNA (for the creation and maintenance of an organism). The contents in human LTM must be altered for learning to happen, and likewise, in evolutionary biology, a species' genome must be altered for evolution by natural selection to exist.

1.3.2 Borrowing and reorganising principle.

The borrowing and reorganising principle refers to mechanisms used for information stores to acquire information when needed. Examples include, when information stored in LTM is insufficient to complete a particular task, and equivalently, when information stored in a genome may be insufficient to conduct a vital activity. In human cognition, information can be acquired by borrowing from another human (either through imitation, listening or reading) and reorganising knowledge by encoding the new information into existing knowledge. Similarly, in the case of biological evolution, a species' genome is changed through asexual (copying of a genetic code) and sexual (a novel combination of female and male genetic codes) reproduction. Inherent to the borrowing principle is the random manner in which old information combines with new information. For example, when borrowing information from another human, information cannot be copied in the exact way because of the way elements are uniquely combined from another LTM in the prevailing LTM. Similarly, novel constructions arise when genetic codes are copied and combined.

1.3.3 Randomness as genesis principle.

The randomness as genesis principle relates to the situation when individuals endeavour to solve a problem by random generation and testing because knowledge through the borrowing and reorganising principle is unavailable. The randomness as genesis principle provides the mechanism for creativity. It allows humans to go beyond their existing knowledge to create new information by using their primary knowledge to generate new, domain-specific, secondary information. Just as humans randomly generate ideas by testing them for effectiveness during problem solving, new genetic codes are randomly generated and tested. In biological evolution, if new information is effective, this results in random mutation (changes in DNA) – similar to the way learners retain successful, novel solution steps when they successfully solve a problem using no prior knowledge. On the other hand, when a randomly-generated genetic code is ineffective after testing, the new information is abandoned – similar to the way learners abandon a particular way of solving a problem after a failed attempt.

1.3.4 Narrow limits of change principle.

The narrow limits of change principle is concerned with the slow, incremental stages by which new information is generated in a natural information processing system. A rapid alteration in human LTM by way of a combinatorial explosion (for example, if a learner were to attempt to simultaneously process a large quantity of unorganised, random elements) would not be possible because of the capacity and duration limitations of WM. For instance, mathematically, if three elements of information are handled, the number of possible permutations (i.e., the number of various ways the three elements can be ordered) is $3! = 3 \times 2 \times 1 = 6$ (i.e., 123; 132; 213; 231; 312; and 321), an amount that is unlikely to overload a natural information system. However, if the number of elements is doubled, the number of possible permutations drastically increases to $6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$, an amount that cannot be easily processed by a natural information system. Limited WM in human cognition ensures information in LTM is not manipulated by a random mechanism (because of the borrowing and reorganising principle, and randomness as genesis principle) that has no function. Only small amounts of novel information can be processed consciously in WM at a given time, in conjunction with information held in LTM. In a similar way, genetic changes develop over numerous generations because the epigenetic system has a similar role in evolutionary biology as WM has in human cognition. The epigenetic system selects the information from the external environment and transfers this information to the genetic system (which has the capability to make a change in an organism's DNA).

1.3.5 Environmental organising and linking principle.

The environmental organising and linking principle describes how stored information is used to generate actions that are appropriate to a particular environment. In addition, the principle explains why the narrow limits of change principle does not function after information is organised in the information store to be used in the environment. In human cognition, WM has no capacity and duration limitations when it uses information that is drawn from LTM and applied to a particular situation. Because accumulated information in LTM has been tested for its effectiveness, it does not require random processing in WM, and instead there are no limitations on what can be reused, thus reducing unnecessary WM load. Correspondingly, in evolutionary biology, vast amounts of information stored in a genome can be used by an epigenetic system to ensure, for example, there is sufficient protein synthesis required by a particular environment. The comparison in this case is the analogy between LTM and the genome, and WM and the epigenetic system.

The abovementioned five principles are all required for a continual information processing system to exist. For instance, without a complex store of information held in LTM, the human cognitive system is unable to accommodate an individual's skilled performance in a domain. In more detail, the extent to which an individual is skilled in a domain depends on the information they acquire in

LTM as a result of the alteration and construction of knowledge. When an individual is unable to solve a problem because knowledge is unavailable, the learner will test randomly generated solution steps as a last resort. Sweller (2009) argued that the absence of learned material gives rise to human creativity as individuals are unable to retrieve acquired knowledge from previous experience.

As previously discussed, central to CLT are the limitations imposed by WM in both capacity (Miller, 1956) and duration (Peterson & Peterson, 1959). Given Miller's (1956) finding that no more than seven items of novel information can be held in STM, and Cowan's (2001) observation that no more than four items can be processed simultaneously in WM, random alterations to existing knowledge in LTM is slow and limited. These narrow limits of change pertain to novice learners because they lack sophisticated schemas associated with a task at hand. Conversely, expert learners, who deal with familiar information, are able to transfer large amounts of organised schematic information in a single, higher level element, from LTM to WM, to assist with carrying out complex problem-solving tasks.

1.4 Summary of Chapter 1

This chapter outlined the properties of human cognitive architecture critical for understanding human cognition. The key organisational and structural properties of this architecture comprised the modal model of human cognition, with particular reference made to the WM and LTM. The conceptualisation of CLT was also discussed in relation to five biological evolutionary principles, showing how they can be used to explain particular cognitive characteristics.

The next chapter will discuss how CLT developed on the basis of human cognitive architecture. It will detail the theoretical framework of CLT, and how our understanding of CLT can significantly impact learning and instruction.

Chapter 2: Cognitive Load Theory

2.1 Introduction to Chapter 2

CLT developed as an instructional theory to explain the relationship between two constructs: cognitive load (the difference between a task demand imposed on a learner and their ability to master the task demand) and learning (the process of acquiring new or adjusting existing knowledge and skills by instruction or study). This chapter will review the theoretical framework of CLT and its contribution to an understanding of human cognitive processes in learning. Categories of cognitive load, subjective measures of cognitive load, and cognitive load effects, which inform many aspects of the present study, will be examined. Specifically, detailed discussions of the worked example effect, the expertise reversal effect, the redundancy effect, the split-attention effect, and the variability effect, will present how these instructional design effects influence cognitive load to improve learning and instruction.

2.2 Overview of Cognitive Load Theory

CLT is an instructional theory primarily associated with complex cognitive tasks, where novel, unorganised information needs to be processed concurrently for meaningful learning (Paas, Renkl, & Sweller, 2003; Sweller, 1988, 1999). Since its initial development in the 1980s by John Sweller, CLT has assumed a human cognitive architecture that focuses on WM limitations, and how these impact learning and instructional design (Sweller et al., 1998). As stated by Paas et al. (2004), CLT suggests that learning is most efficient and effective when instructional designs and procedures are in accordance with major characteristics of human cognitive architecture.

The limitations of WM make it necessary to avoid “cognitive overload” – a situation where competing demands of cognitive processes, induced by a learning task, exceed the processing capacity of WM and impair student learning (Sweller et al., 2011). The potential for cognitive overload makes it necessary to achieve a balance between instructional guidance and the availability of an organised learner knowledge base. Accordingly, the manner in which information is attended to and manipulated in WM during the learning phase, before it can be stored in LTM, necessitates the implementation of appropriate procedures and techniques to optimise cognitive load.

2.2.1 Categories of cognitive load.

Cognitive load is generally defined as WM resources used to process a specific task (Sweller et al., 1998). There will always be cognitive load associated with any learning process because performing a particular task involves a level of intellectual complexity. Therefore, instructional design should ideally avoid imposing unnecessary cognitive load (Chandler & Sweller, 1996). Of essential interest to cognitive load theorists is the ease with which WM processes information to facilitate schema construction within LTM (van Merriënboer & Sweller, 2005).

According to CLT, cognitive load can arise from three sources: *intrinsic*, *extraneous*, and *germane* (Paas, Renkl, & Sweller, 2003; Paas et al., 2004; Sweller, 2010; Sweller et al., 1998; van Merriënboer & Sweller, 2005). Understanding these categories of cognitive load can explain why some learning environments create higher processing load and why some learners can be overwhelmed by the numerous elements involved in a complex learning task.

2.2.1.1 Intrinsic cognitive load.

When learning complex tasks, a high number of new, interacting elements need to be processed simultaneously in WM for learning to begin (Chandler & Sweller, 1991; Sweller & Chandler, 1994; Sweller et al., 1998). Element interactivity occurs when elements must be processed in conjunction with other elements, and without this interconnection, presenting them in isolation would have no meaning (Chandler & Sweller, 1991; Sweller & Chandler, 1994). By estimating the number of interacting elements in a learning task, the level of element interactivity can be determined (Sweller & Chandler, 1994; Tindall-Ford, Chandler, & Sweller, 1997). In mathematics, for instance, a low element interactivity task requires the processing of fewer individual learning elements to reach a solution, compared to a high element interactivity task. For example, consider the equivalent problems, '$18m = 12m + 36$' and '$(6 \times 3)m = 24m - 12m + (6)^2$', where both need to be solved for m. Assuming novice learners in both cases, the first problem is a lower element interactivity task because it requires fewer steps to solve the equation for m, compared to the second problem. The first problem requires at least two steps to reach the solution. One solution method would be to subtract $12m$ from both sides (to arrive at the solution step, $6m = 36$), and then divide both sides by 6 (to obtain the answer, $m = 6$). The second problem requires at least five steps to reach the same solution. One solution method would be to multiply 6 and 3 (to obtain 18), to subtract $12m$ from $24m$ (to obtain $12m$), to square 6 (to obtain 36), to formulate the simpler equation '$18m = 12m + 36$', and then follow the same two steps that were used to solve the first problem.

Since element interactivity is partly dependent on the intrinsic nature of the learning task, element interactivity is the main influencer of intrinsic cognitive load. Intrinsic cognitive load is dependent on the complexity of the material that is to be learned and the expertise of the learner (Ayres, 2006; Kalyuga et al., 2003; Sweller & Chandler, 1994). High intrinsic cognitive load will arise when learning material is characterised by high element interactivity and when a learner lacks rudimentary schemas to complete the task (Sweller & Chandler, 1994). For instance, in relation to the previous mathematics example, the first problem might be regarded as being a high interactivity task for a lower-ability mathematics student, who requires more than two steps to solve the equation for *m*.

Closely related to the previous point, the complexity of learning tasks is dependent on the number of interacting elements that are actively related and controlled in WM during the learning process (van Gog, Paas, & van Merriënboer, 2006). Complex mathematical tasks are often high in intrinsic cognitive load because numerous elements must be dealt with simultaneously and not sequentially. To assist less-knowledgeable learners in a domain, complex learning can be enhanced if information is presented sequentially. For example, by using simple-to-complex scaffolding, van Merriënboer et al. (2003) argued that the intrinsic components of cognitive load could be reduced.

Sequential instruction involves single elements that can be learned independently of each other, and does not impose a high cognitive load because element interactivity is low, irrespective of the number of elements that need to be assimilated (Chandler & Sweller, 1996; Clarke, Ayres, & Sweller, 2005). Pollock et al.'s (2002) study provided evidence that sequencing instruction of complex material in two parts – from isolated elements (initially presenting discrete elements that could only be processed serially) to interacting elements (subsequently presenting the original full set of interacting elements) – could artificially reduce intrinsic cognitive load. By reducing interactions among information elements during the initial stages of learning, learners were able to fully understand the complex material when they were confronted with it during their later learning phase. In relation to the previous mathematics example, for instance, the interactive elements in both equations are presented simultaneously. However, if the elements were presented sequentially, this would help reduce intrinsic cognitive load for a lower ability mathematics student. Breaking down the higher element interactivity task into simpler, sequential tasks, such as '$6 \times 3 = ?$', '$(6 \times 3) \times m = ?$', '$(6 \times 3)m = ?$', would omit many interacting elements in the second problem, and reduce cognitive load to a more manageable level.

Eliminating interacting elements may partially compromise full understanding of the relationships between the elements. However, an advantage of learning from independent (isolated) elements is that students are able to form partial schemas initially, and later form a whole schema after receiving instructions on how the elements interact with each other (Pollock et al., 2002). To reduce the element interactivity of learning materials, Blayney, Kalyuga, and Sweller's (2010) experiment, which tested university students in the domain of accountancy, compared an isolated-elements condition (by presenting information sequentially in an isolated form) with a fully interactive-elements condition. Their results revealed that less-knowledgeable learners benefitted more from an isolated-elements condition, as opposed to more-knowledgeable learners. In contrast, the instructional sequence involving an isolated-elements format interfered with the more-knowledgeable students' learning because extra memory resources were required to integrate the simple, isolated elements with their existing knowledge. These findings indicate that more-experienced learners do not need to process elements serially because they have sufficient information stored in their LTM to handle the fully interacting elements without cognitive overload. On the other hand, when solving high element interactivity problems, presenting isolated elements of information (that can be processed sequentially, rather than simultaneously) facilitates learning for less-experienced learners when intrinsic cognitive load exceeds cognitive capacity.

To avoid the difficulty of comprehending the high element interactivity of complex tasks, an option left for learners is to combine information elements into cognitive schemas so that interacting elements are not considered individually within WM (Sweller et al., 1998). Intrinsic cognitive load reduces when a task containing a collection of elements is organised by a single schema. In relation to the previous example, one way to combine some of the interacting elements in the higher element interactivity task, '$(6 \times 3)m = 24m - 12m + (6)^2$', would be to divide each term by a common factor of six. Applying this problem-solving schema would produce a simpler equation, '$3m = 4m - 2m + 6$', comprising of fewer interacting elements. Applying such a schema would make it easier to accomplish the intellectual task.

Along with the process of schema acquisition, Schneider and Shiffrin (1977) claimed that capacity limitations of WM could be bypassed if learners used automated schemas. By using automated schemas, fewer demands are placed upon the capacity of WM because "an automatic process operates through a relatively permanent set of associative connections" in LTM, and "once learned, an automatic process is difficult to suppress, to modify, or to ignore" (Schneider & Shiffrin, 1977, p. 2). In the previous mathematics example, an expert mathematics student could perform the recurrent task aspects efficiently and effectively if they possessed an automated schema which involved associating a particular characteristic of the higher element interactivity task (e.g., where every term in the equation is a multiple of six) to a particular action (e.g., by dividing every term by six to produce the equation, $3m = 4m - 2m + 6$, which is easier to solve).

An automatic sequence does not consume any WM resources because it has been learned from earlier controlled processing which consumed WM capacity (Schneider & Shiffrin, 1977). The importance of schema construction and automation of pre-existing schemas arose from the understanding of human cognitive architecture. Namely, a procedural activity is rapid and requires low levels of conscious attention and WM resources when automated LTM knowledge structures are used to execute the automatic procedure.

2.2.1.2 Extraneous cognitive load.

Unlike intrinsic cognitive load which is innate and unalterable other than by changing what is learned or changing the expertise of learners, extraneous cognitive load is controllable because it varies exclusively by the way instructional information is presented to learners. Extraneous cognitive load is imposed when effort is needed to process poorly designed or unsuitable instructional procedures (Sweller et al., 1998). This commonly occurs when a problem-solving task, without any guidance, is presented to a learner that has insufficient prior knowledge to build upon. Under such conditions, the novice learner's WM resources are absorbed by random attempts at finding a solution to the problem, rather than being directly involved in learning, since they do not possess developed schemas to problem solve error-free. Consequently, when extraneous cognitive load, imposed by a problem-solving task, interferes with learning, fewer cognitive resources remain in WM that can be devoted to schema construction and automation

(Sweller, 1994). Thus, extraneous cognitive load is considered detrimental for learning because WM resources are involved with following instructions, rather than learning (Paas et al., 2004). As discussed below, differences in extraneous cognitive load are just as dependent on variations in element interactivity as differences in intrinsic cognitive load.

Element interactivity is the major source of WM load and can be used to distinguish between intrinsic and extraneous cognitive load (Sweller, 2010). Reducing extraneous cognitive load lies at the core of CLT, given it is a theory of cognition and instructional design. However, like intrinsic cognitive load that is commonly discussed in terms of element interactivity, it was suggested by Sweller (2010) that extraneous cognitive load should also fall within the concept of element interactivity. For example, a learning task that requires the mental integration of separately presented sources of information would be regarded as a high element interactivity task if learners had to intensively match elements to make any sense of the information. Element interactivity in this case would relate to extraneous cognitive load (because the suboptimal instructions can be improved by physically integrating the elements), rather than intrinsic cognitive load (because learning content remains the same before and after the physical integration of the split-source elements of information).

For effective learning to occur, within the framework of CLT, the total of intrinsic and extraneous cognitive load should not exceed total WM resources. If limited WM capacity is exceeded by high intrinsic cognitive load and high extraneous cognitive load, this will interfere with, and have negative effects on learning (Sweller et al., 1998). Since intrinsic load is relevant to learning (and in some cases can be temporarily altered by instructional interventions such as sequential instruction), it is essential to decrease extraneous load when intrinsic cognitive load is high, in order to avoid cognitive overload. It should be noted that high extraneous cognitive load may not overload WM if intrinsic cognitive load is very low, because total cognitive load is manageable.

2.2.1.3 Germane cognitive load.

Similar to extraneous cognitive load, germane cognitive load varies because it is also imposed by the design of instructional material. However, unlike extraneous load which impedes learning, germane cognitive load is relevant to learning. Germane cognitive load contributes to learning because it comprises WM resources devoted to accommodating intrinsic cognitive load by contributing to the development of schema construction and automation (Sweller et al., 1998; van Merriënboer, Kester, & Paas, 2006). For this reason, germane cognitive load is not considered as an independent source of cognitive load in addition to intrinsic and extraneous cognitive load (Kalyuga, 2011; Sweller, 2010).

The concept of germane cognitive load was initially introduced by Sweller et al. (1998) to explain the effects of variability in learning materials. Processing high-variability material requires the learner to identify variants of the task such as the context and way the task is presented. Studies on variability have shown that high variability yields beneficial effects on schema construction and transfer of learning as demonstrated by superior performance in solving novel problems (Paas & van Merriënboer, 1994; Quilici & Mayer, 1996; Ranzijn, 1991). Determining which schemas are applicable in solving high-variability problems requires the learner to invest more cognitive load because of the highly varied sequence. It is this increase in WM resources, which is devoted to dealing with intrinsic cognitive load (and less WM resources devoted to dealing with extraneous cognitive load), that comprises germane cognitive load.

On the grounds that germane cognitive load is concerned with processes that are relevant for the acquisition of knowledge, this strengthens the idea that germane cognitive load should be considered as *germane resources* (Ayres, 2018; Kalyuga, 2011; Sweller, 2010; Sweller et al., 2011). Given both extraneous and germane cognitive load are imposed by the design of the task, in contrast to intrinsic cognitive load which is imposed by the nature of the information contained in the task, there exists a rational view that a reduction in extraneous factors and a replacement of these with intrinsic factors require mental effort expended in the form of germane cognitive load (van Gog & Paas, 2008). Accordingly, Sweller (2018) inferred that the increase in germane cognitive load from a resultant decrease in extraneous cognitive load produces a decrease in overall cognitive load. Hence, this explains why germane cognitive load should not be classified as an independent source of cognitive load (Sweller, 2010).

2.2.2 Managing cognitive load through instructional design.

On the basis that total cognitive load consists of the addition of intrinsic and extraneous cognitive load, as mentioned, learning can only occur if overall cognitive load does not exceed available resources in WM (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Consequently, most research into CLT has focused on investigating instructional techniques that reduce extraneous load to allow an increase in germane resources. Paas et al. (2004) claimed that the source, and not the level, of cognitive load was of greater importance, providing that total cognitive load associated with the instructional design was maintained at a manageable level and did not exceed WM capacity. In particular, if mental load is imposed by the primary factors of learning (such as the construction or automation of schemas), it will have positive effects on learning. In contrast, if mental load interferes with, and is not necessary for learning (because it is extraneous in nature), it will have negative effects on learning (Paas et al., 2004).

In order for instruction to be effective, intrinsic cognitive load (difficulty of a task) should be adequately aligned with the learner's level of expertise so as not to overburden the learner's WM (e.g., ensuring the task is not too complex) or sub-challenge the learner's WM (e.g., ensuring the task is not too easy) (Schnotz & Kürschner, 2007). This was supported by Kalyuga (2011) who stated that intrinsic cognitive load must be managed to an appropriate level (decreased or increased), so that materials are not too complex or too simple in relation to the learner's level of expertise. In particular, extraneous cognitive load must be reduced, or eliminated where possible, to make available more WM resources for dealing with learning activities that require intrinsic cognitive load (Kalyuga, 2011). This implies that minimising extraneous cognitive load to free up cognitive resources for germane activities is the most beneficial way to improve learning, as the engagement in conscious cognitive processing directly pertains to the construction and automation of schemas.

According to the classical view of CLT, an optimal instructional design should reduce extraneous load and increase germane load, and thereupon allow the learner to invest more effort in essential learning processes such as schema construction and automation. Extraneous cognitive load is not necessarily detrimental when intrinsic element interactivity and consequent cognitive load are low (Sweller & Chandler, 1994; van Merriënboer & Sweller, 2005). On the other hand, many studies containing instructional designs to reduce extraneous cognitive load for tasks that are high in intrinsic load have been successful in fostering effective learning environments (e.g., Brünken, Plass, & Leutner, 2004; Cierniak, Scheiter, & Gerjets, 2009; Florax & Ploetzner, 2010; Gerjets, Scheiter, Opfermann, Hesse, & Eysink, 2009; Mayer & Moreno, 2003; Sweller, 1999).

To re-articulate, the central tenet of CLT is to design instructional techniques that decrease extraneous cognitive load and bolster germane cognitive load by not overloading or underloading the available WM capacity (Sweller et al., 1998). Examples include the processing of solution steps in worked examples in more depth using self-explanation (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Renkl, 1997, 1999; Renkl, Stark, Gruber, & Mandl, 1998) and imagination procedures (Cooper, Tindall-Ford, Chandler, & Sweller, 2001; Ginns, Chandler, & Sweller, 2003) by students who hold pre-requisite schemas; eliciting self-explanations to promote the integration of newly learned information with existing knowledge (Chi, de Leeuw, Chiu, & Lavancher, 1994); tailoring the level of instructional guidance to the level of learner experience (Kalyuga et al., 2003; Kalyuga, Chandler et al., 2001); omitting steps (fading) of a worked example until the worked example becomes a conventional problem (Renkl & Atkinson, 2003; Renkl, Atkinson, Maier, & Staley, 2002); combining a fading procedure with self-explanation prompts (Atkinson, Renkl, & Merrill, 2003; van Merriënboer et al., 2006); and removing process information during instruction when learners gain an understanding of the solution procedure and replacing it with product-oriented worked examples (van Gog, Paas, & van Merriënboer, 2008).

In Atkinson, Derry, Renkl, and Wortham's (2000) review of instructional paradigms, the effectiveness of worked-example instruction was examined alongside pure problem-solving practice, in relation to promoting the acquisition of skills and transferable cognitive structures. The conclusion was that learners benefit from worked examples that minimise excessive cognitive load. In particular, novices, who possess inadequate knowledge schemas, are more likely to experience cognitive overload when they are forced to integrate and cross-reference multiple sources of information.

Research efforts have often focused on identifying strategies to reduce extraneous cognitive load because intrinsic load is considered to be unchangeable by instructional design as it is dependent on the number of essential elements of the given task that need to be processed in WM (Sweller et al., 1998; Sweller, 1999). Despite this assumption, there has been limited research focusing on manipulating this task-related load when learning highly complex information (Ayres, 2006; Gerjets et al., 2004; Pollock et al., 2002; van Merriënboer et al., 2003). Gerjets et al.'s (2004) findings revealed that intrinsic cognitive load could be reduced by changing solution procedures from being "molar", which focus on problem categories, to being "modular", which focus on breaking down the solution procedure into smaller, meaningful, comprehensible pieces. Similarly, Pollock et al.'s (2002) *isolated-interactive elements effect* was observed when element interactivity was reduced in the first part of instruction by not presenting the whole information, and in the second part of instruction, by presenting all of the information at once. The results indicated that learning improved for novice learners when they were initially presented with individual elements which were processed serially (instead of simultaneously) as isolated pieces of information, followed by fully interactive material presented during the later stages of learning, which had to be processed simultaneously. This mixed approach produced better understanding in the second phase of instruction, even though understanding may have decreased during the first phase. These recent studies show that intrinsic cognitive load can be artificially reduced by the manipulation of instructional materials.

The distinction between the three categories of cognitive load appears highly reliable due to the myriad of empirical evidence which has identified factors that inhibit learning, alongside strategies that facilitate learning. The key assumption of CLT that learning becomes more difficult when cognitive load increases has continually driven the theory to improve ways of measuring cognitive load.

2.2.3 Subjective measures of cognitive load.

Given the centrality of cognitive load in instructional design theories such as CLT, the measurement of cognitive load has helped CLT researchers understand why the effectiveness of different learning environments may vary as a function of learner characteristics and experimental treatment. Cognitive load was initially conceptualised as a unidimensional concept. It became multidimensional with Sweller et al.'s (1998) distinction between three different sources of cognitive load: intrinsic, extraneous and germane. Following Sweller et al.'s (1998) publication, ongoing research has seen a growing number of ways for measuring cognitive load, especially instruments that enable researchers to measure the changes in different types of cognitive load. However, the most adopted measure is the subjective rating scale, which was originally developed by Bratfisch, Borg, and Dornic (1972). This scale was subsequently modified by Paas (1992), who was the first to introduce a unidimensional nine-point symmetrical category rating scale in the context of CLT, which was later used by Paas and van Merriënboer (1994) and widely adopted by other researchers.

The basis upon which Paas (1992) designed the subjective rating scale was that learners could directly self-evaluate the amount of mental effort they invest in the learning and/or test phases. As an alternative to a subjective rating scale, difficulty scales have also been used where learners are asked to rate the degree of ease or difficulty they experienced in completing a task. Despite measures of difficulty and measures of mental load (cognitive load) being related to some extent, these different tools measure different constructs. For instance, in van Gog and Paas's (2008) review, they stated that very different outcomes and interpretations may arise in an extreme case, for example, when a learner does not invest any effort in completing a task because they perceive it to be extremely difficult.

Subjective measures of difficulty have been used in many studies to find hypothesised differences in cognitive load (to name a few: Corbalan, Kester, & van Merriënboer, 2006; Hummel, Paas, & Koper, 2004; Kalyuga, Chandler, & Sweller, 2000, 2004; Marcus, Cooper, & Sweller, 1996; Paas, Van Gerven, & Wouters, 2007; van Gog et al., 2008). In the present study, a modified version of the nine-point scale developed by Paas & van Merriënboer (1994) was used to measure the perceived amount of difficulty experienced during learning. Participants reported how easy or difficult they found the learning phase, by circling a number in the range from "1" to "9", where "1" represented "Extremely Easy", "3" represented "Moderately Easy", "5" represented "Neither Easy nor Difficult", "7" represented "Moderately Difficult", and "9" represented "Extremely Difficult".

Paas' (1992) subjective rating-scale technique continues to be widely used because it has repeatedly demonstrated internal consistency with high reliability and sufficient face validity (Ayres, 2006; Kester et al., 2006; Paas, Tuovinen et al., 2003; Paas, van Merriënboer, & Adam, 1994). Although subjective measures are generally reliable, valid, non-intrusive, require very little time to complete, and are easy to implement, collect and analyse, the rationale behind using this type of measure hinges on the assumption that learners are able to retrospectively report on the mental effort they experience (Paas et al., 1994). More so, this assumption presupposes the ability of learners to interpret the wording of items in the way intended by the researcher.

Subjective measures, in many cases, are used to obtain information on an individual cognitive load by varying only one constituent aspect of total cognitive load and holding the others constant (Ayres, 2006; Brünken, Steinbacher, Plaas, & Leutner, 2004; Cierniak et al., 2009; Kalyuga, Chandler, & Sweller, 1998). However, in some cases the single item measure has not been matched with significant treatment group differences (Hummel et al., 2004; Kester, Kirschner, & van Merriënboer, 2005). In Kester et al.'s (2005) study on the comparison of two information presentation formats, there were no differences in subjective ratings of mental effort between the split-source format and the integrated format conditions. Kester et al. (2005) argued the likely cause for the same average amount of mental effort was due to the participants in the split-source group dividing their cognitive capacity over learning and mental integration processes. On the other hand, the participants in the integrated group allocated all of their cognitive capacity solely to learning processes, such as general schema construction. This consequently led to better performance on transfer-test performance for the participants in the integrated group.

In recent years, scales have been developed with the intention of measuring intrinsic, extraneous, and germane cognitive loads separately (Cierniak et al., 2009;

DeLeeuw & Mayer, 2008; Leppink, Paas, van der Vleuten, van Gog, & van Merriënboer, 2013). By analysing the effects of different instructional formats, types of cognitive load and learning outcomes, DeLeeuw and Mayer's (2008) study provided empirical support for the dissociation of three types of cognitive load as defined by CLT. Compared to the cognitive load measures used by DeLeeuw and Mayer (2008) and Cierniak et al. (2009), Leppink et al.'s (2013) ten-item psychometric instrument is regarded as being more cogent in differentiating intrinsic, extraneous, and germane cognitive load.

Leppink et al.'s (2013) instrument was subjected to an online pilot study involving students at a Belgian university. It was administered in four studies in a randomised experiment in statistics using different cohorts of students, and different lectures and subject matter, which may have produced confounding. However, the advantage of the instrument was the applicability of the ten items to any complex knowledge domain, with only a minor adjustment being that the word "statistics" (that appeared in some items) had to be replaced if the instrument was used in another domain. The 10-item questionnaire comprised the measurement of intrinsic cognitive load (items 1-3), extraneous cognitive load (items 4-6), and germane cognitive load (items 7-10). Items 2 and 9 referred to formulae, and items 1, 3, 7, 8, and 10 referred to concepts or definitions or topic content. To avoid potential confounding generated by the questionnaire, the items were counterbalanced in three randomised orders (so the participants sitting next to each other did not respond to the same items in the same order).

The realisation of the association between intrinsic and germane cognitive load led Leppink, Paas, van Gog, van der Vleuten, and van Merriënboer (2014) to modify Leppink et al.'s (2013) psychometric scales. Leppink et al. (2014) made the modification by adding one item to each of the three groups of items representing intrinsic, extraneous and germane cognitive load. Leppink et al.'s (2014) modified multi-item scale was consistent with the reconceptualised idea of germane cognitive load; namely, the idea that germane load represents WM resources allocated to dealing with intrinsic cognitive load. However, their findings indicated that further development of instruments for measuring different types of cognitive load was required given how the participants interpreted the wording of items in multi-scale measures. The positive correlation that was found between intrinsic and extraneous cognitive load – which based on CLT should have been close to zero because these independent cognitive loads are additive (Sweller et al., 1998) – led Leppink et al. (2014) to conclude that it was worthwhile investigating, in new tests, the effect of the wording that was used to distinguish the different types of cognitive load.

Paas (1992) pointedly referred to subjective measures of mental effort as "valuable research tools for assessment of cognitive load in instructional research" (p. 433). Moreover, Ayres (2018) built upon this notion by expressing the relativistic nature of one-item scales. By rating mental effort made or difficulty experienced, subjective measures can measure differences in cognitive load rather than measure it in absolute terms (Ayres (2018). In most studies, the advancement of cognitive load measurement techniques has successfully provided empirical evidence for the effect of different instructional interventions on cognitive load (Paas, Tuovinen et al., 2003). However, some studies have found cognitive load differences with no treatment group differences (Kalyuga et al., 1998; Van Gerven, Paas, van Merriënboer, Hendriks, & Schmidt, 2003).

Furthermore, the points where subjective ratings are collected are usually at the discretion of researchers. To establish the viability of rating scales, Schmeck, Opfermann, van Gog, Paas, and Leutner (2015) investigated whether the time of collection of rating scales affected the ratings and if cognitive load measures, using mental effort and difficulty, were suitable predictors of performance. Their findings replicated those of van Gog, Kirschner, Kester, and Paas (2012) who found a delayed, discrete mental effort rating, made after the completion of tasks, was higher than the corresponding average ratings collected immediately after the completion of each task and independent of the sequence of the tasks. Such a discrepancy in ratings indicated the need for further research to find ways of measuring cognitive load with a higher degree of validity. Valid and reliable measures of cognitive load are fundamental in understanding the wide variety of factors affecting cognitive load and learning caused by different instructional manipulations.

2.2.4 Cognitive load effects.

Over the last several decades, empirical results from CLT studies have led to the demonstration of several cognitive load effects. Cognitive load effects have been extensively tested for their effectiveness in optimising cognitive load and improving learning outcomes to facilitate the development of learner expertise. The following sections discuss the worked example effect, the expertise reversal effect, the redundancy effect, the spilt-attention effect, and the variability effect, all of which are of prime concern to the present study.

2.2.4.1 The worked example effect.

A worked example comprises a full sequence of solution steps that demonstrate how to solve a problem or perform a task (Clark et al., 2006).

Learning from worked examples benefits a novice learner who does not possess the relevant knowledge, by directing their focus to elements that solely represent correct solution steps so that they avoid taking irrelevant steps to complete the task at hand (Tuovinen & Sweller, 1999). When a learner is unable to draw on existing, well-developed domain schemas, studying a worked example provides a substitute for missing schemas (Renkl & Atkinson, 2003). This cognitive-instruction technique emphasises the borrowing and reorganising principle (discussed in section 1.3.2) because a novice is able to borrow and recognise information through imitation (e.g., by studying worked-out solutions because they themselves lack well-structured schemas) which helps them develop knowledge in a particular domain (e.g., using the worked-out solution to solve similar problems).

A worked example, as articulated by Atkinson et al. (2000), is an instructional device comprising a problem statement followed by an expert's problem-solving model that can be studied and emulated. This conveys that worked examples are most critical during the initial stages of cognitive skill acquisition in assisting novice learners to acquire problem schemas so that they can solve other similar problems. Notwithstanding that research on worked examples has frequently taken place in controlled laboratory settings, Atkinson et al. (2000) stipulated that findings from laboratory settings can be applied to real classroom settings because of the similarities that exist between the two settings.

As discussed, the limited capacity of WM plays a significant role for novice learners, especially when they attempt to problem solve. Since problem solving requires rapidly sorting through random ideas from an existing knowledge base (Sweller, 2006), novices are usually unable to pay attention to the essential information required for learning new knowledge. As a consequence, novices will address a problem-solving task by searching for solution paths, using a trial-and-error strategy or a means-ends analysis (Kalyuga, Chandler, & Sweller, 2001; Sweller, 1999). Such a condition inhibits learning because it induces high extraneous cognitive load rather than increasing germane WM resources (Sweller & Chandler, 1994). A more effective alternative is worked-example-based learning because novices can focus their WM resources on dealing with intrinsic element interactivity related to learning new knowledge (relating to the solution concept), which helps reduce any unnecessary extraneous cognitive processes (Sweller, 2010).

The use of worked examples as an instructional tool, within the cognitive load theoretical framework, runs contrary to the constructivist views of instruction. According to CLT, studying worked examples is a more efficient way of acquiring complex skills than solving conventional problems. Research on well-designed worked examples has consistently demonstrated that example-based instruction is more effective than problem solving to support initial skill acquisition (see Kirschner, Sweller, & Clark, 2006, for a review). The worked example effect occurs when learners who study worked examples perform better on subsequent test problems compared to learners who attempt to solve the equivalent problems.

The worked example effect, from a cognitive load perspective, was first described by Sweller and Cooper (1985), and Cooper and Sweller (1987) in the area of mathematics. It has since been replicated by many empirical studies using a large variety of learners in scientific-based domains, such as statistics, computer programming, physics, and engineering, and in arts-based domains, such as foreign language acquisition, athletics, music instruction and English literature (see Atkinson et al., 2000; Sweller et al., 1998; van Gog & Rummel, 2010, for reviews).

Consequently, Sweller (2006) argued that the worked example effect is the best-known and extensively studied cognitive load effect.

Most of all, worked examples are designed to reduce extraneous cognitive load (by scaffolding at the beginning of skill acquisition) to allow novice learners to allocate more WM resources to acquire relevant knowledge (Nievelstein, van Gog, van Dijck, & Boshuizen, 2013; Paas, 1992; van Gog et al., 2006). Using three computer-based training strategies (conventional problems, worked-out problems, and partly worked-out problems), Paas (1992) found that instruction which focused attention on partly or completely worked-out problems resulted in better performance and a lower perceived mental effort in a transfer test, with the completely worked-out problems condition attaining the lowest time on training. Exposure to concrete schemas enabled the participants in the worked-out conditions to invest their limited WM resources in relevant learning processes (such as appropriate schema abstraction). In contrast, cognitive capacity that was allocated to the construction of incorrect solution procedures (during means-ends analysis), by the participants in the problem-solving condition, resulted in WM resources being unavailable for schema acquisition (Paas, 1992). In a similar way, van Gog et al.'s (2006) investigation demonstrated the basic worked example effect. Senior secondary electrotechnics students whose training consisted of worked examples performed better in near and far transfer performance, with less mental exertion and investment of time (in training and the test) than students whose training consisted of solving conventional problems (van Gog et al., 2006).

Although learning from worked examples is a widely used cognitive load-reducing technique, worked examples lose their effectiveness when cognitive capacity is allocated to an activity that is unrelated to schema construction. Tarmizi and Sweller (1988), using geometry, and Ward and Sweller (1990), using physics, confirmed that worked examples were ineffective when they required the mental integration of mutually referring information in order to be understood (e.g., a diagram with an accompanying textual explanation). When neither source of information is intelligible on its own and meaning can only be extracted by matching the sources of information, cognitive resources are allocated to an activity that is unrelated to schema construction. When such an activity is present, the consequences for learning are the same as problem solving. The inappropriate allocation of resources because of the way the material is designed is discussed in more detail in section 2.2.4.3 (the redundancy effect) and in section 2.2.4.4 (the split-attention effect).

Some worked-example-based learning can reduce the potential for conceptual understanding if a worked example fails to provide the learner with the opportunity to integrate their existing knowledge with new principle knowledge. If a worked example does not highlight the main solution concept, the learner will likely gloss over the worked-example material, rather than attend to the new knowledge (Atkinson et al., 2000; van Gog, Paas, & van Merriënboer, 2004). In such cases, the learner will likely perform poorly on far transfer tests because the worked-example-based learning is considered to be a passive learning activity (Moreno, 2006).

Employing example-problem pairs, in which a worked example is immediately followed by a similar problem to solve, is a traditional method that has been used in cognitive load research (Carroll, 1994; Cooper & Sweller, 1987; Kalyuga, Chandler et al., 2001; Sweller & Cooper, 1985). There is research to suggest that studying examples is more effective when they are sequenced such that they are immediately followed by a similar problem. Trafton and Reiser (1993) showed that the most efficient way to present learning material to acquire a skill was to link the worked example to a similar target practice problem rather than present an entire set of worked examples followed by an entire set of practice problems. This method of alternating worked examples and problem solving enabled the learner to derive the full benefit of using applicable prior information from the example (source) to build rules for the problem (target) to be solved.

When a learner is aware that they will solve a problem immediately after studying a similar worked example (example-problem pair instruction), they are likely to study the original example with more incentive. Contrary to problem-solving-only instruction, example-problem pair instruction may strengthen the learner's knowledge acquired from the worked example (Sweller et al., 2011). Consequently, the example-problem pair condition engenders lower cognitive load and generates superior post-test performance of learners compared to the equivalent problem-solving-alone condition (Leppink et al., 2014). Moreover, van Gog (2011) showed that providing novice learners with example-problem pairs was more effective for learning than providing problem-example pairs. Van Gog's (2011) experiment demonstrated that the example-problem pair instruction required less mental effort (in both the learning tasks and transfer post-test) and outperformed the problem-example pair instruction in the transfer post-test, despite the identical learning instructions provided to both instructional conditions. These findings confirmed that initial exposure to worked examples enabled novices to acquire cognitive schemas which assisted with subsequent problem solving.

To refine the CLT explanation of the worked example effect, van Gog, Kester, and Paas (2011) found that the examples-only and the example-problem pairs conditions led to better post-test performance than the problem-solving only and problem-example pairs conditions. Van Gog et al. (2011) established that their findings aligned with the CLT view that worked examples were more efficient (with lower investment of mental effort during the training) and more effective (with higher performance) when provided before problems. This was further exemplified with no significant difference in the test performance between the example-problem pair and the examples only condition (van Gog et al., 2011). This finding by van Gog et al. (2011) was replicated by van Gog and Kester (2012) who demonstrated that the example only and the example-problem pair conditions were equally effective in an immediate post-test.

Given worked examples entail specific solution procedures, Atkinson et al. (2000) pointed out that critics did not view worked examples to be ideal for adaptive learning because exposure to specific procedures made it difficult for learners to solve novel problems which deviated from worked examples. This explanation was evident in Carroll's (1994) discussion with his participants, who reported experiencing difficulties with transfer even with slight differences between worked examples and the practice problems. To foster initial cognitive skill acquisition, Renkl, Atkinson, and Große (2004) claimed it was more effective to use a series of worked examples prior to solving problems.

Gradually fading out worked-solution steps has been shown to generate better learning outcomes compared to the traditional technique that involves example-problem pairs (Renkl et al., 2002; Atkinson et al., 2003). The completion or fading strategy considers the learner's increasing knowledge level of a task by allowing the learner to smoothly transition from example study (during the early stage of skill acquisition) to working on incomplete examples to solving problems (during the later stage of skill acquisition). This effective delivery strategy for learning does not require learners to work out task-specific solutions during the early stages of learning, thus assisting learners with retaining sufficient cognitive capacity to focus on deepening their understanding. This strategy is consistent with findings from Renkl et al.'s (2004) extensive controlled experiments which were conducted in both the classroom and laboratory. Recognising that previous studies on fading out worked-solution steps did not investigate how sequential steps of fading affected learning outcomes, Renkl et al. (2004) conducted further research by examining the series of solution steps across two experiments.

In Renkl et al.'s (2004) study, the findings from Experiment 1 confirmed that learners gained knowledge from the particular type of solution step that was faded rather than the position of the solution step that was faded (backward or forward). In Renkl et al.'s (2004) Experiment 2, which looked at the learning processes more directly with the collection and analyses of thinking-aloud protocols, it was revealed that learners using the fading procedure experienced fewer unproductive impasse-triggered events and more productive learning events compared to those encountered by learners in the example-problem pairs group. These findings show that fading out worked examples, by gradually omitting solution steps, facilitates a smooth transition from learning from a complete worked example, in the earlier stages of skill acquisition, to an incomplete example and finally to conventional problem solving in the later stages. This strategy of providing learners with complete worked examples during the beginning stages of cognitive skill acquisition, to assist them to gain a deeper understanding of domain principles, suggests that effective provision of worked examples are related to the expertise level of the learner.

2.2.4.2 The expertise reversal effect.

CLT provides a framework for conceptual and practical explanations about why instructional design needs to be tailored to the knowledge level of the intended learner (see Kalyuga, 2007; Kalyuga et al., 2003; Kalyuga et al., 2012; Kalyuga & Sweller, 2005; Paas, Renkl, & Sweller, 2003, for reviews). According to Kalyuga et al. (2003), choosing the most appropriate teaching approach depends on what stage the learner is at. Explicit teaching during the early stages of learning is effective as a means of reducing the cognitive load on students. Once learners acquire sufficient knowledge and develop automated skills, they become capable of engaging in relevant problem solving. As mentioned on several occasions, experts are able to bypass their WM capacity limitations because they are able to identify a recognizable arrangement of multiple elements of information as a familiar schema. As learners advance beyond novice status, they may still require some form of guidance to reduce or minimise cognitive load during their learning process. Consequently, it is vital for instructors to provide opportunities for learners to assimilate their prior knowledge with new knowledge by assisting learners to acquire the appropriate skills needed to comprehend the material. The level of prior knowledge in the domain will affect the rate at which problems are solved.

