The Worked Example Effect, the Generation Effect, and Element Interactivity

Ouhao Chen

A thesis in fulfilment of the requirements of the degree of

Doctor of Philosophy

School of Education

Arts and Social Science

February 2016

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: **CHEN**

First name: **OUHAO**

Other name/s:

Abbreviation for degree as given in the University calendar: **PhD**

School: **School of Education**

Faculty: **Arts and Social Science**

Title: **The Worked Example Effect, the Generation Effect, and Element Interactivity**

Abstract 350 words maximum: (PLEASE TYPE)

The worked example effect indicates that showing students worked examples (high guidance) is superior to problem solving (low guidance) which provides no guidance, whereas the generation effect suggests that self-generating items (low guidance) is superior to studying the externally presented answers (high guidance). This obvious contradiction between the two effects was hypothesized to be resolved by suggesting that the materials used had different levels of element interactivity. For the worked example effect, materials may be high in element interactivity, while, for the generation effect, simpler materials are used. With an increase of learner expertise, the worked example effect may be eliminated or reversed because expertise reduces element interactivity, but the generation effect should be still robust. Five 2 (levels of guidance: low and high) x 2 (levels of element interactivity: low and high) mixed factorial experiments were conducted to investigate the hypotheses. In Experiment 1 to 3, the level of learner expertise gradually increased in the domain of geometry with the results supporting hypotheses. The interaction of guidance and element interactivity was obtained with novices but the worked example effect that contributed to the interaction was eliminated or reversed with more knowledgeable students. Experiment 4 and 5 were designed to replicate the results of the first three experiments by testing students in the domain of trigonometry on both immediate and delayed tests. The results were replicated not only on an immediate test, but also on a delayed test. When combined, the results of all five experiments indicated that levels of element interactivity might be a key factor when deciding on levels of instructional guidance.

COPYRIGHT STATEMENT

AUTHENTICITY STATEMENT

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.'

Acknowledgements

I always believe that if I set up a big goal, I may reach a smaller one; if I set up a small goal, I may get nothing. Therefore, when I was in my master degree, I determined to do a PhD after that.

A three-year PhD is really not long enough. Although I did almost the same thing every day, I still felt very happy and satisfied. I believe that I owe my satisfaction to my two helpful and supportive supervisors, Professor Slava Kalyuga and Emeritus Professor John Sweller. On the one hand, my principal supervisor, Professor Slava Kalyuga, gave me a lot of guidance on how to improve my academic writing and English grammar. Because of Slava's comments on my academic writing, I was able to finish my dissertation in time; on the other hand, my second supervisor, Emeritus Professor John Sweller, instructed me how to design experiments and how to analyze data from different aspects. Because of my two supervisors' cooperation, I successfully published two papers and another three papers are being under review during my PhD study, which I hope will make my research internationally recognized.

The second thanks should go to the China Scholarship Council which provided me with a research grant to do my PhD at UNSW. Without their help, I could not pursue my PhD overseas and fulfill my goals.

The last but very important thanks is to my parents who supported my continued study and research. As an international student, I have left home for a long time and missed a lot of chances to hug and kiss my dearest parents, so I must say that I cannot be a Doctor without their support.

Abstract

The worked example effect indicates that showing students worked examples (high guidance) is superior to problem solving (low guidance) which provides no guidance, whereas the generation effect suggests that self-generating items (low guidance) is superior to studying the externally presented answers (high guidance). This obvious contradiction between the two effects was hypothesized to be resolved by suggesting that the materials used had different levels of element interactivity. For the worked example effect, materials may be high in element interactivity, while, for the generation effect, simpler materials are used. With an increase of learner expertise, the worked example effect may be eliminated or reversed because expertise reduces element interactivity, but the generation effect should be still robust. Five 2 (levels of guidance: low and high) x 2 (levels of element interactivity: low and high) mixed factorial experiments were conducted to investigate the hypotheses. In Experiments 1 to 3, the level of learner expertise gradually increased in the domain of geometry with the results supporting hypotheses. The interaction of guidance and element interactivity was obtained with novices but the worked example effect that contributed to the interaction was eliminated or reversed with more knowledgeable students. Experiments 4 to 5 were designed to replicate the results of the first three experiments by testing students in the domain of trigonometry on both immediate and delayed tests. The results were replicated not only on an immediate test, but also on a delayed test. When combined, the results of all five experiments indicated that levels of element interactivity might be a key factor when deciding on levels of instructional guidance.

Table of Contents

List of Tables

List of Figures

Introduction

An important issue in Pedagogical Psychology is how to facilitate learning. Over the past 50 years, discussions of the influence of instructional guidance during teaching have occurred (Ausubel, 1964; Craig, 1956; Keislar & Shulman, 1966; Mayer, 2004). On the one hand, in a traditional classroom setting, a teacher, students and a textbook were the three main constituents of the classroom, with the teacher at the center providing full guidance. On the other hand, when constructivist theory became popular, it was suggested that knowledge should be constructed by students themselves, and that teachers should not be the center of teaching. The next sections will review some contradictory views about traditional teaching and instructions which are based on constructivist theory.

Traditional Teaching and Constructivism-based Instructions

Traditional teaching usually provides direct instructional guidance. This direct instructional guidance contains information which provides full explanations of the concepts and procedures that learners should acquire and the learning strategy provided should be in line with human cognitive architecture (Kirschner, Sweller, & Clark, 2006). Giving lecture-style presentations is a kind of traditional teaching and is often regarded as old-fashioned having a lot of disadvantages (Schwerdt & Wuppermann, 2011): (1) teachers cannot get feedback about students' learning, as all the students follow what the teacher has designed, and it is assumed that students in a lecture study are at the same pace; (2) a long lecture cannot always make students focus on the lecture; (3) as students do not actively participate in the lecture, knowledge learned may have a tendency to be forgotten soon. However, controlled experiments have indicated that providing explicit guidance about the procedures and concepts to learners is necessary for learning novel information (Kirschner et al., 2006).

The purpose of constructivist instructional design is to supply a tool for generative mental construction which requires learners to construct schemas by themselves, embedded in relevant learning environments that facilitate knowledge construction by learners (Kazemi & Ghoraishi, 2012). The minimal guidance approach has different names: e.g., Problem-Based Learning (PBL), experiential learning, inquiry learning, however, the common characteristic of them is to let learners explore and discover basic concepts and principles in the information-rich environment like a professional researcher (Kirschner et al., 2006). Take Problem-Based Learning (PBL) or Inquiry-based learning which has been suggested for effective learning (Kazemi & Ghoraishi, 2012) as an example. Barrows (1980) asserted that PBL is an instructional method in which students learn through solving problems and reflecting on their experiences. In Problem-Based Learning (PBL), new knowledge is acquired in the context of some meaningful problems or situations. Students are actively engaged in problem-based learning in which they build their own understanding with some cues from teachers, but teachers will not help students construct their own knowledge (Lai & Tang, 1999). Therefore, for traditional teaching, high guidance is provided by showing full solutions or concepts, like worked examples (see next section), while, low guidance, like generation (see next section), is offered for constructivist instructional design, which requires learners to self-construct knowledge or schemas.

The Worked Example and Generation Effects

The worked example effect, based on cognitive load theory, does not support the view that explicit instruction is disadvantageous. Worked examples provide learners with full guidance which contains the key steps needed to solve a problem. Sweller and Cooper (1985) indicated that worked examples could result in better performance than conventional problem solving which has no guidance.

The generation effect describes a phenomenon that occurs when items (e.g., words) are generated by learners in the presence of a stimulus and an encoding rule; the items are better remembered than when the same items are simply read by learners (McElroy & Slamecka, 1982; Slamecka & Graf, 1978). This effect has been described as caused by active participation in the learning process producing better retention than passive observation. The effect has been found in cued recall, recognition and free recall tests. The generation effect can be used to support Problem-Based Learning (PBL), with learners actively engaged in the cases and building their own understanding under the guidance of an instructor (or the cue), but the instructor (or the cue) does not do the building for the students (Lai & Tang, 1999).