In cognitive psychology, expertise is generally defined as the possession of a large amount of available and applicable complex knowledge in a domain. The novice-expert difference, according to Chase and Simon (1973a) in their chess-related research, is based on the size of the chunks each recalled in relation to a sequence of moves, where a “chunk” is defined as a unit of knowledge structure. Beyond the fact that experts possess more higher-order chunks which develop with increasing skill, Chi, Glaser and Rees (1982) commented that a salient difference between novices and experts is the rapidity with which a solution is applied during problem solving.

In their findings, Chi et al. (1982) alluded to an important difference between expert learners' knowledge and novice learners' knowledge. Experts' schemas contain more procedural knowledge – knowing how to complete a task by processing or manipulating a knowledge structure; as opposed to novices' schemas which contain declarative (conceptual) knowledge – knowing facts about a learning task but lacking procedural skills which diminish the ability to learn from existing knowledge. The conclusion drawn from Chi et al.'s (1982) study was that the novices' schemata was "impoverished" and this "[could] seriously hinder their problem-solving success" (p. 62), and that novices had a "limited ability to generate inferences and relations not explicitly stated in the problem" (p. 68). These statements explain why novices lack the ability to abstract pertinent knowledge from relevant cues in a problem, and thus provide support for the premise that procedural and declarative knowledge are required for transferring knowledge and skills from one task to a second, target task.

Further research suggests that as learners develop their schematic knowledge in a domain, their problem-solving abilities improve. Experts are able to categorise new problems by abstracting underlying structural features as belonging to a distinct category of problems which require specific operations to attain a solution (Paas, 1992; Quilici & Mayer, 2002). It is clear from the large volume of published studies that acquired schemas can equip experts with analogies in novel problem-solving situations because domain specific knowledge, in the form of schemas, can be used to map processes to obtain solutions for unknown parts of a problem-solving task (Plass, Moreno, & Brünken, 2010).

Research in domains such as mathematics (Schoenfeld & Herrmann, 1982; Silver, 1981) and physics (Chi et al., 1981; Hardiman, Dufresne, & Mestre, 1989) indicates that while experts categorise problems based on structural features, novices categorise problems based on surface features. However, it cannot be assumed that experts focus exclusively on deep structures of problems (by applying the most appropriate principle(s) and methods of solutions), and novices exclusively focus on surface characteristics (by relying on vocabulary and equations of similar, specific problems) for problem solutions. Rather, findings support the view that learners who are able to organise knowledge in terms of general principles, are likely to solve problems more effectively compared to learners who use surface features. This suggests a causal relation between using principles and better problem solving.

When an instructional technique that is effective for less-experienced learners loses its effectiveness, or even becomes ineffective, if used by more-experienced learners, this phenomenon is known as the "expertise reversal effect" (see Kalyuga, 2007, for an overview of the expertise reversal effect). Kalyuga et al. (2012) clear up any potential misunderstanding of the term "reversal" by stating it refers to a reversal in the relative effectiveness of the instruction and not a reversal in relative performance.

Two types of imbalances in cognitive processing underlie the expertise reversal effect: the deficiency of relevant knowledge, and the overlapping of relevant knowledge. When there is an absence of relevant knowledge, especially during the initial phase of learning, less-experienced learners require externally guided information to help them substitute missing, relevant knowledge to build schemas in an efficient way. Dealing with new elements of information without any external instructional support is likely to cause cognitive overload for novices as they randomly engage in an unguided search for general solution strategies. This inefficient process, which compensates for a novice's limited knowledge base, results in the unnecessary consumption of most of their cognitive resources. In contrast, in the same problem situation, when a more-experienced learner is forced to integrate and cross-reference externally guided information with their available cognitive schemas in WM, this redundant activity imposes a higher cognitive load than problem solving (Kalyuga et al., 2003).

Over the past several decades, the expertise reversal effect has been regarded as a well-established empirical phenomenon. Kalyuga (2007) illustrated this by describing multiple empirical investigations that credibly generated the expertise reversal effect across a large range of instructional materials and participants. Additionally, the cognitive mechanisms in van Merriënboer et al.'s (2003) design model reflected the expertise reversal effect. Their design model encapsulated the importance of cognitive load aspects associated with complex skill acquisition by controlling two forms of scaffolding related to levels of learner expertise: intrinsic factors (e.g., scaffolding tasks by using simple-to-complex sequencing), and extraneous factors (e.g., using a fading procedure which involves initially providing learners with worked examples, followed by completion tasks, and finishing with conventional problems). By preventing a heavy cognitive load via the reduction in intrinsic and extraneous cognitive load, this design model makes it possible to present learners with real-life tasks.

Much research has supported the expertise reversal effect, by demonstrating that for instruction to be efficient and for learners to gain optimal benefits, instructional techniques and procedures must be tailored to different levels of expertise (Clarke et al., 2005; Kalyuga et al., 2003; Kalyuga et al., 1998, 2000; Kalyuga, Chandler et al., 2001; Renkl, 1997; Renkl et al., 2004; Renkl et al., 2002; Yeung, Jin, & Sweller, 1998). In particular, Renkl and Atkinson (2003) showed how their fading procedure was able to build a bridge between studying worked examples during the intermediate phase of cognitive skill acquisition, and problem solving in the later phase. Transitioning from example study to problem solving was achieved by giving considerable instructional support to novice learners for the initial learning tasks and then no support for the final tasks (Renkl & Atkinson, 2003).

Other examples of empirical studies demonstrating the expertise reversal effect include the following. Blayney et al.'s (2010) examination of accountancy students' performance on post-session test questions revealed novice learners benefited most from studying isolated elements instructions and more expert learners benefited most from fully interacting elements instruction. Reisslein, Atkinson, Seeling, and Reisslein's (2006) example-based instructional designs, in a computer-based learning environment, found low prior knowledge participants benefited most from example-problem instruction and high prior knowledge participants benefited most from problem-example instruction. Oksa, Kalyuga, and Chandler's (2010) study, using intrinsically difficult Shakespearian play extracts, showed students who possessed no prior knowledge of the text performed better in the explanatory notes group (which integrated Modern English explanatory interpretations for every line of the original, complex Elizabethan English text), while Shakespearian experts performed better in the control group (which used the original, conventional text with no guidance).

Overall, it can be concluded from the research conducted on the expertise reversal effect that guided instructional support, which is optimal for novice learners, is redundant and even detrimental for more expert learners. Kalyuga (2007) exemplified this by claiming the expertise reversal effect is an example of the redundancy effect in a broader sense. Expressed another way, a version of the redundancy effect refers to different sources of external information covering the same area, whereas the expertise reversal effect refers to the overlapping of internal knowledge structures with external information covering the same area.

2.2.4.3 The redundancy effect.

Cognitive load theorists refer to the term "redundancy" when unnecessary information is added to essential information. Sweller (1999) claimed learners are forced to process redundant material, especially when it is integrated with essential material, because it cannot be ignored. As a consequence, this unnecessary processing imposes extraneous cognitive load. Despite this being a cognitively demanding process, it may not have a negative impact on learning if the redundant material has low element interactivity. In contrast, processing redundant material is more likely to impede meaningful learning if the redundant material has high element interactivity and the combined high extraneous and high intrinsic cognitive load overwhelms WM's limited processing capacity (Sweller, 1993, 1994). When instructional material is fully intelligible in isolation, any additional information which duplicates the material in a different format should be removed. The redundancy effect refers to the phenomenon in instruction when the inclusion of redundant, additional information interferes rather than enhances learning.

The redundancy effect has been demonstrated in studies dating back to the 1930s (see Sweller, 1993, for a brief history). In the past several decades, studies have continued to investigate this phenomenon. Ward and Sweller (1990) found the inclusion of additional explanatory text into worked examples was redundant, and needed to be omitted because it interfered with, rather than assisted, learning. Chandler and Sweller (1991) showed integrated instruction to be more favourable only in situations where different sources of information had to be mentally integrated in order to be understood. However, when integrated instructional material contained redundant information, the physical integration of unnecessary information with information that was intelligible on its own was not beneficial to learning (Chandler & Sweller, 1991).

Additionally, the experiments conducted by Mayer, Bove, Bryman, Mars, and Tapangco (1996) provided evidence that suggested a lengthy text explanation was less efficient than a multimedia summary (combining visual and verbal formats). The multimedia summary was found to promote better understanding of a scientific explanation because the smaller amount of text (based on three criteria: conciseness, coherence and coordination) reduced the burden on the cognitive system and consequently promoted better retention and transfer compared to the full text. Also, in Kalyuga et al.'s (2004) study involving technical apprentices within realistic training facilities, it was suggested that the concurrent presentation of written and spoken text (compared to nonconcurrent presentation, exclusive presentation of written text, or exclusive presentation of auditory text) was beneficial to learning if both modes of textual information were required to be integrated in order to be understood.

According to Kalyuga's (2012) evaluation of the potential benefits of instructional presentations delivered through the auditory modality (e.g., spoken words), visual modality (e.g., printed words, on-screen words, animations), or dual-modality (employing both auditory and visual sensory modalities), the notion of redundancy is also dependent on learner expertise, due to the expertise reversal effect (previously discussed in section 2.2.4.2). Kalyuga (2012) clarified this by pointing out that accompanying spoken words which assist a novice to comprehend a diagram (because both sources of information are relied upon for a better understanding) may be redundant for a more expert learner, who understands the diagram on its own (because both sources of information are intelligible on their own). The latter situation demonstrates the redundancy effect whereby the connection between the redundant verbal explanation and already available schemata does not improve learning, but instead requires additional WM resources for processing, and accordingly results in avoidable increased cognitive load.

Liu, Lin, Gao, Yeh, & Kalyuga (2015) were the first to explore the occurrence of the redundancy effect in a virtual classroom which simulated real classroom conditions, using three experimental conditions: audio-visual, audio only and visual only. Liu et al. (2015) obtained a robust reverse redundancy effect and claimed this was due to two possible factors: the presentation of segmented information, and interference in the classroom. Segmentation resulted from the use of common everyday language in spoken form, which imposed a low intrinsic cognitive load, with the provision of written factual information as a visual back-up. Correspondingly, the spoken narrations, supplemented with on-screen textual information, outperformed the other presentation formats that contained either spoken or written form. Liu et al. (2015) concluded that interference or cognitive overload, replicated from real classroom conditions, which resulted in students losing fundamental information from either source (written on-screen text or narrated text) could be easily replaced by the other source. Liu et al.'s (2015) findings demonstrated that the simultaneous provision of two forms of the same information had a complementary effect rather than a redundancy effect. Any lost information from one source was backed by the retrieval of information from the other source, which promoted better learning. Liu et al.'s (2015) study laid the groundwork for future research to consider classroom interference as a significant cause of heavy cognitive demands when students rely on one source of information for learning.

When learners split their attention between different sources of information, for example, within a visual modality containing a diagram and separated explanatory text, the process of mentally integrating the sources unnecessarily increases cognitive load. Kalyuga et al. (1998) clarified this further by distinguishing between sources of information that are intelligible in isolation (e.g., when text explaining a diagram is redundant) and sources of information that cannot be understood in isolation (e.g., when the diagram requires additional, possibly textual, explanatory information). Kalyuga et al. (1998) postulated that a redundancy situation occurred when the sources of information were intelligible in isolation, and a split-attention situation occurred when the sources of information could not be understood in isolation.

2.2.4.4 The split-attention effect.

When information is in a split-source format, disparate sources of visual information require constant changes in the focus of attention to and from each source; for example, diagrams and text, solely textual information, solely diagrammatic information, or text and equations. In a split-attention situation, learning is likely to be interfered with because of the intensive search-and-match process that is required for mental integration. The split-attention effect arises from the advantage of presenting multiple sources of information in an integrated format so the learner is presented with a single source of information (see Sweller et al., 2011, for an overview). In an example of the expertise reversal effect, it was suggested by Yeung et al. (1998) that the efficacy of an integrated format may be moderated by the degree of learner expertise. In Yeung et al.'s (1998) investigation, presenting additional information (e.g., vocabulary definitions and explanatory notes) in an integrated form either facilitated or interfered with performance through a split-attention effect or a redundancy effect respectively.

Multiple studies in split attention within a cognitive load framework found that students learned more effectively when they were not required to simultaneously mentally search and physically integrate disparate sources of information (Cerpa, Chandler, & Sweller, 1996; Chandler & Sweller, 1991, 1992; Sweller & Chandler, 1994; Sweller, Chandler, Tierney, & Cooper, 1990; Tarmizi & Sweller, 1988; Ward & Sweller, 1990). Tarmizi and Sweller (1988), and Ward and Sweller (1990) showed how converting ineffective worked examples which imposed a heavy, extraneous cognitive load because students were required to mentally integrate disparate sources of information, to reformatted worked examples which reduced multiple sources of mutually referring information, resulted in the reintroduction of the worked example effect; by not having to integrate material into a single source, cognitive load reduced and learning from integrated instructions was superior to conventional instructions.

Using engineering and high school coordinate geometry materials, Sweller et al. (1990) found that eliminating split attention by combining the mutually referring units of diagrammatic and written material into a unitary source of information substantially enhanced performance. Similarly, Chandler and Sweller (1991, 1992) demonstrated the spilt-attention effect, showing integrated instructions outperformed conventional instructions in which mutually referring information was separated and so had to be mentally integrated because it was unintelligible in isolation. Additionally, the results from Sweller and Chandler's (1994) experiments showed that a self-contained, modified manual format (having no contact with the apparatus) was superior to the other formats because, despite its high level of element interactivity, extraneous cognitive load was reduced by controlling split-attention.

The findings from Ginns' (2006) review revealed the benefits of reducing split attention between spatially or temporally disparate, related elements of information. By synthesising the results of fifty independent, instructional design experimental studies, Ginns' (2006) meta-analysis showed that learning was more efficient and effective when related elements of information were integrated over space (the spatial contiguity effect, also known as the split-attention effect) or over time (the temporal contiguity effect). A major limitation of the generalisability of Ginns' (2006) meta-analysis is that the results came from experimental studies using learners who were classified as novices and materials that were predominantly high in element interactivity.

To test the interaction of element interactivity with the split-attention effect, Chandler and Sweller's (1996) study, which introduced learners to a new computer application, showed that using low element interactivity materials produced non-significant differences between instructional formats. These results indicated that extraneous cognitive load, imposed by a sub-optimal instructional design containing split attention, only became critical with high element interactivity instructional material. Otherwise, the combination of increased extraneous cognitive load imposed by a sub-optimal instructional design containing split attention and decreased intrinsic cognitive load from low element interactivity material was less likely to overload limited WM resources, leaving adequate WM resources available for schema construction and possibly automation (Chandler & Sweller, 1996).

2.2.4.5 The variability effect.

Research has consistently revealed that multiple examples are more effective than one example for promoting learning (Cooper & Sweller, 1987; Namy & Gentner, 2002; Rittle-Johnson & Star, 2007, 2009; Silver, Ghousseini, Gosen, Charalambous, & Strawhun, 2005; Sweller & Cooper, 1985). Studies have also suggested that multiple examples with the same solution structure and different surface characteristics can foster the acquisition of transferrable knowledge (Catrambone & Holyoak, 1989; Gick & Holyoak, 1983; Quilici & Mayer, 1996). This was exemplified in the work undertaken by Pirolli and Anderson (1985) who found that novice learners were able to solve novel and difficult problems by finding analogies between examples. The results of Holyoak and Koh's (1987) study suggested that transfer could be obtained if the surface features of the source had at least one salient similarity to the features of the target problem.

Additionally, Corbalan, Kester, and van Merriënboer (2011) found that a learner-controlled selection of tasks which differed in their surface features yielded beneficial effects on learning through enhanced transfer test performance, compared to a learner-controlled selection of tasks which differed in their structural features. Since surface features are salient for both novices and experts, Chen and Mo (2004) also found that when tasks differed in their surface features, this helped learners recognise the same solution steps which then improved the construction of broader and flexible schemas. This process, referred to as generalization, enhances schema induction because tasks that are performed in the same way assist learners to see beneath the surface features, and thus create generalised rules for a wider class of related tasks (Gick & Holyoak, 1987).

Furthermore, Reeves and Weisberg (1994) highlighted the need for learners to be explicitly instructed to compare multiple examples with respect to their similarities if the instructional goal is schema construction. This is because the acquisition of schemas from diverse examples was driven by the identification of structural features. Additionally, Gentner and Namy (1999) found that comparing examples helped to highlight a common structure which encouraged conceptual learning. However, not all comparisons may be equally effective. It is well established from a variety of studies that the effectiveness of comparing multiple examples is dependent upon the type of variability of the examples being compared (Paas & van Merriënboer, 1994; Quilici & Mayer, 1996; Renkl et al., 1998).

To better understand the effects of variability, it is important to recognise that much of the literature highlights that examples are generally analysed in terms of surface and structural features (Holyoak & Koh, 1987; Paas & van Merriënboer, 1994; Quilici & Mayer, 1996; Reed, 1989). In line with these two distinct features, the present study distinguished between low-variability tasks (such as worked examples and problem solving) that varied only in surface features, and high-variability tasks that varied in structural features. The low-variability tasks in the present study varied by way of numbers and mathematical terms (i.e., the surface features changed, and the structure remained the same between tasks) and the high-variability tasks varied by way of rules and procedures, as well as numbers and mathematical terms (i.e., the structural features changed between tasks).

Several studies have attempted to explain how students learn to categorise, based on the surface and structural features of tasks, to form generalised problem schemas. Quilici and Mayer’s (2002) study, which investigated word problems in statistics, revealed that direct exposure to structure-emphasising examples enabled students to abstract underlying structural features of a problem statement which led to the successful categorisation of elementary statistics problems. Quilici and Mayer (2002) confirmed that increasing structural awareness allowed students to more successfully engage in creative problem solving after feeling assured that a statistics problem did not fall within a pre-existing category. Correspondingly, Quilici and Mayer (1996) found that structure-emphasising example problems improved students’ performance in problem categorisation within the domain of statistics. By comparing two groups, where one group studied worked examples with varied surface features and the second group studied worked examples with varied structural features, they revealed that the latter group was better at sorting task performance (Quilici & Mayer, 1996). This indicates that the process of schema construction is not an automatic process that occurs when learners are presented with multiple examples that emphasise the same surface instruction. Rather, multiple tasks that emphasise the same structural characteristics enable the learner to distinguish features so that they are able to recognise how one problem may be similar to another. By extracting constant, underlying solution rationales, the learner is able to construct abstract solution schemas that can facilitate accurate transfer schemas to tasks that contain different surface features.

Despite their promising results, Quilici and Mayer (1996) recommended that further research was required to determine whether guided instruction would augment the effectiveness of structure-emphasising examples. Accordingly, Renkl et al. (1998) found that confronting learners with highly variable examples without instructional support (in the form of self-explanation) did not significantly influence near or far transfer, which led to poor learning outcomes. However, Paas and van Merrienboer's (1994) noteworthy study was the first study of variability from a cognitive load perspective to reveal that exposure to a highly varied sequence of worked examples yielded better schema construction and transfer, compared to exposure to conventional problems that had to be solved, followed by worked-example study. Similarly, van Merriënboer and Sweller (2005) postulated that by increasing variability, learners are exposed to a greater range of tasks which facilitates the differentiation between relevant and irrelevant features of worked examples. By assessing new tasks in light of any previous tasks, this process promotes the abstraction of schemas into LTM, so learners know when to apply the concept, thus boosting transfer performance (Clark et al., 2006). Increasing the variability of tasks increases the number of interacting elements related to the task. The effectiveness of this increased element interactivity gives learners the opportunity to invest germane cognitive load by engaging in deeper processing.

In contrast, some findings on the effect of learning from multiple examples have shown that high-variability examples make it difficult for the structural features to be identified which then reduces the likelihood of inducing a schema (Gentner & Namy, 1999; Namy & Gentner, 2002; Richland, Holyoak, & Stigler, 2004; Ross & Kennedy, 1990). However, other findings have shown that low-variability (similar) examples interfere with schema formation because learners are more likely to regard surface features as more important (Paas & van Merriënboer, 1994; Quilici & Mayer, 1996; Ranzijn, 1991). With these contradictory findings in mind, Renkl et al. (1998) hypothesised that the positive impact of high variability in worked examples is dependent on two factors: the type of learning outcome desired; and the prior knowledge in the domain. In relation to learning outcomes, Renkl et al. (1998) considered that high-variability worked examples are particularly important if far-transfer tasks need to be solved. Otherwise, presenting learners with similar worked examples (where solution methods do not need to be modified) are sufficient if near-transfer tasks need to be solved, despite the likelihood of redundant information. With respect to prior knowledge in the domain, Renkl et al. (1998) asserted that more-experienced learners in a domain benefit from studying high-variability examples because they are not overloaded by the complexity of structurally different examples. On the other hand, less-experienced learners are overburdened by high-variability examples because they find it difficult to detect any structural similarities across examples. Renkl et al.'s (1998) hypothesis in relation to prior knowledge in the domain was investigated in Experiment 3 of this thesis.

In reviewing the literature, relatively few studies have been found on the question of how learners' prior knowledge interacts with the variability of examples. Some research has established that learners with less prior knowledge in a domain do not gain from comparing multiple examples, particularly complex examples (Gentner, Loewenstein, & Hung, 2007; Holmqvist, Gustavsson, & Wernberg, 2007; Schwartz & Bransford, 1998). While research performed by

Quilici and Mayer (1996) opposed this by showing that lower-ability learners gained from comparing very different (structure-emphasising) examples and higher-ability learners gained from comparing all types of examples. More specifically, Quilici and Mayer (1996) established that lower-ability learners naturally paid attention to surface features (unless they were conditioned to focus on structural features), whereas higher-ability learners tended to focus on solution procedures that were based on structural features, irrespective of being exposed to a surface-emphasising or structure-emphasising instructional condition.

Previous research has established that novices in a domain effectively learn a new solution method when they compare multiple similar examples of the same solution method which they can easily align to solve transparent problems (Gentner et al. 2007; Kotovsky & Gentner, 1996). This accords with Rittle-Johnson, Star, and Durkin's (2009) findings that novice learners in a domain built competence if they were assigned to learn a single solution method (by comparing problem features of multiple examples) before comparing different solution methods from the start. Of the 236 seventh- and eighth-grade mathematics students who participated in Rittle-Johnson et al.'s (2009) study, students with sufficient prior knowledge of solving algebraic equations benefitted most from comparing different solution methods for solving the same equation, as opposed to students who did not commence the study with equation-solving skills. This suggests that less-experienced learners can build competence by familiarising themselves with one solution method before they compare alternative solution methods.

In their examination of schema construction and transfer of learning, van Merriënboer and Ayres (2005) affirmed that the variability of problem situations increases the likelihood of the learner distinguishing between relevant and irrelevant features, as well as recognising similar features. As a consequence, the intrinsic complexity of a high-variability task requires the learner to “invest more effort in genuine learning” (van Merriënboer & Ayres, 2005, p. 7). One way to make cognitive resources available for genuine learning is to diminish extraneous cognitive load, so this suggests that example variability can be employed as an instructional technique to assist with substituting extraneous load with intrinsic load. However, if the complexity of the learning tasks remains excessively high after the extraneous load is minimised, van Merriënboer and Ayres (2005) advised that simplification of learning tasks may be required during the early stages of learning. This is in line with CLT where it is well established that increasing the complexity of a task increases the number of interacting elements associated with the task. Thus, increasing task variability by exposing the learner to a greater range of tasks, increases element interactivity which in turn increases intrinsic cognitive load, which can potentially overwhelm WM capacity.

In chapter four of their book, van Merriënboer and Kirschner (2018) notably discuss how the variability of practice, through which the use of a varied set of learning tasks representative of real-life tasks, supports the construction of new cognitive schemas and the modification of existing ones in WM. Additionally, in relation to designing a sequence of learning tasks, van Merriënboer & Kirschner (2018) described the importance of *scaffolding*, a process that entails the provision of optimal level of support and guidance to learners during the initial stages of learning so they are able to carry out the required steps, and the fading away of support and guidance when learners become capable of carrying out the required action. Scaffolding is necessary because of the expertise reversal effect, on the grounds that instructional techniques that are effective for novices are likely to lose their effectiveness or even become detrimental when used with more-experienced learners (Kalyuga et al., 2003; Kalyuga et al., 2012; van Merriënboer et al., 2003).

Drawing on research undertaken on task variability, the present study attempted to provide empirical evidence of the impact of exposing learners to high- or low-variability learning material using either worked examples or problem-solving tasks, on test questions that were similar in context or dissimilar in context (transfer tasks) to the learning material (in Experiments 1, 2 and 4). Additionally, the impact of exposing more-experienced or less-experienced learners to high- or low-variability learning material, on test questions that were similar or dissimilar in context to the learning material was also investigated (in Experiment 3).

2.3 Summary of Chapter 2

This chapter established the theoretical framework of CLT and discussed the wide benefits of its use for instructional design in education. A review of cognitive load was examined with respect to the identification of intrinsic, extraneous and germane cognitive load; managing cognitive load through different instructional designs to optimise learning; and subjective measures of cognitive load, as a tool for assessing the mental effort made or level of difficulty experienced by learners during instruction. Finally, there was an examination of five instructional design effects: the worked example effect, the expertise reversal effect, the redundancy effect, the split-attention effect, and the variability effect. All of these relate to the present study. They were reviewed in detail to clarify how they can influence cognitive load and learning outcomes. Specifically, the chapter discussed empirical research relating to the effectiveness of adjusting instructional design methods for learners under worked-example instruction or problem-solving conditions, and learners' levels of prior knowledge in a domain.

The next chapter, which initiates Part II of this thesis, will provide the rationale for the dependent and independent variables, and the theoretical arguments for the cognitive load effects used to develop the hypotheses for the present study.

PART II: EMPIRICAL STUDY

Chapter 3: Design of Empirical Study

3.1 Introduction to Chapter 3

By synthesising the information presented in the previous chapters, this chapter will discuss the dependent and independent variables used to formulate the hypotheses in all four experiments of the present study. In addition, fundamental cognitive load effects – the worked example effect, the variability effect, and the expertise reversal effect – will be discussed in relation to the development of these hypotheses. Particular reference will be made to Paas and van Merriënboer's (1994) study, which provided a major impetus for the present study.

3.2 Dependent Variables

Transfer of learning and subjective ratings of difficulty were the two dependent variables that were measured in the present study. Given their significance for measuring post-test performance and cognitive load respectively, they are reviewed in detail in this section.

3.2.1 Transfer tasks as measures of post-test performance.

3.2.1.1 Overview of transfer of learning.

The fundamental aim of learning is to acquire the ability to judge which skill and knowledge to use in a new context by generalising from past lessons and specific experience (McKeough, Lupart, & Marini, 2015). This ability to spontaneously adapt what has been previously learned is known as transfer of learning. Along the same lines, Brooks and Dansereau (1987) claimed that effective transfer from one task to another occurs when an individual is able to determine which skills and knowledge to apply to a second, target task. They clarified this by arguing that "skills may be substantially content independent and, thus, may be a part of the individual's repertoire of general learning and problem-solving strategies" (Brooks & Dansereau, 1987, p. 125). If these skills are non-existent and "compensatory guidance through the transfer task is not provided by an instructor, supervisor, or someone in a similar capacity" (p. 125), Brooks and Dansereau (1987) claimed that transfer will not take place because the gap between an individual's existing knowledge and the target task cannot be bridged.

Over the last century, educational psychologists have developed several different theoretical perspectives about how to best design educational programs to facilitate transfer. However, educational transfer remains a complex phenomenon (Barnett & Ceci, 2002; Cormier & Hagman, 1987; Gick & Holyoak, 1983, 1987; Hayes & Simon, 1977; Holyoak & Koh, 1987; Nickerson, Perkins, & Smith, 1985; Thorndike & Woodworth, 1901a, 1901b, 1901c). For example, Nickerson et al. (1985) acknowledged that transfer may occur spontaneously. However, it is not always possible for students to directly carry over skills they have acquired in one context to different contexts due to the "significant adjustment" (p. 335) required which creates "another barrier in the way of smooth transfer" (Nickerson et al., 1985, p. 335). Developing a clear understanding of the principles of transfer continuous to be a principal focus amongst educational psychologists since effective transfer is of practical importance in educational settings.

Despite the challenge of designing instruction to effectively teach for transfer, several educational theorists have successfully reported on different ways that transfer can be achieved. Thorndike (1913) posited that increasing the similarity of elements between tasks played a key role in the transfer learning process, arguing that "*successful* responses are due to fruitful connections and analogies" (p. 48). As a basis for analogical transfer, Gick and Holyoak (1983) found that a verbal or spatial summary of a solution schema (single analogue) did not facilitate transfer to a target problem compared to mapping two analogues, which induced more schema abstraction.

Group cooperation in the form of reciprocal teaching has been shown to evoke zones of proximal development (Vygotsky, 1978) whereby all learners, with varying skill and expertise levels, are encouraged to undertake increasing responsibility for more-experienced roles. Palincsar and Brown (1984) notably argued that reciprocal teaching encouraged transfer when students were provided with four reading strategies (questioning, clarifying, summarising, and predicting) to actively enhance reading comprehension. Additionally, Salomon and Perkins (1989) maintained that transfer emerged in two ways, by either “low-road transfer” (p. 113), or “high-road transfer” (p. 113). The former is dependent on considerable practice that leads to automatization of learning. And, although is a quick way of accomplishing transfer, results in less flexibility. The latter is dependent on deep insights into the learning tasks. And while it is a slower way of attaining transfer, it is much wider in scope and adaptability.

In their approach to transfer of learning, Hayes and Simon (1977) employed different sets of problem isomorphs by making changes in the form of problem text representation. Two problems were defined as being isomorphs “if any solution path of one may be translated step by step into a solution path of the other and vice versa” (Hayes & Simon, 1977, p. 22). By presenting virtually the same problem differently, Hayes and Simon (1977) found that increased transfer between two problems occurs when subjects are able to recognise the similarity between the structures of the two problems (isomorphs). These findings indicate that transfer of a solution is more likely to be obtained when a learner’s mental representation of the training and transfer situations is associated with similar goals and processing between both situations. Increased perceived similarity of training and transfer tasks facilitates the retrieval of previously acquired knowledge of the training task and the appropriate application to the transfer task.

Barnett and Ceci (2002) elucidated the nature of a skill being transferred by distinguishing between two dimensions: “near” versus “far”; and “specific” versus “general”. The former dimension “relat[ing] to the similarity of the training and transfer situations” (p. 620), and the latter dimension “relat[ing] to how generally applicable the learned information is [, such as,] specific facts or procedures versus general skills, principles, or strategies” (Barnett & Ceci, 2002, p. 620). Moreover, Barnett and Ceci (2002) specified that the two dimensions should be separated given that far transfer can be defined as specific or general, and that general skills can be transferred to a near or far situation.

A high degree of similarity implies near transfer, whilst a low degree implies far transfer (Mayer, 1975; Royer, 1979). Using a multilevel transfer post-test, Mayer (1975) assessed the learning of binomial probability concepts by including near transfer items that required applying the formula; for example, subjects had to find the value for P(R,N) given values for R, N, and P (which were similar to the items presented in the booklet), and far transfer items were included that required interpretation; for example, using only part of the formula or recognising that the formula couldn’t be applied. Mayer (1975) found that subjects who received pretraining in general concepts were more competent at transferring what they learned to far transfer items (novel situations) compared to subjects who received pretraining in using the formula. In the opposite way, subjects who received pretraining in using the formula were faster at completing near transfer items than subjects who received pretraining in general concepts. This shows that a learner could master a specific skill if they previously learned that same skill in a particular learning situation. However, a learner may not have been able to recognise and correctly apply the specific skill to a far transfer situation which did not share identical elements with the original learning task.

In their research review of transfer, Gick and Holyoak (1987) affirmed that "the interrelationships between task structure, encoding and retrieval processes, and the prior knowledge of the learner" (p. 39-40) are factors related to transfer. Gick and Holyoak (1987) clarified this by maintaining that the structure of a task impacts on the encoding of the task which then fosters the acquisition of rules that can be applied to an array of tasks with structural commonalities that are superficially different. Retrieval of acquired knowledge in the context of transfer is then made possible if the transfer task induces "similar goals and processing mechanisms, or has salient surface resemblances to the training task" (Gick & Holyoak, 1987, p. 40). Additionally, Marini and Genereux (2015) claimed that the basic elements connected with transfer involve: the learner, the instructional and transfer tasks, and the instructional and transfer context; and that a combination of all of these elements are critical when designing instruction. Correspondingly, establishing the degree of difference between the instructed task/context and the transfer task/context respectively must be interpreted with caution when establishing the extent (near or far) and manner (specific or general) in which transferred knowledge is tested.

3.2.1.2 Transfer of learning in CLT research.

Effective and efficient transfer performance has been attained from a variety of CLT studies involving the use of worked examples (Cooper & Sweller, 1987; Hummel et al., 2004; Lim, Reiser, & Olina, 2009; Mawer & Sweller, 1982; Paas, 1992; Paas & van Merriënboer, 1994; Ranzijn, 1991; Robins & Mayer, 1993). This was first evident in Cooper and Sweller's (1987) series of experiments which showed that the reduced cognitive load associated with worked examples facilitated the interaction between schema acquisition (the conscious process whereby acquired schemas were used to directly solve test problems that were similar to previously seen problems) and schema automation (the automatic process which mainly bypassed WM) which was required for solving transfer problems. Schema automation, as defined by Sweller et al. (1998), occurs after extensive practice, which justifies why automaticity develops more slowly than schema acquisition. Since schema automation is carried out with minimal WM resources, available WM capacity makes it possible to complete novel tasks in cases where acquired schemas are of limited use on transfer problems (Kotovsky et al., 1985). In addition, as previously discussed in section 2.2.4.4 (the split-attention effect), better transfer test performance has been shown when two or more sources of information are integrated (to avoid split attention) when mutually referring information cannot be understood in isolation (Chandler & Sweller, 1992; Kester et al., 2005; Ward & Sweller, 1990).

An issue pertaining to undesired instructional design, referred to as the *transfer paradox*, is the phenomenon whereby efficient instructional methods, that are designed to minimise training time, reduce the number of practice items, and reach non-integrated, specific learning objectives, may yield low transfer of learning (Helsdingen, van Gog, & van Merriënboer, 2011a, 2011b). In contrast, less efficient instructional methods, that vary on both surface and structural features, which require more time-on-task and investment of mental effort because they ultimately require learners to work on randomly sequenced tasks, will yield higher transfer of learning. Van Merriënboer and Kirschner (2018) avowed that variability and random sequencing of tasks fostered increased investment in abstraction – a process whereby comparing and contrasting information enhances transfer of learning. When learners work “harder and longer” (p. 67), by accomplishing a set of varied learning tasks in a random order, they are able to construct knowledge that is general and abstract, which then enables them to reach integrated objectives and confront unfamiliar problems (van Merriënboer & Kirschner, 2018). In other words, if the aim of the instructional design is to yield higher transfer of learning, the instructional method should not encourage learners to construct specific knowledge that enables them to deal with tasks that are concordant with isolated, specific objectives.

The practice of random sequencing, which is referred to as *interleaving*, promotes the development of general and abstract schemas that enables learners to deal with unfamiliar features of novel problems (i.e., transfer). Interleaving helps learners to develop a more integrated knowledge base because they are required to practise different versions of the same constituent skills (Birnbaum, Kornell, Bjork, & Bjork, 2013). An example where unrelated tasks are interleaved, for instance, would be where successive 30 min lessons on mathematics, music and science respectively are followed by another set of successive 30 min lessons in the same subject areas, as opposed to presenting a single set of one-hour lessons on each subject. On the other hand, when learners practice the same constituent skills for equivalent tasks, they are less likely to compare and contrast the tasks, and as a result show lower transfer of learning (de Croock & van Merriënboer, 2007; Helsdingen et al., 2011a, 2011b). The present study did not investigate interleaving and its related effects because any reference made to variability in the four experiments specifically referred to changes to surface features and/or structural variations made to the question formats, in a similar fashion to the investigations undertaken by Paas and van Merriënboer (1994), and Quilici and Mayer (1996).

Although worked-example-based learning (as discussed in section 2.2.4.1) and variation between tasks (as discussed in section 2.2.4.5) have a positive effect on the development of proficient task performance, the present study aimed to shed more light on instructional design that enables learners to solve new and unfamiliar tasks after the achievement of initial learning of familiar tasks. In particular, the present study further investigated how the exposure to task variability (as discussed in section 2.2.4.5) can play a critical role in promoting the abstraction of general schemas and facilitate successful transfer to novel problems.

3.2.2 Subjective ratings of difficulty as measures of cognitive load.

When learners process inherently complex material, this imposes high intrinsic cognitive load, and if extraneous cognitive load associated with the instructional design is also high, this will increase the number of interactive elements that must be processed simultaneously in WM in order to complete the task. As discussed in section 2.2.1, the level of element interactivity is determined by the extent to which the material imposes cognitive load by the intrinsic nature of the material (which can vary from low to high depending on the learner's level of prior knowledge) and the extraneous factors caused by the instructional design (which can vary by the way the instructional material is presented without altering what needs to be learned).

Subjective ratings of difficulty were employed in the present study to distinguish between extraneous and intrinsic cognitive load. Based on the cognitive load theoretical framework, material consisting of elements with an intrinsically high degree of interactivity require considerable cognitive resources. Hence, the present study assumed that learners could complete high-variability tasks if extraneous cognitive load was reduced by studying worked examples or if the high-variability tasks were assigned to higher ability learners for completion, who are more likely to have sufficient WM capacity to deal with the complexity of high element interactivity tasks, compared to lower ability learners.

Out of the four experiments in the current study, Experiments 1, 2 and 4 investigated whether studying worked examples increases WM capacity by reducing extraneous cognitive load. Subjective ratings of difficulty were used to compare the cognitive load of participants who attempted to solve problems (without guidance) and participants who studied fully-guided worked examples. In Experiment 3, subjective ratings of difficulty by more-knowledgeable learners and less-knowledgeable learners were compared to investigate whether more-knowledgeable learners would experience less difficulty in completing the learning tasks because of their greater prior knowledge in the domain.

3.3 Independent Variables

Levels of variability and guidance, and levels of variability and learner expertise, were the independent variables used in Experiments 1, 2 and 4, and Experiment 3 respectively. These independent variables each had two levels (high and low) in order to observe the effect they had on transfer of learning and subjective ratings of difficulty.

The present study aimed to expand previous empirical work pertaining to the variability effect, worked example effect, and expertise reversal effect. Comprising four experiments, the present study investigated the effect of high- and low-variability tasks on a learner's capacity to construct schemas and transfer acquired knowledge. The participants' post-test results in each experiment were composed of two parts: similar questions (which were similarly structured to the tasks in the initial training and tested for schema formation); and transfer questions (which required the abstraction of a solution method from the tasks in the initial training). Experiments 1, 2 and 4 explored the effect of high- or low-variability tasks with high guidance (studying fully-guided worked examples) or low guidance (generating problem solutions without any guidance). Experiment 3 examined the effect of high- or low-variability tasks with more-experienced learners (experts) or less-experienced learners (novices).

All four experiments in the present study used high-variability tasks that involved the application of same solution process in a wider variety of contexts. This was achieved by changing not only the surface features (e.g., numbers) but also varying the structure of the problems (i.e. question formats), as opposed to the low-variability tasks, whereby only some surface features changed and the structure of the tasks remained the same.

3.3.1 Levels of variability and guidance.

According to Ranzijn (1991), widely dispersed examples (as opposed to narrowly dispersed examples) improved procedural knowledge. Based on this outcome, Experiments 1, 2 and 4 investigated the likelihood that studying high-variability worked examples would increase learners' ability to identify similar features between tasks and increase their ability to differentiate between relevant and irrelevant features. Using the rationale that students are more likely to achieve learning transfer after studying highly variable examples, the present study further explored the variability of practice.

The low-variability worked examples in Experiments 1, 2 and 4 were analogous to the structure-emphasising worked-out examples in Quilici and Mayer's (1996) study which aimed to foster schema construction. In Experiments 1, 2 and 4, low-variability worked examples were designed by using the solution structures of the source examples (which were presented on the board throughout the general instruction, during the initial part of the Learning Phase) as analogues to the solution structures of the target examples in the learning handout (which were studied during the latter part of the Learning Phase).

Unlike the hypothesis that was explored by Quilici and Mayer (1996) whereby college students who studied structure-emphasising example problems were more likely to categorise and apply the correct statistical test (e.g., t- test or correlation or chi-square) to statistics word problems compared to students who studied surface-emphasising example problems, the aim of Experiments 1, 2 and 4 was to examine the likelihood that participants in the high-variability worked-examples group would be better at solving test problems that required the adjustment of learned solution methods, compared to participants in the low-variability worked-examples group. In other words, studying high-variability worked examples would enhance schema formation and transfer more than studying low-variability worked examples (that were designed to emphasise the same structure). Thus, Experiments 1, 2 and 4 augmented Quilici and Mayer's (1996) study with the incorporation of two problem-solving groups. The high-variability problem-solving and low-variability problem-solving learning handouts provided students with the same problem statements as those in their respective high- and low-variability worked-examples handout but without the inclusion of solution steps.

The four instructional groups in Experiments 1, 2 and 4 were akin to those used by Paas and van Merriënboer (1994), whose investigation involved four computer-based training strategies: a low-variability and high-variability worked condition (which involved studying worked examples); and a low-variability and high-variability conventional condition (which involved solving conventional practice problems). However, unlike the participants in Paas and van Merriënboer's (1994) conventional condition, who worked on six problems (with two attempts to get each answer right within a restricted timeframe) and later studied the solutions to these problems (which were identical to the ones studied by the participants in the worked condition), the participants in the problem-solving groups in Experiments 1, 2 and 4 were provided with problem statements (excluding solution steps), with answers (only in Experiments 2 and 4). However, similar to Paas and van Merriënboer's (1994) study, the worked-examples condition in Experiments 1, 2 and 4 did not contain any subsequent practice problem-solving tasks. That is to say, Experiments 1, 2 and 4 used a pure worked-examples and pure problem-solving procedure for the high- and low-variability conditions. The pure worked-examples and pure problem-solving procedures in Experiments 1, 2 and 4 were in contrast to procedures previously used in some worked-examples studies (Cooper & Sweller, 1987; Leppink et al., 2014; Paas, 1992; Sweller et al., 1990) that showed the worked-example – problem-solving sequence to be superior to the reverse sequence.

3.3.2 Levels of variability and learner expertise.

Experiment 3 investigated the variability effect further. In it, participants were separated into novice or expert groups, according to their levels of prior knowledge. The instructional design of Experiment 3 was unique in comparison to any other CLT empirical investigation reported to date, as it tested for main effects of task variability (i.e., the possibility of a variability effect) and expertise, and an interaction effect between task variability and expertise (i.e., the possibility of a reverse variability effect and consequently, an expertise reversal effect). Unlike Experiments 1, 2 and 4 (which distinctly included worked-examples or problem-solving learning handouts), all the handouts in Experiment 3 contained the same initial worked examples, followed by either low-variability problem-solving tasks (which were of similar structure to the preceding worked examples) or high-variability problem-solving tasks (which were of dissimilar structure to the preceding worked examples).

Since levels of guidance (tested in Experiments 1, 2 and 4) and levels of expertise (tested in Experiment 3) both alter element interactivity, the learning handouts in Experiment 3 (which contained both worked-examples and problem-solving tasks) were designed to negate the need to vary guidance between groups because it was argued that levels of expertise could act as a substitute for levels of guidance.

3.4 Development of the Hypotheses

3.4.1 Worked example effect.

To obtain a worked example effect (as discussed in section 2.2.4.1), the learning material should be high in element interactivity (e.g. a complex task), otherwise, if the learning material is low in element interactivity (e.g., a simple task), the effectiveness of worked examples is reduced. When completing a simple task (where fewer elements need to be processed simultaneously compared to a complex task), more WM resources are available that can be allocated to manage extraneous cognitive load imposed by suboptimal task instructions (e.g., problem-solving tasks which require more effort to complete because of the absence of any instructional guidance, as opposed to studying the solution steps of a worked example for a similar task).

Experiments 1, 2 and 4 in the present study compared a worked-examples condition with a problem-solving condition under high element interactivity conditions because it was assumed that the learners were novices concerning the tasks. It was assumed that any weak problem-solving strategies (used by the learners in the problem-solving condition) would impose high extraneous cognitive load on WM which would not be conducive for learning even if these learners were successful in solving the problems. Since attempting to solve problems contributes less to learning because fewer WM resources are available for constructing cognitive schemas, it was presumed that the learners in the worked-examples condition would devote all their available WM capacity to studying the solution steps in the worked examples, and therefore be able to abstract general rules from the worked examples.

3.4.2 Variability effect.

The variability effect (as discussed in section 2.2.4.5) was investigated in all four experiments of the present study. In order to identify the variability effect, Experiments 1, 2 and 4 compared worked-examples study and problem solving with either high-variability or low-variability tasks, in a similar way to Paas and van Merriënboer (1994). Experiment 3 compared experts and novices completing either high-variability or low-variability tasks.

In their study of instructional effects of variability, Paas and van Merriënboer (1994), assuming that the learners in all condition groups were novices, confirmed that increasing variability would only be effective in the worked-examples condition. In line with the arguments put forward by Paas and van Merriënboer (1994), the presumption made for Experiments 1, 2 and 4 in the present study was that the variability effect could not be obtained in the problem-solving condition because problem solving would impose a higher WM load on novices compared to studying worked examples. Increased WM load associated with high-variability tasks (because more elements are needed to be managed simultaneously in WM compared to low-variability tasks) combined with the WM load associated with problem solving, would place a heavy strain on the learners' limited cognitive capacity. However, presenting learners with high-variability worked examples would reduce unnecessary extraneous cognitive load (because learners would not be faced with unfamiliar problems to solve) and increase intrinsic cognitive load (because learners would direct their attention to the highly variable solution steps), which is important for learning and would improve learning.

In Experiments 1, 2 and 4, and based on Paas and van Merriënboer's (1994) findings, it was assumed that studying high-variability worked examples would likely lead to better transfer of learning, compared to low-variability worked examples, because cognitive load imposed by high-variability tasks is intrinsic and not extraneous. In contrast, it was assumed that high-variability problem-solving tasks would impose intrinsic cognitive load, as well as extraneous cognitive load – due to the material being intrinsically high in element interactivity, and the absence of solution steps making it difficult for the learner to acquire cognitive skills (that otherwise are supported through the design of worked examples).

In Experiment 3, it was hypothesised that increasing variability would only be effective for more-knowledgeable learners (experts), but not for novices. This hypothesis will be discussed in the section below.

3.4.3 Expertise reversal effect.

The expertise reversal effect (as discussed in section 2.2.4.2) was investigated conjointly with the variability effect in Experiment 3. In this experiment, all the participants were placed into a novice or expert group (after initial testing of their prior mathematical knowledge that related to the experimental learning material). This was designed to test Renkl et al.'s (1998) assertion that more-experienced learners were more likely to benefit from studying high-variability examples because their cognitive capacity would be less likely to be overloaded by the structurally different examples, compared to less-experienced learners whose cognitive capacity was more likely to be overburdened. However, contrary to using only examples, all the learning handouts in Experiment 3 comprised the same worked examples followed by either high- or low-variability conventional problem-solving tasks.

It was predicted that presenting experts with high-variability problem-solving tasks and novices with low-variability problem-solving tasks would better induce schema construction to facilitate improved performance on the Post-Test similar questions than presenting experts with low-variability problem-solving tasks and novices with high-variability problem-solving tasks accordingly. Improved performance for the experts and novices for the above conditions would extend to the Post-Test transfer questions if the rules that were learned from studying the worked examples (in the initial part of all the learning handouts) and practised by solving the high- and low-variability problems (which followed the worked examples) respectively, became strengthened (or automated). That is, of primary importance was the prediction that problem-solving high-variability tasks would be more effective only for experts (compared to solving low-variability tasks). The effectiveness of this instruction was predicted to reverse for novices, such that problem solving low-variability tasks would be more effective compared to problem solving high-variability tasks.

The possibility of a reverse variability effect was predicted for novices because practising to solve low-variability problems would require less cognitive load due to the lower level of element interactivity of low-variability tasks compared to high-variability tasks. Simultaneously dealing with fewer elements would make greater cognitive capacity available for novices to deal with aspects of the low-variability tasks that were unfamiliar. In contrast, it was predicted that a variability effect would occur for experts because they had acquired more prior knowledge in the target domain. More-knowledgeable learners in a domain would likely require fewer search strategies (such as means-ends analysis), compared to less-knowledgeable learners, when dealing with high-variability problems. If experts attempted to solve high-variability problems, it was expected that it would be more beneficial for schema construction and for transfer of acquired skills (as opposed to solving low-variability problems) because they would be dealing with a wider range of different problem-solving tasks which would extend the range and applicability of schemas.

3.4.4 Cognitive load.

Comparing high- and low-variability groups in all four experiments of this thesis involved the evaluation of whether more mental effort (cognitive load) would be invested in completing high-variability learning tasks compared to low-variability tasks. The rationale behind this comparison was based on the notion that the higher element interactivity in high-variability learning tasks made them more difficult to learn compared to low-variability tasks. The more difficult it was to complete a task, the greater the processing demands and hence the greater amount of imposed cognitive load. It was assumed that learners exposed to the high-variability condition would experience greater intrinsic cognitive load due to the tasks being higher in element interactivity, compared to the low-variability tasks. Hence, it was hypothesised that subjective ratings of difficulty for completing high-variability tasks would be higher compared to completing low-variability tasks in all the experiments.

In Experiments 1, 2 and 4, it was hypothesised that the problem-solving conditions would impose a high extraneous cognitive load on participants who attempted to problem solve the learning tasks. This would hinder their learning because participants (who were assumed to be novices) would be forced to resort to weak methods such as means-ends analysis, leaving few or no cognitive resources left for useful processes such as schema construction. In contrast, it was hypothesised that the worked-examples conditions would diminish extraneous cognitive load, and studying the solution steps contained in worked examples would efficiently assist the learners to arrive at the correct answers.

In Experiment 3, it was hypothesised that more-knowledgeable learners who had acquired more schema-based knowledge in the target domain, in contrast to less-knowledgeable learners, would be more likely to resort to more efficient schema-based problem-solving strategies, and thus experience less difficulty in completing the learning tasks. On the other hand, it was hypothesised that less-knowledgeable learners would experience more difficulty in completing the learning tasks because they would encounter more unfamiliar elements when attempting to solve problems. This would more likely overwhelm their WM and increase cognitive load.

3.5 Summary of Chapter 3

This chapter provided the theoretical arguments for the dependent and independent variables used to formulate the hypotheses for the four experiments in the present study, as well as a discussion of the development of these hypotheses in relation to the expected cognitive load effects. The next chapter will detail the empirical examination of these four experiments.

Chapter 4: Four Experiments

4.1 Introduction to Chapter 4

This chapter will outline four empirical experiments which comprise the present study. These experiments aimed to examine two main areas: how studying worked examples or attempting to problem solve low- or high-variability mathematics tasks can influence test performance outcomes (Experiments 1, 2 and 4); and how studying the same worked examples followed by low- or high-variability mathematics problem-solving tasks can influence test performance outcomes for less-experienced and more-experienced learners in the domain (Experiment 3). Across all four experiments, subjective ratings of difficulty were used as measures of cognitive load to help identify the factors inhibiting or facilitating learning.

4.2 Experiment 1

4.2.1 Introduction.

Experiment 1 examined whether the variability effect could be obtained by analysing the comparison of studying high-variability and low-variability worked examples during the Learning Phase, and whether a worked example effect could be obtained by analysing the comparison of studying worked examples (full guidance) and solving problems (no guidance) during the Learning Phase, regardless of task variability. Hypotheses were tested in this experiment by using a 2 by 2 design with two levels of variability (high or low) and two levels of guidance (worked examples or problem solving).

Variability in the high- and low-variability learning handouts was achieved by changing the range of tasks for which worked examples were studied or problems had to be solved for the second and third questions (keeping the first question identical). In the case of the worked-examples – high-variability handout (see Appendix A) and corresponding problem-solving handout (see Appendix B), the solution processes for Questions 2 and 3 were applied in a wider variety of contexts, in both surface and structural features to that of Question 1. For example, the expression "4 / [*k* (*k* + 4)]" in Question 1, was replaced with the expression "(-7*k* – 6) / [*k* (*k* + 1) (*k* + 2)]" in Question 2 (i.e., the number "4" in the numerator was replaced with the algebraic expression "-7*k* – 6", and the two factors in the denominator, "k" and "*k* + 4", were replaced with three factors, "*k*", "*k* + 1" and "*k* + 2"). In the case of the worked-examples – low-variability handout (see Appendix C) and corresponding problem-solving handout (see Appendix D), only the surface features varied in Questions 2 and 3, and not the question format, to that of Question 1. For example, the expression "4 / [*k* (*k* + 4)]" in Question 1 was replaced with the expression "5 / [*k* (*k* + 5)]" in Question 2 (i.e., only the number "4" in the numerator and denominator was replaced with the number "5", without making any structural changes to the question).

The worked examples in the worked-examples – high-variability and worked-examples – low-variability handouts required students to study fully-guided tasks that contained step-by-step solutions for how to solve a problem, while the problem-solving statements in the corresponding high- and low-variability problem-solving handouts required students to generate problem solutions without any guidance. Given the difference in instructional design between the worked examples and problem-solving tasks, participants completed a difficulty rating scale; this rating scale was partly used to detect variations in cognitive load due to extraneous cognitive load.

Since instructional formats containing high-variability mathematical tasks are likely to increase mental effort because they are performed under conditions that require a highly varied sequence of solution steps, this experiment also explored whether completing high-variability mathematical tasks would further increase cognitive load because of the increased level of element connectedness. Learning to solve high-variability problems (either by studying worked-example solution steps or attempting to generate solutions without any guidance) requires learners not only to focus on one type of solution procedure but also to recognise which problems require relevant solution procedures. Therefore, increasing variability increases intrinsic cognitive load. Hence, the difficulty rating scale that each participant completed was also partly used to detect variations in cognitive load due to intrinsic cognitive load.

This experiment examined the learners' ability to solve similar and transfer questions in the Post-Test (see Appendix E), on the basis that effective mathematical learning occurs when students understand learned procedures and develop skills that allow them to solve new problems beyond a single context. For example, the expression '1 / [($5k$ – 2) ($5k$ + 3)]" in Question 1 of the Post-Test was similar to the expression "4 / [k (k + 4)" in Question 1 of the learning handout (i.e., the number "4" in the numerator was replaced with "1", and the two factors in the denominator, "k" and "k + 4", were replaced with "$5k$ – 2" and "$5k$ + 3"). The expression "$log_2[1 + (1/k)]$" in Question 2 of the Post-Test, for example, tested the learners' ability to transfer knowledge and skills, by looking for any patterns and relationships from any past lessons/experience that could be generalised and widely applied.

4.2.2 Hypotheses.

Experiment 1 tested the following hypotheses:

1. Learners who study fully-guided worked examples will yield higher post-test performance scores, compared to learners who attempt to solve problems without any guidance, due to a reduction in extraneous cognitive load. Accordingly, subjective ratings of difficulty for worked-example study will be lower compared to problem solving.
2. Learners who study high-variability worked examples will yield higher post-test performance scores, compared to learners who study low-variability worked examples, due to increased intrinsic cognitive load. However, this difference will not be generated under problem-solving conditions. Subjective ratings of difficulty for completing high-variability tasks will be higher compared to completing low-variability tasks.

4.2.3 Method.

4.2.3.1 Participants.

The participants were tertiary students enrolled in the Discrete Mathematics course (MATH1081) at the School of Mathematics and Statistics, University of New South Wales, Sydney. The experiment convenor attended one of the students' lectures to discuss and invite the students to participate in the upcoming experiment. Out of the 421 students enrolled in the course, 106 consented to participate in the experiment, of which 97 were present for the entire experiment. In order to enrol in the course, students needed to demonstrate assumed knowledge by achieving a combined mark of at least 100/150 in their high school advanced and extension mathematics subjects (which they completed in their final year of secondary school, prior to enrolling in their university program). The sample comprised 26 females (27%) and 71 males, aged between 17 and 29 (M_{age} = 20.11, SD_{age} = 2.32). The academic abilities of each participant varied as they were enrolled in diverse programs (mainly from Computer Science, Mathematics Education, Science, Advanced Science, Advanced Mathematics, Engineering and Actuarial Studies) and they were at different stages in their program.

4.2.3.2 Materials.

The material used in the experiment focused on the topic of "telescoping sums" which is taught in the first unit of the Discrete Mathematics course. To measure the amount of prior knowledge of each participant, an average percentage score of two previous class tests (that tested for other mathematical topics) was used to assess the participants' ability to acquire, integrate and apply knowledge.

During the Learning Phase (initial part of the experiment), each participant received a handout in accordance with the assigned experimental group they were in: 'worked-examples – high-variability', 'worked-examples – low-variability', 'problem-solving – high-variability', or 'problem-solving – low-variability'.

When dealing with high-variability tasks, learners were expected to do comparisons within a broader range of tasks. For example, in the high-variability handouts, the expression "$4 / [k (k + 4)]$" in Question 1, was replaced with the expression "$\cos(2k - 1)$" in Question 3, using "$2\sin(1) \cos(2k - 1) = \sin(2k) - \sin(2k - 2)$" (i.e., the basic algebraic expression in Question 1 was replaced with an advanced trigonometric expression and equation in Question 3, which required more solution steps to evaluate). On the other hand, low-variability tasks required learners to repeatedly deal with the same type of tasks within a narrower range. For example, in the low-variability handouts, the expression "$4 / [k (k + 4)]$" in Question 1, was replaced with the expression "$1 / [(k + 4) (k + 5)]$" in Question 3 (i.e., both expressions comprised the same algebraic format with only the number in the numerator of Question 1, “4”, being replaced with “1” in Question 3, and the pronumeral in the denominator of Question 1, “*k*”, being replaced with the expression “$k + 5$” in Question 3). The high-variability handouts contained tasks that were high in element interactivity as these tasks required more solution steps and/or the complexity of each solution step was higher than the tasks in the low-variability handouts.

The worked-examples handouts contained fully-guided worked examples with explicit solution steps. The problem-solving handouts consisted only of the same problem statements, excluding any written instructions.

All participants were given the same single-item, nine-point Likert-type rating scale to complete. The question was: “How difficult was it for you to complete the tasks?” (see Appendix F). The rating scale ranged from “Extremely Easy” (on the far left, with point “1” assigned to the answer) to “Neither Easy nor Difficult” (in the middle, point “5”) and to “Extremely Difficult” (on the far right, point “9”). The participants’ subjective ratings of difficulty were used to measure cognitive load imposed during their completion of the Learning Phase handout.

The Post-Test consisted of three questions that included eight tasks in total. The first four tasks were structurally similar to the questions that were explicitly presented during the participants’ lecture that took place prior to the commencement of the experiment. The remaining four tasks were intended to test for transfer of learning because they required the ability to extend what had been learned (from attending the lecture and completing the learning handout) and to use this knowledge in new contexts. The Post-Test was identical in content for all participants. The questions were internally reliable based on a Cronbach’s Alpha of .75.

4.2.3.3 Procedure.

Prior to the commencement of the experiment, the course lecturer provided the participants with comprehensive explicit instruction on the topic of telescoping sums that reflected standard solution procedures on the board during the participants' normal lecture time. Following the lecture, the experiment convenor met with the participants during two normal mathematics tutorials (which were two weeks apart).

During the first tutorial (30 min), the participants were randomly assigned to one of the four experimental conditions: worked-examples – high-variability group (23 students), worked-examples – low-variability group (27 students), problem-solving – high-variability group (24 students), and problem-solving – low-variability group (23 students). The number of students in the worked-examples – low-variability group contained more students because there were some students in the other three groups that did not participate in the Post-Test (second part of the experiment) and hence were not counted as participating in the overall experiment.

Random assignment was achieved by allowing the students to choose their seat when they entered the classroom and then handing out the different materials in a sequential order so that every fourth student received the same material. The worked-examples – high-variability group and the worked-examples – low-variability group were instructed to study step-by-step worked-out solutions for eight high- or low-variability problems respectively. The problem solving-high variability group and the problem-solving – low-variability group were instructed to generate solutions for the same eight high- or low-variability problems. After the Learning Phase, the handouts were collected and each participant completed the subjective rating of difficulty scale.

During the second tutorial that followed two weeks after the Learning Phase, all participants completed an identical, 30 min Post-Test.

4.2.3.4 Marking procedure.

The marking of the Post-Test, which was an objective test, was undertaken by the experiment convenor. Consistency was achieved by considering all possible solutions. Answers to questions were awarded one mark for every correct solution step, and incorrect solution steps were awarded a mark of zero. Based on this procedure, the highest possible total score for the Post-Test was 38 marks, consisting of 19 marks for the similar questions and 19 marks for the transfer questions. All the raw scores were converted into percentage scores.

4.2.4 Results.

There were two independent variables: level of variability and level of guidance; and three dependent variables: Post-Test (similar questions) scores, Post-Test (transfer questions) scores, and subjective ratings of difficulty. Table 1 shows the descriptive statistics for the participants' performance. Prior mathematical knowledge was also tested to ensure group comparability.

4.2.4.1 Prior knowledge.

To compare the level of prior mathematical knowledge for the four groups, a one-way between-groups analysis of variance was conducted for the average class test scores for two previous class tests that were completed before the commencement of the experiment. It involved one independent variable (the condition group) across four levels (worked-examples – high-variability group, worked-examples – low-variability group, problem-solving – high-variability group, and problem-solving – low-variability group) and one dependent variable

Table 1

Means (and Standard Deviations) of the Average Class Test Scores, Post-Test (Similar Questions) Scores, Post-Test (Transfer Questions) Scores, and Subjective Ratings of Difficulty for Experiment 1

	Experimental condition							
	Worked-examples – high-variability group		Worked-examples – low-variability group		Problem-solving – high-variability group		Problem-solving – low-variability group	
Variable	*M* *(SD)*	*N*	*M* *(SD)*	*N*	*M* *(SD)*	*N*	*M* *(SD)*	*N*
Average class test scores (%)	60.43 (17.70)	23	62.59 (22.38)	27	70.42 (18.76)	24	68.26 (21.41)	23
Post-test (similar questions) scores (%)	65.33 (28.56)	23	55.75 (24.82)	27	60.42 (31.64)	24	58.93 (25.21)	23
Post-test (transfer questions) scores (%)	21.97 (28.06)	23	17.74 (22.62)	27	28.95 (32.02)	24	20.48 (24.90)	23
Subjective ratings of difficulty (1-9)	2.87 (1.66)	23	2.48 (1.67)	27	5.25 (2.07)	24	4.74 (2.78)	23

(average class test score). The results did not reach statistical significance, $F(3, 93) = 1.28$, $MSE = 523.98$, $p = .29$, partial $\eta^2 = .04$. Therefore, the average class test scores were not used to control for any differences between the experimental groups for any Post-Test performance results.

4.2.4.2 Post-test scores.

Two 2 by 2 between-groups analyses of variance were conducted on the Post-Test (similar questions) scores and Post-Test (transfer questions) scores. The results for guidance did not reach statistical significance for the Post-Test (similar questions) scores, $F(1, 93) = .02$, $MSE = 18.24$, $p = .88$, partial $\eta^2 = .000$, and the Post-Test (transfer questions) scores, $F(1, 93) = .78$, $MSE = 570.35$, $p = .38$, partial $\eta^2 = .01$. Likewise, the results for variability did not reach statistical significance on the Post-Test (similar questions) scores, $F(1, 93) = .97$, $MSE = 739.82$, $p = .33$, partial $\eta^2 = .01$, and on the Post-Test (transfer questions) scores, $F(1, 93) = 1.33$, $MSE = 973.13$, $p = .25$, partial $\eta^2 = .01$. The results of these analyses showed no evidence of a relationship between levels of guidance (studying worked examples or problem solving) or levels of task variability (high or low) for the completion of Post-Test tasks in similar and novel situations. Hence the expected worked example effect and variability effect were not obtained.

The variability by guidance interactions were not statistically significant for the Post-Test (similar questions) scores, $F(1, 93) = .52$, $MSE = 395.02$, $p = .47$, partial $\eta^2 = .01$, and the Post-Test (transfer questions) scores, $F(1, 93) = .15$, $MSE = 108.43$, $p = .70$, partial $\eta^2 = .002$. Hence there was no simultaneous effect of the two independent variables (variability and guidance) on any of the dependent variables (similar and transfer Post-Test scores) in which one of the independent variables differed depending on the level of the other independent variable.

4.2.4.3 Subjective ratings of difficulty.

A 2 by 2 between-groups analysis of variance was conducted to assess the effect of the two independent variables on subjective ratings of difficulty (cognitive load). The results showed a statistically significant main effect for guidance, $F(1, 93) = 30.02$, $MSE = 129.86$, $p < .001$, partial $\eta^2 = .24$. Subjective ratings of difficulty were less for the worked-examples groups compared to the problem-solving groups. However, the results for variability were not statistically significant, $F(1, 93) = 1.13$, $MSE = 4.88$, $p = .29$, partial $\eta^2 = .01$ Thus, the subjective ratings were not different for the low- and high-variability groups. This result indicates that the similarity and dissimilarity of the low- and high-variability tasks respectively did not have an impact on cognitive load. On the other hand, the significant main effect of guidance on subjective ratings indicates that higher cognitive load was imposed on learners who generated solutions to problem-solving tasks (without any guidance) compared to learners who studied fully-guided worked examples, regardless of the level of variability that the learners were exposed to. The variability by guidance interaction for cognitive load was not statistically significant, $F(1, 93) = .02$, $MSE = .09$, $p = .89$, partial $\eta^2 = .00$.

4.2.4.4 Discussion.

The post-test performance of learners who studied worked examples was not superior to learners who attempted to solve equivalent problems during the Learning Phase. The failure to obtain a worked example effect, and thus the failure to support the first part of Hypothesis 1, may have been due to the complexity of the material contained in the worked examples. Even though the worked example effect is usually obtained when learners struggle to understand high element interactivity information, the results suggest that the worked examples in the current experiment may have been unable to guide the learners through the problem-solving process so that they could deal with the complex information. Supposing the material studied in the worked examples was difficult to understand, this may have left no processing capacity for the germane cognitive processes that help learners to construct cognitive schemas for solving such problems. If the learners were able to construct cognitive representations from the worked examples, this would have helped them solve similar test problems, and transfer test problems for which general rules had to be abstracted from the examples they studied.

The two-week time frame between the Learning Phase and the Post-Test may not have induced the students to study the worked examples with more incentive. The findings from Sweller et al. (2011) suggest that students' learning in the present experiment may have been enhanced if they recognised that they needed to solve test problems immediately after studying worked examples.

Despite the absence of a worked example effect on the Post-Test scores, the last part of Hypothesis 1 was supported whereby participants in the worked-examples groups found their instructional procedure substantially easier than the participants in the problem-solving groups. The higher level of cognitive efficiency for the worked-examples condition suggests that there was a relative benefit in studying worked examples. This finding is in accordance with the CLT view whereby studying worked examples eliminates the need to resort to weak problem-solving strategies (e.g., means-ends analysis) that impose a high extraneous cognitive load.

Learners who studied high-variability worked examples did not attain higher levels of conceptual understanding and superior transfer skills compared to learners who studied low-variability worked examples and high- or low-variability problem-solving tasks. The failure to obtain a variability effect, and thus the failure to support the first part of Hypothesis 2, meant that the results did not replicate the Paas and van Merriënboer (1994) findings. One reason for this could be due to the absence of a worked example effect arising from the possibility that the prior knowledge of the participants was lower than expected in relation to understanding the solution steps in the worked examples. If studying worked examples did not free sufficient WM resources for schema formation (by reducing extraneous cognitive load), then processing high-variability worked examples would have further reduced the likelihood of fostering the acquisition of conceptual and transferable knowledge. The limited capacity of WM may have made it impossible to deal with the high level of element connectedness in the high-variability tasks which imposed high levels of intrinsic cognitive load. On the other hand, if a worked example effect had been obtained, this would have freed up WM resources for schema formation, thus increasing the benefits of processing high-variability worked examples.

The lower than anticipated prior knowledge of the participants could be the reason why the last part of Hypothesis 2 was not confirmed. Despite the lower level of element interactivity of the low-variability tasks compared to the high-variability tasks, the participants may not have found the low-variability tasks significantly easier than the high-variability tasks. The decrease in intrinsic cognitive load by reducing the level of element interactivity may have been insufficient for the learners dealing with the low-variability tasks.

Overall, the non-significant difference in the subjective ratings of difficulty between the high- and low-variability groups indicate that the element interactivity and consequent intrinsic cognitive load experienced by the learners may have been excessively high, which possibly jeopardised any potential for learning. Consequently, the same hypotheses were tested in Experiment 2 but with learning material that the students were expected to find less difficult.

4.3 Experiment 2

4.3.1 Introduction.

Similar to Experiment 1, Experiment 2 investigated the effect of high- and low-variability conditions on worked examples or problem-solving tasks. Correspondingly, a 2 (worked examples vs. problem solving) by 2 (high- vs. low-variability) design, in line with Paas and van Merriënboer's (1994) study, was used to examine whether the variability effect was more likely to occur under worked-examples rather than problem-solving conditions. This design was based on the assumption that problem-solving conditions increase element interactivity due to extraneous cognitive load beyond the point where working memory can handle increased element interactivity associated with higher variability.

As discussed in Part I of this thesis, task variability increases element interactivity associated with intrinsic cognitive load because of the need for learners to distinguish between and classify different problem types. This increase in intrinsic cognitive load increases learning outcomes, provided there is sufficient capacity in WM to process the increased number of elements. Similar to Experiment 1, the current experiment examined whether the use of high-variability worked examples would allow the increased number of interacting elements associated with intrinsic cognitive load to be processed, because studying the explicit solution steps of the worked examples could help reduce element interactivity due to extraneous cognitive load. When problem solving without any guidance is used as an instructional method, most working memory resources are taken up by random attempts at finding a solution to the problem, rather than learning its essential features.

Since the variability effect was not obtained in Experiment 1 under worked-examples conditions, the current experiment used learning material that was expected to be easier (for a different mathematical topic), with different participants (who were at the secondary mathematics knowledge level), compared to Experiment 1 (who were at the tertiary mathematics knowledge level). Subjective measures of difficulty were used to detect variations in extraneous cognitive load (due to the instructional factors arising from studying worked examples or problem solving equivalent tasks) and intrinsic cognitive load (due to the difference in element interactivity in high- and low-variability tasks).

4.3.2 Hypotheses.

On the basis that the present study re-examined the likelihood of the worked example effect and the variability effect, the hypotheses remained the same as those tested in Experiment 1, namely:

1. Learners who study fully-guided worked examples will yield higher post-test performance scores, compared to learners who attempt to solve problems without any guidance, due to a reduction in extraneous cognitive load. Accordingly, subjective ratings of difficulty for worked-example study will be lower compared to problem solving.

2. Learners who study high-variability worked examples will yield higher post-test performance scores, compared to learners who study low-variability worked examples, due to increased intrinsic cognitive load. However, this difference will not be generated under problem-solving conditions. Subjective ratings of difficulty for completing high-variability tasks will be higher compared to completing low-variability tasks.

4.3.3 Method.

4.3.3.1 Participants.

The participants were 103 mathematics students, aged between 18 and 55 (M_{age} = 26.57, SD_{age} = 7.55), enrolled in a preparation program at the University of New South Wales, Sydney. This post-secondary education program prepares students for admission to university programs. The academic abilities of each participant varied because of individual differences in workplace experience and/or former education. The sample comprised 42 females (41%) and 61 males.

4.3.3.2 Materials.

The material used in the experiment comprised the definition of a quadratic function, the formulae for the roots of a quadratic function, the axis of symmetry and the vertex of a parabola, and the skill of graphing a quadratic function. All participants were regarded as novice learners in relation to quadratic functions as this topic was the next scheduled topic in the mathematics preparation program. To measure the amount of prior knowledge of each participant, an average percentage score of three previous class tests (that tested for other mathematical topics) was used to assess the participants' ability to acquire, integrate and apply knowledge.

During the first half of the Learning Phase (initial part of the experiment), on the board, the experiment convenor provided the participants with explicit instruction on the topic of quadratic functions that reflected standard solution procedures during the participants' normal lecture time. During the second half of the Learning Phase, each participant received a handout in accordance with the assigned experimental group they were in: 'worked-examples – high-variability' (see Appendix G), 'worked-examples – low-variability' (see Appendix H), 'problem-solving – high-variability' (see Appendix I), or 'problem-solving – low-variability' (see Appendix J).

The high-variability handouts contained tasks that differed in the presentation format and the solution procedures from the tasks performed on the board during the explicit instruction part of the Learning Phase. For example, the first question in the high-variability handout, which transitioned into the topic of quadratic functions, consisted of the following tasks: "Given $g(x) = 5 - 2x$, for all real x, find: (i) $g(-3x)$, (ii) $g(1/4)$, [and] (iii) $g(a + 5)$". These high-variability questions required the participants to substitute x with three highly varied expressions, namely: "$-3x$" (a variable with a negative coefficient), "$1/4$" (a fraction), and "$a + 5$" (an algebraic expression).

The low-variability handouts contained tasks that were similar to those performed on the board during the explicit instruction part of the Learning Phase. For example, the first question in the low-variability handout consisted of the following tasks: "Given $f(x) = 4x + 8$, for all real x, find: (i) $f(0)$, (ii) $f(-2)$, [and] (iii) $f(a)$". This question had a similar instructional format to Example 1 which was demonstrated on the board during the explicit instruction part of the Learning Phase: "Consider the function defined by $f(x) = 3x + 6$, for all real x. Find: (i) $f(1)$, (ii) $f(-2)$, [and] (iii) $f(a)$". The name of the function in first question of the low-variability handout, $f(x)$, was the same function name used in Example 1 of the explicit instruction. However, to increase variability in the first question of the high-variability handout, the function name changed from f to g.

The worked-examples handouts contained fully-guided worked examples with explicit solution steps and diagrams. Similar to Experiment 1, the problem-solving handouts consisted only of the same problem statements, excluding any written instructions or diagrams. However, unlike Experiment 1, the last page of the high-variability and low-variability problem-solving handouts contained answers (but not step-by-step worked-out solutions) to each problem task on the previous pages to enable the participants to compare their answers to the correct answers. This was intended to assist students in gauging the accuracy of their attempts, so that whenever a correct answer was obtained for a particular task, they could repeat their solution steps for similar type tasks that followed.

All participants were given the same single-item, nine-point Likert-type rating scale to complete, identical to the one used in Experiment 1 (see Appendix F). The participants' subjective ratings were used to measure cognitive load imposed during their completion of the Learning Phase handout.

The Post-Test consisted of seven questions that included 10 tasks in total (see Appendix K). The first six tasks were structurally similar to the examples demonstrated on the board during the explicit instruction part of the Learning Phase. For example, Question 3(i) in the Post-Test, "Determine the concavity of the following quadratic functions: (i) $y = -4 + 2x - x^2$", was similarly structured to Example 4(iii) that was presented on the board during the direct instruction, "Determine the concavity of the following quadratic functions: $y = -x^2 + 4x + 5$". The last four questions in the Post-Test were structurally different and were intended to test for transfer of learning, because they required the capacity to apply the learned knowledge in new situations. For example, Question 7 in the Post-Test, "Without sketching the graph, determine whether the curve $y = 2x^2 + 4x + 5$ crosses the x-axis", required students to apply and generalise the skills they acquired during the explicit instruction part of the Learning Phase. Participants did not practise answering this type of question prior to the Post-Test. The Post-Test was identical in content for all participants. The questions were internally reliable based on a Cronbach's Alpha of .75.

4.3.3.3 Procedure.

At the start of the experiment, participants were randomly assigned to one of the four experimental conditions: worked-examples – high-variability group (25 students), worked-examples – low-variability group (26 students), problem-solving – high-variability group (26 students), and problem-solving – low-variability group (26 students). The random assignment was achieved by allowing the students to choose their seat when they entered the lecture theatre and then handing out the different materials in a sequential order so that every fourth student received the same material.

The duration of the experiment was 1 hr 30 min, and it was conducted during the participants' normal mathematics lecture and tutorial time. The experiment consisted of a Learning Phase (60 min) and a Post-Test Phase (30 min). In the first half of the Learning Phase (30 min), the experiment convener provided all participants with comprehensive explicit instructions on quadratic functions reflecting standard solution procedures on the board. In the second half of the Learning Phase (30 min), participants were instructed to complete a different handout that assigned them to their respective experimental group.

The worked-examples – high-variability group and the worked-examples – low-variability group were instructed to study step-by-step worked-out solutions for 14 high- or low-variability problems respectively. The problem-solving – high-variability group and the problem-solving – low-variability group were instructed to generate solutions for the same 14 high- or low-variability problems. The last page of the high-variability and low-variability problem-solving handouts contained answers to each problem task on the previous pages to enable the participants to compare their answers to the correct answers.

Similar to Experiment 1, after the Learning Phase, the handouts were collected and each participant completed a subjective rating of difficulty. Following the subjective rating questionnaire, all participants completed the Post-Test.

4.3.3.4 Marking procedure.

The marking of the Post-Tests was undertaken by the experiment convenor. Similar to Experiment 1, consistency was achieved by considering all possible solutions. Answers to questions were awarded one mark for every correct solution step, and incorrect solution steps were awarded a mark of zero. Based on this procedure, the highest possible total score for the Post-Test was 38 marks, consisting of 25 marks for the similar questions and 13 marks for the transfer questions. All the raw scores were converted into percentage scores.

4.3.4 Results.

There were two independent variables: level of variability and level of guidance; and three dependent variables: Post-Test (similar questions) scores, Post-Test (transfer questions) scores, and subjective ratings of difficulty. Table 2 shows the descriptive statistics for the participants' performance. Prior mathematical knowledge was also tested to ensure group comparability.

4.3.4.1 Prior knowledge.

The prior mathematical knowledge of each participant was measured by averaging scores for three previous class tests that were completed before the commencement of the experiment. A one-way between-groups analysis of variance was conducted for the average class test scores to compare the level of prior

Table 2

Means (and Standard Deviations) of the Average Class Test Scores, Adjusted Post-Test (Similar Questions) Scores, Adjusted Post-Test (Transfer Questions) Scores, and Subjective Ratings of Difficulty for Experiment 2

	Experimental Condition							
	Worked-examples – high-variability group		Worked-examples – low-variability group		Problem-solving – high-variability group		Problem-solving – low-variability group	
Variable	*M* *(SD)*	*N*	*M* *(SD)*	*N*	*M* *(SD)*	*N*	*M* *(SD)*	*N*
Average class test scores (%)	61.12 (22.45)	25	53.67 (19.40)	26	68.94 (20.66)	26	69.73 (19.34)	26
Adjusted post-test (similar questions) scores (%)	45.91 (18.31)	25	37.98 (18.81)	26	38.16 (18.45)	26	33.10 (18.51)	26
Adjusted post-test (transfer questions) scores (%)	39.94 (16.79)	25	32.06 (17.25)	26	37.22 (16.92)	26	31.09 (16.97)	26
Subjective ratings of difficulty (1-9)	3.64 (2.20)	25	3.92 (2.23)	26	6.69 (2.13)	26	6.85 (1.57)	26

mathematical knowledge for the four groups: worked-examples – high-variability group, worked-examples – low-variability group, problem-solving – high-variability group, and problem-solving – low-variability group. The dependent variable was average class test score. The results showed a statistically significant difference, $F(3, 99) = 3.52$, $MSE = 1477.44$, $p = .02$, partial $\eta^2 = .10$ (medium effect size). Bonferroni post-hoc pairwise comparisons indicated a significant difference between the worked-examples – low-variability group and problem-solving – low-variability group only. Even though the previous class tests comprised questions that differed from the topic content used in the current experiment, the average class test scores were used as a covariate for all Post-Test performance results to control for the unexpected differences between the experimental groups in the prior knowledge analyses.

4.3.4.2 Post-test scores.

Two 2 by 2 between-groups analyses of covariance were conducted on the Post-Test (similar questions) scores and Post-Test (transfer questions) scores. The results for variability on the Post-Test (similar questions) scores did not reach statistical significance, $F(1, 98) = 3.23$, $MSE = 1080.06$, $p = .08$, partial $\eta^2 = .03$. The results showed a statistically significant main effect for variability on the Post-Test (transfer questions) scores, $F(1, 98) = 4.46$, $MSE = 1254.08$, $p = .04$, partial $\eta^2 = .04$. The results of these analyses showed that increasing variability effectively boosted transfer of learning: the ability to solve problems that have not been solved before. In addition, there was evidence of a marginal relationship (a possible effect) between high-variability tasks and the application of knowledge and skills in completing similar questions.

The results for guidance were not statistically significant for the Post-Test (similar questions) scores, $F(1, 98) = 2.83$, $MSE = 943.92$, $p = .10$, partial $\eta^2 = .03$, and the Post-Test (transfer questions) scores, $F(1, 98) = .29$, $MSE = 80.53$, $p = .59$, partial $\eta^2 = .003$. The results of these analyses showed no evidence of a relationship between guidance and the application of knowledge and skills in completing tasks in similar and novel situations. These results indicate that the expected worked example effect was not obtained.

The variability by guidance interactions were not statistically significant for the Post-Test (similar questions) scores, $F(1, 98) = .16$, $MSE = 52.31$, $p = .69$, partial η^2 = .002, and the Post-Test (transfer questions) scores, $F(1, 98) = .07$, $MSE = 19.48$, $p = .79$, partial η^2 = .001. These results showed that levels of guidance did not alter the effects of variability on the Post-Test questions.

4.3.4.3 Subjective ratings of difficulty.

A 2 by 2 between-groups analysis of variance was conducted to assess the effect of the two independent variables on subjective ratings of difficulty (cognitive load). The results showed a statistically significant main effect for guidance, $F(1, 99) = 54.88$, $MSE = 229.79$, $p < .001$, partial η^2 = .36. Subjective ratings of difficulty were less for the worked-examples groups compared to the problem-solving groups. However, the results for variability were not statistically significant, $F(1, 99) = .29$, $MSE = 1.23$, $p = .59$, partial η^2 = .003. Thus, the subjective ratings were not different for the high- and low-variability groups. This result indicates that the similarity and dissimilarity of the high- and low-variability tasks respectively did not have an impact on cognitive load. On the other hand, the significant main effect of guidance on subjective ratings of difficulty indicates that higher cognitive load was imposed on learners who generated solutions to problem-solving tasks (without any guidance) compared to learners who studied fully-guided worked examples, regardless of the level of variability that the learners were exposed to. The variability by guidance interaction for cognitive load was not statistically significant, $F(1,99) = .03$, $MSE = .11$, $p = .87$, partial η^2 = .00.

4.3.4.4 Discussion.

The findings of this experiment showed that when students are provided with high-variability tasks, they will attain higher test scores under worked-examples conditions. Thus, the first part of Hypothesis 2 was supported (for the

Post-Test transfer questions, and marginally for the Post-Test similar questions). As part of this hypothesis, this effect would not be obtained under problem-solving conditions but in fact, the same effect was obtained under both worked-examples and problem-solving conditions. This result provides support for the conceptual premise that high-variability tasks promote transfer learning; that is, understanding gained through learning to distinguish a high variety of tasks assists learners in abstracting common structures of problems, instead of common surface features confined to a specific problem.

The results did not support the first part of Hypothesis 1, that worked examples would facilitate superior test performance. There were no significant differences between the worked-examples and problem-solving conditions. Neither did the results replicate the Paas and van Merriënboer (1994) findings, where the variability effect was only obtainable using worked examples but not problem solving, even though the worked-examples group found their instructional procedure substantially easier than the problem-solving group. Notwithstanding, our results are consistent and coherent with the Paas and van Merriënboer (1994) findings.

Whether or not a worked example effect can be obtained is determined by the participants' knowledge levels. The worked example effect is just as susceptible to the expertise reversal effect as any other cognitive load effect. If knowledge levels are sufficiently high, the effect will not be obtained or may even be reversed (Kalyuga, Chandler et al., 2001). In the current experiment, there was no evidence of improved test performance following instruction using worked examples rather than problem-solving tasks. With no difference in test performance between worked-examples and problem-solving conditions, we should not expect any differences in the effect of variability associated with worked examples or problem solving. If test performance scores following problem solving are just as high as following worked examples, there should be sufficient working memory resources available to handle the element interactivity associated with problem solving and variability. Accordingly, our failure to replicate the Paas and van Merriënboer (1994) results following problem solving is to be expected given the failure to obtain a worked example effect on the test results. Both the worked-examples and problem-solving conditions should be susceptible to a variability effect as indicated by our results.

The last part of Hypothesis 1, whereby subjective ratings of difficulty would be lower for the worked-examples conditions compared to the problem-solving conditions, was supported. This indicates that the search-based problem-solving conditions imposed a heavy extraneous cognitive load, which supports the assumptions based on CLT. However, the last part of Hypothesis 2, whereby subjective ratings of difficulty would be higher for the high-variability conditions compared to the low-variability conditions, was not supported. Contrary to CLT expectations, this may have been due to a reversal of the effectiveness of instructional techniques. There may have been significant differences in the participants' level of prior knowledge in the target domain. The level of difficulty in completing the learning tasks (due to the level of element interactivity), may not have been high for all the participants in the high-variability conditions or low for all the participants in the low-variability conditions. This result provided the impetus for Experiment 3.

Experiment 3 initially tested the prior knowledge of each participant by having them all complete a Pre-Test that assessed how much they already knew about the new material. The results of the Pre-Test in Experiment 3, which determined whether the participants were placed in the expert or novice groups, were used to test variability further, by comparing the post-test performance of more- and less-experienced learners in the domain.

4.4 Experiment 3

4.4.1 Introduction.

The variability effect was obtained in Experiment 2 following both worked-examples and problem-solving conditions. This was arguably due to the participants' knowledge levels being too high to obtain a worked example effect. Given this, Experiment 3 tested the variability effect further in connection with levels of learner prior knowledge by comparing the test performance of more-experienced (expert) and less-experienced (novice) learners. This experiment examined whether more-experienced learners would demonstrate the variability effect (i.e. high-variability tasks resulting in increased test performance), while less-experienced learners, who are unable to effectively process high-variability tasks, would demonstrate a reverse variability effect.

For more-experienced learners, element interactivity should be relatively low, allowing them to readily process the additional element interactivity associated with increased variability. A conventional variability effect should result. For less-experienced learners, element interactivity should be higher, and adding additional element interactivity by increasing variability may result in cognitive load exceeding working memory capacity (Chen, Castro-Alonso, Paas, & Sweller, 2018). Accordingly, reduced rather than increased test performance due to increased variability should be the result, with a reverse variability effect being obtained. The current experiment investigated whether a conventional variability effect for more expert learners associated with a reverse variability effect for less able learners would provide an example of an expertise reversal effect.