At least on the surface, the worked example and generation effects have an apparent contradiction: worked examples used in the worked example effect provide full guidance to learners, resulting better performance than problem solving which provides no guidance, whereas the generation effect encourages students to generate solutions by themselves but produces better performance compared to presentation which provides full guidance.

High-element Interactivity and Low-element Interactivity Materials

Element interactivity is an index of the nature of learning material within the framework of cognitive load theory. Interacting elements are defined as elements that must be processed simultaneously in working memory because they are logically related (Sweller, Ayres, & Kalyuga, 2011). Some learning materials can be processed individually, such as learning the English words for second language learners or Chemical symbols in the periodic table. These materials can be learned independently and without referring to any other content, so they are low-element interactivity materials. However, if the materials cannot be learned independently but must be processed simultaneously in working memory, they belong to the high-element interactivity category. An example is learning to solve a problem such as: x+5=8, solve for x. In order to solve this problem, learners should first hold a series of single mathematic elements (such as x, +, =) in their limited working memory. If they only hold these single mathematic symbols, there is no chance to solve this problem. Therefore, learners also need to process the relations between different symbols in order to finally understand this question and finally find the value of x. Therefore, for this simple algebra problem, we need to process a lot of interactive elements in our limited working memory, resulting in high-element interactivity.

The obvious contradiction between the worked example and generation effects may be due to the consequences of various element-interactivity materials. The real difference between the worked example and generation effects may be the nature of the materials used in demonstrating these two effects. In this thesis, the main hypothesis was that the materials used to demonstrate the generation effect were low in element interactivity, whereas for the worked example effect, the materials used were all high in element interactivity. The experiments described in the thesis tested this hypothesis.

This thesis contains three parts. The first part is a theoretical part which contains seven Chapters to provide a general theoretical framework of this dissertation. Chapter 1 discusses human cognitive architecture which is the base of cognitive load theory and five key principles relevant to this cognitive architecture. Chapter 2 discusses cognitive load theory itself, including different types of cognitive load and different methods for measuring cognitive load. Chapter 3 provides some research results about the worked example effect, especially introducing types of worked examples. Chapter 4 reveals some design rules for constructing a worked example based on other cognitive load effects which are relevant to a worked example design. Chapter 5 discusses the expertise reversal effect. Chapter 6 presents empirical results testing the generation effect which has neither been generated nor explained by cognitive load theory, in contrast to the worked example effect which was generated by cognitive load theory. The last Chapter of the first part (Chapter 7) develops research questions and hypotheses based on discussions about element interactivity, the worked example effect and generation effect in previous Chapters.

The second part of this thesis is empirical study (Chapter 8 to Chapter 12) which contains five experiments. The last part of this thesis (Chapter 13) includes a summary of the main findings, a general discussion of the results obtained from the five experiments, some instructional implications and future studies.

PART 1 LITERATURE REVIEW

Chapter 1 Human Cognitive Architecture

This chapter will discuss human cognitive architecture which provides a base for cognitive load theory. Instructional design issues and human cognitive architecture are inseparably intertwined (Sweller, 2004). The science of instruction will be random if we do not have knowledge of human cognitive architecture and in turn, instructional design can heavily influence the development of our knowledge of human cognitive architecture. Therefore, knowing how students learn and solve problems informs us how we should organize their learning environment (Sweller, 2004).

1.1 Human Cognitive Architecture

Sweller (2003) indicated that cognitive structures and their relations were either discovered or emphasized as individual structures by various researchers since the early 1930s and had been conceptualized into a unified architecture since the early 1970s with some controversies. Human Cognitive Architecture refers to the way in which the components, such as working memory and long-term memory, are organized (Sweller et al., 2011). Nature has different kinds of information processing systems, from simple ones to the most complex ones, such as biological evolution and the human cognition. In this chapter, I will only focus on the most sophisticated ones: biological evolution and the human cognition. Human cognition has four general characteristics: First, in order to live and adapt to a varied environment, human cognition should be able to create novel information; second, human cognition should be able to remember information which has been created and is effective; third, human cognition should have a knowledge base to store that knowledge to manage human activities; lastly, the system should be able to spread effective information via space and time. The next sections will discuss human cognitive architecture through the aspect of biological evolution. However, before introducing the human cognitive architecture in details, the categories of knowledge which are relevant to the human cognitive architecture and cognitive load theory may need to be clarified first.

1.2 Biologically Primary and Secondary Knowledge

In this chapter, categories of knowledge will be based on Geary's theory. Geary (2007, 2008) suggested a distinction between biologically primary and secondary knowledge, which is relevant to instructional design. Biologically primary knowledge is learnable but not teachable, whereas biologically secondary knowledge is learnable as well as teachable. So, an informal way to judge whether this knowledge belongs to biologically primary or secondary knowledge is to see whether it requires some instructional supports for its acquisition.

1.2.1 Biologically Primary Knowledge

Living in a complex environment, we acquire a lot of biologically primary knowledge. All of that kind of knowledge is obtained naturally and effortlessly, as we have evolved to have this kind of knowledge, therefore, we do not need any special courses or explicit instructions to teach biologically primary knowledge.

There are many cognitive skills that are likely to be biologically primary and most are learned when we were very young. For example, we evolved to learn how to talk in our mother language without any special courses, but in order to talk in our first language; we need to store some information such as the manipulation of our lips, tongue, and voice in different situations. We do not need to be taught this kind of skill, as it is learned automatically when we are learning to talk. Geary (2007, 2008) sorted such instinctively obtained skills as biologically or evolutionary primary skills requiring biologically primary knowledge. Some general problem solving strategies can also be categorized as biologically primary knowledge. Newell and Simon (1972) and Sweller (1988) discussed a well-known problem solving strategy: Means-ends analysis. This strategy requires problem solvers to consider the present problem state and the final problem state, and then find problem solving operators to reduce the differences between the present and final states. Once the operator is applied to the present state, it will change to another state which is closer to the final state, and problem solvers are required to find other operators for this new state until they reach the final state. This kind of problem solving strategy belongs to biologically primary knowledge, as we do not need any explicit instructions for learning it. Therefore, there are no successful examples demonstrating that problem-solve performance will be improved after teaching this kind of skill.

To sum up, biologically primary knowledge is acquired without any kind of special courses, and human beings have evolved over many generations to acquire this kind of knowledge which is basic and irrelevant to our specific culture. Most importantly, biologically primary knowledge can be learned but not taught.

1.2.2 Instructional Consequences of Biologically Primary Knowledge

Besides we do not need to teach learners how to use a means-ends strategy in problem solving, which requires learners to identify the goal and then work backwards or forwards to reduce differences between where they are and the goal, because we have evolved to acquire this kind of knowledge. Another example (Polya, 1957) is that when learners face a problem without a solution in mind, they should think of a similar problem for which they know the solution to help to solve the original problem by analogy. In fact, teaching learners to find a similar problem is as ineffective as teaching them to use a means-ends strategy because if we know a similar problem which can be used to solve the original problem, we will always use that similar problem automatically. The knowledge that similar problems with similar solutions are helpful belongs to biologically primary knowledge. Therefore, there is no need to teach this kind of knowledge or skill. The same point can be applied to all general cognitive techniques, such as making decisions and thinking.

1.2.3 Biologically Secondary Knowledge

With the development and change of human culture, human knowledge increases accordingly, and the nature of that new knowledge is quite different from biologically primary knowledge. Geary (2007, 2008) called this kind of knowledge as biologically secondary knowledge. This category of knowledge needs to be taught via formal instructions and is relevant to a specific culture. We have evolved to obtain each category of primary knowledge through its own acquisition system. We also have one, possibly unique system to acquire biologically secondary knowledge (Geary, 2007, 2008). As indicated in the biologically primary knowledge section, we evolved to speak, but we do not obtain writing skills naturally, therefore, we need special courses or explicit instruction to acquire knowledge of how to write. This learning process is also different from biologically primary knowledge in that it is effortful and conscious. We can learn speak if we are in a speaking society, but we cannot automatically learn how to write if we are in a writing society. As biologically secondary knowledge requires formal or informal explicit instructions, it indicates that we should pay attention to the design of teaching this kind of knowledge.

1.2.4 Instructional Consequences with Biologically Secondary Knowledge

With the development of educational psychology, different kinds of instructional procedures which are based on new educational psychology theories have appeared. One of the most prominent theories is constructivist learning, which includes discovery learning and problem based learning. The instructional procedures discussed in constructivist learning require students to obtain knowledge by themselves with some hints from teachers. Sweller

(2009) indicated that these instructional procedures assume that information should be withheld from learners during instruction. On this view, withholding information from learners should be beneficial, because explicit information is not required for learners to acquire much of the information we need to function in our society (Kirschner et al., 2006; Mayer, 2004). Compared to traditional instruction in which teachers give prepared knowledge to students with full guidance, constructivist learning encourages students to discover and build knowledge structures by themselves. In addition, constructivist learning gives rise to another important question: whether the quality of knowledge constructed by learners themselves is better than the knowledge given by teachers? On this question, Bruner (1961) suggested that discovered knowledge should be qualitatively better than directly taught knowledge. However, this result was rejected by Klahr and Nigam (2004). They found that there was no difference between the quality of knowledge of science learners who discovered a science principle and those to whom knowledge was explicitly presented. They further pointed out that the only difference in the knowledge obtained between these two kinds of procedure was the time required to acquire knowledge. Constructivist learning requires more time than explicit instruction. Another important question is whether learners should be taught how to construct knowledge first and then to gain more knowledge. For survival, we have evolved to construct knowledge. This activity is biologically primary, so we do not need to learn how to construct the knowledge. But the specific concepts and procedures of one discipline belong to secondary knowledge that needs to be taught.

1.3 Human Memory Systems

Human memory is regarded as an information processing system, the general structure of human memory system has two kinds of versions, one version distinguishes the permanent, structure of memory system and the control processes which are selected and varied according to instructions and subjects' background; the second version which was proposed by Atkinson and Shiflrin (1968) has three parts: sensory registers, short-term memory and long-term memory (See Figure 1.1).

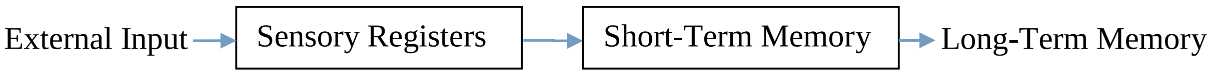


Figure 1.1. Structure of the Working Memory System

1.3.1 Sensory Registers

Human has different kinds of sensory registers, such as visually, aurally sensory registers. When a stimulus is presented, this stimulus will trigger the relevant sensory register to register, and stays very short time (about several hundred milliseconds) at sensory register before disappears. Elements stored in sensory registers are then scanned (with the information which is retrieved in long-term memory) and then transferred to short term memory. Among those sensory registers, visual register has been well investigated about its function (Sperling, 1960), others have not.

1.3.2 Working Memory

Initially, working memory was called short-term memory (STM) by Miller (1956), but later, researchers usually referred it as working memory (WM) (e.g., Baddeley & Hitch, 1974) . The change from short-term memory to working memory was because of a change in emphasis from a holding store to the processing engine of the cognitive system (Sweller, 2003). In addition, working memory can be regarded as our consciousness, as we only are conscious of information held in working memory.

Baddeley (1992) defined working memory as a system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks as language comprehension, learning, and reasoning. Initially, working memory was conceptualized as a unitary concept, but Baddeley (e.g., Baddeley, 1992; Baddeley & Hitch, 1974) divided working memory into three sub-systems (see Figure 1.2): the visuospatial sketch pad, which manipulates visual images, such as two-dimensional diagrams or three-dimensional information; another sub-system is the phonological loop, which stores and rehearses speech-based information and is necessary for the acquisition of both native and second-language vocabulary; and finally a central executive works as a coordinating processor.

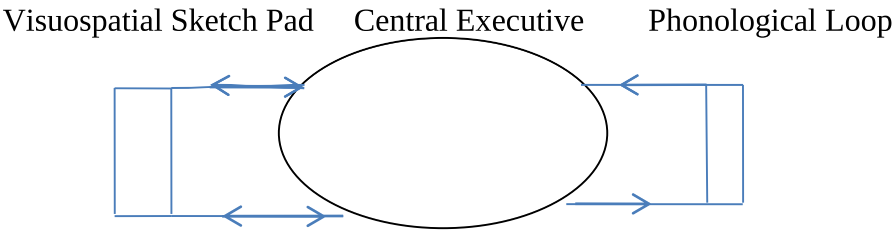


Figure 1.2. Working Memory (according to Baddeley, 1992)

Later, Baddeley (2000) extended this model of working memory by adding an episodic buffer as a temporary storage which can store multidimensional episodes or chunks.

Working memory has two unique characteristics: firstly, it has very limited capacity (e.g., Cowan, 2001; Miller, 1956); secondly, it has short duration time (Peterson & Peterson, 1959). But these two characteristics are only for novel information held in working memory; for organized, well-structured information, working memory does not have those two limitations and large amounts of information can be retrieved from long-term memory and transferred to working memory. Therefore, in order to process more information and store it longer, long-term memory is needed.

1.3.3 Long-term Memory

When information is temporarily stored in working memory, transfer of information to long-term memory happens throughout this period (Atkinson & Shiflrin, 1968). Therefore, our final knowledge store is long-term memory. Knowledge stored in long-term memory is unconscious, unless the information is transferred to working memory. Therefore, researchers took a long time to realize that long-term memory is not just used to recognize or recall information but also is an integral component of all cognitive activities, including activities such as high-level problem solving (Sweller, 2003).

Research indicating that long-term memory stores large amounts of biologically secondary knowledge is provided by the difference between chess grandmasters and chess novices. De Groot (1965) and Chase and Simon (1973) used the example of chess grandmasters and chess novices to demonstrate that chess grandmasters had stored in long-term memory tens of thousands of board configurations along with best moves associated with those configurations. Initially, De Groot (1965) wanted to find out why chess grandmaster always defeats weekend players. The first hypothesis was that chess grandmasters engaged in greater search for the best moves than chess novices or weekend players. However, De Groot (1965) found that the only difference between chess grandmasters and chess novices was related to their memory of chess board configurations taken from real games rather than their problem-solving skills. There was no difference in remembering non-standard chess configurations between chess grandmasters and chess novices (Chase & Simon, 1973). De Groot (1965) and Chase and Simon (1973) suggested that chess grandmasters stored large amounts of standard chess board configurations with suitable moves associated with those configurations in their long-term memory. Similar findings have been obtained in many other areas, such as algebra (Sweller & Cooper, 1985), electronic engineering (Egan & Schwartz, 1979) and programming (Jeffries, Turner, Polson, & Atwood, 1981). De Groot's findings let us know that the function of long-term memory is not only for recalling events but it is also central to those aspects of cognition that are seen as representing the apex of the human mind (Sweller et al., 2011). Finally, unlike the working memory, long-term memory has unlimited storage capacity for biologically primary and secondary knowledge and the information stored in long-term memory can last for a very long time.