In essence, the current experiment attempted to reproduce the results obtained by Paas and van Merriënboer (1994) by changing levels of expertise rather than changing levels of guidance. Increased levels of guidance via worked examples should reduce element interactivity just as increased expertise should reduce element interactivity. Both should allow the variability effect to occur provided element interactivity is sufficiently low to allow working memory to process the increased elements associated with variability. In contrast, if element interactivity is too high, increasing it further by increasing variability should eliminate the variability effect or even reverse it.

4.4.2 Hypotheses.

Experiment 3 tested the following hypotheses:

1. Providing more-knowledgeable learners with high-variability tasks, compared to low-variability tasks, will yield higher post-test performance scores.
2. Providing less-knowledgeable learners with low-variability tasks, compared to high-variability tasks, will yield higher post-test performance scores.
3. Subjective ratings of difficulty by less-knowledgeable learners will be higher compared to more-knowledgeable learners.
4. Subjective ratings of difficulty for completing high-variability tasks will be higher compared to completing low-variability tasks.

4.4.3 Method.

4.4.3.1 Participants.

The participants were 56 mathematics students enrolled in the same university preparation program to the one in Experiment 2. The academic abilities of each participant varied because of individual differences in workplace experience and/or former education. The participants were aged between 18 and 55 (M_{age} = 26.27, SD_{age} = 8.20), comprising of 21 females (37.5%) and 35 males.

4.4.3.2 Materials.

The material used in the experiment comprised the definition of a logarithm, logarithmic laws, and solving logarithmic equations. The domain-specific prior knowledge of each participant was measured using a Pre-Test that consisted of 22 tasks (see Appendix L). The first 16 tasks (in Questions 1-4) evaluated baseline pre-requisite knowledge required for learning logarithmic equations. These tasks were simpler than the tasks in the experimental Learning Phase (and accordingly, in the Post-Test) as they required knowledge of individual components and isolated procedures involved in the learning tasks. A combination of these components and procedures was required to work out the Learning Phase and Post-Test tasks. For example, the first tasks in Questions 1-4 were respectively: "In the expression 2^3: What is the exponent?"; "Simplify the following, writing the answers in index form with positive indices: $8a^4 \times 2a^6$"; "Find the exact value of the following without using a calculator: 8^{-2}"; and "Solve the following equations for *x*: $4^x = 1/16$". The final six tasks (in Questions 5-7) were analogous to typical test questions that met the expected learning outcomes for logarithmic equations. For example, the second tasks in Questions 5-7 were respectively: "In the logarithmic equation $\log_3 9 = 2$: What is the base of the log?"; "Make *x* the subject for the following: $\log_b (x+2) = 3$"; and "Solve the following equations for *x*: $\log_3 1 = x$". These six logarithmic tasks were included at the end of the Pre-Test to gauge if the participants had any prior understanding of the new topic that was going to be taught during the first half of the Learning

Phase. These six logarithmic tasks were tested again in the Post-Test (without notifying the participants). The Pre-Test questions were identical in content for all participants and were internally reliable based on a Cronbach's Alpha of .75.

During the second half of the Learning Phase, each participant received either a low-variability handout (see Appendix M) or a high-variability handout (see Appendix N), regardless of whether they were designated as an expert or novice, respectively. The first page of the handout, which was identical for the high- and low-variability handouts, contained four worked examples that consisted of fully-guided written instructions. The remaining three pages of the handout contained sixteen problem-solving tasks which differed according to the experimental condition (high or low variability). The problem-solving tasks in the low-variability handouts contained tasks with numbers that changed in the same part of each question (without changing the position of the variable). For example, the first problem-solving question in the low-variability handout consisted of the following tasks: "Write the following in logarithmic form, <u>without</u> solving for x: (a) $3^x = 9$; (b) $4^x = 1/4$; (c) $125^x = 5$; [and] (d) $32^x = 4$". On the other hand, the high-variability handout contained the equivalent problem-solving questions as those in the low-variability handout with the exception that the position of the variable changed in some tasks for each question (i.e., the question format changed for some tasks to increase variability). For example, the first problem-solving question in the high-variability handout consisted of the following tasks: "Write the following in logarithmic form, <u>without</u> solving for x: (a) $3^x = 9$; (b) $4^{-1} = x$; (c) $x^{1/3} = 5$; [and] (d) $32^x = 4$".

All participants were given the same single-item, nine-point Likert-type rating scale to complete, identical to the one used in Experiments 1 and 2 (see

Appendix F). The participants' subjective ratings of difficulty were used to measure cognitive load imposed during their completion of the Learning Phase handout.

Unlike Experiments 1 and 2, in which the participants completed one Post-Test (that contained similar and transfer questions), the participants in the current experiment completed two Post-Tests to ensure that an equal amount of time was allocated to answering similar and transfer questions. Post-Test 1 comprised 16 similar tasks (see Appendix O) and Post-Test 2 comprised seven transfer tasks (see Appendix P). The questions were internally reliable based on a Cronbach's Alpha of .82 for the similar questions and .69 for the transfer questions.

In Post-Test 1, the first six tasks (in Questions 1-3) were identical to the last six tasks of the Pre-Test, which were structurally similar, both in context and concept, to the examples that were demonstrated on the board by the experiment convenor during the explicit instruction part of the Learning Phase. These tasks met the expected learning outcomes for the topic of logarithms that would normally be covered during the participants' usual lecture and tutorial time. The remaining 10 tasks (in Questions 4-6) were also structurally similar to the examples that were demonstrated on the board by the experiment convenor. For instance, Question 5(a) in Post-Test 1, "Evaluate … $\log_8 2$", had the same question format as Example 4, which was presented on the board during the direct instruction, "Evaluate … $\log_8 4$".

In Post-Test 2, all seven transfer tasks were structurally different, both in context and concept, from the examples that were demonstrated on the board by the experiment convenor during the explicit instruction part of the Learning Phase. These questions required the application of acquired knowledge in relatively new task situations. For instance, Question 1(b) in Post-Test 2, "Evaluate … $(\log_{10} 25)/(\log_{10} 5) \times 10^{\log_{10} 3}$", required the application of numerous index laws and logarithm laws (which were presented at different stages on the board during the direct instruction), in order to arrive at the following solution steps: $\log_{10} 25 = \log_{10}(5^2) = 2\log_{10} 5$; if $x = 10^{\log_{10} 3}$, then $\log_{10} x = \log_{10} 3$, thus $x = 3$; so $[(\log_{10} 25)/(\log_{10} 5)] \times 10^{\log_{10} 3} = [(2\log_{10} 5)/(\log_{10} 5)] \times 3 = 2 \times 3 = 6$. In contrast, the examples presented on the board during the direct instruction required the application of fewer combinations of these laws.

4.4.3.3 Procedure.

The experiment was conducted during the participants' normal mathematics lecture and tutorial time. Its duration was 30 min for the Pre-Test followed one week later by a 1 hr 30 min block. Participants who scored in the top half of the Pre-Test were designated as experts (more-experienced learners). Participants who scored in the bottom half of the Pre-Test were designated as novices (less-experienced learners). Participants were evenly apportioned with 28 participants in the expert groups and 28 participants in the novice groups.

During the 1 hr 30 min block, the experiment comprised of a Learning Phase (60 min) and a Post-Test Phase (30 min). At the start of the 1 hr 30 min block, expert participants were randomly assigned to one of two experimental conditions: 'expert – low-variability' or 'expert – high-variability'. And likewise, novice participants were randomly assigned to one of two experimental conditions: 'novice – low-variability' or 'novice – high-variability'. This gave 14 participants in each of the four groups. The random assignment was achieved by allowing the expert and novice participants to choose their seat when they entered the lecture theatre.

In the first half of the Learning Phase (30 min), the experiment convener provided all participants with comprehensive explicit instructions on the board by demonstrating step-by-step solution procedures for the new topic of logarithmic equations. In the second half of the Learning Phase, the expert and novice participants received different handouts assigning them to their respective experimental groups (high or low variability). The first page of the handout, which was identical for all groups, contained four worked examples that consisted of fully-guided written instructions. The remaining three pages of the handout contained 16 problem solving tasks which differed according to the experimental condition (high or low variability). The purpose of the same worked examples on the first page of the high- and low-variability handouts, was to provide an identical summary of what had been taught on the board by the experiment convenor during the first half of the Learning Phase and to provide participants with a general guide on how to generate solutions for the problem-solving tasks that followed on the remaining pages.

Participants were given 30 min to complete their Learning Phase handouts. Immediately following, the answers to each of the tasks in both conditions (high and low variability) were presented on the board. Participants were given a few minutes to check their work by comparing the correct answers on the board with their written answers, enabling them to determine which problem-solving tasks they answered correctly.

Similar to Experiments 1 and 2, after the Learning Phase, the handouts were collected and each participant completed a subjective rating of difficulty. Following the completion of the subjective rating questionnaire, all participants completed Post-Test 1 in 15 min, and then proceeded to complete Post-Test 2 in 15 min.

4.4.3.4 Marking procedure.

The marking of the Pre-Tests and Post-Tests, both of which were objective tests, was undertaken by the experiment convenor. Similar to Experiments 1 and 2, consistency was achieved by considering all possible solutions. Answers to questions were awarded one mark for every correct solution step, and incorrect solution steps were awarded a mark of zero. Based on this procedure, the highest possible score for the Pre-Test was 50 marks, for Post-Test 1 was 34 marks and for Post-Test 2 was 34 marks. As in Experiments 1 and 2, the raw scores were converted to percentage scores.

4.4.4 Results.

There were two independent variables: level of variability and level of expertise; and four dependent variables: Pre-Test scores, Post-Test 1 (similar questions) scores, Post-Test 2 (transfer questions) scores, and subjective ratings of difficulty. Table 3 shows the descriptive statistics for the participants' performance**.**

4.4.4.1 Pre-test scores.

In order to evaluate the level of prior knowledge of logarithmic equations for the four groups, a 2 (expert vs. novice) by 2 (high- vs. low-variability) between-groups analysis of variance was conducted on the Pre-Test score. As expected, the results showed a statistically significant main effect for expertise on the Pre-Test scores, $F(3, 52) = 111.15$, $MSE = 9884.57$, $p < 0.001$, partial $\eta^2 = 0.68$. The results did not show a main effect for variability on the Pre-Test scores, $F(1, 52) = 2.70$, $MSE = 240.29$, $p = 0.11$, partial $\eta^2 = 0.05$, indicating that the high- and low-variability groups had similar levels of prior knowledge.

Table 3

Means (and Standard Deviations) of the Pre-Test Scores, Adjusted Post-Test 1 (Similar Questions) Scores, Adjusted Post-Test 2 (Transfer Questions) Scores, and Subjective Ratings of Difficulty for Experiment 3

	Experimental Condition							
	Expert – low-variability group		Expert – high-variability group		Novice – low-variability group		Novice – high-variability group	
Variable	*M (SD)*	*N*	*M (SD)*	*N*	*M (SD)*	*N*	*M (SD)*	*N*
Pre-test scores (%)	52.00 (8.56)	14	62.14 (13.53)	14	31.43 (5.68)	14	29.57 (8.20)	14
Adjusted post-test 1 (similar questions) scores (%)	49.22 (12.94)	14	73.91 (16.76)	14	63.70 (14.26)	14	35.86 (14.97)	14
Adjusted post-test 2 (transfer questions) scores (%)	29.86 (14.08)	14	54.08 (18.23)	14	51.55 (15.51)	14	19.13 (16.28)	14
Subjective ratings of difficulty (1-9)	2.43 (1.60)	14	3.07 (1.33)	14	4.93 (1.64)	14	7.86 (.95)	14

The expertise by variability interaction effect on the Pre-Test scores was significant, $F(1, 52) = 5.67$, $MSE = 504.00$, $p = 0.02$, partial $\eta^2 = 0.10$, indicating that the magnitude of the difference between the low- and high-variability groups' scores was different at different levels of expertise. Given the statistically significant interaction, follow-up analyses were performed to determine whether there were any simple effects. For expert learners, the high-variability group scored higher in the Pre-Test than the low-variability group, $F(1, 26) = 5.62$, $MSE = 720.14$, $p = 0.03$, partial $\eta^2 = 0.18$, indicating that the more able learners in the high-variability group were more knowledgeable than those in the low-variability group. For novice learners, there was no simple effect, $F(1, 26) = .49$, $MSE = 24.14$, $p = 0.49$, partial $\eta^2 = 0.02$. To control for the unexpected differences between the experimental groups in the Pre-Test analyses, the Pre-Test scores were used as a covariate in all Post-Test performance analyses.

4.4.4.2 Post-test scores.

Two 2 by 2 between-groups analyses of covariance were conducted on the Post-Test 1 (similar questions) scores and Post-Test 2 (transfer questions) scores. The results showed that the expert learner groups produced significantly higher scores than the novice learner groups for the Post-Test 1 (similar questions) scores, $F(1, 51) = 4.46$, $MSE = 619.53$, $p = 0.04$, partial $\eta^2 = 0.08$. However, the results did not show a main effect for expertise on Post-Test 2 (transfer questions) scores, $F(1, 51) = 1.19$, $MSE = 196.14$, $p = 0.28$, partial $\eta^2 = 0.02$. The results did not show a main effect for variability on the Post-Test 1 (similar questions) scores, $F(1, 51) = 0.24$, $MSE = 32.89$, $p = 0.63$, partial $\eta^2 = 0.01$, or on Post-Test 2 (transfer questions) scores, $F(1, 51) = 1.36$, $MSE = 223.68$, $p = 0.25$, partial $\eta^2 = 0.03$.

The variability by expertise interaction was statistically significant on the Post-Test 1 (similar questions) scores, $F(1, 51) = 62.66$, $MSE = 8708.87$, $p < 0.001$, partial $\eta^2 = 0.55$, and the Post-Test 2 (transfer questions) scores, $F(1, 51) = 61.51$, $MSE = 10119.81$, $p < 0.001$, partial $\eta^2 = 0.55$. Figures 1 and 2 graphically depict the dis-ordinal interactions that existed between levels of expertise and levels of variability for the Post-Test 1 and Post-Test 2 scores respectively.

To test the degree to which variability is differentially effective at the expert and novice levels, simple effects were tested following the significant interactions.

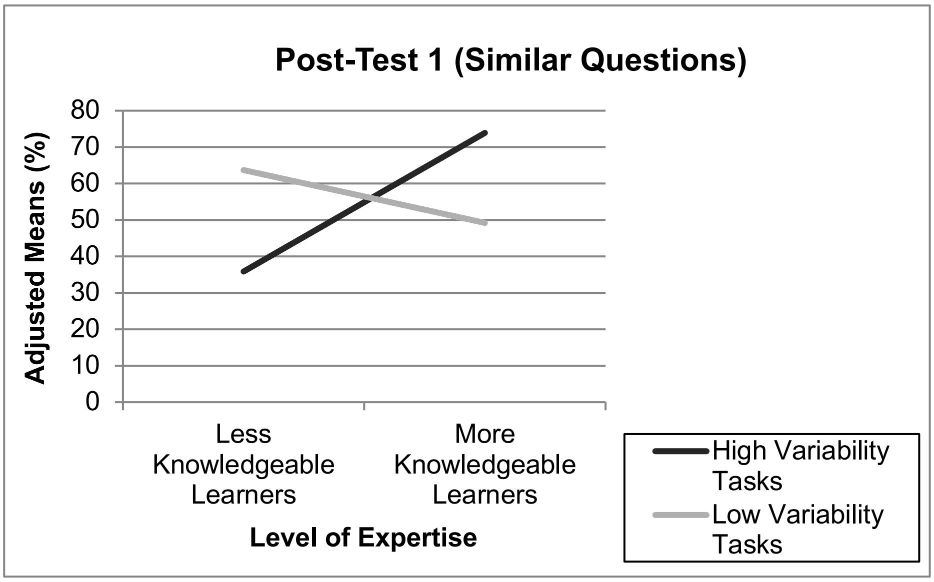


Figure 1. Variability by expertise interaction using adjusted means for Post-Test 1 in Experiment 3.

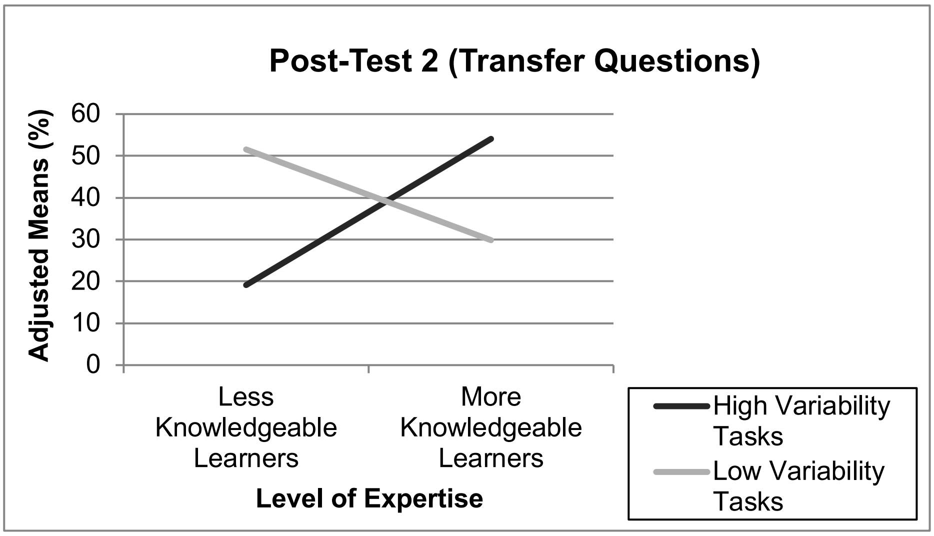


Figure 2. Variability by expertise interaction using adjusted means for Post-Test 2 in Experiment 3.

For expert learners, high-variability tasks led to higher scores in Post-Test 1 (similar questions) compared to low-variability tasks, $F(1, 25) = 21.64$, $MSE = 3312.79$, $p < 0.001$, partial $\eta^2 = 0.46$. For novice learners, low-variability tasks led to higher scores in Post-Test 1 (similar questions) compared to high-variability tasks, $F(1, 25) = 42.54$, $MSE = 5452.47$, $p < 0.001$, partial $\eta^2 = 0.63$. For expert learners, high-variability tasks led to higher scores in Post-Test 2 (transfer questions) compared to low-variability tasks, $F(1, 25) = 16.18$, $MSE = 2924.96$, $p < 0.001$, partial $\eta^2 = 0.39$. For novice learners, low-variability tasks led to higher scores in Post-Test 2 (transfer questions) compared to high-variability tasks, $F(1, 25) = 53.38$, $MSE = 7574.65$, $p < 0.001$, partial $\eta^2 = 0.68$. These results show that the effect of levels of variability (high and low) differed significantly depending on levels of learner expertise (expert or novice). In particular, superior performance scores for expert learners were associated with high-variability tasks, while superior performance scores for novice learners were associated with low-variability tasks.

4.4.4.3 Subjective ratings of difficulty.

A 2 by 2 between-groups analysis of variance was conducted to assess the effect of the two independent variables on subjective ratings of difficulty (cognitive load). The results showed a statistically significant main effect for expertise, $F(1, 52) = 93.80$, $MSE = 185.79$, $p < 0.001$, partial $\eta^2 = 0.64$, and for variability, $F(1, 52) = 22.54$, $MSE = 44.64$, $p < 0.001$, partial $\eta^2 = 0.30$. The lower subjective ratings of difficulty experienced by the expert groups shows that the more-knowledgeable learners were able to attempt the learning tasks with less mental effort, given their greater prior knowledge in the domain, compared to the less-knowledgeable learners. The lower subjective ratings of difficulty experienced by the participants in the low-variability groups indicates that the low-variability tasks were lower in element interactivity, compared to the high-variability tasks.

The variability by expertise interaction was statistically significant, $F(1, 52) = 9.23$, $MSE = 18.29$, $p = 0.004$, partial $\eta^2 = 0.15$. Figure 3 depicts the ordinal interaction between levels of expertise and levels of variability for cognitive load during the Learning Phase. Analyses of simple effects were conducted following the significant interaction between the levels of expertise and the levels of variability. For expert learners, there was no significant difference between cognitive load for high-variability tasks compared to low-variability tasks during the Learning Phase, $F(1, 26) = 1.34$, $MSE = 2.89$, $p = 0.26$, partial $\eta^2 = 0.05$. For novice learners, cognitive load was higher for high-variability tasks compared to low-variability tasks, $F(1, 26) = 33.47$, $MSE = 60.04$, $p < 0.001$, partial $\eta^2 = 0.56$. These simple effect analyses show that the interaction between expertise and variability was due to the difficulty less-knowledgeable learners had dealing with high-variability problems compared to low-variability problems. That difference was reduced for more-knowledgeable learners.

4.4.4.4 Discussion.

This experiment tested the hypothesis that the variability effect could be obtained if learners had sufficient working memory capacity associated with higher levels of expertise to enable them to process the increased levels of element interactivity. Problem sets with higher degrees of variability include more interactive elements of information than problem sets with lower degrees of

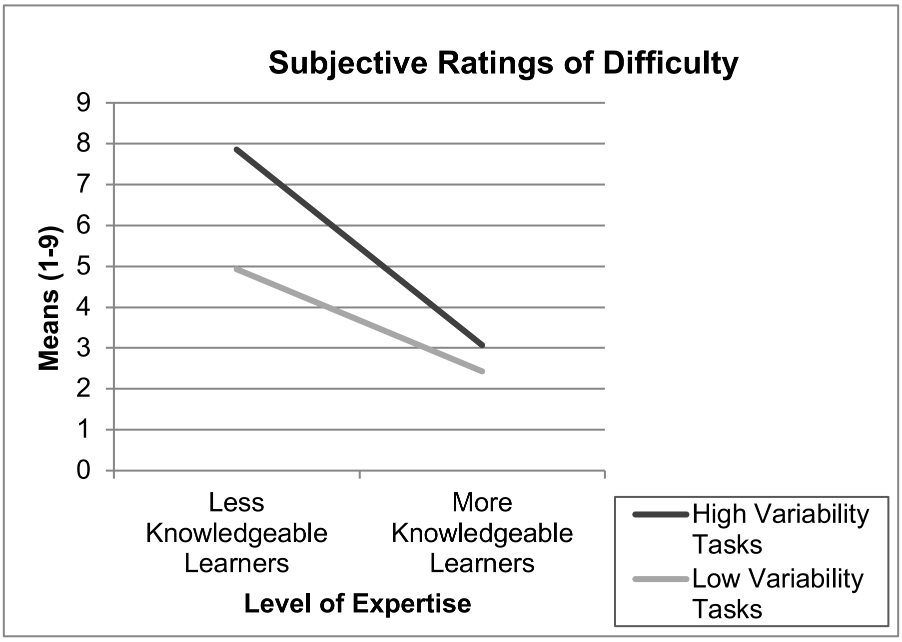


Figure 3. Variability by expertise interaction using means for Subjective Ratings of Difficulty in Experiment 3.

variability. This increased element interactivity of information increases intrinsic cognitive load. If, because of insufficient levels of expertise, learners do not have sufficient spare working memory capacity to handle the increased intrinsic cognitive load associated with high-variability information, it was hypothesised that the variability effect would be reduced or even reversed.

Three of the four hypothesised results were obtained. A standard variability effect was obtained using more-knowledgeable learners who learned more from high rather than low-variability problems. Thus, Hypothesis 1 was supported. High-variability learning tasks improved performance on both similar and transfer tasks for these learners. For the less-knowledgeable learners, the reverse result was obtained. Thus, Hypothesis 2 was supported. Low-variability learning tasks improved performance on both similar and transfer tasks for these learners.

Furthermore, in the case of less-knowledgeable learners, the improvement in scores using low-variability problems was associated with a significant reduction in cognitive load. It appears that the novice learners experienced less cognitive load when solving low-variability tasks as they were able to identify a surface match between the similarly structured questions without the need to go any further. In contrast, these learners found processing high-variability problems relatively difficult. In order to comprehend dissimilar features of high-variability tasks, whereby all the questions did not share any common surface features with the examples studied during the explicit instruction part of the Learning Phase, more mental effort was required to process the deeper features until the underlying common features were found. In addition, the results demonstrated significantly lower cognitive load experienced by the more-knowledgeable learners, compared to the less-knowledgeable learners. Thus, Hypothesis 3 was supported.

Finally, Hypothesis 4 was not confirmed since there was no difference in subjective ratings of difficulty for completing high-variability tasks compared to low-variability tasks. Possibly, the level of difficulty experienced by the learners may have been primarily influenced by their level of experience in the domain. Since the novice groups found the low- and high-variability learning tasks more difficult compared to the expert groups (as predicted by Hypothesis 3), subjective ratings of difficulty for both variability conditions were a mixture of higher and lower levels of subjective ratings of difficulty respectively because the low- and high-variability groups consisted equally of novices and experts in the domain. Hence, neither variability condition could produce a significantly lower (in the case of low-variability tasks) or higher (in the case of high-variability tasks) level of difficulty.

4.5 Experiment 4

4.5.1 Introduction.

Experiment 4 re-examined whether the worked example effect and the variability effect could be achieved by using the same problem statements from Experiment 2. Given that the worked example effect was not obtained in Experiment 2, the worked-examples handouts that were used in Experiment 2 were modified by removing any redundant information and inserting arrows to assist learners with physically integrating disparate sources of mutually-referring information. By redesigning the worked-examples – high-variability and worked-examples – low-variability handouts, it was envisaged that extraneous cognitive activities that interfered with learning, such as processing redundant and split-source information, would be minimised and this would facilitate the worked example effect. The present experiment used a 2 by 2 design with two levels of variability (high or low) and two levels of guidance (worked examples or problem solving). This experimental design was identical to the one used in Experiments 1 and 2.

4.5.2 Hypotheses.

On the basis that the present study re-examined the likelihood of the worked example and variability effects, the hypotheses remained the same as those in Experiments 1 and 2, namely:

1. Learners who study fully-guided worked examples will yield higher post-test performance scores, compared to learners who attempt to solve problems without any guidance, due to a reduction in extraneous cognitive load. Accordingly, subjective ratings of difficulty for worked-example study will be lower compared to problem solving.

2. Learners who study high-variability worked examples will yield higher post-test performance scores, compared to learners who study low-variability worked examples, due to increased intrinsic cognitive load. However, this difference will not be generated under problem-solving conditions. Subjective ratings of difficulty for completing high-variability tasks will be higher compared to completing low-variability tasks.

4.5.3 Method.

4.5.3.1 Participants.

The participants were 68 mathematics students, aged between 18 and 65 (M_{age} = 26.68, SD_{age} = 8.40), enrolled in the same post-secondary education program indicated in Experiments 2 and 3. However, the participants were part of a different cohort. Similar to Experiments 2 and 3, the academic abilities of each participant varied because of individual differences in workplace experience and/or former education. The sample comprised of 24 females (35%) and 44 males.

4.5.3.2 Materials.

The material used in Experiment 4 focused on the same topic area as in Experiment 2, namely: the definition of a quadratic function, the formulae for the roots of a quadratic function, the axis of symmetry and the vertex of a parabola, and the skill of graphing a quadratic function. All participants were regarded as novice learners in relation to quadratic functions as this topic was the next scheduled topic in the mathematics preparation program. To measure the amount of prior knowledge of each participant, an average percentage score of four previous class tests (that tested for other mathematical topics) was used to assess the participants' ability to acquire, integrate and apply knowledge.

During the first half of the Learning Phase, the experiment convenor provided the same explicit instructions that were used in Experiment 2 whereby solutions of the relevant tasks were demonstrated on the board. During the second half of the Learning Phase, each participant received a handout in accordance with the same experimental groups as those in Experiment 2: 'worked-examples – high-variability', 'worked-examples – low-variability', 'problem-solving – high-variability', or 'problem-solving – low-variability'. The range of tasks, for which worked examples were studied or problems had to be solved, were identical to those used in Experiment 2.

Since the high-variability and low-variability problem-solving tasks were identical to those used in Experiment 2, the format and content of the high-variability – problem-solving handout and the low-variability – problem-solving handout was reused (see Appendices I and J). However, as previously stated, the worked-examples handouts were redesigned (to eliminate any redundancy or split-attention) with the aim of re-testing for the worked examples effect (see Appendices Q and R for the modified worked-examples – high-variability and work-examples – low-variability handouts respectively). In particular, arrows were inserted to reduce extraneous cognitive load by reducing split-source information. The physical integration of information, through the use of arrows in the worked examples of the current experiment, was designed to facilitate more learning by imposing less cognitive load compared to studying the split-source worked examples in Experiment 2, which required the mental integration of disparate sources of information. Additionally, any redundant information in the worked examples that were studied in Experiment 2 was removed. This was to avoid any undue cognitive load that was imposed by processing unnecessary information. For example, written text such as the following, which appeared in Question 3 of the high-variability – worked-examples handout for Experiment 2, was removed in Question 3 of the high-variability – worked-examples handout for the current experiment: “STEP 1”, “STEP 2”, “STEP 3”, “we have to factorise before we can complete the square”, “[h]alve the co-efficient of *x* and square it”, and “remember to add this number to the constant term”.

All participants were given the same single-item, nine-point Likert-type rating scale to complete, identical to the one used in Experiments 1, 2 and 3 (see Appendix F). The participants’ subjective ratings of difficulty were used to measure cognitive load imposed during their completion of the Learning Phase handout. Equivalently, the Post-Test was identical to the one used in Experiment 2 (see Appendix K). The Post-Test was identical in content for all participants. The questions were internally reliable based on a Cronbach’s Alpha of .79.

4.5.3.3 Procedure.

Similar to Experiment 2, the participants were randomly assigned to one of the four experimental conditions at the start of the experiment, namely: worked-examples – high-variability group (17 students), worked-examples – low-variability group (17 students), problem-solving – high-variability group (17 students), and problem-solving – low-variability group (17 students). Correspondingly, the duration of the experiment, which was 1 hr 30 min, was conducted during the participants’ normal mathematics lecture and tutorial time, and consisted of a Learning Phase (60 min) and a Post-Test Phase (30 min), identical to Experiment 2.

Similar to Experiments 1, 2 and 3, after the Learning Phase, the handouts were collected and each participant completed a subjective rating of difficulty.

Following the subjective rating questionnaire, all participants completed the Post-Test.

4.5.3.4 Marking procedure.

The marking of the Post-Tests was undertaken by the experiment convenor. Similar to Experiments 1, 2 and 3, consistency was achieved by considering all possible solutions. Answers to questions were awarded one mark for every correct solution step, and incorrect solution steps were awarded a mark of zero. Based on this procedure, the highest possible total score for the Post-Test was 38 marks, consisting of 25 marks for the similar questions and 13 marks for the transfer questions. All the raw scores were converted into percentage scores.

4.5.4 Results.

There were two independent variables: level of variability and level of guidance; and three dependent variables: Post-Test (similar questions) scores, Post-Test (transfer questions) scores, and subjective ratings of difficulty. Table 4 shows the descriptive statistics for the participants' performance. Prior mathematical knowledge also was tested to ensure group comparability.

4.5.4.1 Prior knowledge.

In an equivalent way to Experiment 2, the prior mathematical knowledge of each participant was measured by averaging scores for four previous class tests that were completed before the commencement of the experiment. A one-way between-groups analysis of variance was conducted for the average class test scores to compare the level of prior mathematical knowledge for the four groups: worked-examples – high-variability group, worked-examples – low-variability group, problem-solving – high-variability group, and problem-solving – low-

Table 4

Means (and Standard Deviations) of the Average Class Test Scores, Post-Test (Similar Questions) Scores, Post-Test (Transfer Questions) Scores, and Subjective Ratings of Difficulty for Experiment 4

	Experimental Condition							
	Worked-examples – high-variability group		Worked-examples – low-variability group		Problem-solving – high-variability group		Problem-solving – low-variability group	
Variable	*M* *(SD)*	*N*	*M* *(SD)*	*N*	*M* *(SD)*	*N*	*M* *(SD)*	*N*
Average class test scores (%)	83.77 (35.00)	17	80.63 (41.50)	17	75.78 (20.00)	17	83.65 (23.25)	17
Post-test (similar questions) scores (%)	42.59 (29.39)	17	45.41 (33.91)	17	30.12 (27.35)	17	46.59 (28.18)	17
Post-test (transfer questions) scores (%)	33.94 (29.68)	17	31.67 (25.35)	17	30.77 (16.54)	17	47.06 (22.73)	17
Subjective ratings of difficulty (1-9)	3.65 (1.37)	17	1.76 (1.09)	17	6.82 (1.55)	17	6.06 (2.30)	17

variability group. The dependent variable was average class test score. The results did not reach statistical significance, $F(3, 64) = .45$, $MSE = 204.73$, $p = .72$, partial $\eta^2 = .02$. Therefore, the average class test scores were not used to control for any differences between the experimental groups for any Post-Test performance results.

4.5.4.2 Post-test scores.

Two 2 by 2 between-groups analyses of variance were conducted on the Post-Test (similar questions) scores and Post-Test (transfer questions) scores. The results for guidance did not reach statistical significance on the Post-Test (similar questions) scores, $F(1, 64) = .61$, $MSE = 542.12$, $p = .44$, partial $\eta^2 = .01$, or on the Post-Test (transfer questions) scores, $F(1, 64) = 1.10$, $MSE = 634.22$, $p = .30$, partial $\eta^2 = .02$. Likewise, the results for variability did not reach statistical significance on the Post-Test (similar questions) scores, $F(1, 64) = 1.78$, $MSE = 1582.12$, $p = .19$, partial $\eta^2 = .03$, or on the Post-Test (transfer questions) scores, $F(1, 64) = 1.45$, $MSE = 836.29$, $p = .23$, partial $\eta^2 = .02$. The results of these analyses showed no evidence of a relationship between levels of guidance (worked examples or problem solving) or levels of task variability (high or low) for the completion of Post-Test tasks in similar and novel situations. Hence the expected worked example effect and variability effect were not obtained.

The variability by guidance interactions were not statistically significant for the Post-Test (similar questions) scores, $F(1, 64) = .89$, $MSE = 791.53$, $p = .35$, partial $\eta^2 = .01$, or the Post-Test (transfer questions) scores, $F(1, 64) = 2.53$, $MSE = 1462.62$, $p = .12$, partial $\eta^2 = .04$. Hence, there was no indication of an interaction between the two independent variables of level of guidance and level of variability.

4.5.4.3 Subjective ratings of difficulty.

A 2 by 2 between-groups analysis of variance was conducted to assess the effect of the two independent variables on subjective ratings of difficulty (cognitive load). The results showed a statistically significant main effect for guidance, $F(1, 64) = 88.08$, $MSE = 237.19$, $p < .001$, partial $\eta^2 = .58$, and for variability, $F(1, 64) = 11.06$, $MSE = 29.78$, $p = .001$, partial $\eta^2 = .15$. As anticipated, the subjective ratings of difficulty were less for the worked-examples groups compared to the problem-solving groups, and less for the low-variability groups compared to the high-variability groups. These results firstly indicate that a lower cognitive load was imposed on learners who studied fully-guided worked examples compared to learners who generated solutions to problem-solving tasks (without any guidance) and secondly, that learners who worked on low-variability tasks experienced a lower cognitive load compared to learners who worked on high-variability tasks. The variability by guidance interaction for cognitive load was not statistically significant, $F(1,64) = 1.97$, $MSE = 5.31$, $p = .17$, partial $\eta^2 = .03$.

4.5.4.4 Discussion.

Experiment 4 tested the hypotheses that when learners study worked examples (compared to generating problem solutions without any guidance), and when learners are provided with high-variability tasks (compared to low-variability tasks) under worked-examples conditions, they will attain higher levels of conceptual understanding and superior transfer skills. Neither hypothesis was supported since the worked example effect (the first part of Hypothesis 1) and variability effect (the first part of Hypothesis 2) were not obtained. These analyses showed no effect on the Post-Test performance scores generated by the provision or absence of guided instruction, and the completion of high- or low-variability learning tasks. The lack of significance on the Post-Test performance scores may have been due to the sample size being too low.

The higher group averages of the class test scores (which measured prior knowledge) in Experiments 4 (83.77%, 80.63%, 75.78%, and 83.65%), compared to Experiment 2 (61.12%, 53.67%, 68.94%, and 69.73%), provide a possible explanation as to why the worked example effect was not obtained in Experiment 4. More specifically, reference can be made to the expertise reversal effect: as expertise increases, the worked example effect decreases. This effect suggests why the worked example effect was not obtained in Experiment 4 – many of the participants may have been too knowledgeable to demonstrate a worked example effect.

Despite the failure to obtain a worked example effect, similar to Experiment 2, there was a benefit to studying worked examples compared to solving problems. The significant difference in cognitive load experienced by the participants in the worked-examples groups compared to the problem-solving groups demonstrated a significant advantage for participants in the worked-examples groups compared to the problem-solving groups. Learners could understand the high- and low-variability tasks more easily by studying worked examples, compared to problem solving, as a learning device. On the other hand, learners may have found it difficult to simultaneously deal with the extra WM resources required when generating potential solution steps. Hence, the last part of Hypothesis 1 was supported.

Despite the absence of a worked example effect, it seems that the modified worked examples in Experiment 4 successfully reduced redundant and split-source information. A higher effect size of .58 for the subjective ratings of difficulty in the worked-examples condition was obtained, compared to .36 for the worked-examples condition in Experiment 2.

A major observation to emerge from the findings in Experiment 4 was the significant difference in cognitive load experienced by the participants in the high-variability groups compared to the low-variability groups. This outcome is contrary to the findings in Experiments 1 and 2, where the similarity and dissimilarity of the low- and high-variability tasks did not have an impact on cognitive load. A possible explanation for obtaining the last part of Hypothesis 2 in the current experiment could be attributed to the participants having more prior mathematical knowledge (according to the higher group averages of the class test scores in the current experiment, as discussed previously). Having greater prior knowledge suggests there are fewer unfamiliar elements needed to be incorporated into existing schemas, thus making more cognitive resources available to deal with the low-variability tasks which consisted of a lower level of element interactivity, compared to the high-variability tasks which may have caused cognitive overload. That is, learners in the low-variability – problem-solving group were less likely to use means-ends analysis compared to the learners in the high-variability – problem-solving group, and schema acquisition and automation may have occurred more quickly for the learners in the worked-examples – low-variability group compared to the worked-examples – high-variability group.

The hypotheses in relation to subjective ratings of difficulty were confirmed. The results for subjective ratings of difficulty demonstrated a significant advantage for participants in the worked-examples groups and the low-variability groups. This is in line with CLT which argues the superiority of worked examples to problem solving when learners do not have sufficient prior knowledge in the domain, and the lower mental effort required to identify a surface match between similarly structured (low-variability) tasks without the need to go any further, compared to processing high-variability tasks that require more mental effort to process the deeper features until underlying common features are found.

4.6 Summary of Chapter 4

This chapter explained the method of analysis, reported the results for the method used, and discussed whether the hypotheses were supported, for each of the four experiments comprising the present study. Experiments 1, 2 and 4 investigated the variability effect under worked-examples conditions, and Experiment 3 investigated the variability effect according to learners' level of prior knowledge in the domain.

The next chapter will summarise the objectives and major findings of the present study, discuss the limitations of the present study, share the educational implications of the present findings, and provide a scope for future research on the variability effect. Finally, the general conclusion, which brings this thesis to a close, will focus on how the present study contributes to existing research on the variability effect.

Chapter 5: General Discussion

5.1 Introduction to Chapter 5

The first part of this chapter summarises the empirical findings of the present study in relation to its testing of the worked example effect, the variability effect, the expertise reversal effect, and the subjective rating of difficulty scale used to measure cognitive load. It then identifies some of the limitations of the study, clarifies the study's major educational implications and suggests potential future research. Finally, the significance of the study is emphasised with a focus on the insights gained into the variability effect.

5.2 Summary of the Study

The present study investigated if a relationship exists between high-variability or low-variability mathematical tasks and post-test performance. In particular, if the optimal level of variability of tasks alters according to levels of instructional guidance or levels of learner expertise.

The conclusion reached in Experiments 1, 2 and 4 was that learners' exposure to the worked-examples conditions did not result in higher post-test performance compared to the problem-solving conditions. Despite the lower subjective ratings of difficulty for the worked-examples groups in Experiment 1, suggesting lower extraneous cognitive load imposed by studying the worked examples compared to problem-solving equivalent tasks, the failure to obtain a worked example effect may have been due to the participants not having sufficient cognitive capacity to engage in germane activities to improve learning. In contrast, the findings in Experiment 2 suggest that all participants may have been more knowledgeable in the domain than expected because of the variability effect that was obtained for both the worked-examples and problem-solving groups. It seems that the participants in Experiment 2 had sufficient germane resources available to deal with the higher element interactivity of the high-variability tasks (under the worked-examples and problem-solving conditions), compared to the low-variability tasks, despite the students rating the study of the explicit solution steps in the worked examples as being easier to deal with compared to the unguided problem-solving of the equivalent tasks. Experiment 4 also failed to obtain the worked example effect, despite using worked examples with less redundant and split-attention information in the worked examples that were used in Experiment 2. It seems that reducing the extraneous cognitive load that interfered with learning (by removing the redundant and split-source information contained in the Experiment 2 worked examples) did not result in the load being replaced by germane load which would in turn have resulted in a significant improvement in learning.

In Experiment 2, the level of variability proved to be a major mediator between the tasks during the Learning Phase and the subsequent performance on the Post-Test transfer tasks. Completing high-variability tasks during the Learning Phase, in the worked-examples or problem-solving groups, effectively boosted the ability of learners to solve novel problems in the Post-Test. The variability effect that was obtained in Experiment 2 may have been due to most learners having the ability to cope with the larger number of interacting pieces of information in the high-variability tasks in order to allow for efficient learning. The diverse Post-Test results in the worked-examples and problem-solving conditions of Experiment 2 within the high-variability condition suggested that ability levels were not given due consideration in Experiment 2, and hence were the driving force for separating participants into novice and expert groups in Experiment 3.

In Experiment 3, the participants' prior knowledge in the target domain was pre-tested and used as a measure of expertise. Using a median split, by transforming the Pre-Test scores (continuous variables) into categorical variables (expert or novice groups), the Post-Test performance scores did not show a main effect for variability. Rather, the effectiveness of the high- or low-variability tasks during the Learning Phase was dependent on the participants' prior knowledge levels. The findings revealed that learners' exposure to high- or low-variability tasks resulted in different Post-Test performance scores based on their level of expertise. In particular, a classic expertise reversal effect was obtained in Experiment 3. When novices and experts were exposed to high- or low-variability tasks, it became evident that low-variability tasks worked better for novices and had no, or possibly adverse, effects for experts. In the opposite manner, exposure to high-variability tasks worked better for experts because the tasks were higher in element interactivity for novices but lower in element interactivity for experts.

Experiment 3 revealed that providing novices with high-variability tasks during the Learning Phase seemed to overwhelm their WM, possibly due to their lack of schema-based knowledge in the domain. Novices benefitted more from completing low-variability tasks during the Learning Phase than from high-variability tasks because the low-variability tasks assisted the novices to create long-term schemas. In contrast, experts were able to complete high-variability tasks more effectively than low-variability tasks because they had acquired more schema-based knowledge in the domain compared to the novices. Perhaps completing low-variability tasks had negative consequences for experts because these tasks contained redundant information. In particular, it seems that the experts' higher level of prior knowledge in the domain was critical in helping them solve the Post-Test transfer questions because these novel questions required a more flexible representation of knowledge. As a result, the reversal of effectiveness of providing high- or low-variability instruction produced a reverse variability effect because of the differing levels of prior knowledge of more-experienced and less-experienced learners. Under these circumstances, a more substantial conclusion was reached regarding the variability effect in Experiment 3. Two levels of expertise, the novice condition and the expert condition, made it possible to identify that high-variability tasks were a better instructional strategy for teaching experts, and low-variability tasks were a better instructional strategy for teaching novices.