1.4 Schema Theory

1.4.1 Schema Characteristics

As described above, information is stored in long-term memory. A question is in what form is this information stored in long-term memory? Schema theory provides an answer. A schema can be regarded as a cognitive construct which lets us categorize multiple elements of information into a single element based on the way in which the multiple elements are used (Chi, Glaser, & Rees, 1982). Tuddenham (1966) regarded a schema as a flexible mental structure, the primary unit of mental organization. For example, in order to solve the algebra problem, $x+5=8$, solve for x, people who have studied mathematics in high school with a basic knowledge of equations know that we just need to subtract 5 from both sides of this equation. If someone has formed a schema for this problem, all similar problems and the numerical system relevant to this problem will be regarded as an entity or one schema. When learners need to solve a similar problem later, they will use the same procedure to solve that problem. As another example, if we have a schema for the letter "A", irrespective of its format, we will recognize that the letter is "A" rather than other letters. Therefore, a schema helps us to react similarly to face a similar situation.

Schema theory became important with the work of Piaget (1928) and Bartlett (1932). Bartlett described an experiment which clearly indicated the nature and function of schemas. He asked a person to write down what they remembered from one passage and then the next one read what the first person wrote and wrote out what they remembered. This process continued for 10 persons. Bartlett (1932) found two effects: familiar things were emphasized, while unfamiliar things disappeared. The conclusion is that long-term memory holds countless numbers of schemas and those schemas determine how we process incoming information.

1.4.2 Schema Acquisition

Tuddenham (1966) indicated that all particular experiences were incorporated into an already present schema and so schemas could be built and developed. He indicated that Piaget had used the terms *assimilation* and *accommodation* to describe changes to schemas. If we already have a suitable schema, the incoming information can be directly assimilated into this existing schema, however, if we do not have a suitable schema, the novel information changes the existing schema to a new one which can assimilate this novel information. This process is called accommodation. Through this building, lower-level schemas can be incorporated in a new, but sophisticated schema which contains more information, resulting in an improvement of our relevant skills. This acquisition of schemas is an active, constructive process.

Sweller, Van Merrienboer and Paas (1998) indicated that the functions of schema acquisition are firstly, to provide the mechanism for knowledge organization and storage as indicated above and secondly, to reduce working memory load. The second function originates from the mechanism of schema acquisition. When a schema is formed, a number of individual elements of information can be processed in working memory as a single entity

(Sweller et al., 1998) rather than a number of individual, smaller schemas. Therefore, working memory space is released to engage in other activities reducing working memory load.

The work of de Groot (1965), Chase and Simon (1973), Gick and Holyoak (1980, 1983), Larkin, McDermott, Simon, and Simon (1980) and Chi et al. (1982) demonstrated the critical role of schemas in expert problem solving. Most researchers have accepted that a high level of problem solving expertise in complex areas requires the acquisition of thousands of domain-specific schemas, and those schemas will allow problem solvers to recognize problem states according to the moves which are associated with them. Accordingly, skill in any area is dependent on the acquisition of specific schemas which are stored in long-term memory (Sweller, 2003).

1.4.3 Schema Automation

All information can be processed either consciously or automatically (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). For newly acquired schemas, they must be processed using working memory resources. However, with increasing practice, especially extensive practice, those schemas will be used with less and less conscious processing. Automation is therefore an important factor in schema construction. The main function of automation is to release the resources of working memory. The reason has been described at the beginning of this section: with increasing practice, especially extensive practice, those schemas will be used with less and less conscious processing, which will require less and less working memory resources. Kotovsky, Hayes, and Simon (1985) demonstrated the benefits of automated processing to problem-solving skill. Problem solvers using automated rules to solve a problem were 16 times faster than those who solved problems with conscious effort. With automation, familiar tasks are likely to be performed accurately and fluidly, and unfamiliar tasks can be learned with maximum efficiency because maximum working memory capacity is available. Instructional designs should not only encourage the construction of schemas, but also the automation of schemas (Van Merriënboer, 1997; van Merriënboer, Jelsma, & Paas, 1992).

1.5 Five Basic Principles of Human Cognitive Architecture

Under the framework of Cognitive Load Theory, an evolutionary perspective has been used to consider human cognition. The basic principles of human cognitive architecture have been assumed to be the analogue of biologically evolutionary principles. In this chapter, the five principles will be discussed according to the Figure 1.3 below, from the left to right and then reverse the order:

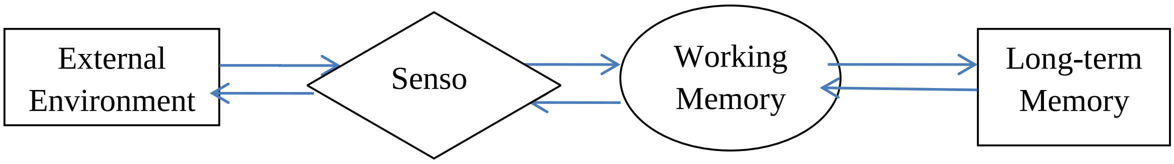


Figure 1.3. Information Transmission Process

Generally, according to Figure 1.3, the five principles indicate the relationships between working memory, long-term memory and the external environment. In the next sections, the five principles will be discussed in details from the perspective of biological evolution and human cognition.

1.5.1 The Borrowing and Reorganizing Principle

From Figure 1.3, it can be seen that information is transferred from the environment to working memory via sensors. The first section will discuss how human cognition acquires information.

Biological Evolution. Sexual and asexual reproduction is the way information is obtained by a genome. In the case of asexual reproduction, the exact information is copied and transmitted to all the offspring via cell splitting, namely, one cell splits into two identical cells, and the information of parents is exactly copied and transferred to offspring. For sexual reproduction, the information held by the genome of offspring is unique and different from ancestors' genomes, but all of the information is obtained from one male and one female and then combined to give a unique genome. Sexual reproduction can provide novel and creative information held by the genome of offspring, but asexual reproduction copies and transfers the exact information from parents. The two kinds of reproduction provide two different kinds of mechanism: asexual reproduction only borrows information from ancestors, but sexual reproduction involves borrowing as well as reorganizing information from parents and the mix of borrowed and reorganized information results in the creation of unique individuals.

Human Cognition. The analogy is quite straightforward when we discuss how the human cognition system obtains information. It has two similar ways to acquire information: borrowing well-organized information from other people and creating novel information via reorganizing.

Almost all of the secondary knowledge in our long-term memory is borrowed from other people. As described above, the human information processing system can disseminate effective information across time and space, so it provides the basic foundation for borrowing, and what's more, the basic skill of communicating information belongs to the biologically primary knowledge which has evolved to receive such knowledge from others and transmit it to others. Humans naturally imitate others. This tendency of imitating can be attributed to the mirror neuron system which is newly discovered. When we make a movement, mirror neurons become active, and when we see another person doing the same movement, the same neurons will be active as well (Grafton, Arbib, Fadiga, & Rizzolatti, 1996; Iacoboni et al., 1999; Tettamanti et al., 2005). Therefore, we can use the characteristics of the mirror neuron system to understand this basic ability of imitation. Similarly, Van Gog and Paas (2008) used the biological theory of both Geary (2007) and Sweller and Sweller (2006) to argue that humans have evolved to learn by imitation and observation and they suggested that observation was similar to studying worked examples.

This borrowing ability was addressed by Bandura's imitation study (1986). In Bandura's theory, people study via observation, which has four stages. 1. Attention. Learners who study via observation should firstly pay attention to the observed persons' behavior. 2. Sustainability. This stage requires learners to hold on what they have observed from the person being imitated. 3. Representation. Learners should change these symbols and perceptual images held in mind to real actions or movements. 4. Motivation. Whether learners can keep doing this movement is dependent on the consequences of this behavior. If learners get good results after imitating this movement, this movement will be sustained. So in a real classroom setting, we assume that when we demonstrate something to learners that they will assimilate the knowledge associated with that demonstration (Sweller et al., 2011).