5.3 Limitations of the Study

Although the study generated significant results, these results need to be interpreted with caution because they are somewhat limited by a relatively small sample size. Experiment 2, which produced the variability effect, comprised 103 participants, and Experiment 3, which produced the reverse variability effect, comprised 56 participants. If the sample size of Experiment 3 was as large as Experiment 2, the generalisability of the findings in Experiment 3 would hold for a better theoretical understanding of when high- or low-variability tasks enhance understanding for more-knowledgeable and less-knowledgeable learners.

Another limitation of this study is that the findings may not hold for learners in subject domains other than mathematics and for younger learners (e.g., elementary school and junior high school) because the current research was conducted with older mathematics learners (at secondary and tertiary levels). This is particularly important because it cannot be assumed that the present results are generalisable across all subject learning areas and at all learner age levels.

Despite the successful use of a subjective, one-item scale measure of cognitive load in over 100 well-documented CLT experiments, this measurement tool can be considered questionable in the current study. The participants did not consistently present with subjective ratings of difficulty that support the assumptions based on CLT in all four experiments. Subjective ratings of difficulty were lower for all the worked-examples groups compared to the problem-solving groups (in Experiments 1, 2 and 4), as anticipated. However, the ratings were lower for the low-variability groups compared to the high-variability groups only for the novice groups and not the expert groups in Experiment 3, and for both the worked-examples and problem-solving groups in Experiment 4 (and not in Experiments 1 and 2 as anticipated). To explain this conflicting theoretical position, we can draw on Ayres (2018), who postulated that a subjective measure "is a self-reflection on a cognitive process and is somewhat relativistic in nature, being highly dependent upon prior knowledge in the domain" (p. 21). Perhaps using a multi-item scale could produce responses that enable learners to self-reflect more accurately, despite their level of expertise in the domain.

Furthermore, the cognitive load construct does not consider the psychological impact that learners' beliefs and goals have on their cognitive load perceptions (Bannert, 2002; Moreno, 2006). Up until now, CLT has mainly focused on cognitive principles of learning, and as a result, cognitive load research has not considered possible psychological effects arising from learners' interest, task engagement and emotions.

The reverse variability effect which prominently appeared in the findings of Experiment 3 aligned with the expertise reversal effect. However, the measure used to place the participants in either a novice or expert group is questionable. Using the median Pre-Test score to split the participants into two groups may not have been the most valid measure of expertise due to the Pre-Test scores that were closely positioned above and below the median. More specifically, median splits become problematic when the score just above the median is considered the same as the scores further away on the higher end, but considered different to the score just below the median. However, in their perspective of the literature on artificial categorisation, DeCoster, Gallucci, and Iselin (2011) argued that choosing to artificially categorise continuous variables was commonly used by researchers who examined interaction effects. Language used to describe findings when using artificial categorisation should express the “relations between the abstract constructs” (DeCoster et al., 2011, p. 205). Accordingly, this prescription was adhered to in Experiment 3 where novice and expert groups reflected the relation between their prior knowledge in the domain, by referring to the participants as being less- and more-experienced learners respectively. Furthermore, in Iacobucci, Posavac, Kardes, Schneider, and Popovich’s (2015) analysis of the statistical properties of a median split, they argued that using a median split in a factorial experimental design (followed by the use of ANOVA to model and report the findings) was “perfectly legitimate” (p. 662). However, despite the possibility of producing Type 1 errors, a median split was used in Experiment 3 given Iacobucci et al.’s (2015) recommendation that this practice is suited for research in group differences.

5.4 Educational Implications

The findings from this study displayed the benefits of task variability to improve mathematical educational practice. In particular, the findings proved useful in expanding our understanding of the expertise reversal effect by identifying a reverse variability effect. This innovative aspect suggests that high and low task variability can act as a powerful method for fostering better learning for experts and novices, respectively. For more-experienced learners in a domain, there is a cost when they are presented with low-variability tasks because they do not require repetitive tasks to learn a particular solution method. On the other hand, less-experienced learners in a domain benefit from low-variability tasks because they require more practice to assist with the understanding of the underlying solution method.

Since processing by experts is more abstract than it is for novices, an important practical implication from the current findings is that processes that underlie experts' superior performance must be considered when selecting the level of task variability. High-variability tasks must be presented to more-experienced learners because they have available schemas to modify solution methods to deal with tasks that differ from one another on all dimensions. Elements that vary between high-variability tasks should highlight meaningful and extensive concepts to enable experts to compare and apply solution methods more broadly, to help develop their competence. In contrast, low-variability tasks must be presented to less-experienced learners to assist them to capitalise on familiar solution methods which can help them identify similar problem features. Hence, the current findings suggested that greater effort is needed for mathematics educators to design learning tasks that align with learners' existing mathematical abilities in order to facilitate efficient and effective learning.

5.5 Future Research

The present study contributes to the current literature by providing deeper insight into the effectiveness of altering levels of variability with levels of guidance and levels of expertise. However, to extend the current literature about the variability effect, further replication studies are necessary that involve different instructional materials and learners of different ages. Even though the present study aimed to be the first to thoroughly examine the relationship between levels of variability and expertise, a natural progression of the present empirical investigation would be to repeat the study with a larger sample to potentially provide more conclusive evidence of the reverse variability effect.

Since the majority of previous studies on variability were published over a decade ago, more work is required to explore whether altering the level of task variability can provide a powerful way to foster transfer of learning. It seems reasonable to assert that greater attention to the mechanism of transfer will help reveal its association to human cognition. Not only is fostering successful transfer associated with learning but it can also reveal other aspects of human cognition such as memorising, reasoning, categorising, and problem solving. More specifically, there can be further exploration of the theoretical questions about what elements of the task should vary and what elements should remain the same to promote learners to spontaneously solve novel tasks. Moreover, research on the robustness of transfer after an extended period of time is required to distinguish between temporary and permanent transfer effects. This will shed some light on how long transfer can persist for experts and novices completing high- or low-variability tasks after training tasks have been mastered.

Given that most CLT research has been conducted in mathematics and science-based domains (such as computing, physics and engineering), and far less in non-scientific content domains (such as foreign language acquisition, English literature and music instruction), further research on the variability effect should include these non-scientific domains. Based on the present findings, carrying out experimental investigations in non-scientific content domains could yield clearer practical implications about the positive effects of altering the variability of learning tasks to benefit learners with different levels of expertise in specific domains.

Whilst the current findings confirm that to generate higher post-test performance, it is advisable to decrease or increase the variability of learning tasks depending on learners' levels of expertise, several questions remain unanswered in relation to randomly sequencing those learning tasks (interleaving). The effect of interleaving can be made clearer if learners are presented with adjacent high- or low-variability learning tasks that require them to practise different versions of the constituent skills. For example, the use of a random schedule of learning tasks that requires students to apply skills at the same level of difficulty across different mathematics topics could boost their ability to learn the critical features of concepts and skills. Studying high- or low-variability tasks across different mathematics topics could potentially strengthen categorisation and problem-solving skills because the learner is required to make associations by differentiating between concepts. Even though there are numerous studies that have assessed the efficacy of interleaving, further research could usefully explore if combining random sequencing with high- or low-variability tasks, tailored for experts and novices respectively, can improve the transfer of learning.

5.6 General Conclusion

Part I of this thesis explored the key areas of human cognitive architecture, in particular WM, where the temporary storage of information is used to process complex cognitive tasks; and LTM, where unlimited amounts of information, in the form of schemas, are stored. This was followed by an explanation of how CLT began as an instructional theory, and how the basic theoretical assumptions underlying CLT are based on our knowledge of human cognitive architecture. An overview of the extensive empirical development of CLT over the past three decades was provided. This elucidated how CLT has become an influential theoretical framework within educational psychology that provides a powerful tool for synthesising learning, cognition and instructional design, and successfully generating an array of cognitive load effects. As argued by Kirschner (2002) in the special issue on the instructional implications of CLT, following CLT guidelines can assist with presenting information in ways that encourage learners to optimise their intellectual performance.

Part I concluded with a detailed discussion of the categories of cognitive load (intrinsic, extraneous and germane cognitive load), managing cognitive load through instructional design, using subjective measures of cognitive load, and five cognitive load effects: the worked example effect, the expertise reversal effect, the redundancy effect, the split-attention effect, and the variability effect. These five instructional design effects were reviewed in detail in terms of how they influence cognitive load to improve learning and instruction. These effects were revisited in Part II of this thesis which comprised the present study.

Part II provided further empirical evidence about the nature of the variability effect. Within the framework of CLT, the study investigated how the capacity to acquire knowledge and skills could be developed for mathematics learners by altering the level of task variability. Four experiments were conducted to examine the effect of low- and high-variability tasks for learning, solving problems, and transferring skills. Experiments 1, 2 and 4 explored the effect between low- and high-variability tasks when studying worked examples or attempting to solve problems; and Experiment 3 further explored the effect between low- and high-variability tasks with less- and more-experienced learners. The significant findings of the present study revealed that variability should only be increased once learners' levels of knowledge have advanced sufficiently to allow them to process the increased element interactivity associated with increased variability.

From a practical perspective, the major educational implication of these findings are that less-experienced learners in a domain should initially be presented with low-variability tasks. Low-variability tasks with their low levels of element interactivity can assist in the acquisition of essential problem-solving concepts and procedures associated with a given area of study. Once these concepts and procedures have been acquired, it is appropriate to acquaint learners with the various types of problems to which the concepts and procedures apply. At this point, with the basic knowledge stored in LTM, rather than having to be processed in WM, learners should have sufficient spare WM capacity to process the elements associated with variability. Until this point is reached, presenting learners with the basic concepts and procedures, and the various conditions to which they apply simultaneously, may overload WM. Further studies are needed to test this hypothesis.

References

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*(2), 181-214. doi:10.3102/00346543070002181

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology*, *95*, 774-783. doi:10.1037/0022-0663.95.4.774

Ayres, P. (2006). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology*, *20*, 287-298. doi:10.1002/acp.1245

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, *16*, 389-400. doi:10.1016/j.learninstruc.2006.09.001

Ayres, P. (2018). Subjective measures of cognitive load. In R. Z. Zheng (Ed.), *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (1st ed., pp. 9-28). New York, NY: Routledge. doi:10.4324/9781315296258

Ayres, P., & Sweller, J. (1990). Locus of difficulty in multistage mathematics problems. *The American Journal of Psychology*, *103*(2), 167-193. doi:10.2307/1423141

Baddeley, A. (1986). *Working memory*. New York, NY: Oxford University Press.

Baddeley, A. (1992). Working memory. *Science*, *255*, 556-559. doi:10.1126/science.1736359

Baddeley, A. D. (2001). Is working memory still working? *American Psychologist*, *56*, 851-864. doi:10.1037/0003-066X.56.11.851

Baddeley, A. D., & Dale, H. C. A. (1966). The effect of semantic similarity on retro-active interference in long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *5*, 417-420. doi:10.1016/S0022-5371(66)80054-3

Baddeley, A.D., & Hitch, G. J. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47-89. doi:10.1016/S0079-7421(08)60452-1

Bannert, M. (2002). Managing cognitive load – Recent trends in cognitive load theory. *Learning and Instruction*, *12*, 139-146. doi:10.1016/S0959-4752(01)00021-4

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612-637. doi:10.1037/0033-2909.128.4.612

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The role of discrimination and retrieval. *Memory and Cognition*, *41*, 392-402. doi:10.3758/s13421-012-0272-7

Blayney, P., Kalyuga, S., & Sweller, J. (2010). Interactions between the isolated-interactive elements effect and levels of learner expertise: Experimental evidence from an accountancy class. *Instructional Science*, *38*(3), 277-287. doi:10.1007/s11251-009-9105-x

Bratfisch, O., Borg, G., & Dornic, S. (1972). *Perceived item-difficulty in three tests of intellectual performance capacity* (*Tech. Rep. No. 29*). Stockholm, Sweden: Institute of Applied Psychology.

Broadbent, D. E. (1957). Immediate memory and simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, *9*(1), 1-11. doi:10.1080/17470215708416214

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, *10*(1), 12-21. doi:10.1080/17470215808416249

Brooks, L. W., & Dansereau, D. F. (1987). Transfer of information: An instructional perspective. In S. M. Cormier & J. D. Hagman (Eds.), *Transfer of learning*: *Contemporary research and applications* (pp. 121-150). San Diego, CA: Academic Press.

Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science*, *32*, 115-132. doi:10.1023/B:TRUC.0000021812.96911.c5

Brünken, R., Steinbacher, S., Plaas, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology*, *49*(2), 109-119. doi:10.1027//1618-3169.49.2.109

Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational & Psychology*, *86*(3), 360-367. doi:10.1037/0022-0663.86.3.360

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1147-1156. doi:10.1037/0278-7393.15.6.1147

Cerpa, N., Chandler, P., & Sweller, J. (1996). Some conditions under which integrated computer-based training software can facilitate learning. *Journal of Educational Computing Research*, *15*(4), 345-367. doi:10.2190/MG7X-4J8N-CKYR-P06T

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*, 293-332. doi:10.1207/s1532690xci0804_2

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, *62*(2), 233-246. doi:10.1111/j.2044-8279.1992.tb01017.x

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*, *10*(2), 151-170. doi:10.1002/(SICI)1099-0720(199604)10:2<151::AID-ACP380>3.0.CO;2-U

Chase, W. G., & Simon, H. A. (1973a). Perception in chess. *Cognitive Psychology*, *4*, 55-81. doi:10.1016/0010-0285(73)90004-2

Chase, W. G., & Simon, H. A. (1973b). The mind’s eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York, NY: Academic Press.

Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2018). Undesirable difficulty effects in the learning of high-element interactivity materials. *Frontiers in Psychology*, *9*(1483), 1-7. doi:10.3389/fpsyg.2018.01483

Chen, Z., & Mo, L. (2004). Schema induction in problem solving: A multidimensional analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 583-600. doi:10.1037/0278-7393.30.3.583

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145-182. doi:10.1016/0364-0213(89)90002-5

Chi, M. T. H., de Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439-477. doi:10.1016/0364- 0213(94)90016-7

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152. doi:10.1207/s15516709cog0502_2

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7-75). Hillsdale, NJ: Lawrence Erlbaum.

Chi, M. T. H., & Ohlsson, S. (2005). Complex declarative learning. In K. J. Holyoak, & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 371-399). New York, NY: Cambridge University Press.

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior*, *25*(2), 315-324. doi:10.1016/j.chb.2008.12.020

Clark, R. C., Nguyen, F., & Sweller, J. (2006). Efficiency in learning: Evidence-based guidelines to manage cognitive load. San Francisco, CA: Pfeiffer. doi:10.1002/pfi.4930450920

Clarke, T., Ayres, P., & Sweller, J. (2005). The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications. *Educational Technology Research and Development*, *53*(3), 15-24. doi:10.1007/BF02504794

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, *79*, 347-362. doi:10.1037/0022-0663.79.4.347

Cooper, G., Tindall-Ford, S., Chandler, P., & Sweller, J. (2001). Learning by imagining. *Journal of Experimental Psychology: Applied*, *7*(1), 68-82. doi:10.1037///1076-898X.7.1.68

Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2006). Towards a personalized task selection model with shared instructional control. *Instructional Science*, *34*, 399-422. doi:10.1007/s11251-005-5774-2

Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2011). Learner-controlled selection of tasks with different surface and structural features: Effects on transfer and efficiency. *Computers in Human Behavior*, *27*(1), 76-81. doi:10.1016/j.chb.2010.05.026

Cormier, S. M., & Hagman, J. D. (Eds.). (1987). *Transfer of learning*: *Contemporary research and applications*. San Diego, CA: Academic Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87-114. doi:10.1017/S0140525X01003922 de Croock, M. B. M., & van Merriënboer, J. J. G. (2007). Paradoxical effects of information presentation formats and contextual interference on transfer of a complex cognitive skill. *Computers in Human Behavior*, *23*(4), 1740-1761. doi:10.1016/j.chb.2005.10.003

de Groot, A. D. (1965). *Thought and choice in chess*. The Hague, Netherlands: Mouton.

DeCoster, J., Gallucci, M., & Iselin, A.-M. R. (2011). Best practices for using median splits, artificial categorization, and their continuous alternatives. *Journal of Experimental Psychopathology*, *2*(2), 197-209. doi:10.5127/jep.008310

DeLeeuw, K. E., & Mayer, R. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, *100*(1), 223-234. doi:10.1037/0022-0663.100.1.223

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*(4448), 1181-1182. doi:10.1126/science.7375930

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211-245. doi:10.1037/0033-295X.102.2.211

Florax, M., & Ploetzner, R. (2010). What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. *Learning and Instruction*, *20*(3), 216-224. doi:10.1016/j.learninstruc.2009.02.021

Geary, D. C. (2008). An evolutionary informed education science. *Educational Psychologist*, *43*(4), 179-195. doi:10.1080/00461520802392133

Gentner, D., Loewenstein, J., & Hung, B. (2007). Comparison facilitates children's learning of names for parts. *Journal of Cognition and Development*, *8*(3), 285-307. doi:10.1080/15248370701446434

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development*, *14*(4), 487-513. doi:10.1016/S0885-2014(99)00016-7

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science*, *32*, 33-58. doi:10.1023/B:TRUC.0000021809.10236.71

Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., & Eysink, T. H. S. (2009). Learning with hypermedia: The influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior*, *25*(2), 360-370. doi:10.1016/j.chb.2008.12.015

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1-38. doi:10.1016/0010-0285(83)90002-6

Gick, M. L., & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In S. M. Cormier & J. D Hagman (Eds.), *Transfer of learning*: *Contemporary research and applications* (pp. 9-46). San Diego, CA: Academic Press.

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, *16*(6), 511-525. doi:10.1016/j.learninstruc.2006.10.001

Ginns, P., Chandler, P., & Sweller, J. (2003). When imagining information is effective. *Contemporary Educational Psychology*, *28*(2), 229-251. doi:10.1016/S0361-476X(02)00016-4

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Science*, *5*(6), 236-243. doi:10.1016/S1364-6613(00)01662-4

Hardiman, P. T., Dufresne, R., & Mestre, J. P. (1989). The relation between problem categorization and problem solving among experts and novices. *Memory & Cognition*, *17*(5), 627-638. doi:10.3758/BF03197085

Hayes, J. R., & Simon, H. A. (1977). Psychological differences among problem isomorphs. In N. J. Castellan, Jr., D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory* (Vol. 2, pp. 21-41). Hillsdale, NJ: Lawrence Erlbaum.

Healy, A. F., & McNamara, D. S. (1996). Verbal learning and memory: Does the modal model still work? *Annual Review of Psychology*, *47*, 143-172. doi:10.1146/annurev.psych.47.1.143

Helsdingen, A. S., van Gog, T., & van Merriënboer, J. J. G. (2011a). The effects of practice schedule on learning a complex judgement task. *Learning and Instruction*, *21*(1), 126-136. doi:10.1016/j.learninstruc.2009.12.001

Helsdingen, A., van Gog, T., & van Merriënboer, J. (2011b). The effects of practice schedule and critical thinking prompts on learning and transfer of a complex judgement task. *Journal of Educational Psychology*, *103*(2), 383-398. doi:10.1037/a0022370

Holmqvist, M., Gustavsson, L., & Wernberg, A. (2007). Generative learning: Learning beyond the learning situation. *Educational Action Research*, *15*(2), 181-208. doi:10.1080/09650790701314684

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, *15*(4), 332-340. doi:10.3758/BF03197035

Hummel, H. G. K., Paas, F., & Koper, E. J. R. (2004). Cueing for transfer in multimedia programmes: Process worksheets vs. worked-out examples. *Journal of Computer Assisted Learning*, *20*(5), 387-397. doi:10.1111/j.1365-2729.2004.00098.x

Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015). Toward a more nuanced understanding of the statistical properties of a median split. *Journal of Consumer Psychology*, *25*(4), 652-665. doi:10.1016/j.jcps.2014.12.002

James, W. (1890). *The principles of psychology*, Vols. 1 and 2. London: Macmillan.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*, 509-539. doi:10.1007/s10648-007-9054-3

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, *23*(1), 1-19. doi:10.1007/s10648-010-9150-7

Kalyuga, S. (2012). Instructional benefits of spoken words: A review of cognitive load factors. *Educational Research Review*, *7*(2), 145-159. doi:10.1016/j.edurev.2011.12.002

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*, 23-31. doi:10.1207/S15326985EP3801_4

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors*, *40*(1), 1-17. doi:10.1518/001872098779480587

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, *92*(1), 126-136. doi:10.1037/0022-0663.92.1.126

Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology*, *21*(1), 5-23. doi:10.1080/01443410124681

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors*, *46*, 567-581. doi:10.1518/hfes.46.3.567.50405

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, *93*, 579-588. doi:10.1037//0022-0663.93.3.579

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science*, *38*, 209-215. doi:10.1007/s11251-009-9102-0

Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review*, *24*(2), 313-337. doi:10.1007/s10648-012-9195-x

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, *96*(3), 558-568. doi:10.1037/0022-0663.96.3.558

Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development*, *53*(3), 83-93. doi:10.1007/BF02504800

Kester, L., Kirschner, P. A., & van Merriënboer, J. J. G. (2005). The management of cognitive load during complex cognitive skill acquisition by means of computer-simulated problem solving. *British Journal of Educational Psychology*, *75*, 71-85. doi:10.1348/000709904X19254

Kester, L., Kirschner, P. A., & van Merriënboer, J. J. G. (2006). Just-in-time information presentation: Improving learning a troubleshooting skill. *Contemporary Educational Psychology*, *31*, 167-185. doi:10.1016/j.cedpsych.2005.04.002

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction*, *12*(1), 1-10. doi:10.1016/S0959-4752(01)00014-7 I

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. doi:10.1207/s15326985ep4102_1

Kotovsky, K., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, *67*(6), 2797-2822. doi:10.1111/j.1467-8624.1996.tb01889.x

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*(2), 248-294. doi:10.1016/0010-0285(85)90009-X

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Models of competence in solving physics problems. *Cognitive Science*, *4*, 317-345. doi:10.1016/S0364-0213(80)80008-5

Leppink, J., Paas, F., van der Vleuten, C. P. M., van Gog, T., & van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*, 1058-1072. doi:10.3758/s13428-013-0334-1

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction*, *30*, 32-42. doi:10.1016/j.learninstruc.2013.12.001

Lim, J., Reiser, R. A., & Olina, Z. (2009). The effects of part-task and whole-task instructional approaches on acquisition and transfer of a complex cognitive skill. *Educational Technology, Research and Development*, *57*(1), 61-77. doi:10.1007/s11423-007-9085-y

Liu, T. C., Lin, Y. C., Gao, Y., Yeh, S. C., & Kalyuga, S. (2015). Does the redundancy effect exist in electronic slideshow assisted lecturing? *Computers & Education*, *88*, 303-314. doi:10.1016/j.compedu.2015.04.014

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology*, *88*(1), 49-63. doi:10.1037/0022-0663.88.1.49

Marini, A., & Genereux, R. (2015). The challenge of teaching for transfer. In A. McKeough, J. Lupart, & A. Marini (Eds.), *Teaching for transfer: Fostering generalization in learning* (pp. 1-19). New York, NY: Routledge.

Mawer, R. F., & Sweller, J. (1982). Effects of subgoal density and location on learning during problem solving. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *8*(3), 252-259. doi:10.1037/0278-7393.8.3.252

Mayer, R. E. (1975). Information processing variables in learning to solve problems. *Review of Educational Research*, *45*(4), 525-541. doi:10.3102/00346543045004525

Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology*, *88*(1), 64-73. doi:10.1037/0022-0663.88.1.64

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43-52. doi:10.1207/S15326985EP3801_6

McKeough, A., Lupart, J., & Marini, A. (Eds.). (2015). *Teaching for transfer: Fostering generalization in learning*. New York, NY: Routledge.

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

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York, NY: Cambridge University Press.

Moreno, R. (2006). When worked examples don't work: Is cognitive load theory at an impasse? *Learning and Instruction*, *16*, 170-181. doi:10.1016/j.learninstruc.2006.02.006

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of*

Experimental Psychology: *General*, *131*(1), 5-15. doi:10.1037//0096-3445.131.1.5

Nickerson, R. S., Perkins, D. N., & Smith, E. E. (1985). *The teaching of thinking*. Hillsdale, NJ: Lawrence Erlbaum.

Nievelstein, F., van Gog, T., van Dijck, G., & Boshuizen, H. P. A. (2013). The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. *Contemporary Educational Psychology*, *38*(2), 118-125. doi:10.1016/j.cedpsych.2012.12.004

Oksa, A., Kalyuga, S., & Chandler, P. (2010). Expertise reversal effect in using explanatory notes for readers of Shakespearean text. *Instructional Science*, *38*(3), 217-236. doi:10.1007/s11251-009-9109-6

Owen, E., & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Educational Psychology*, *77*(3), 272-284. doi:10.1037/0022-0663.77.3.272

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429-434. doi:10.1037/0022-0663.84.4.429

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*, 1-4. doi:10.1207/S15326985EP3801_1

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, *32*, 1-8. doi:10.1023/B:TRUC.0000021806.17516.d0

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

Paas, F., Van Gerven, P. W. M., & Wouters, P. (2007). Instructional efficiency of animation: Effects of interactivity through mental reconstruction of static key frames. *Applied Cognitive Psychology*, *21*, 783-793. doi:10.1002/acp.1349

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*(1), 122-133. doi:10.1037/0022-0663.86.1.122

Paas, F. G. W. C., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual & Motor Skills*, *79*, 419-430. doi:10.2466/pms.1994.79.1.419

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*(2), 117-175. doi:10.1207/s1532690xci0102_1

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193-198. doi:10.1037/h0049234

Pirolli, P. L., & Anderson, J. R. (1985). The role of learning from examples in the acquisition of recursive programming skills. *Canadian Journal of Psychology*, *39*(2), 240-272. doi:10.1037/h0080061

Plass, J. L., Moreno, R., & Brünken, R. (Eds.). (2010). *Cognitive load theory*. Cambridge, UK: Cambridge University Press.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, *12*, 61-86. doi:10.1016/S0959-4752(01)00016-0

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, *88*(1), 144-161. doi:10.1037/0022-0663.88.1.144

Quilici, J. L., & Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology*, *16*, 325-342. doi:10.1002/acp.796

Ranzijn, F. J. A. (1991). The number of video examples and the dispersion of examples as instructional design variables in teaching concepts. *The Journal of Experimental Education*, *59*(4), 320-330. doi:10.1080/00220973.1991.10806570

Reed, S. K. (1989). Constraints on the abstraction of solutions. *Journal of Educational Psychology*, *81*, 532-540. doi:10.1037/0022-0663.81.4.532

Reeves, L. M., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin*, *115*(3), 381-400. doi:10.1037/0033-2909.115.3.381

Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and Instruction*, *16*(2), 92-103. doi:10.1016/j.learninstruc.2006.02.008

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*, 1-29. doi:10.1207/s15516709cog2101_1

Renkl, A. (1999). Learning mathematics from worked-out examples: Analyzing and fostering self-explanations. *European Journal of Psychology of Education*, *14*(4), 477-488. doi:10.1007/BF03172974

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skills acquisition: A cognitive load perspective. *Educational Psychologist*, *38*, 15-22. doi:10.1207/S15326985EP3801_3

Renkl, A., Atkinson, R. K., & Große, C. S. (2004). How fading worked solution steps works – A cognitive load perspective. *Instructional Science*, *32*, 59-82. doi:10.1023/B:TRUC.0000021815.74806.f6

Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *The Journal of Experimental Education*, *70*(4), 293-315. doi:10.1080/00220970209599510

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology*, *23*, 90-108. doi:10.1006/ceps.1997.0959

Richland, L. E., Holyoak, K. J., & Stigler, J. W. (2004). Analogy use in eighth-grade mathematics classrooms. *Cognition and Instruction*, *22*(1), 37-60. doi:10.1207/s1532690Xci2201_2

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, *99*(3), 561-574. doi:10.1037/0022-0663.99.3.561

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology*, *101*(3), 529-544. doi:10.1037/a0014224

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, *101*(4), 836-852. doi:10.1037/a0016026

Robins, S., & Mayer, R. E. (1993). Schema training in analogical reasoning. *Journal of Educational Psychology*, *85*(3), 529-538. doi:10.1037/0022-0663.85.3.529

Ross, B., & Kennedy, P. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning*, *Memory*, *and Cognition*, *16*(1), 42-55. doi:10.1037/0278-7393.16.1.42

Royer, J. M. (1979). Theories of the transfer of learning. *Educational Psychologist*, *14*(1), 53-69. doi:10.1080/00461527909529207

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist*, *24*(2), 113-142. doi:10.1207/s15326985ep2402_1

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science*, *43*(1), 93-114. doi:10.1007/s11251-014-9328-3

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychological Review*, *84*, 1-66. doi:10.1037/0033-295X.84.1.1

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(5), 484-494. doi:10.1037/0278-7393.8.5.484

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*(4), 469-508. doi:10.1007/s10648-007-9053-4

Schraw, G. (2006). Knowledge: Structures and processes. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 245-263). New York, NY: Routledge.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475-522. doi:10.1207/s1532690xci1604_4

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology*, *22*, 261-273. doi:10.1080/00335557043000203

Shiffrin, R. M. (1975). Short-term store: The basis for a memory system. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (Vol. 1, pp. 193-218). Hillsdale, NJ: Lawrence Erlbaum.

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review*, *76*, 179-193. doi:10.1037/h0027277

Silver, E. A. (1981). Recall of mathematical problem information: Solving related problems. *Journal for Research in Mathematics Education*, *12*(1), 54-64. doi:10.2307/748658

Silver, E. A., Ghousseini, H., Gosen, D., Charalambous, C., & Strawhun, B. T. F. (2005). Moving from rhetoric to praxis: Issues faced by teachers in having students consider multiple solutions for problems in the mathematics classroom. *Journal of Mathematical Behavior*, *24*, 287-301. doi:10.1016/j.jmathb.2005.09.009

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257-285. doi:10.1016/0364-0213(88)90023-7

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology*, *81*(4), 457-466. doi:10.1037/0022-0663.81.4.457

Sweller, J. (1993). Some cognitive processes and their consequences for the organisation and presentation of information. *Australian Journal of Psychology*, *45*(1), 1-8. doi:10.1080/00049539308259112

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295-312. doi:10.1016/0959-4752(94)90003-5

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Victoria: Australian Council for Educational Research.

Sweller, J. (2003). Evolution of human cognitive architecture. *Psychology of Learning and Motivation*, *43*, 215-266. doi:10.1016/S0079-7421(03)01015-6

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science*, *32*, 9-31. doi:10.1023/B:TRUC.0000021808.72598.4d

Sweller, J. (2006). Discussion of 'emerging topics in cognitive load research: Using learner and information characteristics in the design of powerful learning environments'. *Applied Cognitive Psychology*, *20*, 353-357. doi:10.1002/acp.1251

Sweller, J. (2008). Instructional implications of David C. Geary's evolutionary educational psychology. *Educational Psychologist*, *43*(4), 214-216. doi:10.1080/00461520802392208

Sweller, J. (2009). Cognitive bases of human creativity. *Educational Psychology Review*, *21*(1), 11-19. doi:10.1007/s10648-008-9091-6

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123-138. doi:10.1007/s10648-010-9128-5

Sweller, J. (2018). The role of independent measures of load in cognitive load theory. In R. Z. Zheng (Ed.), *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (1st ed., pp. 3-8). New York, NY: Routledge. doi:10.4324/9781315296258

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer. doi:10.1007/978-1-4419-8126-4

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*(3), 185-233. doi:10.1207/s1532690xci1203_1

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General*, *119*(2), 176-192. doi:10.1037/0096-3445.119.2.176

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*, 59-89. doi:10.1207/s1532690xci0201_3

Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology*, *4*, 434-458. doi:10.1177/147470490600400135

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251-296. doi:10.1023/A:1022193728205

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology*, *80*(4), 424-436. doi:10.1037/0022-0663.80.4.424

Thorndike, E. L. (1913). *Educational Psychology, Volume II: The psychology of learning*. New York, NY: Teachers College, Columbia University. doi:10.1037/13051-000

Thorndike, E. L., & Woodworth, R. S. (1901a). The influence of improvement in one mental function upon the efficiency of other functions. (I). *Psychological Review*, *8*(3), 247-261. doi:10.1037/h0074898

Thorndike, E. L., & Woodworth, R. S. (1901b). The influence of improvement in one mental function upon the efficiency of other functions. (II). The estimation of magnitudes. *Psychological Review*, *8*(4), 384-395. doi:10.1037/h0071280

Thorndike, E. L., & Woodworth, R. S. (1901c). The influence of improvement in one mental function upon the efficiency of other functions. (III). Functions involving attention, observation and discrimination. *Psychological Review*, *8*(6), 553-564. doi:10.1037/h0071363

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied*, *3*(4), 257-287. doi:10.1037/1076- 898X.3.4.257

Trafton, J. G., & Reiser, B. J. (1993). The contributions of studying examples and solving problems to skill acquisition. In M. Polson (Ed.), *Proceedings of the fifteenth annual conference of the cognitive science society* (pp. 1017-1022). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tulving, E., & Craik, F. I. M. (2000). *The oxford handbook of memory*. New York, NY: Oxford University Press.

Tulving, E., & Patterson, R. D. (1968). Functional units and retrieval processes in free recall. *Journal of Experimental Psychology*, *77*, 239-248. doi:10.1037/h0025788

Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology*, *91*, 334-341. doi:10.1037/0022-0663.91.2.334

Van Gerven, P. W. M., Paas, F., van Merriënboer, J. J. G., Hendriks, M., & Schmidt, H. G. (2003). The efficiency of multimedia learning into old age. *British Journal of Educational Psychology*, *73*, 489-505. doi:10.1348/000709903322591208 van Gog, T. (2011). Effects of identical example-problem and problem-example pairs on learning. *Computers & Education*, *57*, 1775-1779. doi:10.1016/j.compedu.2011.03.019

van Gog, T., & Kester, L. (2012). A test of the testing effect: Acquiring problem-solving skills from worked examples. *Cognitive Science*, *36*(8), 1532-1541. doi:10.1111/cogs.12002

van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, *36*(3), 212-218. doi:10.1016/j.cedpsych.2010.10.004

van Gog, T., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and frequency of mental effort measurement: Evidence in favor of repeated measures. *Applied Cognitive Psychology*, *26*(6), 833-839. doi:10.1002/acp.2883

van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, *43*(1), 16-26. doi:10.1080/00461520701756248

van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2004). Process-oriented worked examples: Improving transfer performance through enhanced understanding. *Instructional Science*, *32*, 83-98. doi:10.1023/B:TRUC.0000021810.70784.b0

van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction*, *16*(2), 154-164. doi:10.1016/j.learninstruc.2006.02.003

van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction*, *18*, 211-222. doi:10.1016/j.learninstruc.2007.03.003

van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, *22*(2), 155-174. doi:10.1007/s10648-010-9134-7

van Merriënboer, J. J. G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development*, *53*(3), 5-13. doi:10.1007/BF02504793

van Merriënboer, J. J. G., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: Balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology*, *20*, 343-352. doi:10.1002/acp.1250

van Merriënboer, J. J. G., & Kirschner, P. A. (2018). *Ten steps to complex learning: A systematic approach to four-component instructional design* (3rd ed.). New York, NY: Routledge. doi:10.1007/s11528-018-0254-0

van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist*, *38*(1), 5-13. doi:10.1207/S15326985EP3801_2

van Merriënboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147-177. doi:10.1007/s10648-005-3951-0

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. and Trans.). Cambridge, MA: Harvard University Press.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*(1), 1-39. doi:10.1207/s1532690xci0701_1

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, *72*(2), 89-104. doi:10.1037/h0021797

Yeung, A. S., Jin, P., & Sweller, J. (1998). Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology*, *23*(1), 1-21. doi:10.1006/ceps.1997.0951

Appendices

Appendix A

Experiment 1: Worked-Examples – High-Variability Handout

WORKED EXAMPLES - HIGH VARIABILITY

School of Mathematics and Statistics, UNSW Sydney

Investigator: Vicki Likourezos | Experiment 1, 2017

1. (i) Examine the following where a and b are real numbers:

$$\frac{4}{k(k+4)} = \frac{a}{k} + \frac{b}{k+4}, \text{ where } k \neq -4, 0.$$

To find a and b, we multiply the above identity by $k(k+4)$ to obtain:

$$4 = a(k+4) + bk.$$

By putting $k = 0$, we see that $4 = 4a \Rightarrow a = 1$
and by putting $k = -4$, we see that $4 = -4b \Rightarrow b = -1$.
Hence,

$$\frac{4}{k(k+4)} = \frac{1}{k} - \frac{1}{k+4}. \quad (1)$$

(ii) A telescoping sum is a sum which is written in such a way where most of the terms cancel and the rest are easy to add up.

To evaluate $\sum_{k=1}^{n} \frac{4}{k(k+4)}$, where $k \geq 1$, the sum may be written as a telescoping sum by applying the identity in (1) and writing out the terms as follows:

$$\begin{aligned}
\sum_{k=1}^{n} \frac{4}{k(k+4)} &= \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+4}\right) \\
&= \left(\frac{1}{1} - \frac{1}{5}\right) + \left(\frac{1}{2} - \frac{1}{6}\right) + \left(\frac{1}{3} - \frac{1}{7}\right) + \left(\frac{1}{4} - \frac{1}{8}\right) \\
&+ \left(\frac{1}{5} - \frac{1}{9}\right) + \left(\frac{1}{6} - \frac{1}{10}\right) + \left(\frac{1}{7} - \frac{1}{11}\right) + \left(\frac{1}{8} - \frac{1}{12}\right) \\
&+ \ ... + \left(\frac{1}{n-4} - \frac{1}{n}\right) + \left(\frac{1}{n-3} - \frac{1}{n+1}\right) \\
&+ \left(\frac{1}{n-2} - \frac{1}{n+2}\right) + \left(\frac{1}{n-1} - \frac{1}{n+3}\right) + \left(\frac{1}{n} - \frac{1}{n+4}\right)
\end{aligned}$$

By cancelling out the common terms, the sum simplifies to:

$$\sum_{k=1}^{n} \frac{4}{k(k+4)} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} - \frac{1}{n+1} - \frac{1}{n+2} - \frac{1}{n+3} - \frac{1}{n+4}.$$

(iii) The sum $\sum_{k=1}^{n} \frac{1}{k+4}$, where n is an integer, $n \geq 1$, can be transformed by shifting the summation index. By substituting $j = k+4$ we have:

$$\sum_{k=1}^{n} \frac{1}{k+4} = \sum_{j=c}^{d} \frac{1}{j}, \text{ where } c, d \text{ are integers and } c \leq d.$$

Using this substitution we have
Lower limit: when $k = 1$, we get $j = 1 + 4 = 5$.
Upper limit: when $k = n$, we get $j = n + 4$.
Hence $\sum_{k=1}^{n} \frac{1}{k+4} = \sum_{j=5}^{n+4} \frac{1}{j}$.
By replacing the variable j with the variable k we have

$$\sum_{k=1}^{n} \frac{1}{k+4} = \sum_{k=5}^{n+4} \frac{1}{k}. \quad (2)$$

(iv) Shifting the summation index is another useful technique that can be used to evaluate $\sum_{k=1}^{n} \frac{4}{k(k+4)}$, where $k \geq 1$.

By applying the identity in (1) and the substitution in (2) we have

$$\begin{aligned}
\sum_{k=1}^{n} \frac{4}{k(k+4)} &= \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+4} \right) \\
&= \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=1}^{n} \frac{1}{k+4} \\
&= \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=5}^{n+4} \frac{1}{k} \\
&= \left(\sum_{k=1}^{4} \frac{1}{k} + \sum_{k=5}^{n} \frac{1}{k} \right) - \left(\sum_{k=5}^{n} \frac{1}{k} + \sum_{k=n+1}^{n+4} \frac{1}{k} \right)
\end{aligned}$$

By cancelling out the common sums, the sum simplifies to:

$$\begin{aligned}
\sum_{k=1}^{n} \frac{4}{k(k+4)} &= \sum_{k=1}^{4} \frac{1}{k} - \sum_{k=n+1}^{n+4} \frac{1}{k} \\
&= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} - \frac{1}{n+1} - \frac{1}{n+2} - \frac{1}{n+3} - \frac{1}{n+4}.
\end{aligned}$$

2. (i) To evaluate $\sum_{k=1}^{n} \frac{-7k-6}{k(k+1)(k+2)}$, where $k \geq 1$, we can follow the same process we used to evaluate $\sum_{k=1}^{n} \frac{4}{k(k+4)}$ in parts 1(i) and (ii).

Firstly, we need to examine the following where a, b and c are real numbers:

$$\frac{-7k-6}{k(k+1)(k+2)} = \frac{a}{k} + \frac{b}{k+1} + \frac{c}{k+2}, \text{ where } k \neq -2, -1, 0.$$

To find a, b and c, we multiply the above identity by $k(k+1)(k+2)$ to obtain:

$$-7k - 6 = a(k+1)(k+2) + bk(k+2) + ck(k+1).$$

By putting $k = 0$, we see that $-6 = 2a \Rightarrow a = -3$.
By putting $k = -1$, we see that $1 = -b \Rightarrow b = -1$.
By putting $k = -2$, we see that $8 = 2c \Rightarrow c = 4$.
Hence,

$$\frac{-7k-6}{k(k+1)(k+2)} = -\frac{3}{k} - \frac{1}{k+1} + \frac{4}{k+2}. \quad (3)$$

To evaluate $\sum_{k=1}^{n} \frac{-7k-6}{k(k+1)(k+2)}$, where $k \geq 1$, the sum may be written as a telescoping sum by applying the identity in (3) and writing out the terms as follows:

$$\begin{aligned}
\sum_{k=1}^{n} \frac{-7k-6}{k(k+1)(k+2)} &= \sum_{k=1}^{n} \left(-\frac{3}{k} - \frac{1}{k+1} + \frac{4}{k+2} \right) \\
&= \left(-\frac{3}{1} - \frac{1}{2} + \frac{4}{3} \right) \\
&+ \left(-\frac{3}{2} - \frac{1}{3} + \frac{4}{4} \right) \\
&+ \left(-\frac{3}{3} - \frac{1}{4} + \frac{4}{5} \right) \\
&+ \left(-\frac{3}{4} - \frac{1}{5} + \frac{4}{6} \right) \\
&+ \ \ldots \\
&+ \left(-\frac{3}{n-2} - \frac{1}{n-1} + \frac{4}{n} \right) \\
&+ \left(-\frac{3}{n-1} - \frac{1}{n} + \frac{4}{n+1} \right) \\
&+ \left(-\frac{3}{n} - \frac{1}{n+1} + \frac{4}{n+2} \right)
\end{aligned}$$

By cancelling out the common terms, the sum simplifies to:

$$\begin{aligned}
\sum_{k=1}^{n} \frac{-7k-6}{k(k+1)(k+2)} &= -3 - \frac{1}{2} - \frac{3}{2} + \frac{4}{n+1} - \frac{1}{n+1} + \frac{4}{n+2} \\
&= -5 + \frac{3}{n+1} + \frac{4}{n+2}.
\end{aligned}$$

(ii) To evaluate $\sum_{k=1}^{n} \frac{-7k-6}{k(k+1)(k+2)}$, where $k \geq 1$, we can follow the same process we used to evaluate $\sum_{k=1}^{n} \frac{4}{k(k+4)}$ in parts 1(iii) and (iv). By using the technique of shifting the summation index, we can transform the sums $\sum_{k=1}^{n} \frac{1}{k+1}$ and $\sum_{k=1}^{n} \frac{1}{k+2}$, where n is an integer, $n \geq 1$.