However, borrowing is not enough. When information is acquired from the environment, it must be combined with old information which has been stored in long-term memory, so reorganization of the information occurs. However, the result of this reorganizing can be beneficial or detrimental, so we need to test it for effectiveness. If the combined information is effective, it will be retained, or it will disappear if it is ineffective. This kind of reorganizing is supported by the experiment of Bartlett (1932) mentioned above. Different people remembered and wrote down different things. Only things which are relevant to the information stored in long-term memory were remembered and written down. Therefore, a schema borrowed from other people and the process of schema acquisition involves some degree of reorganization, as a new schema constructed is likely to be different from a schema borrowed from other people.

1.5.2 Randomness as Genesis Principle

After borrowing information from other people, and after the information has been reorganized, the next step is to check whether the newly reorganized information is effective or not. However, when we plan to check the effectiveness of novel information, we have no prior knowledge available. Therefore, if knowledge concerning the effectiveness of a potential reorganization is unavailable prior to acquiring a schema, a random generation and test process will be required as a concomitant of the reorganization process and this process relates to our next principle: Randomness as Genesis Principle.

The borrowing and reorganizing principle is about how information is communicated in natural information systems, but the randomness as genesis principle is about how the information is initially created.

Biological Evolution. All the genetic variation between individual organisms can be regarded as having been initiated by a series of random mutations (Sweller et al., 2011). The test following the random generation process determines the number of offspring and makes sure that effective information can be retained and ineffective information will be eliminated. Random mutation of a genome is the major source of variation and creating novel information and it is critical to biological evolution.

Human Cognition. The role of random mutation in biology provides a particular example of random generation and test. Random generation and test during solving problem may play the same role in human cognition. When we solve a problem, if the relevant knowledge has been stored in our long-term memory, then we can retrieve it directly and use it to solve this problem. However, if we do not have this schema, we need to randomly generate moves and test the effectiveness of those moves.

We have two lines of evidences to support this principle: the first line of evidence is empirical. When solvers solve a novel and complex problem, they usually get to dead ends at various points. Arriving at a dead end indicates that the previous moves were incorrect. These dead ends may be a consequence of a random generation and test process.

The second line of evidence comes from logic. Generally, we have three processes that can be used to solve a novel problem: 1. Using stored information directly; 2. Randomly generating moves; 3. Combining 1 and 2. Most problem solving requires the third process. Analogical problem solving provides an example. Firstly, we need to decide which analogical problem can be used via the superficial characteristics as well as the deep structure features; secondly, we retrieve the analogical problem from long-term memory, but we cannot decide whether this analogical problem is effective or a dead end after putting it into practice. Therefore, this whole process must be accompanied with a random generating and testing process. Using the method of analogical problem solving combines using stored knowledge from long-term memory as well as random generation.

Based on the analysis above, the randomness as genesis principle is a principle which is very dependent on knowledge stored in long-term memory. The human architecture used in cognitive load theory does not have an independent central executive (Sweller, 2003), but assumes that knowledge held in long-term memory acts as an independent central executive which organizes and controls cognitive processes.

Similar to random mutation in a genome, human cognition creates novel information via a random generation and test process. Random mutation provides the base of genetic variation while randomly generating problem solving moves acts in an identical role in human cognition. Therefore, human creativity may be just as reliant on a random generating and testing process as evolution by natural selection (Sweller et al., 2011).

1.5.3 Narrow Limits of Change Principle

The previous two principles tell us how information is acquired and created. Randomly generated information is not organized and there are limitations to the amount of unorganized information that a processing system can handle. Mathematically, if the system must handle 3 elements of information, by the logic of permutation, there are 3! =6 possible permutations, which will not overload a natural information system, however, if the number of element goes up to 10, the possible permutations are 10! = 3,628,800 and a natural information system cannot easily process this amount of information. Therefore, we can assume that the amount of randomly generated information which can be processed by a natural information system is limited.

Biological Evolution. When considering this principle in evolution, the epigenetic system is relevant (Jablonka & Lamb, 2005; West-Eberhard, 2003). The epigenetic system is a chemical system which can manage the location and rate of mutation, and it can also turn on or off special genes based on stimuli from the external environment. This system is the bridge between the external environment and the genetic system. It has two functions: Firstly, it selectively processes and transfers information from the external environment to the DNA-based genetic system which can result in genetic changes. Secondly, it can use environment information to determine which parts of a genetic system will function. There are severe limitations to the number of mutations that can occur and successful mutations are rare even under conditions where the rate of mutation is increased by the epigenetic system.

Human Cognition. Working memory is regarded as the system which has the same role as the epigenetic system, namely, it is the channel between the information store and the external environment, which is a reason why Figure 1.3 has links both from left to right and from right to left. Information from the external environment will first be processed in working memory. We are conscious of the information which is in working memory and so working memory can be equated with consciousness. There is a relationship between working-memory and long-term memory. We know that there is a large amount of knowledge held in long-term memory, but we are only aware of a small part of that knowledge which is transferred to working memory from long-term memory.

Using Atkinson and Shiffirin's (1968) architecture (see Figure 1.3), information firstly is received and very briefly processed by different parts of the sensory system (such as visual and auditory senses), then some elements of information are transferred to working memory to be processed consciously in conjunction with information held in long-term memory. This information can be stored for a longer time only when it has been transferred to long-term memory and then will be used to govern further behavior.

Based on the introduction of working memory above, working memory has two surprising characteristics when it processes novel information that are critical for instructional design: it is very limited in capacity (e.g., Cowan, 2001; Miller, 1956) and in duration (Peterson & Peterson, 1959).

For the capacity limitation, Miller (1956) suggested that the limitation was about 7 items but Cowan (2001) suggested that the number of items was about 4. Working memory characteristically is not used to store information but to process information. Sweller et al. (2011) suggested that 2-3 items of novel information can be processed by working memory at a given time. If more information must be processed by working memory, it tends to be broken down. Therefore, the capacity of working memory is very limited when it deals with new information.

Peterson and Peterson (1959) indicated that most novel information only could be held by working memory for a few seconds before we lost almost all of information. We can repeat this information to avoid losing it and this rehearsal have the advantage of assisting in the transfer of information to long-term memory.

However, these characteristics of working memory only apply to novel information rather than the organized, old information which has been stored in long-term memory. This point is very important for the fifth principle below.

1.5.4 The Information Store Principle

In this section, we will move to the final part of Figure 1.3 (from left to right). The first three principles tell us how we obtain and create novel information from the external environment and how working memory limits the amount of novel information created or borrowed. The forth principle is going to reveal where and how the created or borrowed information stored.

Biological Evolution. In order to survive in a complex environment, all genomes have a huge amount of DNA-based information that determines most biological activities (Portin, 2002; Stotz & Griffiths, 2004). A simple small information system may not be able to adapt to a complex environment. There are no consistent measures of genomic complexity or size, but one genome does consist of thousands of information elements and organisms rely on that large store of information. If all natural information systems have a large amount of information stored, it naturally allows us to assume that human cognition also must have a system or function for storing large amount of information.

Human Cognition. Like a genome in organisms, long-term memory is a place to store information, and the form of stored information is a schema which has been discussed above. As mentioned above, De Groot's (1965) and Chase and Simon's (1973) work on chess can be used to indicate that long-term memory stores an immense amount of secondary biologically information. Of course much of biologically primary knowledge is also stored in long-term memory. Long-term memory has a possibly unlimited capacity to store information transferred from working memory as a major characteristic.

The information stored in long-term memory can be retrieved by working memory to be processed. This feature is also relevant to our fifth principle and it is a reason why Figure 1.3 can continue from right to left.

1.5.5 Environmental Organizing and Linking Principle

The previous four principles are concerned with the processes associated with the transfer of information from the external environment to long-term memory. Figure 1.3 above depicts the information flow of those four principles from left to right. However, the final purpose of natural information processing systems storing information is to permit that stored information to have appropriate functions when dealing with the natural environment, so the information flow from right to left indicates the procedures by which natural information systems interact with the environment.

Biological Evolution. As indicated above, the epigenetic system is central to the narrow limits of change principle. Under the environment organizing and linking principle, the epigenetic system is used as well to provide a link between DNA-based information and the environment, but with quite different characteristics. When dealing with the novel information, the epigenetic system has a limited capacity to process novel information, however, under the environment organizing and linking principle, the epigenetic system can handle unlimited amounts of DNA-based information.

The importance of the epigenetic system is as a bridge between the environment and the genome. This feature can be seen in the narrow limits of change principle and the environment organizing and linking principle. The epigenetic system is triggered by the environment to determine the location of mutations and the rate of mutations under the narrow limits of change principle. Similarly, under the environment organizing and linking principle, the epigenetic system is activated by the environment to determine which stored genetic information will be used or ignored to make sure that activity is appropriate for the environment. But under the two principles, the same system has different features: from right to left of Figure 1.3 above, the epigenetic system can handle huge amounts of information as this information has been well organized in a genome. From left to right in figure 1.3, very little information can be handled at any given time.

Human Cognition. Working memory provides the bridge which coordinates activity between the external environment and long-term memory. Working memory also has the same features as the epigenetic system when working memory deals with well-organized information held in long-term memory. When working memory processes information from long-term memory, there is no limitations of capacity and duration. Ericsson and Kintsch (1995) suggested a new processor when working memory deals with information retrieved from long-term memory, which they call “long-term working memory”. Its characteristics are that it has no limitations of capacity and duration.

As the function of working memory is similar to the epigenetic system, therefore, information in long-term memory does not become active until it has been triggered by cues from the environment that induce working memory to choose which set of schemas to be used. Once the environmental information triggers working memory to choose a particular set of schemas held in long-term memory, those schemas can be used to govern complex behavior that is suitable for that environment. Finally, if we have a huge amount of information held in long-term memory, then we are likely to have a large number of environmental circumstances that we can deal with under the environmental organizing and linking principle.

1.6 Summary of Chapter 1

In this chapter, human cognitive architecture was discussed using a framework of five principles. The first part of this chapter discussed biologically primary and biologically secondary knowledge with their instructional implications. Biologically primary knowledge is learnable but not teachable while biologically secondary knowledge is learnable as well as teachable. Following those characteristics of biologically primary and secondary knowledge, we have some instructional implications: 1. For general problem-solving strategies, such as means-ends strategy, we do not need to give learners a special curriculum, as this kind of skill belongs to biologically primary knowledge. There is no evidence we can teach people to use a means-ends strategy. 2. For domain-specific knowledge (biologically secondary knowledge), such as learning mathematics, we need explicit instructions.

The second part of this chapter included the introduction of working memory and long-term memory, as these two systems are the main constituents of human cognitive architecture within cognitive load theory. Working memory has limited capacity and duration for new information but long-term memory has an unlimited capacity and duration.

The next section was about schema theory. From this theory, we knew that our knowledge stored in long-term memory was in the form of schemas. Schemas can be built and developed via practice. In order to reduce working memory load based on the features of schemas, two processes are needed: schema acquisition and schema automation.

The final section concerned the five principles of human cognition architecture discussed within cognitive load theory: The Borrowing and Reorganizing principle; Randomness as Genesis Principle; Narrow Limits of Change Principle; The Information Store Principle; Environmental Organizing and Linking Principle. These principles were discussed using biological evolution as an analogy. The five principles reveal the relationships between working memory and long-term memory, and most importantly, these five principles give us an outline of how our cognition system interacts with the external environment. In Chapter 2, I will discuss cognitive load theory in details.

Chapter 2 Cognitive Load Theory

Chapter 1 introduced human cognition architecture which is the basis of cognitive load theory. In Chapter 1, two kinds of knowledge were introduced: biologically primary knowledge and biologically secondary knowledge. In order to teach biologically secondary knowledge, explicit instruction is needed that follows the principles of human cognition and most importantly, takes into consideration the limited capacity of working memory when dealing with novel information learned from an external environment.

Accordingly, the main research focus of cognitive load theory should be biologically secondary rather than biologically primary knowledge. When dealing with biologically secondary knowledge which requires conscious, controlled processing, the characteristics of working memory and long-term memory should be considered, and they are exactly what this theory pays attention to. As for biologically primary knowledge, we have evolved to acquire this knowledge mostly implicitly, so no special instructions are usually required. The introduction of cognitive load theory in this Chapter will begin with the explanations of types of cognitive load, followed by how to measure types of cognitive load as the end.

2.1 Types of Cognitive Load

Cognitive load is generally considered as a construct representing the load that performing a particular task imposes on the cognitive system (Sweller et al., 1998). This load can be considered in two dimensions: mental load (task-based dimension) and mental effort (learner-based dimension). The former dimension reveals that the load is imposed by the task demands and the latter one is the load that learners actually use to accommodate the task demands (Paas & Van Merriënboer, 1994a). According to its functions, cognitive load within the framework of cognitive load theory was divided into intrinsic load, extraneous load and germane load (Sweller et al., 1998; Paas, Renkl, & Sweller, 2003, 2004; Van Merriënboer & Sweller, 2005). Even though most current descriptions of Cognitive Load Theory also consider these three types of cognitive load, this conceptualization of cognitive load has been recently challenged (see the discussion below for more details). In this thesis, three types of cognitive load will still be applied and discussed as the central concept of element interactivity within cognitive load theory can be the base to distinguish all three types of cognitive load, but the dual model of cognitive load (Kalyuga, 2011; Sweller, 2010; Sweller et al., 2011) will be used to discuss the additivity of types of cognitive load.

2.1.1Element Interactivity

Element interactivity is the basic concept for determining types of cognitive load, and all of them can be defined based on this concept. Interactive elements are defined as elements that must be processed simultaneously in working memory as they are logically related (Sweller et al., 2011). An element which should be processed in working memory can be a symbol or a concept, and it is characteristically a schema. Before a schema forms, its components must be processed in working memory as individual elements, but after the acquisition of the schema, these individual elements can be incorporated into this more complex structure to be a single entity processed in working memory. Element interactivity levels can be determined by estimating the number of interacting elements in learning materials (Sweller & Chandler, 1994; Tindall-Ford, Chandler, & Sweller, 1997).

2.1.2 Element Interactivity and Intrinsic Cognitive Load

Intrinsic load is determined by the nature of information that learners must deal with (Sweller, 1994), or it can be said that intrinsic load is caused by the interaction between the nature of information and the expertise of the learner (Van Merriënboer, Kester, & Paas, 2006). More specifically, it is determined by the levels of element connectedness which determines the nature of information and the learners' knowledge level. This load is directly relevant to learning through the acquisition of schemas, and the working memory resources needed to acquire schemas. Therefore, the resources of working memory used to process this type of load will contribute to schema acquisition and automation resulting in learning.

According to the degree of element interactivity, instructional materials can be divided into high-element-interactivity and low-element-interactivity materials. This division also corresponds to high-intrinsic-load and low-intrinsic-load materials. For example, when students learn the symbols in the periodic table of Chemistry, each chemical symbol stands for one element which should be processed in working memory. Most importantly, students can study each symbol individually with no reference to other symbols. For example, when students try to learn the symbol for iron, *Fe*, they can do so independently of learning the symbol for copper, *Cu* and students do not need to pay attention to the relationship between them. This kind of material has a low degree of element interactivity and therefore, it belongs to low-intrinsic-load material. The important characteristic of this kind of learning material is that it occupies little working memory capacity. Another example illustrating low-element-interactivity materials is when foreigners study English vocabulary: each English word which should be remembered belongs to low-element-interactivity material, as learners can process a single word in working memory without referring to other words. Therefore, the whole learning task is a low- intrinsic-load task.