By substuting $j = k+1$ we have

$$\sum_{k=1}^{n} \frac{1}{k+1} = \sum_{j=d}^{e} \frac{1}{j} \text{ where } d, e \text{ are integers and } d \leq e.$$

Using this substitution we have
Lower limit: when $k = 1$, we get $j = 1 + 1 = 2$.
Upper limit: when $k = n$, we get $j = n + 1$.
Hence $\sum_{k=1}^{n} \frac{1}{k+1} = \sum_{j=2}^{n+1} \frac{1}{j}$.
By replacing the variable j with the variable k we have

$$\sum_{k=1}^{n} \frac{1}{k+1} = \sum_{k=2}^{n+1} \frac{1}{k}. \quad (4)$$

By substuting $m = k+2$ we have

$$\sum_{k=1}^{n} \frac{1}{k+2} = \sum_{m=f}^{g} \frac{1}{m} \text{ where } f, g \text{ are integers and } f \leq g.$$

Using this substitution we have
Lower limit: when $k = 1$, we get $m = 1 + 2 = 3$.
Upper limit: when $k = n$, we get $m = n + 2$.
Hence $\sum_{k=1}^{n} \frac{1}{k+2} = \sum_{m=3}^{n+2} \frac{1}{m}$.
By replacing the variable m with the variable k we have

$$\sum_{k=1}^{n} \frac{1}{k+2} = \sum_{k=3}^{n+2} \frac{1}{k}. \quad (5)$$

By applying the identity in (3) and the substitutions in (4) and (5) we have

$$\begin{aligned}
\sum_{k=1}^{n} \frac{-7k-6}{k(k+1)(k+2)} &= \sum_{k=1}^{n}\left(-\frac{3}{k}-\frac{1}{k+1}+\frac{4}{k+2}\right)\\
&= -3\sum_{k=1}^{n}\frac{1}{k}-\sum_{k=1}^{n}\frac{1}{k+1}+4\sum_{k=1}^{n}\frac{1}{k+2}\\
&= -3\sum_{k=1}^{n}\frac{1}{k}-\sum_{k=2}^{n+1}\frac{1}{k}+4\sum_{k=3}^{n+2}\frac{1}{k}\\
&= -3\left(\frac{1}{1}+\frac{1}{2}+\sum_{k=3}^{n}\frac{1}{k}\right)-\left(\frac{1}{2}+\sum_{k=3}^{n}\frac{1}{k}+\frac{1}{n+1}\right)+4\left(\sum_{k=3}^{n}\frac{1}{k}+\frac{1}{n+1}+\frac{1}{n+2}\right)\\
&= -3\left(1+\frac{1}{2}\right)-\frac{1}{2}-3\sum_{k=3}^{n}\frac{1}{k}-\sum_{k=3}^{n}\frac{1}{k}+4\sum_{k=3}^{n}\frac{1}{k}-\frac{1}{n+1}+4\left(\frac{1}{n+1}+\frac{1}{n+2}\right)
\end{aligned}$$

By cancelling out the common sums, the sum simplifies to:

$\sum_{k=1}^{n} \frac{-7k-6}{k(k+1)(k+2)} = -5+\frac{3}{n+1}+\frac{4}{n+2}.$

3. (i) To evaluate $\sum_{k=1}^{n}\cos(2k-1)$, it is useful to use the products-to-sums formula $2\sin 1\cos(2k-1)=\sin(2k)-\sin(2k-2)$, which can be rewritten as

$$\cos(2k-1)=\frac{1}{2\sin 1}[\sin(2k)-\sin(2k-2)] \quad (6)$$

To evaluate $\sum_{k=1}^{n}\cos(2k-1)$, where $k\geq 1$, the sum may be written as a telescoping sum by applying the identity in (6) and writing out the terms as follows:

$$\begin{aligned}
\sum_{k=1}^{n}\cos(2k-1) &= \sum_{k=1}^{n}\frac{1}{2\sin 1}[\sin(2k)-\sin(2k-2)]\\
&= \frac{1}{2\sin 1}\sum_{k=1}^{n}[\sin(2k)-\sin(2k-2)]\\
&= \frac{1}{2\sin 1}[(\sin 2-\sin 0)+(\sin 4-\sin 2)+(\sin 6-\sin 4)+\\
&\quad \ldots+\sin[2(n-1)]-\sin[2(n-1)-2]+\sin(2n)-\sin(2n-2)]
\end{aligned}$$

By cancelling out the common terms, the sum simplifies to:

$$\begin{aligned}
\sum_{k=1}^{n}\cos(2k-1) &= \frac{1}{2\sin 1}(-\sin 0+\sin(2n))\\
&= \frac{\sin(2n)}{2\sin 1}.
\end{aligned}$$

(ii) To evaluate $\sum_{k=1}^{n} \cos(2k-1)$, we can follow the same process we used to evaluate $\sum_{k=1}^{n} \frac{4}{k(k+4)}$ in parts 1(iii) and (iv).
By using the technique of shifting the summation index, we can transform the sum $\sum_{k=1}^{n} \sin(2k-2)$ where n is an integer, $n \geq 1$.

By substituting $2k-2 = 2j$ we have

$$\sum_{k=1}^{n} \sin(2k-2) = \sum_{j=a}^{b} \sin(2j) \text{ where } a, b \text{ are integers and } a \leq b.$$

By simplifying this substitution from $2k-2 = 2j$ to $k-1 = j$ we have
Lower limit: when $k = 1$, we get $j = 1 - 1 = 0$.
Upper limit: when $k = n$, we get $j = n - 1$.
Hence $\sum_{k=1}^{n} \sin(2k-2) = \sum_{j=0}^{n-1} \sin(2j)$.
By replacing the variable j with the variable k we have

$$\sum_{k=1}^{n} \sin(2k-2) = \sum_{k=0}^{n-1} \sin(2k). \quad (7)$$

By applying the identity in (6) and the substitution in (7) we have

$$\begin{aligned}
\sum_{k=1}^{n} \cos(2k-1) &= \sum_{k=1}^{n} \frac{1}{2\sin 1}[\sin(2k) - \sin(2k-2)] \\
&= \frac{1}{2\sin 1} \sum_{k=1}^{n} [\sin(2k) - \sin(2k-2)] \\
&= \frac{1}{2\sin 1} \left[\sum_{k=1}^{n} \sin(2k) - \sum_{k=0}^{n-1} \sin(2k) \right] \\
& \frac{1}{2\sin 1} \left[\sum_{k=1}^{n-1} \sin(2k) + \sin(2n) - \left(\sin 0 + \sum_{k=1}^{n-1} \sin(2k) \right) \right]
\end{aligned}$$

By cancelling out the common terms, the sum simplifies to:

$$\begin{aligned}
\sum_{k=1}^{n} \cos(2k-1) &= \frac{1}{2\sin 1}(\sin(2n) - \sin 0) \\
&= \frac{\sin(2n)}{2\sin 1}.
\end{aligned}$$

Appendix B

Experiment 1: Problem-Solving – High-Variability Handout

PROBLEM SOLVING - HIGH VARIABILITY

School of Mathematics and Statistics, UNSW Sydney

Investigator: Vicki Likourezos — Experiment 1, 2017

1. (i) Find real numbers a and b so that
$\dfrac{4}{k(k+4)} = \dfrac{a}{k} + \dfrac{b}{k+4}$, where $k \neq -4, 0$.

(ii) Use the identity from (i) to complete a telescoping sum for
$\sum\limits_{k=1}^{n} \dfrac{4}{k(k+4)}$, where $k \geq 1$.

(iii) If n is an integer where $n \geq 1$, find integers c and d where $c \leq d$ so that $\sum_{k=1}^{n} \frac{1}{(k+4)} = \sum_{k=c}^{d} \frac{1}{k}$, where $k \geq 1$.

(iv) Use the identity from (i) and the change of summation index from (iii) to evaluate $\sum_{k=1}^{n} \frac{4}{k(k+4)}$, where $k \geq 1$.

2. Evaluate $\sum_{k=1}^{n} \frac{-7k-6}{k(k+1)(k+2)}$, where $k \geq 1$, by:

(i) completing a telescoping sum;

2. Evaluate $\sum_{k=1}^{n} \frac{-7k-6}{k(k+1)(k+2)}$, where $k \geq 1$, by:

(ii) changing the summation index.

3. Evaluate $\sum_{k=1}^{n} \cos(2k-1)$, using $2\sin(1)\cos(2k-1) = \sin(2k) - \sin(2k-2)$, where $k \geq 1$, by:

(i) completing a telescoping sum;

3. Evaluate $\sum_{k=1}^{n} \cos(2k-1)$, using $2\sin(1)\cos(2k-1) = \sin(2k) - \sin(2k-2)$, where $k \geq 1$, by:

(ii) changing the summation index.

Appendix C

Experiment 1: Worked-Examples – Low-Variability Handout

WORKED EXAMPLES - LOW VARIABILITY

School of Mathematics and Statistics, UNSW Sydney

Investigator: Vicki Likourezos — Experiment 1, 2017

1. (i) Examine the following where a and b are real numbers:

$$\frac{4}{k(k+4)} = \frac{a}{k} + \frac{b}{k+4}, \text{ where } k \neq -4, 0.$$

To find a and b, we multiply the above identity by $k(k+4)$ to obtain:

$$4 = a(k+4) + bk.$$

By putting $k = 0$, we see that $4 = 4a \Rightarrow a = 1$
and by putting $k = -4$, we see that $4 = -4b \Rightarrow b = -1$.
Hence,

$$\frac{4}{k(k+4)} = \frac{1}{k} - \frac{1}{k+4}. \quad (1)$$

(ii) A telescoping sum is a sum which is written in such a way where most of the terms cancel and the rest are easy to add up.

To evaluate $\sum_{k=1}^{n} \frac{4}{k(k+4)}$, where $k \geq 1$, the sum may be written as a telescoping sum by applying the identity in (1) and writing out the terms as follows:

$$\begin{aligned}
\sum_{k=1}^{n} \frac{4}{k(k+4)} &= \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+4}\right) \\
&= \left(\frac{1}{1} - \frac{1}{5}\right) + \left(\frac{1}{2} - \frac{1}{6}\right) + \left(\frac{1}{3} - \frac{1}{7}\right) + \left(\frac{1}{4} - \frac{1}{8}\right) \\
&+ \left(\frac{1}{5} - \frac{1}{9}\right) + \left(\frac{1}{6} - \frac{1}{10}\right) + \left(\frac{1}{7} - \frac{1}{11}\right) + \left(\frac{1}{8} - \frac{1}{12}\right) \\
&+ \ldots + \left(\frac{1}{n-4} - \frac{1}{n}\right) + \left(\frac{1}{n-3} - \frac{1}{n+1}\right) \\
&+ \left(\frac{1}{n-2} - \frac{1}{n+2}\right) + \left(\frac{1}{n-1} - \frac{1}{n+3}\right) + \left(\frac{1}{n} - \frac{1}{n+4}\right)
\end{aligned}$$

By cancelling out the common terms, the sum simplifies to:

$$\sum_{k=1}^{n} \frac{4}{k(k+4)} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} - \frac{1}{n+1} - \frac{1}{n+2} - \frac{1}{n+3} - \frac{1}{n+4}.$$

(iii) The sum $\sum_{k=1}^{n} \frac{1}{k+4}$, where n is an integer, $n \geq 1$, can be transformed by shifting the summation index. By substituting $j = k+4$ we have:

$$\sum_{k=1}^{n} \frac{1}{k+4} = \sum_{j=c}^{d} \frac{1}{j}, \text{ where } c, d \text{ are integers and } c \leq d.$$

Using this substitution we have
Lower limit: when $k = 1$, we get $j = 1 + 4 = 5$.
Upper limit: when $k = n$, we get $j = n + 4$.
Hence $\sum_{k=1}^{n} \frac{1}{k+4} = \sum_{j=5}^{n+4} \frac{1}{j}$.
By replacing the variable j with the variable k we have

$$\sum_{k=1}^{n} \frac{1}{k+4} = \sum_{k=5}^{n+4} \frac{1}{k}. \quad (2)$$

(iv) Shifting the summation index is another useful technique that can be used to evaluate $\sum_{k=1}^{n} \frac{4}{k(k+4)}$, where $k \geq 1$.

By applying the identity in (1) and the substitution in (2) we have

$$\begin{aligned}
\sum_{k=1}^{n} \frac{4}{k(k+4)} &= \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+4} \right) \\
&= \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=1}^{n} \frac{1}{k+4} \\
&= \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=5}^{n+4} \frac{1}{k} \\
&= \left(\sum_{k=1}^{4} \frac{1}{k} + \sum_{k=5}^{n} \frac{1}{k} \right) - \left(\sum_{k=5}^{n} \frac{1}{k} + \sum_{k=n+1}^{n+4} \frac{1}{k} \right)
\end{aligned}$$

By cancelling out the common sums, the sum simplifies to:

$$\begin{aligned}
\sum_{k=1}^{n} \frac{4}{k(k+4)} &= \sum_{k=1}^{4} \frac{1}{k} - \sum_{k=n+1}^{n+4} \frac{1}{k} \\
&= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} - \frac{1}{n+1} - \frac{1}{n+2} - \frac{1}{n+3} - \frac{1}{n+4}.
\end{aligned}$$

2. (i) To evaluate $\sum_{k=1}^{n} \frac{5}{k(k+5)}$, where $k \geq 1$, we can follow the same process we used to evaluate $\sum_{k=1}^{n} \frac{4}{k(k+4)}$ in parts 1(i) and (ii).
Firstly, we need to examine the following where a and b are real numbers:

$$\frac{5}{k(k+5)} = \frac{a}{k} + \frac{b}{k+5}, \text{ where } k \neq -5, 0.$$

To find a and b, we multiply the above identity by $k(k+5)$ to obtain:

$$5 = a(k+5) + bk.$$

By putting $k = 0$, we see that $5 = 5a \Rightarrow a = 1$
and by putting $k = -5$, we see that $5 = -5b \Rightarrow b = -1$.
Hence,

$$\frac{5}{k(k+5)} = \frac{1}{k} - \frac{1}{k+5}. \quad 3$$

To evaluate $\sum_{k=1}^{n} \frac{5}{k(k+5)}$, where $k \geq 1$, the sum may be written as a telescoping sum by applying the identity in 3 and writing out the terms as follows:

$$\begin{aligned}
\sum_{k=1}^{n} \frac{5}{k(k+5)} &= \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+5}\right) \\
&= \left(\frac{1}{1} - \frac{1}{6}\right) + \left(\frac{1}{2} - \frac{1}{7}\right) + \left(\frac{1}{3} - \frac{1}{8}\right) + \left(\frac{1}{4} - \frac{1}{9}\right) + \left(\frac{1}{5} - \frac{1}{10}\right) \\
&+ \left(\frac{1}{6} - \frac{1}{11}\right) + \left(\frac{1}{7} - \frac{1}{12}\right) + \left(\frac{1}{8} - \frac{1}{13}\right) + \left(\frac{1}{9} - \frac{1}{14}\right) + \left(\frac{1}{10} - \frac{1}{15}\right) \\
&+ \ldots + \left(\frac{1}{n-5} - \frac{1}{n}\right) + \left(\frac{1}{n-4} - \frac{1}{n+1}\right) + \left(\frac{1}{n-3} - \frac{1}{n+2}\right) \\
&+ \left(\frac{1}{n-2} - \frac{1}{n+3}\right) + \left(\frac{1}{n-1} - \frac{1}{n+4}\right) + \left(\frac{1}{n} - \frac{1}{n+5}\right)
\end{aligned}$$

By cancelling out the common terms, the sum simplifies to:

$$\sum_{k=1}^{n} \frac{5}{k(k+5)} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} - \frac{1}{n+1} - \frac{1}{n+2} - \frac{1}{n+3} - \frac{1}{n+4} - \frac{1}{n+5}.$$

(ii) To evaluate $\sum_{k=1}^{n} \frac{5}{k(k+5)}$, where $k \geq 1$, we can follow the same process we used to evaluate $\sum_{k=1}^{n} \frac{4}{k(k+4)}$ in parts 1(iii) and (iv). By using the technique of shifting the summation index, we can transform the sum $\sum_{k=1}^{n} \frac{1}{k+5}$, where n is an integer, $n \geq 1$.

By substuting $j = k+5$ we have

$$\sum_{k=1}^{n} \frac{1}{k+5} = \sum_{j=c}^{d} \frac{1}{j} \text{ where } c, d \text{ are integers and } c \leq d.$$

Using this substitution we have
Lower limit: when $k = 1$, we get $j = 1 + 5 = 6$.
Upper limit: when $k = n$, we get $j = n + 5$.
Hence $\sum_{k=1}^{n} \frac{1}{k+5} = \sum_{j=6}^{n+5} \frac{1}{j}$.
By replacing the variable j with the variable k we have

$$\sum_{k=1}^{n} \frac{1}{k+5} = \sum_{k=6}^{n+5} \frac{1}{k}. \quad (4)$$

By applying the identity in (3) and the substitution in (4) we have

$$\begin{aligned}
\sum_{k=1}^{n} \frac{5}{k(k+5)} &= \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+5}\right) \\
&= \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=1}^{n} \frac{1}{k+5} \\
&= \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=6}^{n+5} \frac{1}{k} \\
&= \left(\sum_{k=1}^{5} \frac{1}{k} + \sum_{k=6}^{n} \frac{1}{k}\right) - \left(\sum_{k=6}^{n} \frac{1}{k} + \sum_{k=n+1}^{n+5} \frac{1}{k}\right)
\end{aligned}$$

By cancelling out the common sums, the sum simplifies to:

$$\begin{aligned}
\sum_{k=1}^{n} \frac{5}{k(k+5)} &= \sum_{k=1}^{5} \frac{1}{k} - \sum_{k=n+1}^{n+5} \frac{1}{k} \\
&= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} - \frac{1}{n+1} - \frac{1}{n+2} - \frac{1}{n+3} - \frac{1}{n+4} - \frac{1}{n+5}.
\end{aligned}$$

3. (i) To evaluate $\sum_{k=1}^{n} \frac{1}{(k+4)(k+5)}$, where $k \geq 1$, we can follow the same process we used to evaluate $\sum_{k=1}^{n} \frac{4}{k(k+4)}$ in parts 1(i) and (ii). Firstly, we need to examine the following where a and b are real numbers:

$$\frac{1}{(k+4)(k+5)} = \frac{a}{k+4} + \frac{b}{k+5}, \text{ where } k \neq -5, -4.$$

To find a and b, we multiply the above identity by $(k+4)(k+5)$ to obtain:

$$1 = a(k+5) + b(k+4).$$

By putting $k = -4$, we see that $a = 1$
and by putting $k = -5$, we see that $b = -1$.
Hence, $\frac{1}{(k+4)(k+5)} = \frac{1}{k+4} - \frac{1}{k+5}$ 7

To evaluate $\sum_{k=1}^{n} \frac{1}{(k+4)(k+5)}$, where $k \geq 1$, the sum may be written as a telescoping sum by applying the identity in 7 and writing out the terms as follows:

$$\begin{aligned}
\sum_{k=1}^{n} \frac{1}{(k+4)(k+5)} &= \sum_{k=1}^{n} \left(\frac{1}{k+4} - \frac{1}{k+5}\right) \\
&= \left(\frac{1}{5} - \frac{1}{6}\right) + \left(\frac{1}{6} - \frac{1}{7}\right) + \left(\frac{1}{7} - \frac{1}{8}\right) + \left(\frac{1}{8} - \frac{1}{9}\right) \\
&+ \ \ldots + \left(\frac{1}{n+2} - \frac{1}{n+3}\right) + \left(\frac{1}{n+3} - \frac{1}{n+4}\right) + \left(\frac{1}{n+4} - \frac{1}{n+5}\right)
\end{aligned}$$

By cancelling out the common terms, the sum simplifies to:

$$\sum_{k=1}^{n} \frac{1}{(k+4)(k+5)} = \frac{1}{5} - \frac{1}{n+5}.$$

(ii) To evaluate $\sum_{k=1}^{n} \frac{1}{(k+4)(k+5)}$, where $k \geq 1$, we can follow the same process we used to evaluate $\sum_{k=1}^{n} \frac{4}{k(k+4)}$ in parts 1(iii) and (iv).
By using the technique of shifting the summation index, we can apply the identity in (7) and the substitutions in (2) and (4) to get

$$\begin{aligned}
\sum_{k=1}^{n} \frac{1}{(k+4)(k+5)} &= \sum_{k=1}^{n} \left(\frac{1}{k+4} - \frac{1}{k+5} \right) \\
&= \sum_{k=1}^{n} \frac{1}{k+4} - \sum_{k=1}^{n} \frac{1}{k+5} \\
&= \sum_{k=5}^{n+4} \frac{1}{k} - \sum_{k=6}^{n+5} \frac{1}{k} \\
&= \left(\frac{1}{5} + \sum_{k=6}^{n+4} \frac{1}{k} \right) - \left(\sum_{k=6}^{n+4} \frac{1}{k} + \frac{1}{n+5} \right)
\end{aligned}$$

By cancelling out the common sums, the sum simplifies to:
$\sum_{k=1}^{n} \frac{1}{(k+4)(k+5)} = \frac{1}{5} - \frac{1}{n+5}$.

Appendix D

Experiment 1: Problem-Solving – Low-Variability Handout

PROBLEM SOLVING - LOW VARIABILITY

School of Mathematics and Statistics, UNSW Sydney

Investigator: Vicki Likourezos — Experiment 1, 2017

1. (i) Find real numbers a and b so that

$$\frac{4}{k(k+4)} = \frac{a}{k} + \frac{b}{k+4}\ , \text{ where } k \neq -4, 0.$$

(ii) Use the identity from (i) to complete a telescoping sum for

$$\sum_{k=1}^{n} \frac{4}{k(k+4)}\ , \text{ where } k \geq 1.$$

(iii) If n is an integer where $n \geq 1$, find integers c and d where $c \leq d$ so that $\sum_{k=1}^{n} \frac{1}{(k+4)} = \sum_{k=c}^{d} \frac{1}{k}$, where $k \geq 1$.

(iv) Use the identity from (i) and the change of summation index from (iii) to evaluate $\sum_{k=1}^{n} \frac{4}{k(k+4)}$, where $k \geq 1$.

2. Evaluate $\sum_{k=1}^{n} \frac{5}{k(k+5)}$, where $k \geq 1$, by:

(i) completing a telescoping sum;

2. Evaluate $\sum_{k=1}^{n} \frac{5}{k(k+5)}$, where $k \geq 1$, by:

(ii) changing the summation index.

3. Evaluate $\sum_{k=1}^{n} \frac{1}{(k+4)(k+5)}$, where $k \geq 1$, by:

(i) completing a telescoping sum;

3. Evaluate $\sum_{k=1}^{n} \frac{1}{(k+4)(k+5)}$, where $k \geq 1$, by:

(ii) changing the summation index.

Appendix E

Experiment 1: Post-Test

POST-TEST

School of Mathematics and Statistics, UNSW Sydney

Investigator: Vicki Likourezos | Experiment 1, 2017

1. (i) Find real numbers a and b so that

$$\frac{1}{(5k-2)(5k+3)} = \frac{a}{5k-2} + \frac{b}{5k+3}\text{ , where } k \neq -\tfrac{3}{5}, \tfrac{2}{5}.$$

(ii) Use the identity from (i) to complete a telescoping sum for

$$\sum_{k=1}^{n} \frac{1}{(5k-2)(5k+3)}\text{ , where } k \geq 1.$$

(iii) If n is an integer where $n \geq 1$, find integers c and d so that
$\sum_{k=1}^{n} \frac{1}{(5k-2)} = \sum_{k=c}^{d} \frac{1}{(5k+3)}$, where $k \geq 1$.

(iv) Use the identity from (i) and the change of summation index from (iii) to evaluate
$\sum_{k=1}^{n} \frac{1}{(5k-2)(5k+3)}$, where $k \geq 1$.

2. Evaluate the following sum by completing a telescoping sum:

(i) $\sum_{k=1}^{n} \log_2(1 + \frac{1}{k})$, where $k \geq 1$.

2. Evaluate the following sum by completing a telescoping sum:

(ii) $\sum_{k=0}^{n} k.k!$, where $k \geq 0$.

3. Evaluate the following sum by changing the summation index:

(i) $\sum\limits_{k=2}^{n} \dfrac{2}{k(k^2-1)}$, where $k \geq 2$.

3. Evaluate the following sum by changing the summation index:

(ii) $\sum_{k=1}^{\infty} \frac{pq(p-q)}{k(k+p)(k+q)}$, where p and q are integers such that $q > p \geq 1$, and $k \geq 1$.

Appendix F

Subjective Rating of Difficulty Scale

Questionnaire

Student ID: ______________________________ Date: ____ /___ /____

Student name: ____

	Extremely Easy	*Very Easy*	*Moderately Easy*	*Slightly Easy*	*Neither Easy nor Difficult*	*Slightly Difficult*	*Moderately Difficult*	*Very Difficult*	*Extremely Difficult*
How difficult was it for you to complete the tasks?	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Appendix G

Experiment 2: Worked-Examples – High-Variability Handout

WORKED EXAMPLES - HIGH VARIABILITY
University Preparation Program, UNSW Sydney

Investigator: Vicki Likourezos — Experiment 1, 2017

(1) Given $g(x) = 5 - 2x$, for all real x,
to find: (i) $g(-3x)$

Substitute $x = -3x$ into $g(x)$
we get $g(-3x) = 5 - 2(-3x)$
$= 5 + 6x$

(ii) $g(\frac{1}{4})$

Substitute $x = \frac{1}{4}$ into $g(x)$
we get $g(\frac{1}{4}) = 5 - 2(\frac{1}{4})$
$= 4\frac{1}{2}$

(iii) $g(a+5)$

Substitute $x = a+5$ into $g(x)$
we get $g(a+5) = 5 - 2(a+5)$
$= 5 - 2a - 10$
$= -5 - 2a$

(2) Given $h(x) = -x^2 + 6x + 7$, for all real x,
to find: (i) $h(-1)$

Substitute $x = -1$ into $h(x)$
we get $h(-1) = -(-1)^2 + 6(-1) + 7$
$= -1 - 6 + 7$
$= 0$

(ii) $h(\frac{1}{2})$

Substitute $x = \frac{1}{2}$ into $h(x)$
we get $h(\frac{1}{2}) = -(\frac{1}{2})^2 + 6(\frac{1}{2}) + 7$
$= -\frac{1}{4} + 6(\frac{1}{2}) + 7$
$= 9\frac{3}{4}$

(iii) $h(m-2)$

Substitute $x = m-2$ into $h(x)$
we get $h(m-2) = -(m-2)^2 + 6(m-2) + 7$
$= -(m^2 - 4m + 4) + 6m - 12 + 7$
$= -m^2 + 4m - 4 + 6m - 5$
$= -m^2 + 10m - 9$

(3) Using the method of completing the square, we can find the maximum or minimum value of the quadratic function, $h(x) = -x^2 + 6x + 7$

To complete the square, we follow the steps below:

STEP 1: Consider the x^2 and x terms. Since the coefficient of x^2 is -1, we have to factorise first before we can complete the square.

So we write $h(x) = -x^2 + 6x + 7 = -(x^2 - 6x) + 7$

STEP 2: Halve the co-efficient of x and square it.

We get $\left(-\frac{6}{2}\right)^2 = 9$

STEP 3: Add this square number to the x^2 and x terms to complete a perfect square and remember to add this square number to the constant term.

$$\begin{aligned} \text{We get } h(x) &= -(x^2 - 6x + 9) + 7 + 9 \\ &= -(x-3)^2 + 16 \end{aligned}$$

For any real value of x, $-(x-3)^2 \leq 0$.
This means when $x = 3$, the maximum function value is 16.

(4) To determine the concavity of the following quadratic functions below we need to consider the co-efficient of the x^2 term:

(i) $y = -2x + x^2 - 3$

The coefficient of x^2 is 1, so $a = 1$.
Since $a > 0$ the parabola is concave up.

(ii) $y = x(2 - x)$

Firstly we need to expand the expression to get $y = 2x - x^2$.
The coefficient of x^2 is -1, so $a = -1$.
Since $a < 0$ the parabola is concave down.

(iii) $y = (-2x - 1)(5 - x)$

Firstly we need to expand the expression, to get $y = 2x^2 - 9x - 5$
The coefficient of x^2 is 2, so $a = 2$.
Since $a > 0$ the parabola is concave up.

(5) Sketch the following parabolas showing all essential features:

(i) $y = -x^2 - 1$

1. The coefficient of x^2 is -1, so $a = -1$.
   Since $a < 0$ the parabola is concave down.

2. The axis of symmetry can be found using $x = -\dfrac{b}{2a}$

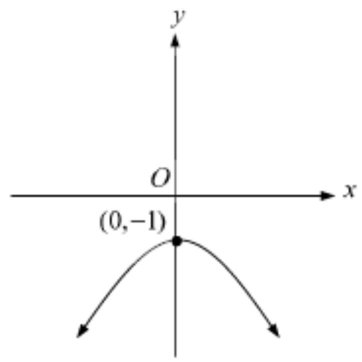
   so $x = -\dfrac{0}{2(1)} = 0$.

3. To find the vertex we substitute $x = 0$ into $y = -x^2 - 1$
   to get $y = -(0)^2 - 1$
   $= -1$
   $\therefore$ Vertex at $(0, -1)$.

4. To find the y-intercept, we let $x = 0$
   to get $y = -(0)^2 - 1$
   $= -1$
   $\therefore (0, -1)$ is the y-intercept.

5. To find the x-intercepts, we let $y = 0$ to get $-x^2 - 1 = 0$.
   But this quadratic equation cannot be factorised. So we check the discriminant.
   Since $\Delta = (0)^2 - 4(-1)(-1) = -4$ and $\Delta < 0$
   $\therefore$ there are no are the x-intercepts.

6. We can sketch the parabola showing all the essential features.



(ii) $y = 2(2x+1)^2$

Firstly we start by expanding the brackets

$$y = 2(4x^2 + 4x + 1)$$
$$= 8x^2 + 8x + 2$$

1. The coefficient of x^2 is 8, so $a = 8$.
   Since $a > 0$ the parabola is concave up.

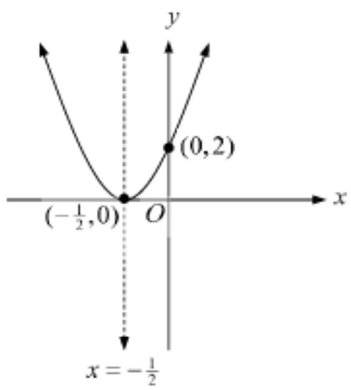
2. The axis of symmetry can be found using $x = -\dfrac{b}{2a}$
   so $x = -\dfrac{8}{2(8)} = -\tfrac{1}{2}$.

3. To find the vertex we substitute $x = -\tfrac{1}{2}$ into $y = 2(2x+1)^2$
   to get $y = 2(2(-\tfrac{1}{2})+1)^2$
   $= 2(-1+1)^2$
   $= 0$
   $\therefore$ Vertex at $(-\tfrac{1}{2}, 0)$.

4. To find the y-intercept, we let $x = 0$
   to get $y = 2[2(0)+1]^2$
   $= 2$
   $\therefore (0,2)$ is the y-intercept.

5. To find the x-intercept, we let $y = 0$ to get $2(2x+1)^2 = 0$
   so $x = -\tfrac{1}{2}$.
   $\therefore (-\tfrac{1}{2}, 0)$ is the x-intercept.

6. We can sketch the parabola showing all the essential features.



(iii) $y = 3x^2 - 4x + 5$

1. The coefficient of x^2 is 3, so $a = 3$.
   Since $a > 0$ the parabola is concave up.

2. The axis of symmetry can be found using $x = -\frac{b}{2a}$

   so $x = -\frac{-4}{2(3)} = \frac{2}{3}$.

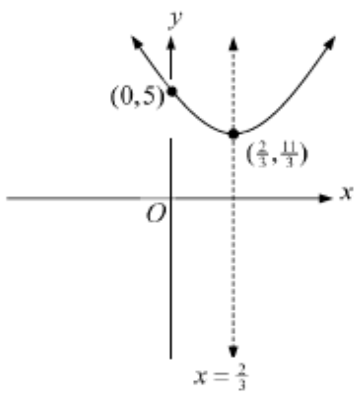
3. To find the vertex we substitute $x = \frac{2}{3}$ into $y = 3x^2 - 4x + 5$

   to get $y = 3(\frac{2}{3})^2 - 4(\frac{2}{3}) + 5$
   $= \frac{12}{9} - \frac{8}{3} + 5$
   $= \frac{11}{3}$
   $\therefore$ Vertex at $(\frac{2}{3}, \frac{11}{3})$.

4. To find the y-intercept, we let $x = 0$
   to get $y = 3(0)^2 - 4(0) + 5$
   $= 5$
   $\therefore (0,5)$ is the y-intercept.

5. To find the x-intercepts, we let $y = 0$ to get $3x^2 - 4x + 5 = 0$.
   But this quadratic equation cannot be factorised. So we check the discriminant.
   Since $\Delta = (-4)^2 - 4(3)(5) = -44$ and $\Delta < 0$
   $\therefore$ there are no x-intercepts.

6. We can sketch the parabola showing all the essential features.



(iv) $y=-2(x+4)(6-x)$

Firstly we start by expanding the brackets

$$\begin{aligned} y &= -2(6x-x^2+24-4x) \\ &= -2(-x^2+2x+24) \\ &= 2x^2-4x-48 \end{aligned}$$

1. The coefficient of x^2 is 2, so $a=2$.
   Since $a>0$ the parabola is concave up.

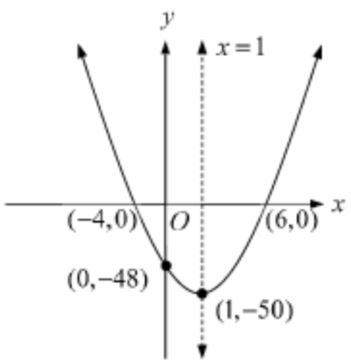
2. The axis of symmetry can be found using $x=-\dfrac{b}{2a}$
   so $x=-\dfrac{-4}{2(2)}=1$.

3. To find the vertex we substitute $x=1$ into $y=2x^2-4x-48$
   to get $y=2(1)^2-4(1)-48$
   $=2-4-48$
   $=-50$
   $\therefore$ Vertex at $(1,-50)$.

4. To find the y-intercept, we let $x=0$
   to get $y=2(0)^2-4(0)-48$
   $=-48$
   $\therefore (0,-48)$ is the y-intercept.

5. To find the x-intercepts, we let $y=0$ to get $-2(x+4)(6-x)=0$
   so $x=-4, 6$.
   $\therefore (-4,0)$ and $(6,0)$ are the x-intercepts.

6. We can sketch the parabola showing all the essential features.



Appendix H

Experiment 2: Worked-Examples – Low-Variability Handout

WORKED EXAMPLES – LOW VARIABILITY
University Preparation Program, UNSW Sydney

Investigator: Vicki Likourezos Experiment 1, 2017

(1) Given $f(x)=4x+8$, for all real x,
to find: (i) $f(0)$

Substitute $x=0$ into $f(x)$
we get $f(0)=4(0)+8$
$=8$

(ii) $f(-2)$

Substitute $x=-2$ into $f(x)$
we get $f(-2)=4(-2)+8$
$=-8+8$
$=0$

(iii) $f(a)$

Substitute $x=a$ into $f(x)$
we get $f(a)=4(a)+8$
$=4a+8$

(2) Given $f(x)=x^2-6x-16$, for all real x,
to find: (i) $f(0)$

Substitute $x=0$ into $f(x)$
we get $f(0)=(0)^2-6(0)-16$
$=-16$

(ii) $f(-2)$

Substitute $x=-2$ into $f(x)$
we get $f(-2)=(-2)^2-6(-2)-16$
$=4+12-16$
$=0$

(iii) $f(a)$

Substitute $x=a$ into $f(x)$
we get $f(a)=(a)^2-6(a)-16$
$=a^2-6a-16$

(3) Using the method of completing the square, we can find the maximum or minimum value of the quadratic function, $f(x) = x^2 - 6x - 16$

To complete the square, we follow the steps below:

STEP 1: Consider the x^2 and x terms. Since the coefficient of x^2 is 1, we can complete the square.

STEP 2: Halve the co-efficient of x and square it.

We get $\left(-\frac{6}{2}\right)^2 = 9$

STEP 3: Add this square number to the x^2 and x terms to complete a perfect square and remember to subtract this square number from the constant term.

$$\begin{aligned} \text{We get } f(x) &= (x^2 - 6x + 9) - 16 - 9 \\ &= (x-3)^2 - 25 \end{aligned}$$

For any real value of x, $(x-3)^2 \geq 0$.
This means when $x = 3$, the minimum function value is -25.

(4) To determine the concavity of the following quadratic functions below we need to consider the co-efficient of the x^2 term:

(i) $y = x^2 - 2x - 3$

The coefficient of x^2 is 1, so $a = 1$.
Since $a > 0$ the parabola is concave up.

(ii) $y = (x-4)(x+1)$

Firstly we need to expand the expression to get $y = x^2 - 3x - 4$.
The coefficient of x^2 is 1, so $a = 1$.
Since $a > 0$ the parabola is concave up.

(iii) $y = -x^2 + 6x + 7$

The coefficient of x^2 is -1, so $a = -1$.
Since $a < 0$ the parabola is concave down.

(5) Sketch the following parabolas showing all essential features:

(i) $y = x^2 + 12x + 35$

1. The coefficient of x^2 is 1, so $a = 1$.
   Since $a > 0$ the parabola is concave up.

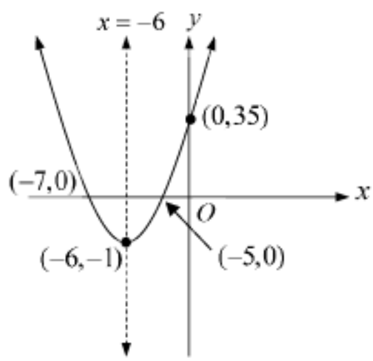
2. The axis of symmetry can be found using $x = -\dfrac{b}{2a}$
   so $x = -\dfrac{12}{2(1)} = -6$.

3. To find the vertex we substitute $x = -6$ into $y = x^2 + 12x + 35$
   to get $y = (-6)^2 + 12(-6) + 35$
   $= 36 - 72 + 35$
   $= -1$
   $\therefore$ Vertex at $(-6, -1)$.

4. To find the y-intercept, we let $x = 0$
   to get $y = (0)^2 + 12(0) + 35$
   $= 35$
   $\therefore (0, 35)$ is the y-intercept.

5. To find the x-intercepts, we let $y = 0$ to get $x^2 + 12x + 35 = 0$.
   By factorising the quadratic equation, we get $(x+7)(x+5) = 0$
   so $x = -7, -5$.
   $\therefore (-7, 0)$ and $(-5, 0)$ are the x-intercepts.

6. We can sketch the parabola showing all the essential features.



(ii) $y = x^2 - 10x + 25$

1. The coefficient of x^2 is 1, so $a = 1$.
   Since $a > 0$ the parabola is concave up.

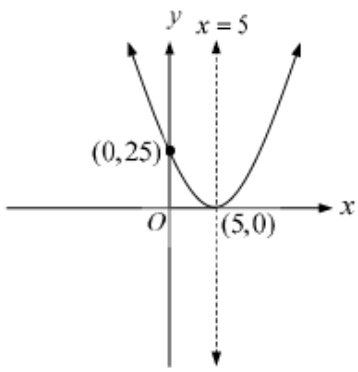
2. The axis of symmetry can be found using $x = -\dfrac{b}{2a}$
   so $x = -\dfrac{-10}{2(1)} = 5$.

3. To find the vertex we substitute $x = 5$ into $y = x^2 - 10x + 25$
   to get $y = (5)^2 - 10(5) + 25$
   $= 25 - 50 + 25$
   $= 0$
   $\therefore$ Vertex at $(5,0)$.

4. To find the y-intercept, we let $x = 0$
   to get $y = (0)^2 - 10(0) + 25$
   $= 25$
   $\therefore (0,25)$ is the y-intercept.

5. To find the x-intercepts, we let $y = 0$ to get $x^2 - 10x + 25 = 0$.
   By factorising the quadratic equation, we get $(x-5)(x-5) = 0$
   so $x = 5$.
   $\therefore (5,0)$ is the x-intercept.

6. We can sketch the parabola showing all the essential features.



(iii) $y = x^2 + 4x + 6$

1. The coefficient of x^2 is 1, so $a = 1$.
   Since $a > 0$ the parabola is concave up.

2. The axis of symmetry can be found using $x = -\dfrac{b}{2a}$

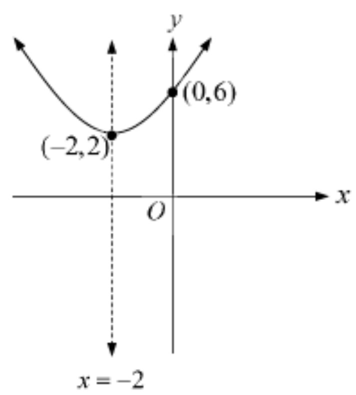
   so $x = -\dfrac{4}{2(1)} = -2$.

3. To find the vertex we substitute $x = -2$ into $y = x^2 + 4x + 6$
   to get $y = (-2)^2 + 4(-2) + 6$
   $= 4 - 8 + 6$
   $= 2$
   $\therefore$ Vertex at $(-2, 2)$.

4. To find the y-intercept, we let $x = 0$
   to get $y = (0)^2 + 4(0) + 6$
   $= 6$
   $\therefore (0,6)$ is the y-intercept.

5. To find the x-intercepts, we let $y = 0$ to get $x^2 + 4x + 6 = 0$.
   But this quadratic equation cannot be factorised. So we check the discriminant.
   Since $\Delta = (4)^2 - 4(1)(6) = -8$ and $\Delta < 0$
   $\therefore$ there are no x-intercepts.

6. We can sketch the parabola showing all the essential features.



(iv) $y=-x^2+6x+7$

1. The coefficient of x^2 is -1, so $a=-1$.
   Since $a<0$ the parabola is concave down.

2. The axis of symmetry can be found using $x=-\dfrac{b}{2a}$

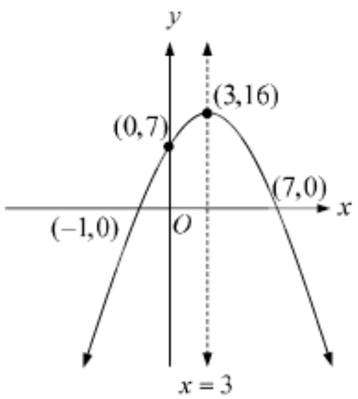
   so $x=-\dfrac{6}{2(-1)}=3$.

3. To find the vertex we substitute $x=3$ into $y=-x^2+6x+7$
   to get
   $$\begin{aligned} y&=-(3)^2+6(3)+7\\ &=-9+18+7\\ &=16 \end{aligned}$$
   $\therefore$ Vertex at $(3,16)$.

4. To find the y-intercept, we let $x=0$
   to get
   $$\begin{aligned} y&=-(0)^2+6(0)+7\\ &=7 \end{aligned}$$
   $\therefore (0,7)$ is the y-intercept.

5. To find the x-intercepts, we let $y=0$ to get $-x^2+6x+7=0$.
   By multiplying the quadratic equation by -1, we get $x^2-6x-7=0$.
   By factorising the quadratic equation, we get $(x-7)(x+1)=0$
   so $x=-1, 7$.
   $\therefore (-1,0)$ and $(7,0)$ are the x-intercepts.

6. We can sketch the parabola showing all the essential features.



Appendix I

Experiment 2: Problem-Solving – High-Variability Handout

PROBLEM SOLVING - HIGH VARIABILITY
University Preparation Program, UNSW Sydney

Investigator: Vicki Likourezos | Experiment 1, 2017

(1) Given $g(x)=5-2x$, for all real x,
find: (i) $g(-3x)$

(ii) $g(\tfrac{1}{4})$

(iii) $g(a+5)$

(2) Given $h(x)=-x^2+6x+7$, for all real x,
find: (i) $h(-1)$

(ii) $h(\tfrac{1}{2})$

(iii) $h(m-2)$

(3) Using the method of completing the square, find the maximum or minimum value of the quadratic function, $h(x) = -x^2 + 6x + 7$

(4) Determine the concavity of the following quadratic functions:

(i) $y = -2x + x^2 - 3$

(ii) $y = x(2 - x)$

(iii) $y = (-2x - 1)(5 - x)$

(5) Sketch the following parabolas showing all essential features:

(i) $y = -x^2 - 1$

(ii) $y = 2(2x+1)^2$

(5) Sketch the following parabolas showing all essential features:

(iii) $y = 3x^2 - 4x + 5$

(iv) $y = -2(x+4)(6-x)$

ANSWERS

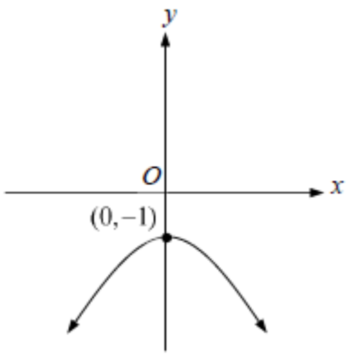
(1) (i) $5+6x^2$
(ii) $4\frac{1}{2}$
(iii) $-5-2a$

(2) (i) 0
(ii) $9\frac{3}{4}$
(iii) $-m^2+10m-9$

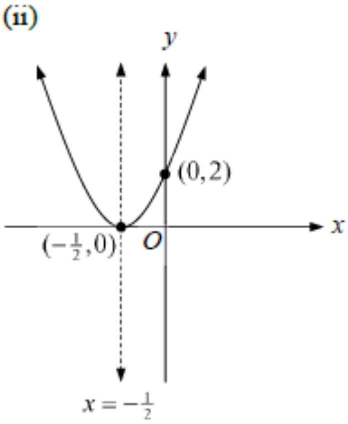
(3) Maximum value of 16.

(4) (i) Concave up
(ii) Concave down
(iii) Concave up

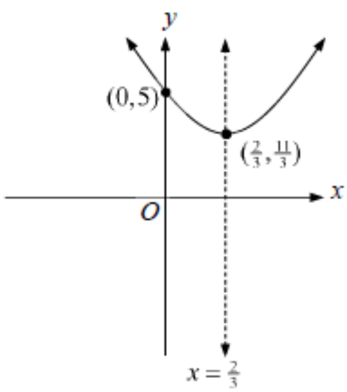
(5) (i)



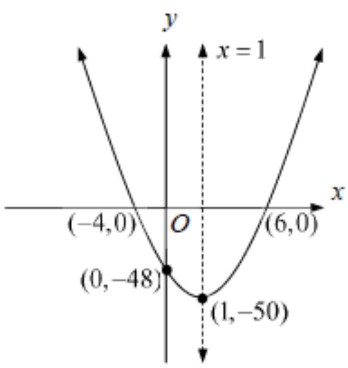
(ii)



(iii)



(iv)



Appendix J

Experiment 2: Problem-Solving – Low-Variability Handout

PROBLEM SOLVING – LOW VARIABILITY
University Preparation Program, UNSW Sydney

Investigator: Vicki Likourezos | Experiment 1, 2017

(1) Given $f(x) = 4x + 8$, for all real x,

find: (i) $f(0)$

(ii) $f(-2)$

(iii) $f(a)$

(2) Given $f(x) = x^2 - 6x - 16$, for all real x,

find: (i) $f(0)$

(ii) $f(-2)$

(iii) $f(a)$

(3) Using the method of completing the square, find the maximum or minimum value of the quadratic function, $f(x) = x^2 - 6x - 16$

(4) Determine the concavity of the following quadratic functions:

(i) $y = x^2 - 2x - 3$

(ii) $y = (x-4)(x+1)$

(iii) $y = -x^2 + 6x + 7$

(5) Sketch the following parabolas showing all essential features:

(i) $y = x^2 + 12x + 35$

(ii) $y = x^2 - 10x + 25$

(5) Sketch the following parabolas showing all essential features:

(iii) $y = x^2 + 4x + 6$

(iv) $y = -x^2 + 6x + 7$

ANSWERS

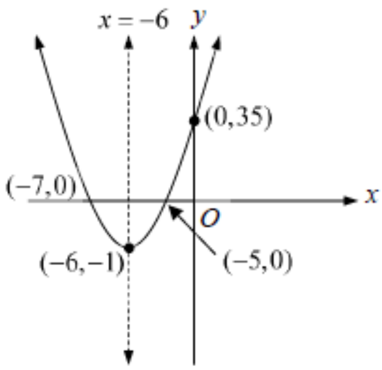
(1) (i) 8
(ii) 0
(iii) $4a+8$

(2) (i) -16
(ii) 0
(iii) $a^2-6a-16$

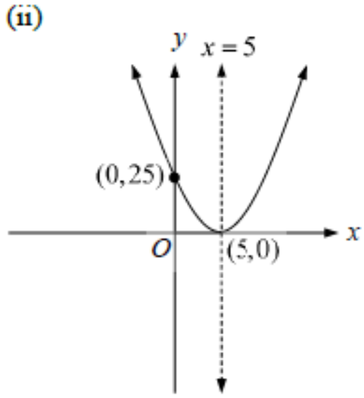
(3) Minimum value of -25.

(4) (i) Concave up
(ii) Concave up
(iii) Concave down

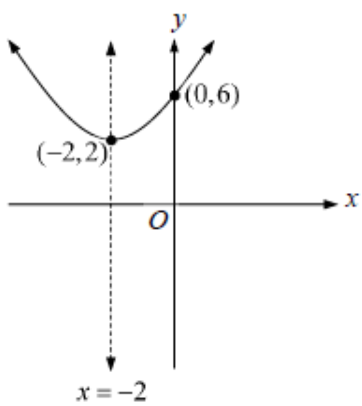
(5) (i)



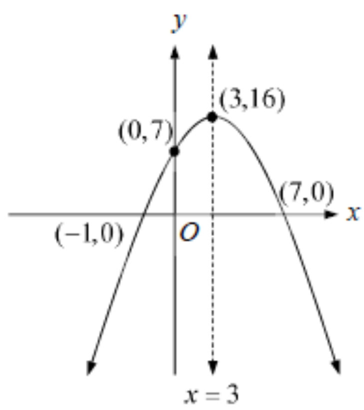
(ii)



(iii)



(iv)



Appendix K

Experiment 2: Post-Test

POST TEST

University Preparation Program, UNSW Sydney

Investigator: Vicki Likourezos Experiment 1, 2017

(1) Given $f(x) = 3x - 3x^2$, for all real x,

find: (i) $f(\frac{1}{2})$

(ii) $f(-3)$

(iii) $f(a-1)$

(2) Using the method of completing the square, find the maximum or minimum value of $f(x) = 5x^2 + 10x - 25$.

(3) Determine the concavity of the following quadratic functions:

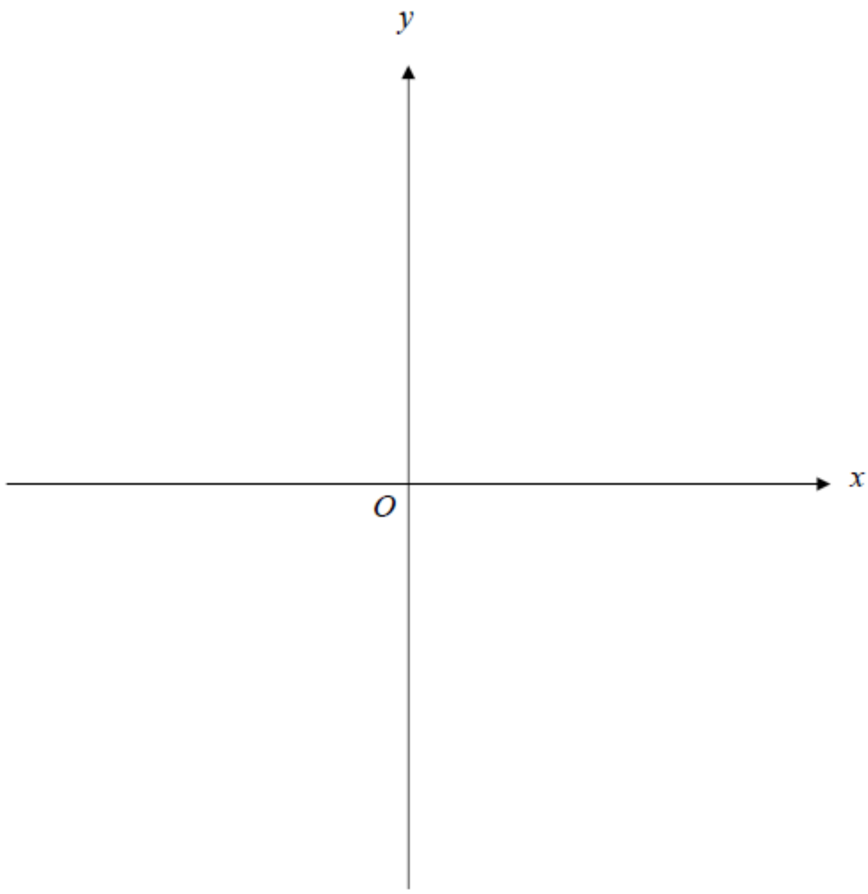
(i) $y = -4 + 2x - x^2$

(ii) $y = -3(2 - x)(-2x + 3)$

(4) Find the axis of symmetry of the graph $f(x) = -2x^2 - x - 2$.

(5) Find the vertex of the graph $h(x) = 3x^2 + 6x - 10$.

(6) Sketch the graph of $y = 15 + 2x - x^2$ on the axes below.
Show all essential features on your sketch.



(7) Without sketching the graph, determine whether the curve $y = 2x^2 + 4x + 5$ crosses the x-axis.

Appendix L

Experiment 3: Pre-Test

PRE-TEST

1. In the expression 2^3:

(a) What is the exponent?

(b) What is the base?

(c) What is the index?

2. Simplify the following, writing the answers in index form with positive indices:

(a) $8a^4 \times 2a^6$

(b) $12n^6 \div 3n^2$

(c) $\frac{(2x^2y^3)^2}{4x^5y^7}$

3. Find the exact value of the following without using a calculator:

(a) 8^{-2}

(b) $(\frac{2}{3})^{-3}$

(c) $27^{\frac{-1}{3}}$

(d) $(a^6)^{\frac{2}{3}}$

(e) $\left(\frac{4}{9}\right)^{-\frac{3}{2}}$

4. Solve the following equations for *x*:

(a) $4^{x} = \frac{1}{16}$

(b) $100^{x} = 1$

(c) $81^{x} = 243$

(d) $2^{x-1} = \frac{\sqrt{2}}{32}$

(e) $3^{3-x} = 27^{x-1}$

5. In the logarithmic equation $\log_3 9 = 2$:

(a) What is the argument of the log?

(b) What is the base of the log?

6. Make x the subject for the following:

(a) $5^x = y$

(b) $\log_b (x + 2) = 3$

7. Solve the following equations for x:

(a) $\log_5 x = 1$

(b) $\log_3 1 = x$

Appendix M

Experiment 3: Low-Variability Handout

LEARNING PHASE – LOW VARIABILITY

The relationship connecting an exponential with a logarithm is:

$$\boldsymbol{x = b^n} \text{ is equivalent to } \mathbf{log}_{\boldsymbol{b}}\, \boldsymbol{x = n}$$

Here are some examples:

(a) $3^4 = 81$ has base 3 and index 4

So we can write, $\mathbf{3^4 = 81}$ is equivalent to $\mathbf{\log_3 81 = 4}$

(b) $3^{-4} = \frac{1}{81}$ has base 3 and index (– 4)

So we can write, $\mathbf{3^{-4} = \frac{1}{81}}$ is equivalent to $\mathbf{\log_3(\frac{1}{81}) = -4}$

(c) $81^{(\frac{1}{4})} = 3$ has base 81 and index $\frac{1}{4}$

So we can write, $\mathbf{81^{(\frac{1}{4})} = 3}$ is equivalent to $\mathbf{\log_{81} 3 = \frac{1}{4}}$

(d) $81^{(\frac{3}{4})} = 27$ has base 81 and index $\frac{3}{4}$

So we can write, $\mathbf{81^{(\frac{3}{4})} = 27}$ is equivalent to $\mathbf{\log_{81} 27 = \frac{3}{4}}$

1. Write the following in logarithmic form, <u>without</u> solving for x:

(a) $3^x = 9$

(b) $4^x = \frac{1}{4}$

(c) $125^x = 5$

(d) $32^x = 4$

2. Write the following in index form, <u>without</u> solving for x:

(a) $\log_{10} 1000 = x$

(b) $\log_4 2 = x$

(c) $\log_7(\frac{1}{49}) = x$

(d) $\log_3\sqrt{3} = x$

3. What number is n?

(a) $\log_n 16 = 2$

(b) $\log_n 3 = \frac{1}{3}$

(c) $\log_n 8 = \frac{3}{2}$

(d) $\log_n(\frac{1}{25}) = -2$

4. Using the appropriate logarithmic laws, solve the following for *x*:

(a) $\log_{10} 25 + \log_{10} 4 = x$

(b) $\log_3 54 - \log_3 18 = x$

(c) $\dfrac{\log_{10} 25}{\log_{10} 5} = x$

(d) $2\log_{10} 5 + \frac{1}{2}\log_{10} 16 = x$

Appendix N

Experiment 3: High-Variability Handout

LEARNING PHASE – HIGH VARIABILITY

The relationship connecting an exponential with a logarithm is:

$\boldsymbol{x = b^n}$ is equivalent to $\mathbf{log}_{\boldsymbol{b}}\, \boldsymbol{x = n}$

Here are some examples:

(a) $3^4 = 81$ has base 3 and index 4

So we can write, $\mathbf{3^4 = 81}$ is equivalent to $\mathbf{log_3\, 81 = 4}$

(b) $3^{-4} = \frac{1}{81}$ has base 3 and index (– 4)

So we can write, $\mathbf{3^{-4} = \frac{1}{81}}$ is equivalent to $\mathbf{log_3\, (\frac{1}{81}) = -4}$

(c) $81^{(\frac{1}{4})} = 3$ has base 81 and index $\frac{1}{4}$

So we can write, $\mathbf{81^{(\frac{1}{4})} = 3}$ is equivalent to $\mathbf{log_{81}\, 3 = \frac{1}{4}}$

(d) $81^{(\frac{3}{4})} = 27$ has base 81 and index $\frac{3}{4}$

So we can write, $\mathbf{81^{(\frac{3}{4})} = 27}$ is equivalent to $\mathbf{log_{81}\, 27 = \frac{3}{4}}$

1. Write the following in logarithmic form, without solving for x:

(a) $3^x = 9$

(b) $4^{-1} = x$

(c) $x^{\frac{1}{3}} = 5$

(d) $32^x = 4$

2. Write the following in index form, without solving for x:

(a) $\log_{10} 1000 = x$

(b) $\log_x 2 = \frac{1}{2}$

(c) $\log_7 x = -2$

(d) $\log_3 \sqrt{3} = x$

3. What number is n?

(a) $\log_n 16 = 2$

(b) $\log_{27} n = \frac{1}{3}$

(c) $\log_4 8 = n$

(d) $\log_n \left(\frac{1}{25}\right) = -2$

4. Using the appropriate logarithmic laws, solve the following for *x*:

(a) $\log_{10} 25 + \log_{10} 4 = x$

(b) $\log_3 54 - \log_3 x = 1$

(c) $\dfrac{\log_{10} x}{\log_{10} 5} = 2$

(d) $2\log_{10} 5 + \frac{1}{2}\log_{10} 16 = x$

Appendix O

Experiment 3: Post-Test 1 (Similar Tasks)

POST-TEST 1

1. In the logarithmic equation $\log_3 9 = 2$:

(a) What is the argument of the log?

(b) What is the base of the log?

2. Make x the subject for the following:

(a) $5^x = y$

(b) $\log_b (x + 2) = 3$

3. Solve the following equations for x:

(a) $\log_5 x = 1$

(b) $\log_3 1 = x$

4. Express the following in logarithmic form:

(a) $4^{-2} = \frac{1}{16}$

(b) $5^{125} = p$

5. Evaluate the following:

(a) $\log_8 2$

(b) $\log_{10}(\frac{1}{10000})$

(c) $\log_5 50 - \log_5 2$

(d) $\log_3 81 \times \log_5 125$

6. Solve the following for *x*:

(a) $\log_2 \left(\frac{x}{2}\right) = 4$

(b) $\log_x \left(\frac{1}{64}\right) = 3$

(c) $\log_{10}(x^3) = 15$

(d) $\frac{1}{2} \log_6 \left(\frac{1}{36}\right) = x$

Appendix P

Experiment 3: Post-Test 2 (Transfer Tasks)

POST-TEST 2

1. Evaluate the following:

(a) $\log_2 12 - (\log_2 2 + \log_2 3)$

(b) $\dfrac{\log_{10} 25}{\log_{10} 5} \times 10^{\log_{10} 3}$

2. Solve the following for *x*:

(a) $\log_{100} x = 0.5$

(b) $\log_2 (3x - 4) = 5$

(c) $\log_2 (\log_{10} x) = 0$

(d) $\log_{\sqrt{2}} (8\sqrt{2}) = x$

3. If $\log_b 81 + \log_b 243 - \log_b 27 = k \log_b 3$ **find *k*.**

Appendix Q

Experiment 4: Worked-Examples – High-Variability Handout

WORKED EXAMPLES – HIGH VARIABILITY
University Preparation Program, UNSW Sydney

Investigator: Vicki Likourezos — **Experiment 1, 2018**

(1) Given $g(x) = 5 - 2x$, for all real x,

(i)
Find $g(-3x)$:

$$g(-3x) = 5 - 2(-3x)$$
$$= 5 + 6x$$

(ii)
Find $g\left(\frac{1}{4}\right)$:

$$g\left(\frac{1}{4}\right) = 5 - 2\left(\frac{1}{4}\right)$$
$$= 5 - \left(\frac{1}{2}\right)$$
$$= 4\frac{1}{2}$$

(iii)
Find $g(a+5)$:

$$g(a+5) = 5 - 2(a+5)$$
$$= 5 - 2a - 10$$
$$= -5 - 2a$$

(2) Given $h(x) = -x^2 + 6x + 7$, for all real x,

(i)
Find $h(-1)$:

$$h(-1) = -(-1)^2 + 6(-1) + 7$$
$$= -1 - 6 + 7$$
$$= 0$$

(iii)
Find $h(m-2)$:

$$h(m-2) = -(m-2)^2 + 6(m-2) + 7$$
$$= -(m^2 - 4m + 4) + 6m - 12 + 7$$
$$= -m^2 + 4m - 4 + 6m - 5$$
$$= -m^2 + 10m - 9$$

(ii)
Find $h\left(\frac{1}{2}\right)$:

$$h\left(\frac{1}{2}\right) = -\left(\frac{1}{2}\right)^2 + 6\left(\frac{1}{2}\right) + 7$$
$$= -\frac{1}{4} + 6\left(\frac{1}{2}\right) + 7$$
$$= -\frac{1}{4} + 3 + 7$$
$$= 9\frac{3}{4}$$

(3) To find the maximum or minimum value of a quadratic function, follow the steps below:

Consider $h(x) = -x^2 + 6x + 7$

$$= -(x^2 - 6x) + 7$$

coefficient is 1 | coefficient is – 6 | constant is 7

monic quadratic

completing the square: $\left(-\frac{6}{2}\right)^2 = 9$

$$h(x) = -(x^2 - 6x + 9) + 9 + 7$$

$$= -(x - 3)^2 + 16$$

$$-(x - 3)^2 \le 0$$

The maximum value of $-(x - 3)^2$ occurs when $x = 3$.

Therefore, the maximum value of the quadratic function is 16.

(4) To determine the concavity of the following quadratic functions, we need to consider the coefficient of the x^2 term:

(i) $y = -2x + x^2 - 3$

coefficient is 1

$\therefore a = 1$

Since $a > 0$, the parabola is concave up.

(ii) $y = x(2 - x)$

$= 2x - x^2$

coefficient is – 1

$\therefore a = -1$

Since $a < 0$, the parabola is concave down.

(iii) $y = (-2x - 1)(5 - x)$

$= -10x + 2x^2 - 5 + x$

$= 2x^2 - 9x - 5$

coefficient is 2

$\therefore a = 2$

Since $a > 0$, the parabola is concave up.

(5) Sketch the following parabolas showing all essential features:

(i)

$$y = -x^2 - 1$$
$$= -x^2 + 0x - 1$$

coefficient is -1   coefficient is 0   constant is – 1

$\therefore a = -1$ $\therefore b = 0$ $\therefore c = -1$

Since $a < 0$, the parabola is concave down.

The axis of symmetry can be found using $x = -\dfrac{b}{2a}$

$$x = -\frac{0}{2(-1)} = 0$$

To find the vertex, we substitute $x = 0$:

$y = -(0)^2 - 1$

$= -0 - 1$

$= -1$

$\therefore$ the vertex is $(0, -1)$.

To find the x-intercept(s),

we substitute $y = 0$:

$0 = -x^2 - 1$

$0 = x^2 + 1$

This quadratic equation cannot be factorised, so we check the discriminant:

$\Delta = b^2 - 4ac$

$\Delta = (0)^2 - 4(-1)(-1)$

$= 0 - 4$

$= -4$

since $\Delta < 0$

$\therefore$ there are no x-intercepts.

To find the y-intercept,

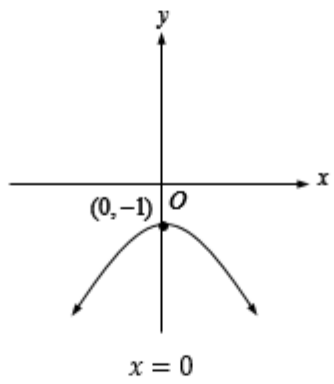
we substitute $x = 0$:

$y = -(0)^2 - 1$

$= -0 - 1$

so $y = -1$

$\therefore$ the y-intercept is $(0, -1)$.



(ii)

$$\begin{aligned} y &= 2(2x+1)^2 \\ &= 2(4x^2+4x+1) \\ &= 8x^2+8x+2 \end{aligned}$$

coefficient is 8 — coefficient is 8 — constant is 2

$\therefore a = 8$ — $\therefore b = 8$ — $\therefore c = 2$

Since $a > 0$, the parabola is concave up.

The axis of symmetry can be found using $x = -\dfrac{b}{2a}$

$$x = -\frac{8}{2(8)} = -\tfrac{1}{2}.$$

To find the vertex, we substitute $x = -\frac{1}{2}$:

$$\begin{aligned} y &= 2(2(-\tfrac{1}{2})+1)^2 \\ &= 2(-1+1)^2 \\ &= 0 \end{aligned}$$

$\therefore$ the vertex is $\left(-\frac{1}{2}, 0\right)$.

To find the x-intercept(s),

we substitute $y = 0$:

$$\begin{aligned} 0 &= 2(2x+1)^2 \\ 0 &= (2x+1)^2 \\ 0 &= 2x+1 \\ -1 &= 2x \end{aligned}$$

so $x = -\frac{1}{2}$

$\therefore$ the x-intercept is $\left(-\frac{1}{2}, 0\right)$.

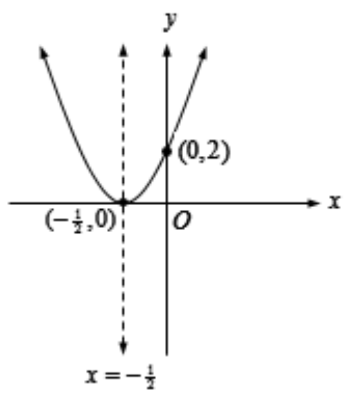
To find the y-intercept,

we substitute $x = 0$:

$$\begin{aligned} y &= 2[2(0)+1]^2 \\ &= 2(0+1)^2 \\ &= 2(1) \end{aligned}$$

so $y = 2$

$\therefore$ the y-intercept is $(0,2)$.



(iii) $y = 3x^2 - 4x + 5$

coefficient is 3 — coefficient is -4 — constant is 5

$\therefore a = 3$ $\quad \therefore b = -4$ $\quad \therefore c = 5$

Since $a > 0$, the parabola is concave up.

The axis of symmetry can be found using $x = -\dfrac{b}{2a}$

$$x = -\frac{-4}{2(3)} = \tfrac{2}{3}.$$

To find the x-intercept(s),

we substitute $y = 0$:

$0 = -3x^2 - 4x + 5$

This quadratic equation cannot be factorised, so we check the discriminant:

$\Delta = b^2 - 4ac$

$\Delta = (-4)^2 - 4(3)(5)$

$\quad = 16 - 60$

$\quad = -44$

since $\Delta < 0$

$\therefore$ there are no x-intercepts.

To find the vertex, we substitute $x = \frac{2}{3}$:

$y = 3(\tfrac{2}{3})^2 - 4(\tfrac{2}{3}) + 5$

$\quad = \frac{12}{9} - \frac{8}{3} + 5$

$\quad = \frac{12}{9} - \frac{24}{9} + \frac{45}{9}$

$\quad = \frac{33}{9}$

$\quad = \frac{11}{3}$

$\therefore$ the vertex is $\left(\frac{2}{3}, \frac{11}{3}\right)$.

To find the y-intercept,

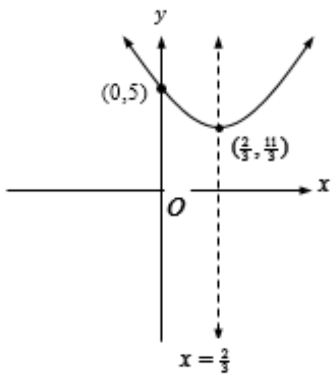
we substitute $x = 0$:

$y = 3(0)^2 - 4(0) + 5$

$\quad = 0 - 0 + 5$

so $y = 5$

$\therefore$ the y-intercept is $(0,5)$.



(iv)

$$y = -2(x+4)(6-x)$$
$$= -2(6x - x^2 + 24 - 4x)$$
$$= -2(-x^2 + 2x + 24)$$
$$= 2x^2 - 4x - 48$$

coefficient is 2 | coefficient is -4 | constant is -48

$\therefore a = 2$ | $\therefore b = -4$ | $\therefore c = -48$

Since $a > 0$, the parabola is concave up.

The axis of symmetry can be found using $x = -\dfrac{b}{2a}$

$$x = -\frac{-4}{2(2)} = 1.$$

To find the x-intercept(s),

we substitute $y = 0$:

$0 = -2(x+4)(6-x)$

$0 = (x+4)(6-x)$

so $x = -4, 6$

$\therefore$ the x-intercepts are $(-4, 0)$ and $(6, 0)$.

To find the vertex, we substitute $x = 1$:

$y = 2(1)^2 - 4(1) - 48$

$= 2 - 4 - 48$

$= -50$

$\therefore$ the vertex is $(1, -50)$.

To find the y-intercept,

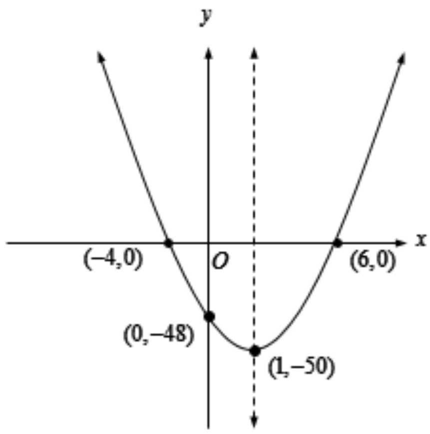
we substitute $x = 0$:

$y = 2(0)^2 - 4(0) - 48$

$= 0 - 0 - 48$

so $y = -48$

$\therefore$ the y-intercept is $(0, -48)$.



Appendix R

Experiment 4: Worked-Examples – Low-Variability Handout

WORKED EXAMPLES – LOW VARIABILITY

University Preparation Program, UNSW Sydney

Investigator: Vicki Likourezos | Experiment 1, 2018

(1) Given $f(x) = 4x + 8$, for all real x,

(i)
Find $f(0)$:

$$f(0) = 4(0) + 8$$
$$= 0 + 8$$
$$= 8$$

(ii)
Find $f(-2)$:

$$f(-2) = 4(-2) + 8$$
$$= -8 + 8$$
$$= 0$$

(iii)
Find $f(a)$:

$$f(a) = 4(a) + 8$$
$$= 4a + 8$$

(2) Given $f(x) = x^2 - 6x - 16$, for all real x,

(i)
Find $f(0)$:

$$f(0) = (0)^2 - 6(0) - 16$$
$$= 0 - 0 - 16$$
$$= -16$$

(ii)
Find $f(-2)$:

$$f(-2) = (-2)^2 - 6(-2) - 16$$
$$= 4 + 12 - 16$$
$$= 0$$

(iii)
Find $f(a)$:

$$f(a) = (a)^2 - 6(a) - 16$$
$$= a^2 - 6a - 16$$

(3) To find the maximum or minimum value of a quadratic function, follow the steps below:

Consider $f(x) = x^2 - 6x - 16$

coefficient is 1

coefficient is – 6

constant is – 16

monic quadratic

completing the square: $\left(-\frac{6}{2}\right)^2 = 9$

$$f(x) = \underbrace{x^2 - 6x + 9}\ \underbrace{- 9 - 16}$$

$$= (x-3)^2 \quad - 25$$

$$(x-3)^2 \geq 0$$

The minimum value of $(x-3)^2$ occurs when $x = 3$.

Therefore, the minimum value of the quadratic function is – 25.

(4) To determine the concavity of the following quadratic functions, we need to consider the coefficient of the x^2 term:

(i) $y = x^2 - 2x - 3$

$\downarrow$

coefficient is 1

$\downarrow$

$\therefore a = 1$

Since $a > 0$, the parabola is concave up.

(ii) $y = (x-4)(x+1)$

$= x^2 + x - 4x - 4$

$= x^2 - 3x - 4$

$\downarrow$

coefficient is 1

$\downarrow$

$\therefore a = 1$

Since $a > 0$, the parabola is concave up.

(iii) $y = -x^2 + 6x + 7$

$\downarrow$

coefficient is -1

$\downarrow$

$\therefore a = -1$

Since $a < 0$, the parabola is concave down.

(5) Sketch the following parabolas showing all essential features:

(i) $y = x^2 + 12x + 35$

coefficient is 1 — $\therefore a = 1$

coefficient is 12 — $\therefore b = 12$

constant is 35 — $\therefore c = 35$

Since $a > 0$, the parabola is concave up.

The axis of symmetry can be found using $x = -\dfrac{b}{2a}$

$x = -\frac{12}{2(1)} = -6$

To find the x-intercept(s), we substitute $y = 0$:

$0 = x^2 + 12x + 35$

$0 = (x + 7)(x + 5)$

so $x = -7, -5$

$\therefore$ the x-intercepts are $(-7,0)$ and $(-5,0)$.

To find the vertex, we substitute $x = -6$:

$y = (-6)^2 + 12(-6) + 35$

$= 36 - 72 + 35$

$= -1$

$\therefore$ the vertex is $(-6, -1)$.

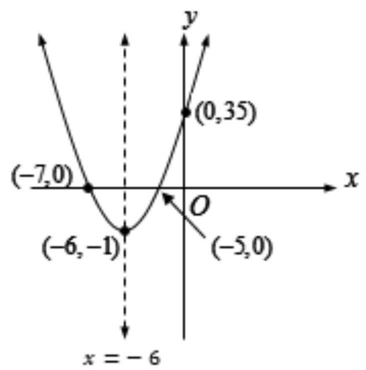
To find the y-intercept, we substitute $x = 0$:

$y = (0)^2 + 12(0) + 35$

$= 0 + 0 + 35$

so $y = 35$

$\therefore$ the y-intercept is $(0,35)$.



(ii) $y = x^2 - 10x + 25$

coefficient is 1 — coefficient is -10 — constant is 25

$\therefore a = 1$ $\quad \therefore b = -10$ $\quad \therefore c = 25$

Since $a > 0$, the parabola is concave up.

The axis of symmetry can be found using $x = -\dfrac{b}{2a}$

$$x = -\frac{-10}{2(1)} = 5.$$

To find the x-intercept(s), we substitute $y = 0$:

$0 = x^2 - 10x + 25$

$0 = (x-5)(x-5)$

so $x = 5$

$\therefore$ the x-intercept is $(5,0)$.

To find the vertex, we substitute $x = 5$:

$y = (5)^2 - 10(5) + 25$

$= 25 - 50 + 25$

$= 0$

$\therefore$ the vertex is $(5,0)$.

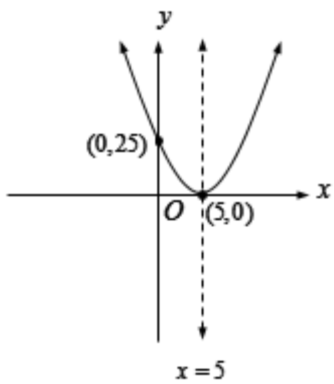
To find the y-intercept, we substitute $x = 0$:

$y = (0)^2 - 10(0) + 25$

$= 0 - 0 + 25$

so $y = 25$

$\therefore$ the y-intercept is $(0,25)$.



(iii) $y = x^2 + 4x + 6$

coefficient is 1 — coefficient is 4 — constant is 6

$\therefore a = 1$ $\quad \therefore b = 4$ $\quad \therefore c = 6$

Since $a > 0$, the parabola is concave up.

The axis of symmetry can be found using $x = -\dfrac{b}{2a}$

$$x = -\frac{4}{2(1)} = -2.$$

To find the vertex, we substitute $x = -2$:

$$y = (-2)^2 + 4(-2) + 6$$
$$= 4 - 8 + 6$$
$$= 2$$

$\therefore$ the vertex is $(-2, 2)$.

To find the x-intercept(s),

we substitute $y = 0$:

$0 = x^2 + 4x + 6$

This quadratic equation cannot be factorised, so we check the discriminant:

$\Delta = b^2 - 4ac$

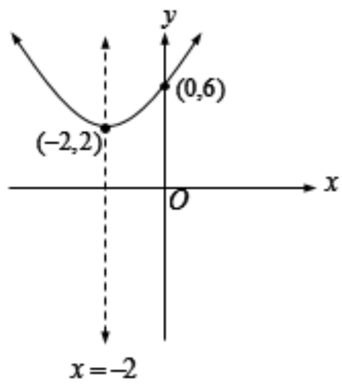
$\Delta = (4)^2 - 4(1)(6)$

$\quad = 16 - 24$

$\quad = -8$

since $\Delta < 0$

$\therefore$ there are no x-intercepts.



To find the y-intercept,

we substitute $x = 0$:

$$y = (0)^2 + 4(0) + 6$$
$$= 0 + 0 + 6$$

so $y = 6$

$\therefore$ the y-intercept is $(0, 6)$.

(iv) $y = -x^2 + 6x + 7$

coefficient is – 1   coefficient is 6   constant is 7

$\therefore a = -1$ $\therefore b = 6$ $\therefore c = 7$

The axis of symmetry can be found using $x = -\dfrac{b}{2a}$

$x = -\dfrac{6}{2(-1)} = 3.$

Since a < 0, the parabola is concave down.

To find the x-intercept(s),

we substitute $y = 0$:

$0 = -x^2 + 6x + 7$

$0 = -(x^2 - 6x - 7)$

$0 = x^2 - 6x - 7$

$0 = (x - 7)(x + 1)$

so $x = 7, -1$

$\therefore$ the x-intercepts are $(7,0)$ and $(-1,0)$.

To find the vertex, we substitute $x = 3$:

$y = -(3)^2 + 6(3) + 7$

$= -9 + 18 + 7$

$= 16$

$\therefore$ the vertex is $(3,16)$.

To find the y-intercept,

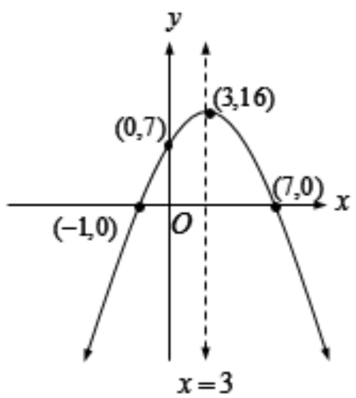
we substitute $x = 0$:

$y = -(0)^2 + 6(0) + 7$

$= -0 + 0 + 7$

so $y = 7$

$\therefore$ the y-intercept is $(0,7)$.



Appendix S

Publication (from the thesis)

Educational Psychology Review (2019) 31:479–497
https://doi.org/10.1007/s10648-019-09462-8

INTERVENTION STUDY

The Variability Effect: When Instructional Variability Is Advantageous

Vicki Likourezos [1] · **Slava Kalyuga** [1] · **John Sweller** [1]

Published online: 12 January 2019

Abstract
Based on cognitive load theory, this paper reports on two experiments investigating the variability effect that occurs when learners' exposure to highly variable tasks results in superior test performance. It was hypothesised that the effect was more likely to occur using high rather than low levels of guidance and testing more knowledgeable than less knowledgeable learners. Experiment 1, which tested 103 adults studying pre-university mathematics, showed no interaction between levels of variability (high vs. low) and levels of instructional guidance (worked examples vs. unguided problem solving). The significant main effect of variability indicated a variability effect regardless of levels of instructional guidance. Experiment 2, which tested another group of 56 adults enrolled in the same mathematics program, showed an interaction between levels of variability (high vs. low) and levels of learner expertise (novices vs. experts). More experienced learners learned more from high rather than low variability tasks demonstrating the variability effect, while less experienced learners learned more from low rather than high variability tasks demonstrating a reverse variability effect. It was suggested that more experienced learners had sufficient available working memory capacity to process high variability information while less experienced learners were overwhelmed by high variability and learned more using low variability information. Subjective ratings of difficulty supported the assumptions based on cognitive load theory. The major educational implication is that learners should initially be presented with low variability or easier tasks, and as they gain more experience in the task domain, variability or task difficulty should increase.

Keywords Cognitive load theory · Worked example effect · Expertise reversal effect · Variability effect

✉ Vicki Likourezos
v.likourezos@unsw.edu.au

[1] School of Education, University of New South Wales, Sydney, Level 1, John Goodsell Building, Library Road, Kensington, NSW 2052, Australia

The variability effect and its interactions with the worked example and expertise reversal effects are examined in this paper. According to *cognitive load theory*, providing novice learners with appropriately designed worked examples during the initial learning phase is more effective than using problem-solving strategies (*worked example effect*, e.g., Atkinson et al. 2000; Carroll 1994; Cooper and Sweller 1987; Paas and van Merriënboer 1994; Schwonke et al. 2009; Sweller and Cooper 1985). This effect can be eliminated or even reversed using easier materials or more expert learners (*expertise reversal effect,* e.g., Bokosmaty et al. 2015; Chen et al. 2017; Kalyuga 2007; Kalyuga et al. 2003; Kalyuga et al. 2001; Kalyuga et al. 2012). The *variability effect* occurs when increasing variability exposes learners to a greater range of related tasks designed to teach a specific concept or procedure, which yields increased test performance (Paas and van Merriënboer 1994; Quilici and Mayer 1996).

Variability in this case refers to surface and structural changes in the question format that learners need to recognise in order to deal appropriately with the information. For example, learners need to recognise that the problem *Steve has 2 pencils and Soula has 3. How many do they have altogether?* is arithmetically the same problem as *How many books do Bella and Bob have between them if Bella has 3 and Bob has 2?* The variability effect does not refer to other forms of variability such as the interleaving effect, where unrelated tasks are interleaved by presenting, for example, successive 30-min lessons on each mathematics, music and science followed by another set of successive 30-min lessons in the same subject areas, as opposed to presenting a single set of one-hour lessons on each subject.

This paper begins with a brief description of the basic theoretical assumptions underlying cognitive load theory, and a discussion of the main empirical findings related to the worked example, expertise reversal and variability effects. The theoretical part is followed by a report on two experiments that investigated the interactions between levels of variability of learning tasks and levels of instructional guidance, as well as levels of learner expertise. Experiment 1 explored the relation between low and high variability tasks and two levels of guidance - high guidance (studying fully-guided worked examples) and low guidance (generating problem solutions without any guidance). Experiment 2 explored the relation between low and high variability tasks and two levels of learner expertise - more experienced learners and less experienced learners.

Principles of Cognitive Load Theory

Cognitive load theory (CLT) was developed as an instructional theory based on our understanding of human cognitive architecture (see Sweller et al. 2011, for a recent overview). Sweller and Sweller (2006) conceptualised CLT in evolutionary terms by aligning it with biological evolutionary principles and Geary's (2008) work on biologically primary and secondary knowledge. Biologically primary knowledge is acquired unconsciously rather than by explicit instruction. Learning to speak and listen in a native language provides an example. Because we have evolved to acquire such knowledge automatically, it cannot be taught. On the other hand, biologically secondary knowledge does need to be explicitly taught and requires conscious effort, with most subject areas taught in educational institutions belonging to this category. CLT is strongly associated with the acquisition of biologically secondary knowledge.

The manner in which human cognition acquires and uses knowledge is based on five fundamental principles (Sweller 2004, 2008; Sweller and Sweller 2006). The *information store principle* refers to the large store of encoded information in human long-term memory that manages most human cognitive activity. The *borrowing and reorganising principle* relates to the transfer of information to long-term memory to facilitate the building of the information store. Information is borrowed from other people and reorganised before it is stored in long-term memory. The *randomness as genesis principle* is used when individuals endeavour to solve a problem by random generation and test without knowledge of solution procedures acquired through the borrowing and reorganising principle. The *narrow limits of change principle* refers to the working memory's limitations in capacity and duration. The *environmental organising and linking principle* indicates that when dealing with organised information from long-term memory, working memory's capacity and duration have no known limits. Based on environmental signals, a large amount of organised information stored in long-term memory can be brought into working memory to govern action appropriate to that environment.

These five principles provide a cognitive architecture able to deal with biologically secondary information. Based on these principles, the aim of instruction should be to facilitate the acquisition of information to be stored in long-term memory, taking into account that novel information must be processed by the limited capacity and duration of working memory. Once stored in long-term memory, information can be retrieved by working memory without any known limits.

Categories of Cognitive Load

There are two types of cognitive load imposed by learning materials: *intrinsic* and *extraneous* load (Sweller 1994, 2010; Sweller and Chandler 1994). Intrinsic cognitive load relates to the innate complexity of the material that must be learned. Intrinsic load is fixed for a given learning task and given level of learner expertise in the domain and can only be altered by changing the nature of the task or by changing the learner's knowledge level. Extraneous cognitive load, on the other hand, is not inherent to the learning material and can be altered by changing the instructional design of the learning material to optimise learning outcomes. Thus, extraneous load is considered unnecessary cognitive load because it is imposed by a non-optimal instructional design. Working memory resources that are used to deal with intrinsic cognitive load and so lead to learning are referred to as germane resources or germane cognitive load, although it is not an independent type of cognitive load (Sweller 2010).