However, a simple algebra equation like x+5=8, solve for x belongs to high-element-interactivity material. For this problem, each mathematical symbol is an element which should be processed in working memory. If we treat this equation as low-element-interactivity material and process each symbol (e.g., x, 5, =, 8) individually, there will be no meaningful learning. In order to understand and solve this problem, students should consider not only the individual symbols, but also the logical relations among them. Therefore, more elements, especially interactive elements, will need to be processed in working memory at a time, and naturally intrinsic load will be increased. As a result, this simple algebra problem belongs to the high-element-interactivity (high-intrinsic-load) category.

Another factor mentioned before which affects intrinsic load is the expertise of learners. If a learner presented with the above example has no relevant schemas for the equation $x+5=8$, solve for x, he/she needs to process each mathematic symbol in this equation as well as the relations among them, therefore, this problem will impose a high intrinsic load. However, a more experienced learner who has acquired relevant schemas about this problem may treat the whole equation $x+5=8$ as an entity in working memory, thus reducing the intrinsic load.

2.1.3 Ways of Managing Intrinsic Cognitive Load

According to Cognitive Load Theory (Sweller et al., 1998), intrinsic cognitive load cannot be altered by instructional interventions because it is determined by the interaction between the nature of the materials being learned and the expertise of the learner (Van Merriënboer & Sweller, 2005). However, Ayres (2006) noted that there were two methods which could be used to alter intrinsic load based on a change of the learning material itself and the expertise of learners. Firstly, according to the simple-to-complex sequencing (van Merrienboer et al., 2006), intrinsic cognitive load could be reduced if the complex task is preceded by simple tasks first, and then the whole complex task is provided. Secondly, with a pre-training technique (Mayer & Moreno, 2003), learners initially are presented some basic concepts to develop specific prior knowledge before the final materials are presented. The next sections will discuss these methods in more details.

Pollock, Chandler and Sweller (2002) conducted the first study of the isolated-interactive elements effect which is associated with simple-to-complex sequencing within the Cognitive Load Theory framework. In this study, Pollock et al. (2002) used complex learning materials (electrical safety tests) with high levels of element interactivity. In order to understand this complex learning material, participants need to process all the interactive elements in working memory simultaneously. However, the limited capacity of working memory makes this impossible, especially for novices. In order to process a large amount of interactive elements at a time, schema construction is required. Within the framework of human cognitive architecture, a schema can incorporate many elements as a single entity. Then a limited number of these schemas can be processed in working memory simultaneously. In Pollock et al. (2002) experiment, participants were randomly assigned to two groups. The first group initially was given instructions involving only very basic, isolated components of the material and then, in the second stage, they received the full task. The second group received all elements of the task in both stages. The isolated-elements group (the first group) significantly outperformed the full element-interactivity group (the second group). The results of this experiment indicated that learners who were provided with materials containing reduced element interactivity initially (isolated elements group) outperformed students who were always provided only fully interacting elements materials.

Ayres (2006, 2013) discussed the effect of segmenting the complex task (associated with simple-to-complex sequencing) to reduce the intrinsic load by using mathematics problems requiring expanding the brackets in the domain of algebra. Ayres (2006) randomly assigned students to three groups. The isolated tasks group received tasks which only contained individual calculations; the integrated tasks group was based on instruction which provided the whole task; and the mixed group switched from the conditions of the isolated group to the conditions of the integrated group. The results showed that the isolated tasks group reduced intrinsic load only for the students with lower level of prior knowledge, but for high-prior knowledge students, it was beneficial to use the integrated task method. Furthermore, Ayres (2013) studied students' errors in similar algebra problems according to different levels of cognitive load. The study investigated whether segmenting a complex task could be used to reduce errors made by students at the solution stages which imposed high levels of cognitive load. In Ayres (2006) investigation, each group received equal practice on each component of the algebra question, finding that errors associated with levels of cognitive load still occurred, even though the learners in the isolated group experienced reduced intrinsic load. Therefore, Ayres (2013) suggested that more practice on the key components which produced more errors with the isolated format of the task might produce further improvements. Students were randomly assigned to three groups: an isolated task with extra practice on key components which might generate more errors; an isolated task with equal practice on each component; and a full task condition containing all calculations. The results indicated that lower-prior-knowledge students benefited from the isolated task, but with equal practice on each component. However, generally, the isolated task group which used the method of segmenting a complex task was superior to the full task group.

In relation to the pre-training technique which can be regarded as pre-learning, Sweller et al. (2011) mentioned that the act of learning itself can reduce intrinsic load. This point can be explained by schema acquisition. For example, learning to read the word "cat" may cause high element interactivity initially, as children recognize it by each letter: c, a, t. However, after learning the word, they incorporate the three single letters into a single schema that would allow them to treat it as a whole word thus reducing working memory load. Therefore, the function of pre-training, which is similar to the function of learning itself, is to let learners firstly study some basic concepts and their relations to form some relevant schemas and then present students with a complex task containing those initial basic concepts.

2.1.4 Task Difficulty

Intrinsic load reflects the nature of the task. With high-intrinsic-load tasks, their interactive elements should be processed in working memory simultaneously. In contrast, for low-intrinsic-load tasks, their elements could be learned individually. However, it does not necessarily make the sequences of such tasks always easy to learn. Therefore, the difficulty of a task needs to be taken into account.

In fact, the task that includes only low-intrinsic-load sub-tasks can also be very difficult. For example, when learning chemical symbols in the periodic table, although the task of learning each symbol has a low intrinsic load, the number of chemical symbols is very large, so the whole task is still very difficult. However, for a task that itself is naturally high in element interactivity and therefore associated with a high intrinsic load, learning that task would inevitably be difficult (e.g., the algebra equation problem above).

2.1.5 Understanding

Element interactivity can also be used to define "understanding" (Marcus, Cooper, & Sweller, 1996). If we understand information, it means that we can process the whole set of interacting elements of this information in working memory (Sweller et al., 2011). For low-element-interactivity material, we cannot use understanding to explain the failure to learn this material. For example, if a learner cannot recognize that *Fe* stands for Iron, this failure is caused by lack of knowledge or memory. However, the failure to solve an algebra problem, such as x+5=8, indicates a lack of understanding. Therefore, understanding only can be used when dealing with high-element-interactivity material rather than low-element-interactivity material. This distinction is also relevant to the difference between learning with understanding and learning by rote.

No matter whether materials are high or low in element interactivity, they can always be learned by remembering individual elements, but only high-element-interactivity materials can be learned with understanding. Learning with understanding requires increased amounts of interactive elements to be processed in working memory. Learning by rote could be regarded as a necessarily initial stage of learning by understanding, as learners can remember some individual and basic elements (concepts) first and then gradually can be presented with more complex materials to learn with understanding (based on simple-to-complex sequencing mentioned above). For example, when we want to understand an algebra equation, such as x+5=8, we firstly need to remember what is x, the addition sign and some other simple rules and mathematical symbols (learning by rote) to finally understand the whole meaning of this equation which contains high interactivity elements (learning by understanding) and then solve a problem based on this equation successfully. Therefore, learning by rote can be a basic step of learning with understanding and may be unavoidable (Pollock et al., 2002).