Levels of cognitive load, whether intrinsic or extraneous, are determined by levels of element interactivity. Interacting elements need to be processed simultaneously in working memory for learning to occur (Sweller 1994). Learning new material that consists of a high number of interacting elements is usually difficult because it imposes a high working memory load. If the elements are intrinsic to the instructional goals, the load can be reduced by constructing knowledge in long-term memory that then can be retrieved as a single element using the environmental organising and linking principle. If the high level of element interactivity is caused by instructional design, extraneous cognitive load can be reduced by changing the design, which constitutes the major function of CLT (Sweller 2010).

Cognitive Load Effects

Worked Example Effect

The worked example effect occurs when learners who study worked examples learn more and perform better on tests compared to learners who problem solve the equivalent tasks (Sweller et al.

2011). Worked examples emphasise the borrowing and reorganising principle rather than the randomness by genesis principle. They explicitly provide information that learners need by allowing them to borrow and reorganise new knowledge rather than randomly generating ideas and testing them for effectiveness (Sweller 2006). Compared to problem solving, worked examples reduce element interactivity associated with extraneous cognitive load (Sweller 2010). Searching for solution steps requires learners to process more elements than being presented with the appropriate steps. That reduction in element interactivity when learning from worked examples can be important when dealing with the increase in element interactivity associated with the variability effect.

Variability Effect

The variability effect occurs when high variability tasks result in enhanced learning compared to tasks with less variability (Paas and van Merriënboer 1994). High variability tasks are higher in element interactivity compared to low variability tasks because more elements must be processed simultaneously in working memory to understand the task at hand. As variability increases, the number of interacting elements increases because learners must learn to distinguish the variety of tasks that require similar solutions. Thus, a set of high variability tasks is intrinsically more difficult to complete compared to a similar set of low variability tasks. Provided that intrinsic cognitive load does not exceed working memory capacity, learning and problem solving with high variability tasks are expected to improve because the quality of constructed knowledge is enhanced. In contrast, if the increased intrinsic cognitive load associated with increased variability does exceed working memory limits, increased variability should result in decreases in learning and problem-solving performance (Chen et al. 2018).

The results of studies on variability reveal that exposure to highly variable, example-based instruction results in enhanced transfer performance (Paas and van Merriënboer 1994; Quilici and Mayer 1996). In a 2 (worked examples vs. problem solving) by 2 (high vs. low variability) design, Paas and van Merriënboer (1994) showed that worked examples led to better transfer compared to problem-solving tasks but also showed that for the worked example condition, high variability was superior to low variability. This advantage disappeared under problem-solving conditions.

The use of worked examples may allow the increased number of interacting elements associated with intrinsic cognitive load to be processed because worked examples reduce element interactivity due to extraneous cognitive load (Sweller 2010). In contrast, because of the high element interactivity related to extraneous cognitive load that is associated with problem solving, an increase in element interactivity associated with variability may increase the overall cognitive load beyond the point where working memory can adequately deal with the information. Accordingly, the variability effect was obtained by Paas and van Merriënboer (1994) using worked examples with their low levels of extraneous cognitive load but not with problem-solving tasks with their high levels of extraneous cognitive load. Provided the increase in intrinsic cognitive load does not exceed working memory capacity, learning should improve under high variability conditions due to the learner successfully processing the essential interacting elements that are important to the task.

Expertise Reversal Effect

The expertise reversal effect occurs when the advantage of one instructional procedure over another is observed for novices but disappears or even reverses with increased expertise.

Similar to other cognitive load effects, this effect can also be explained in terms of element interactivity which depends on the learner's level of expertise (Chen et al. 2017; Kalyuga 2007). An instructional situation that constitutes high element interactivity for a novice learner (e.g., an unguided problem-solving task) may involve low element interactivity for a more knowledgeable learner. Since novices may lack the necessary knowledge to process complex, high element interactivity material in working memory, if instructional guidance is provided, this guidance can effectively substitute for the non-existent knowledge, thus effectively reducing the level of element interactivity for these learners. Conversely, if unnecessary instructional guidance is provided for experts, this redundant information may impose an additional extraneous cognitive load, thus effectively increasing the level of element interactivity. These learners will need to consume additional working memory resources to deal with information consisting of additional elements that they already know. This subsequently impedes their learning compared to situations without any guidance, and results in the expertise reversal effect (Kalyuga et al. 2003).

All cognitive load effects are theoretically susceptible to the expertise reversal effect, and there is every reason to hypothesise that the variability effect will disappear or even reverse with changes in element interactivity caused by changes in levels of expertise (Chen et al. 2018). In order for the effect to occur, element interactivity needs to be sufficiently low to ensure that the increase in element interactivity due to increases in variability does not exceed working memory capacity. If element interactivity is too high, the introduction of variability may overwhelm working memory. A lower level of variability may enhance learning compared to a higher level resulting in a reverse variability effect.

Present Study

The present study explored whether providing learners with high or low variability mathematical tasks would facilitate learning. We investigated the effect of high or low variability conditions on worked examples or problem-solving tasks (Experiment 1), and the effect of high or low variability conditions provided to more experienced (expert) or less experienced (novice) learners (Experiment 2).

In Experiment 1 we hypothesised, in line with Paas and van Merriënboer (1994), that the variability effect was more likely to occur under worked example than problem-solving conditions assuming that problem-solving conditions would increase element interactivity due to extraneous cognitive load beyond the point where working memory could handle the increase in element interactivity associated with higher variability. In Experiment 2 we hypothesised that more expert learners would demonstrate the variability effect because the learners' expertise would reduce element interactivity to the point where they had sufficient working memory capacity to be able to process the increased number of interacting elements associated with the increase in intrinsic cognitive load due to higher variability. For novice learners, we hypothesised that any further increase in intrinsic cognitive load due to variability may result in element interactivity exceeding working memory limits resulting in a reverse variability effect with low variability proving superior to high variability.

Experiment 1 This experiment examined whether the variability effect could be obtained with different levels of instructional guidance that involved studying worked examples or attempting to solve problems. The experiment used a 2 by 2 design with two levels of variability (high or low) and two levels of guidance (worked examples or problem solving). Variability was achieved by changing the range of tasks for which worked examples were studied or problems had to be solved. In the case of high variability, this was done via applying the same solution process in a wider variety of contexts by changing not only the surface features (i.e. numbers) but also varying the structure of the problems (i.e. question formats), as opposed to the low variability problems whereby only some surface features changed. The worked examples condition required students to study fully-guided tasks that contained step-by-step solutions on how to solve a problem, while the problem-solving condition required students to generate problem solutions on their own without any guidance. Since instructional formats containing high variability mathematical tasks are likely to increase mental effort because they are performed under conditions that require a highly varied sequence of solution steps, this experiment also explored whether completing high variability mathematical tasks would increase cognitive load. Given the importance of enhancing learner transfer performance, the objective of the experiment was to examine which variability-guidance combination would yield superior transfer performance outcomes.

Hypotheses

1. Learners who study worked examples that provide explicit solution steps will yield better post-test performance, compared to learners that generate problem solutions without the provision of any solution steps.
2. Providing learners with high variability tasks under worked example conditions will generate a variability effect with better post-test performance, compared to using low variability tasks, while problem-solving conditions will not generate this difference.
3. Subjective ratings of difficulty for attempting to solve problems (without guidance) will be higher compared to studying fully-guided worked examples, irrespective of the level of variability of the task.
4. Subjective ratings of difficulty for completing high variability tasks will be higher compared to completing low variability tasks, irrespective of the level of guidance provided.

Method

Participants The participants were 103 mathematics students, aged between 18 and 55 (M_{age} = 26.57, SD_{age} = 7.55), enrolled in a preparation program at an Australian university. This post-secondary education program prepares students for admission to university programs. The sample comprised of 42 females (41%) and 61 males.

Materials The material used in the experiment focused on the definition of a quadratic function; the roots of a quadratic function; the axis of symmetry and the vertex of a parabola; and how to draw the graph of a quadratic function. All participants were regarded as novice learners in relation to quadratic functions as this topic was the next scheduled topic in the mathematics preparation program.

During the first part of the Learning Phase, the experiment convenor provided explicit instruction for the topic by demonstrating solutions of the relevant tasks on the board. During the second half of the Learning Phase, each participant received a handout in accordance with the assigned experimental group they were in: 'worked examples-high variability'; 'worked examples-low variability'; 'problem solving-high variability'; or 'problem solving-low variability'. The high variability handouts contained tasks that differed in the presentation format and the solution procedures from the tasks performed on the board during the explicit instruction part of the Learning Phase. The low variability handouts, on the other hand, contained tasks that were similar to those performed on the board. The worked examples handouts contained fully-guided worked examples with explicit solution steps and diagrams. The problem-solving handouts consisted only of problem statements, excluding any written instructions or diagrams. For example, the first task in the high variability handout consisted of the following questions: *Given* $g(x) = 5 - 2x$, *for all real x, find: (i)* $g(-3x)$; *(ii)* $g(\frac{1}{4})$; and *(iii)* $g(a + 5)$. These high variability questions required the participants to substitute three highly varied expressions into the function $g(x)$, namely: '$-3x$' (a negative number and a variable); '$\frac{1}{4}$' (a fraction); and '$a + 5$' (an algebraic expression). The first task in the low variability handout, on the other hand, consisted of the following questions: *Given* $f(x) = 4x + 8$, *for all real x, find: (i)* $f(0)$; *(ii)* $f(-2)$; and *(iii)* $f(a)$. These low variability questions contained only whole numbers (0 and −2) and a simple variable a. In addition, the name of the function in the low variability tasks, $f(x)$, was the name of the function used during the explicit instruction on the board by the experiment convenor. To increase variability in the high variability handout, the name of the function was changed from f to g.

All participants were given a single-item, nine-point Likert-type rating scale to complete: 'How difficult was it for you to complete the tasks?' ranging from 'Extremely Easy' (on the far left with point 1 assigned to the answer) to 'Neither Easy nor Difficult' (in the middle, point 5) and to 'Extremely Difficult' (on the far right, point 9). The participants' subjective ratings were used to measure cognitive load experienced during their completion of the Learning Phase handout.

The Post-Test consisted of seven questions that included ten tasks in total. The first six tasks were structurally similar to the questions used during the explicit instruction part of the Learning Phase. The remaining four tasks were structurally different and intended to test for transfer of learning, because they required the capacity to apply the learned knowledge in new situations. The Post-Test was identical in content for all participants. The questions were internally reliable based on a Cronbach's Alpha of 0.75.

Procedure At the start of the experiment, participants were randomly assigned to one of the four experimental conditions: worked examples-high variability group (25 students); worked examples-low variability group (26 students); problem solving-high variability group (26 students); and problem solving-low variability group (26 students). The random assignment was achieved by allowing the students to choose their seat when they entered the lecture theatre and then handing out the different materials in a sequential order so that every fourth student received the same material.

The duration of the experiment was one and a half hours, and it was conducted during the participants' normal mathematics lecture and tutorial time. The experiment consisted of a Learning Phase (60 minutes) and a Post-Test Phase (30 minutes). In the first half of the Learning Phase (30 minutes), the experiment convener provided all participants with comprehensive explicit instructions on quadratic functions reflecting standard solution procedures on the board. In the second half of the Learning Phase (30 minutes), participants were instructed to complete a different handout that assigned them to their respective experimental group. The worked examples-high variability group and the worked examples-low variability group were instructed to study step-by-step worked-out solutions for fourteen high or low variability problems respectively. The problem solving-high variability group and the problem solving-low variability group were instructed to generate solutions for the same fourteen high or low variability problems. The last page of the high variability and low variability problem-solving handouts contained answers (but not step-by-step worked-out solutions) to each problem task on the previous pages to enable the participants to compare their answers to the correct answers.

After the Learning Phase handouts were collected, each participant completed a subjective rating of difficulty. Following the subjective rating questionnaire, all participants completed the Post-Test.

Marking Procedure The marking of the Post-Tests, which was an objective test, was undertaken by the experiment convenor. Consistency was achieved by considering all possible solutions and applying the same rubric throughout the marking process. Answers to questions were awarded one mark for every correct solution step, and incorrect solution steps were awarded a mark of zero. Based on this procedure, the highest possible total score for the Post-Test was 38 marks, consisting of 25 marks for the similar questions and 13 marks for the transfer questions. All the raw scores were converted into percentage scores.

Results and Discussion

There were two independent variables: level of variability and level of guidance, and three dependent variables: Post-Test (similar questions) scores; Post-Test (transfer questions) scores; and subjective ratings of difficulty. Table 1 shows the descriptive statistics for the participants' performance.

Table 1 Means (standard deviations) of the average class test scores, adjusted post-test (total) scores, adjusted post-test (similar questions) scores, adjusted post-test (transfer questions) scores and subjective ratings of difficulty

	Experimental conditions							
	Worked examples-high variability group		Worked examples-low variability group		Problem solving-high variability group		Problem solving-low variability group	
	M (SD)	*N*	*M* (SD)	*N*	*M* (SD)	*N*	*M* (SD)	*N*
Average class test scores (%)	61.12 (22.45)	25	53.67 (19.40)	26	68.94 (20.66)	26	69.73 (19.34)	26
Adjusted post-test (total) scores (%)	43.87 (15.71)	25	35.95 (16.14)	26	37.84 (15.83)	26	32.41 (15.88)	26
Adjusted post-test (similar questions) scores (%)	45.91 (18.31)	25	37.98 (18.81)	26	38.16 (18.45)	26	33.10 (18.51)	26
Adjusted post-test (transfer questions) scores (%)	39.94 (16.79)	25	32.06 (17.25)	26	37.22 (16.92)	26	31.09 (16.97)	26
Subjective ratings of difficulty (1–9)	3.64 (2.20)	25	3.92 (2.23)	26	6.69 (2.13)	26	6.85 (1.57)	26

Prior Knowledge The prior mathematical knowledge of each participant was measured by averaging scores for three previous class tests that were completed before the commencement of the experiment. Even though these class tests comprised questions that differed from the topic content used in the present experiment, the average class test scores were used as a covariate.

To compare the level of prior mathematical knowledge for the four groups, a one-way between-groups analysis of variance was conducted for the average class test scores. It involved one independent variable (the condition group) across four levels (worked examples-high variability group, worked examples-low variability group, problem solving-high variability group and problem solving-low variability group) and one dependent variable (average class test score). The results showed a statistically significant difference, $F(3,99) = 3.52$, $\mathrm{MSE} = 1477.44$, $p = 0.02$, partial $\eta^2 = 0.10$ (medium effect size). Bonferroni post hoc pairwise comparisons indicated a significant difference between the worked example-low variability group and problem solving-low variability group only. To control for the unexpected differences between the experimental groups in the prior knowledge analyses, the average class test score was used as a covariate for all Post-Test performance results.

Post-Test Scores Two 2 by 2 between-groups analyses of covariance were conducted on the Post-Test (similar questions) scores and Post-Test (transfer questions) scores. The results for variability on the Post-Test (similar questions) scores did not reach statistical significance, $F(1,98) = 3.23$, $\mathrm{MSE} = 1080.06$, $p = 0.08$, partial $\eta^2 = 0.03$. The results showed a statistically significant main effect for variability on the Post-Test (transfer questions) scores, $F(1,98) = 4.46$, $\mathrm{MSE} = 1254.08$, $p = 0.04$, partial $\eta^2 = 0.04$. The results of these analyses showed that increasing variability effectively boosted transfer of learning - the ability to solve problems that have not been solved before. In addition, there is evidence of a marginal relationship (a possible effect) between high variability tasks and the application of knowledge and skills in completing similar questions.

The results for guidance were not statistically significant for the Post-Test (similar questions) scores, $F(1,98) = 2.83$, $\mathrm{MSE} = 943.92$, $p = 0.10$, partial $\eta^2 = 0.03$ or the Post-Test (transfer questions) scores, $F(1,98) = 0.29$, $\mathrm{MSE} = 80.53$, $p = 0.59$, partial $\eta^2 = 0.003$. The results of these analyses show there is no evidence of a relationship between guidance and the application of knowledge and skills in completing tasks in similar and novel situations. These results indicate that the expected worked example effect was not obtained.

The variability by guidance interactions were not statistically significant for the Post-Test (similar questions) scores, $F(1,98) = 0.16$, $\mathrm{MSE} = 52.31$, $p = 0.69$, partial $\eta^2 = 0.002$ or the Post-Test (transfer questions) scores, $F(1,98) = 0.07$, $\mathrm{MSE} = 19.48$, $p = 0.79$, partial $\eta^2 = 0.001$. These results show that levels of guidance did not alter the effects of variability on the Post-Test questions.

Subjective Ratings of Difficulty A 2 by 2 between-groups analysis of variance was conducted to assess the effect of the two independent variables on subjective ratings of difficulty (cognitive load). The results showed a statistically significant main effect for guidance, $F(1,99) = 54.88$, $\mathrm{MSE} = 229.79$, $p < 0.001$, partial $\eta^2 = 0.36$. Subjective ratings of difficulty were less for the worked examples groups compared to the problem-solving groups. However, the results for variability were not statistically significant, $F(1,99) = 0.29$, $\mathrm{MSE} = 1.23$, $p = 0.59$, partial $\eta^2 = 0.003$. Thus, the subjective ratings were not different for the low and high variability groups. This result indicates that the level of similarity of the low and high variability tasks respectively did not have an impact on cognitive load. On the other hand, the significant main effect of guidance on subjective ratings indicates that higher cognitive load was imposed on learners who generated solutions to problem-solving tasks (without any guidance) compared to learners who studied the fully-guided worked examples, regardless of the level of variability that the learners were exposed to. The variability × guidance interaction for cognitive load was not statistically significant, $F(1,99) = 0.03$, $\mathrm{MSE} = 0.11$, $p = 0.87$, partial $\eta^2 = 0.00$.

This experiment tested the assumption that when students are provided with high variability tasks, they will attain higher test scores under worked example conditions. This hypothesis was supported. We hypothesised that this effect would not be obtained under problem-solving conditions but in fact, the same effect was obtained under both worked example and problem-solving conditions.

The results did not support the hypothesis that worked examples would facilitate superior test performance. There were no significant differences between worked example and problem-solving conditions. Neither did the results replicate the Paas and van Merriënboer (1994) findings that the variability effect was only obtainable using worked examples, but not problem solving, even though the worked example group found their instructional procedure substantially easier than the problem-solving group. Notwithstanding, our results are entirely consistent and coherent with the Paas and van Merriënboer (1994) findings.

Whether or not a worked example effect can be obtained is determined by the participants' knowledge levels. The worked example effect is just as susceptible to the expertise reversal effect as any other cognitive load effect. If knowledge levels are sufficiently high, the effect will not be obtained or may even be reversed (Kalyuga et al. 2001). In the current experiment, there was no evidence of improved test performance following instruction using worked examples rather than problems. With no difference in test performance between worked example and problem-solving conditions, we should not expect any differences in the effect of variability associated with worked examples or problem solving. If test performance scores following problem solving are just as high as following worked examples, there should be sufficient working memory resources available to handle the element interactivity associated with problem solving and variability. Accordingly, our failure to replicate the Paas and van Merriënboer (1994) results following problem solving is to be expected given the failure to obtain a worked example effect on the test results. Both the worked example and problem-solving conditions should be susceptible to a variability effect as indicated by our results.

Experiment 2

The variability effect was attained in Experiment 1 following both worked examples and problem solving, arguably due to the participants' knowledge levels being too high to obtain a worked example effect. Experiment 2 tested the variability effect further in connection with levels of learner prior knowledge by comparing the test performance of more and less experienced learners. This experiment examined whether more experienced learners would demonstrate the variability effect (i.e. high variability tasks resulting in increased test performance), while less experienced learners, who are unable to effectively process high variability tasks, would demonstrate a reverse variability effect.

For more experienced learners, element interactivity should be relatively low allowing them to readily process the additional element interactivity associated with increased variability. A conventional variability effect should result. For less experienced learners, element interactivity should be higher, and adding additional element interactivity by increasing variability may result in cognitive load exceeding working memory capacity (Chen et al. 2018). Accordingly, reduced rather than increased test performance due to increased variability should be the result, with a reverse variability effect being obtained. A conventional variability effect for more expert learners associated with a reverse variability effect for less able learners would provide an example of an expertise reversal effect.

In effect, in Experiment 2 we were attempting to reinstate the results obtained by Paas and van Merriënboer (1994) by changing levels of expertise rather than changing levels of guidance. Increased levels of guidance via worked examples should reduce element interactivity just as increased expertise should reduce element interactivity. Both should allow the variability effect to occur provided element interactivity is sufficiently low to allow working memory to process the increased elements associated with variability. In contrast, if element interactivity is too high, increasing it further by increasing variability should eliminate the variability effect or even reverse it.

Hypotheses

1. Providing more knowledgeable learners with high variability tasks, compared to low variability tasks, will result in superior post-test performance on the high variability tasks.
2. Providing less knowledgeable learners with low variability tasks, compared to high variability tasks, will result in superior post-test performance on the low variability tasks.
3. Subjective ratings of difficulty by less knowledgeable learners will be higher compared to more knowledgeable learners.
4. Subjective ratings of difficulty for completing high variability tasks will be higher compared to completing low variability tasks.

Method

Participants The participants were 56 mathematics students enrolled in the same Australian university preparation program to the one in Experiment 1. The participants were aged between 18 and 55 (M_{age} = 26.27, SD_{age} = 8.20), comprising of 21 females (37.5%) and 35 males.

Materials The material used in the experiment comprised the definition of a logarithm; logarithmic laws; and solving logarithmic equations. The domain-specific prior knowledge of each participant was measured using a Pre-Test that consisted of 22 tasks. The first 16 tasks (in Questions 1–4) evaluated baseline pre-requisite knowledge required for learning logarithmic equations. These tasks were simpler than the tasks in the experimental Learning Phase materials (and accordingly, in the Post-Test) as they required knowledge of individual components and isolated procedures involved in the learning tasks. A combination of these components and procedures was required to work out the Learning Phase and Post-Test tasks. For example, the first tasks in Questions 1–4 were respectively: *In the expression* $\mathbf{2^3}$*: What is the exponent?*; *Simplify the following, writing the answers in index form with positive indices:* $8a^4 \times 2a^6$; *Find the exact value of the following without using a calculator:* 8^{-2}; and *Solve the following equations for x:* $4^x = \frac{1}{16}$. The final 6 tasks (in Questions 5–7) were analogous to typical test questions that met the expected learning outcomes for logarithmic equations. For example, the second tasks in Questions 5–7 were respectively: *In the logarithmic equation* $log_3 9 = 2$*: What is the base of the log?*; *Make x the subject for the following:* $log_b (x+2) = 3$; and *Solve the following equations for x:* $log_3 1 = x$. These six logarithmic tasks were included at the end of the Pre-Test to gauge if the participants had any prior understanding of the new topic that was going to be taught during the first half of the Learning Phase. These six logarithmic tasks were included again in the Post-Test. The Pre-Test questions were identical in content for all participants and were internally reliable based on a Cronbach's Alpha of 0.75.

During the second half of the Learning Phase, each participant received a handout in accordance with the assigned experimental group they were in: 'expert-low variability', 'expert-high variability', 'novice-low variability' or 'novice-high variability'. The low variability groups received handouts that contained problem-solving tasks with variables that changed in the same part of each question. For example, the first task in the low variability handout consisted of the following question: *Write the following in logarithmic form, without solving for x: (a)* $3^x = 9$; *(b)* $4^x = \frac{1}{4}$; *(c)* $125^x = 5$; and *(d)* $32^x = 4$. The high variability groups received handouts that contained the same problem-solving tasks as those in the low variability handout with the exception that the variables changed in different parts of each question. For example, the first task in the high variability handout consisted of the following question: *Write the following in logarithmic form, without solving for x: (a)* $3^x = 9$; *(b)* $4^{-1} = x$; *(c)* $x^{\frac{1}{3}} = 5$; and *(d)* $32^x = 4$.

All participants were given a single-item, nine-point Likert-type rating scale to complete, identical to that used in experiment 1. The participants' subjective ratings were used to measure cognitive load exerted during their completion of the Learning Phase handout.

Unlike Experiment 1, in which the participants completed one Post-Test (that contained similar and transfer questions), the participants in Experiment 2 completed two Post-Tests to ensure that an equal amount of time was allocated to answering similar and transfer questions. Post-Test 1 comprised similar tasks and Post-Test 2 comprised transfer tasks. In Post-Test 1, the first six tasks (Questions 1–3) were identical to the last six tasks of the Pre-Test. The remaining 10 tasks (Questions 4–6) were structurally similar, both in context and concept, to the tasks that were performed on the board by the experiment convenor during the explicit instruction part of the Learning Phase. In Post-Test 2, the seven transfer tasks were structurally different, both in context and concept, from the tasks that were performed on the board by the experiment convenor during the explicit instruction part of the Learning Phase. These questions required the application of acquired knowledge in relatively new task situations. The questions were internally reliable based on a Cronbach's Alpha of 0.82 for the similar questions and 0.69 for the transfer questions in Post-Tests 1 and 2, respectively.

Procedure The experiment was conducted during the participants' normal mathematics lecture and tutorial time. Its duration was 30 minutes for the Pre-Test followed 1 week later by a one-and-a-half-hour block. Participants who scored in the top half of the Pre-Test were designated as experts (more experienced learners). Participants who scored in the bottom half of the Pre-Test were designated as novices (less experienced learners). Participants were evenly apportioned with 28 participants in the expert groups and 28 participants in the novice groups.

During the one-and-a-half-hour block, the experiment comprised of a Learning Phase (60 minutes) and a Post-Test Phase (30 minutes). At the start of the one-and-a-half-hour block, expert participants were randomly assigned to one of two experimental conditions: expert-low variability or expert-high variability, and likewise novice participants were randomly assigned to one of two experimental conditions: novice-low variability or novice-high variability, giving fourteen participants in each of the four groups. The random assignment was achieved by allowing the expert and novice participants to choose their seat when they entered the lecture theatre.

In the first half of the Learning Phase (30 minutes), the experiment convener provided all participants with comprehensive explicit instructions on the board by demonstrating step-by-step solution procedures for the new topic of logarithmic equations. In the second half of the Learning Phase, the expert and novice participants received different handouts assigning them to their respective experimental groups (high or low variability). The first page of the handout, which was identical for all groups, contained four worked examples that consisted of fully-guided written instructions. The remaining three pages of the handout contained sixteen problem solving tasks which differed according to the experimental condition (high or low variability). The purpose of the worked examples on the first page was to provide a summary of what had been taught on the board by the experiment convenor during the first half of the Learning Phase and to provide participants with a general guide on how to generate solutions for the problem-solving tasks that followed on the remaining pages. Participants were given 30 minutes to complete their Learning Phase handouts. Immediately following, the answers to each of the tasks in both conditions (high or low variability) were presented on the board. Participants were given a few minutes to check their work by comparing the correct answers on the board with their written answers, enabling them to determine how much they had learnt.

After the Learning Phase handouts were collected, each participant completed a subjective rating of difficulty. Following the completion of the subjective rating questionnaire, all participants completed Post-Test 1 in 15 minutes, and then proceeded to completing Post-Test 2 in 15 minutes.

Marking Procedure The marking of the Pre-Tests and Post-Tests, both of which were objective tests, was undertaken by the experiment convenor. Consistency was achieved by considering all possible solutions and applying the same rubric throughout the marking process. Answers to questions were awarded one mark for every correct solution step, and incorrect solution steps were awarded a mark of zero. Based on this procedure, the highest possible score for the Pre-Test was 50 marks, for Post-Test 1 was 34 marks and for Post-Test 2 was 34 marks. As in Experiment 1, the raw scores were converted to percentage scores.

Results and Discussion

There were two independent variables: level of variability and level of expertise, and four dependent variables: Pre-Test scores; Post-Test 1 (similar questions) scores; Post-Test 2 (transfer questions) scores; and subjective ratings of difficulty. Table 2 shows the descriptive statistics for the participants' performance.

Pre-Test Scores In order to evaluate the level of prior knowledge of logarithmic equations for the four groups, a 2 (expert vs. novice) by 2 (high vs. low variability) between-groups analysis of variance was conducted on the Pre-Test score. As expected, the results showed a statistically significant main effect for expertise on the Pre-Test scores, $F(3,52) = 111.15$, $MSE = 9884.57$,

Table 2 Means (standard deviations) of the pre-test scores, adjusted combined post-tests scores, adjusted post-test 1 scores, adjusted post-test 2 scores and subjective ratings of difficulty

	Experimental conditions							
	Expert-low variability group		Expert-high variability group		Novice-low variability group		Novice-high variability group	
	M (SD)	*N*	*M* (SD)	*N*	*M* (SD)	*N*	*M* (SD)	*N*
Pre-test scores (%)	52.00 (8.56)	14	62.14 (13.53)	14	31.43 (5.68)	14	29.57 (8.20)	14
Adjusted combined post-tests scores (%)	39.54 (10.02)	14	63.99 (12.97)	14	57.62 (11.04)	14	27.50 (11.59)	14
Adjusted post-test 1 (similar questions) scores (%)	49.22 (12.94)	14	73.91 (16.76)	14	63.70 (14.26)	14	35.86 (14.97)	14
Adjusted post-test 2 (transfer questions) scores (%)	29.86 (14.08)	14	54.08 (18.23)	14	51.55 (15.51)	14	19.13 (16.28)	14
Subjective ratings of difficulty (1–9)	2.43 (1.60)	14	3.07 (1.33)	14	4.93 (1.64)	14	7.86 (.95)	14

$p < 0.001$, partial $\eta^2 = 0.68$. The results did not show a main effect for variability on the Pre-Test scores, $F(1,52) = 2.70$, MSE = 240.29, $p = 0.11$, partial $\eta^2 = 0.05$, indicating that the high and low variability groups had similar levels of prior knowledge.

The expertise by variability interaction effect on the Pre-Test scores was significant, $F(1,52) = 5.67$, MSE = 504.00, $p = 0.02$, partial $\eta^2 = 0.10$, indicating that the magnitude of the difference between the low and high variability groups' scores was different at different levels of expertise. Given the statistically significant interaction, follow-up analyses were performed to determine whether there were any simple effects. For expert learners, the high variability group scored higher in the Pre-Test than the low variability group, $F(1,26) = 5.62$, MSE = 720.14, $p = 0.03$, partial $\eta^2 = 0.18$, indicating that the more able learners in the high variability group were more knowledgeable than those in the low variability group. For novice learners, there was no simple effect, $F(1,26) = .49$, MSE = 24.14, $p = 0.49$, partial $\eta^2 = 0.02$. To control for the unexpected differences between the experimental groups in the Pre-Test analyses, the Pre-Test scores were used as a covariate in all Post-Test performance analyses.

Post-Test Scores Two 2 by 2 between-groups analyses of covariance were conducted on the Post-Test 1 (similar questions) scores and Post-Test 2 (transfer questions) scores. The results showed that the expert learner groups produced significantly higher scores than the novice learner groups for the Post-Test 1 (similar questions) scores, $F(1,51) = 4.46$, MSE = 619.53, $p = 0.04$, partial $\eta^2 = 0.08$. However, the results did not show a main effect of expertise on Post-Test 2 (transfer questions) scores, $F(1,51) = 1.19$, MSE = 196.14, $p = 0.28$, partial $\eta^2 = 0.02$. The results did not show a main effect of variability on the Post-Test 1 (similar questions) scores, $F(1,51) = 0.24$, MSE = 32.89, $p = 0.63$, partial $\eta^2 = 0.01$ or on Post-Test 2 (transfer questions) scores, $F(1,51) = 1.36$, MSE = 223.68, $p = 0.25$, partial $\eta^2 = 0.03$.

The variability by expertise interaction was statistically significant on the Post-Test 1 (similar questions) scores, $F(1,51) = 62.66$, MSE = 8708.87, $p < 0.001$, partial $\eta^2 = 0.55$ and the Post-Test 2 (transfer questions) scores, $F(1,51) = 61.51$, MSE = 10,119.81, $p < 0.001$, partial $\eta^2 = 0.55$. Figure 1 graphically depicts the dis-ordinal interaction that exists between levels of expertise and levels of variability for the combined means of Post-Test 1 and Post-Test 2 scores.

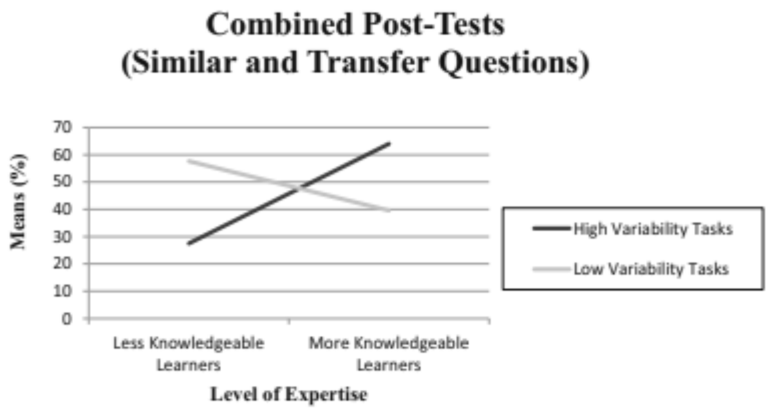


Fig. 1 Variability by expertise interaction using adjusted means on the combined post-tests in Experiment 2

To test the degree to which variability is differentially effective at the expert and novice levels, simple effects were tested following the significant interactions. For expert learners, high variability tasks in Post-Test 1 (similar questions) led to higher scores compared to low variability tasks, $F(1,25) = 21.64$, MSE = 3312.79, $p < 0.001$, partial $\eta^2 = 0.46$. For novice learners, low variability tasks in Post-Test 1 (similar questions) led to higher scores compared to high variability tasks, $F(1,25) = 42.54$, MSE = 5452.47, $p < 0.001$, partial $\eta^2 = 0.63$. For expert learners, high variability tasks in Post-Test 2 (transfer questions) led to higher scores compared to low variability tasks, $F(1,25) = 16.18$, MSE = 2924.96, $p < 0.001$, partial $\eta^2 = 0.39$. For novice learners, low variability tasks in Post-Test 2 (transfer questions) led to higher scores compared to high variability tasks, $F(1,25) = 53.38$, MSE = 7574.65, $p < 0.001$, partial $\eta^2 = 0.68$. These results show that the effect of levels of variability (high and low) differed significantly depending on levels of learner expertise (expert or novice). In particular, superior performance scores for expert learners were associated with high variability tasks, while superior performance scores for novice learners were associated with low variability tasks.

Subjective Ratings of Difficulty A 2 by 2 between-groups analysis of variance was conducted to assess the effect of the two independent variables on subjective ratings of difficulty (cognitive load). The results showed a statistically significant main effect for expertise, $F(1,52) = 93.80$, MSE = 185.79, $p < 0.001$, partial $\eta^2 = 0.64$; a statistically significant main effect for variability, $F(1,52) = 22.54$, MSE = 44.64, $p < 0.001$, partial $\eta^2 = 0.30$ and a significant interaction effect between the level of variability and the level of expertise, $F(1,52) = 9.23$, MSE = 18.29, $p = 0.004$, partial $\eta^2 = 0.15$. Figure 2 depicts the ordinal interaction that exists between levels of expertise and levels of variability for cognitive load during the Learning Phase.

Analyses of simple effects were conducted following the significant interaction between the levels of expertise and the levels of variability. For expert learners, there was no significant difference between cognitive load for high variability tasks compared to low variability tasks during the Learning Phase, $F(1,26) = 1.34$, MSE = 2.89, $p = 0.26$, partial $\eta^2 = 0.05$. For novice learners, cognitive load was higher for high variability tasks compared to low variability tasks, $F(1,26) = 33.47$, MSE = 60.04, $p < 0.001$, partial $\eta^2 = 0.56$. These simple effect analyses show that the interaction between expertise and variability was due to the difficulty less knowledgeable learners had dealing with high variability problems compared to low variability problems. That difference was reduced for more knowledgeable learners.

Fig. 2 Variability by expertise interaction on subjective ratings of difficulty in Experiment 2

This experiment tested the hypotheses that the variability effect could be obtained if learners had sufficient working memory capacity associated with higher levels of expertise to enable them to process the increased levels of element interactivity. Problem sets with higher degrees of variability include more interactive elements of information than problem sets with lower degrees of variability. That increased element interactivity of information increases intrinsic cognitive load. If, because of insufficient levels of expertise, learners do not have sufficient spare working memory capacity to handle the increased intrinsic cognitive load associated with high variability information, it was hypothesised that the variability effect would be reduced or even reversed.

Most hypothesised results were obtained. A standard variability effect was obtained using more knowledgeable learners who learned more from high rather than low variability problems. High variability tasks improved performance on both similar and transfer tasks for these learners. There was no difference in subjective ratings of difficulty for completing high variability tasks, compared to low variability tasks, for expert learners. For the less knowledgeable learners, the reverse result was obtained. Low variability learning tasks improved performance on both similar and transfer tasks for these learners. Furthermore, in the case of less knowledgeable learners, the improvement in scores using low variability problems was associated with a significant reduction in cognitive load. It appears that the novice learners experienced less cognitive load when solving low variability tasks as they were able to identify a surface match between the similarly structured questions without the need to go any further. In contrast, these learners found processing high variability problems relatively difficult. In order to comprehend dissimilar features of high variability tasks, whereby all the questions did not share any common surface features with the tasks studied during the explicit instruction part of the Learning Phase, more mental effort was required to process the deeper features until the underlying common features were found. In addition, the results demonstrated significantly lower cognitive load experienced by the more knowledgeable learners, compared to the less knowledgeable learners.

General Discussion

Both experiments reported in this paper investigated the variability effect. Experiment 1 focused on the level of guidance in conjunction with the level of variability. Experiment 2 focused on the level of expertise in conjunction with the level of variability.

Based on the results of Experiment 1, it was found that providing learners with high variability tasks generated better post-test performance, compared to low variability tasks, irrespective of the level of guidance provided to the learners (worked example or problem-solving condition). Additionally, there was no evidence of a relation between guidance and the post-test performance.

These results are not directly consistent with Paas and van Merriënboer's (1994) study of variability which demonstrated the worked example effect and the variability effect only for the worked examples condition, and not for the problem-solving condition. Nevertheless, our results are coherent with Paas and van Merriënboer (1994) who found a worked example effect using test score performance while we did not. A failure to find a worked example effect should lead to similar results with respect to variability after studying both worked examples and solving problems. That result was obtained in Experiment 1.

The failure to obtain a significant main effect of variability on cognitive load in Experiment 1 left open the possibility that learner levels of expertise needed to be examined conjointly with levels of variability—an inducement for the design of Experiment 2. This was based on the suggestion that high variability tasks involve a relatively higher intrinsic cognitive load generated by a higher degree of element interactivity, compared to low variability tasks, because they demand that multiple elements of information are drawn upon and then integrated to complete the task. Consequently, Experiment 2 explored the effect of high or low variability tasks on more and less knowledgeable learners.

Experiment 2 demonstrated a classic expertise reversal effect. High variability tasks were advantageous for more experienced learners who had sufficient working memory capacity associated with expertise to handle the problem variations. Novice learners, on the other hand, learnt more from low variability tasks as high variability tasks may have exceeded their working memory capacity.

A limitation of this study is that while Experiment 1 varied levels of guidance and their interaction with variability, and Experiment 2 varied levels of expertise and their interaction with variability, neither experiment varied both guidance and expertise. It was argued that levels of guidance and levels of expertise both alter element interactivity and so can act as a substitute for each other. Ideally, both guidance and expertise should be varied along with variability in a single, 2 (levels of guidance) × 2 (levels of expertise) × 2 (levels of variability) experiment. Such an experiment would yield three two-way interactions and one three-way interaction providing more details of the interactions between the three factors.

From a practical perspective, the major educational implication of these findings suggests that learners should either be initially presented with low variability problems or very simple problems with higher levels of variability (van Merriënboer and Kirschner 2018). Low variability tasks with their low levels of element interactivity can assist in the acquisition of essential problem-solving concepts and procedures associated with a given area of study. Once those concepts and procedures have been acquired, it is appropriate to acquaint learners with the various types of problems to which the concepts and procedures apply. At that point, with the basic knowledge stored in long-term memory rather than having to be processed in working memory, learners should have sufficient spare working memory capacity to process the elements associated with variability. Until that point is reached, presenting learners with the basic concepts and procedures and the various conditions to which they apply simultaneously may overload working memory. Variability should only be increased once learners' levels of knowledge have advanced sufficiently to allow them to process the increased element interactivity associated with increased variability. An initial reduction in variability in this manner should simplify more complex tasks. Reducing initial complexity by other means should have a similar effect (van Merriënboer and Kirschner 2018). Further studies are needed to test these hypotheses.

Compliance with Ethical Standards

Ethics Approval All procedures performed in the study involving human participants were in accordance with the ethical standards of the Human Research Ethics Approval Board of the University of New South Wales, Sydney (Approval number 14074).

References

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181–214. https://doi.org/10.3102/00346543070002181.

Bokosmaty, S., Sweller, J., & Kalyuga, S. (2015). Learning geometry problem solving by studying worked examples: effects of learner guidance and expertise. *American Educational Research Journal, 52*(2), 307–333. https://doi.org/10.3102/0002831214549450.

Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology, 86*(3), 360–367. https://doi.org/10.1037/0022-0663.86.3.360.

Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2018). Undesirable difficulty effects in the learning of high-element interactivity materials. *Frontiers in Psychology, 9*(1483), 1–7. https://doi.org/10.3389/fpsyg.2018.01483.

Chen, O., Kalyuga, S., & Sweller, J. (2017). The expertise reversal effect is a variant of the more general element interactivity effect. *Educational Psychology Review, 29*(2), 393–405. https://doi.org/10.1007/s10648-016-9359-1.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*(4), 347–362. https://doi.org/10.1037/0022-0663.79.4.347.

Geary, D. C. (2008). An evolutionary informed education science. *Educational Psychologist, 43*(4), 179–195. https://doi.org/10.1080/00461520802392133.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*(3), 579–588. https://doi.org/10.1037//0022-0663.93.3.579.

Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review, 24*(2), 313–337. https://doi.org/10.1007/s10648-012-9195-x.

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122–133. https://doi.org/10.1037/0022-0663.86.1.122.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88*(1), 144–161. https://doi.org/10.1037/0022-0663.88.1.144.

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: not an artefact of lousy control conditions. *Computers in Human Behavior, 25*(2), 258–266. https://doi.org/10.1016/j.chb.2008.12.011.

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5.

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science, 32*(1–2), 9–31. https://doi.org/10.1023/B:TRUC.0000021808.72598.4d.

Sweller, J. (2006). Discussion of 'emerging topics in cognitive load research: using learner and information characteristics in the design of powerful learning environments'. *Applied Cognitive Psychology, 20*(3), 353–357. https://doi.org/10.1002/acp.1251.

Sweller, J. (2008). Instructional implications of David C. Geary's evolutionary educational psychology. *Educational Psychologist, 43*(4), 214–216. https://doi.org/10.1080/00461520802392208.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory.* New York, NY: Springer. https://doi.org/10.1007/978-1-4419-8126-4.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185–233. https://doi.org/10.1207/s1532690xci1203_1.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59–89. https://doi.org/10.1207/s1532690xci0201_3.

Sweller, J., & Sweller, S. (2006). Natural information processing systems. *EvolutionaryPsychology, 4*(1), 434–458. https://doi.org/10.1177/147470490600400135.

Van Merriënboer, J. J. G., & Kirschner, P. A. (2018). *Ten steps to complex learning: a systematic approach to four-component instructional design* (3rd ed.). New York: Routledge. https://doi.org/10.1007/s11528-018-0254-0.