2.1.6 Instructional Implications of Intrinsic Load

Intrinsic load is directly relevant to learning or schema acquisition, therefore, with the limited capacity of working memory, more resources of working memory should be allocated to deal with intrinsic load. Theoretically, intrinsic load cannot be altered as it is determined by the level of element interactivity of learning materials relevant to learner levels of expertise. However, intrinsic load could be reduced by pre-training, segmenting the task and simple-to-complex task sequencing. When segmenting a complex task, it is important to ensure that the resulting smaller and simpler tasks are meaningful for learners. It is also important to keep in mind that a low-element-interactivity task is not necessarily easy to learn, and that the concept of understanding can only be applied to high-element-interactivity materials.

2.2 Element Interactivity and Extraneous Cognitive Load

Extraneous cognitive load is a load that is not necessary for learning (i.e., schema construction and automation) and can be altered by instructional interventions, as this load is influenced by the way instructional materials are presented (Sweller et al., 1998). This load can be imposed by suboptimal teaching methods, such as those requiring mental integration of separately presented sources of information, or by an instructionally ineffective problem solving method, such as means-ends strategy. This load must be reduced or even eliminated (Kalyuga, 2011) to make more working memory resources available for dealing with intrinsic load, which enhances learning (i.e., schema acquisition or automation).

Extraneous load can also be discussed under the concept of element interactivity. Sweller (2010) suggested that element interactivity was the major source of working memory load underlying extraneous as well as intrinsic cognitive load. If element interactivity can be reduced without altering what is learned, this load belongs to extraneous load. And from the view of element interactivity and cognitive load theory, this load entirely relies on what cognitive activities instruction requires learners to be involved in. For example, if instruction requires learners to learn relations between multiple elements by using intensive up-and-down scanning and traversing between these elements that could split learner's attention, such activities are irrelevant to learning and therefore would lead to a high extraneous load. In contrast, the instruction that physically integrates such elements would eliminate this source of extraneous cognitive load.

2.2.1 Instructional Implications of Extraneous Load

As indicated in the discussion above, extraneous load is irrelevant to learning, therefore, this type of cognitive load should be controlled and should be the focus when we design our instructions. Compared to intrinsic load, increasing extraneous load is never advantageous, as learners need to use the limited resources of working memory to deal with extraneous load leaving few resources of working memory to deal with intrinsic load which is relevant to learning. Therefore, understanding may be increased when extraneous load is reduced. To sum up, when teachers design instructions, they should consider how to reduce extraneous load imposed by their instructions to free more resources of working memory for schema acquisition and automation.

2.3 Element Interactivity and Germane Cognitive Load

According to cognitive load theory, the resources of working memory can be used for dealing with intrinsic load which is imposed by the nature of information and the extraneous load which is caused by suboptimal instructions. The resources allocated to deal with intrinsic load are relevant to learning, namely, to learner’s construction of cognitive structures that enhance performance (Van Merriënboer et al., 2006). These resources are called ‘germane resources’ or germane load. It is assumed that under conditions of low extraneous load, low intrinsic load or the combination of both, the available working memory resources would be sufficient to process information that is relevant to schema acquisition or automation (i.e., sufficient germane resources). Of course, in this condition, the total amount of intrinsic and extraneous load should not exceed the capacity of working memory. If it is excessive, reducing extraneous load could potentially allow an increase in germane resources.

Germane cognitive load also can be specified in terms of element interactivity (Sweller, 2010). However, compared to the other two types of cognitive load that rely on the combination of material and learners characteristics, germane cognitive load is more based on learners characteristics, as this cognitive load refers to the working memory resources actually allocated by learners to deal with the element interactivity which is associated with intrinsic load. The ability to allocate sufficient resources is determined by learner's motivation to learn. If intrinsic load is high, but extraneous load is low, more working memory resources will be used to deal with essential element interactivity associated with intrinsic cognitive load and create more germane load; if extraneous cognitive load is high, learning will be reduced, as fewer resources of working memory can be used to deal with intrinsic cognitive load, then less germane load will be produced.

2.4 Additivity of Types of Cognitive Load

According to the dual model of cognitive load in recent descriptions of cognitive load theory (Kalyuga, 2011; Sweller, 2010; Sweller et al., 2011), two independent types of cognitive load - intrinsic and extraneous - are additive, and the total load formed by intrinsic and extraneous loads indicates required working memory resources. If the total load exceeds the available capacity of working memory, learning will be inhibited. As the capacity of working memory can be regarded as constant for a given learner (relevant to her/his domain-specific knowledge structures), if most of this capacity is used for dealing with extraneous, irrelevant load, fewer resources will be available for dealing with essential, intrinsic load.

Accordingly, instructional design should eliminate (ideally) or reduce extraneous load as this kind of load has nothing to do with learning. As for intrinsic load, it should be managed by selecting appropriate learning tasks (Kalyuga, 2011). The learning task should not be too complex in order not to impose an extremely high intrinsic load and make working memory break down, however, it should not be too simple in order to be sufficiently cognitive challenging and motivating (Schnotz & Kürschner, 2007). The resources of working memory actually allocated to deal with intrinsic load which is relevant to learning and schema acquisition (germane resources) need to be maximized, while resources allocated to deal with extraneous load should be reduced.

2.5 The Construct of Cognitive Load

Paas and van Merriënboer (1994b) suggested a model of cognitive load, in which learner characteristics, learning-task characteristics and their interactions were regarded as the three factors causing cognitive load. Reviewing previous research of cognitive load theory, the main focus was mostly about learner characteristics, learning-task characteristics and their interactions. Therefore, Paas and van Merriënboer (1994b) suggested a model of cognitive load which revealed the interactions between learner characteristics and learning-task characteristics.

However, with the development of cognitive load theory, other research studies have appeared, such as the gaze aversion effect (Doherty-Sneddon & Phelps, 2005), the positive relationship between arterial blood oxygen saturation and quality of cognitive performance (Scholey, Moss, Neave, & Wesnes, 1999) and the effect of emotional state.as a mediator of the relationship between the physical learning environment and learning performance (Erez & Isen, 2002; Uline & Tschannen-Moran, 2008). Choi, van Merriënboer and Paas (2014) indicated that the physical environment should be separated from learning-task characteristics as another factor which may cause cognitive load. Therefore, they modified the construct of cognitive load used in 1994. In this new model, the physical environment embraces the other two factors (learner characteristics and learning-task characteristics). In a real classroom setting, these three factors should interact rather than being totally independent. Therefore, they are intertwined (Choi et al., 2014).

2.6 Instructional Implications of Different Types of Cognitive Load

In the previous sections, intrinsic cognitive load, extraneous cognitive load and germane cognitive load and their relationships with the resources of working memory and the concept of element interactivity were introduced. As mentioned above, the total of intrinsic and extraneous load should not exceed the total resources of working memory for effective learning to occur. For given learning material and a given learner, intrinsic load cannot be altered. If intrinsic load is very high, the management of extraneous load becomes very critical. If extraneous load is very high in this situation, working memory will be overloaded, and learning will be inhibited. However, if intrinsic load is very low, a high extraneous load may have little influence on the total working memory load and considerations of managing cognitive load may not be essential for instructional design. To sum up, cognitive load theory concerns reducing extraneous load and managing intrinsic load (Sweller et al., 2011).

2.7 Measurement of Types of Cognitive Load

Measuring different types of cognitive load is another important issue of cognitive load theory. If cognitive load can be reflected through concrete numbers, it could directly reveal the amount of total load imposed on working memory and show the change of total load between the original instruction and the instruction designed according to cognitive load theory. The quantified total amount of cognitive load can be used to indicate whether the instruction designed under cognitive load theory is indeed more effective.

2.7.1 Indirect Measures of Cognitive Load

Computational Models. Computational models were the first attempt to provide independent evidence that cognitive load was an important factor in instructional design. This measure is based on the hypothesis that more working memory load will be imposed in high problem-solving search than low problem-solving search (Sweller et al., 2011). Therefore, procedures which reduce problem-solving search will reduce cognitive load. Sweller (1988) and Ayres and Sweller (1990) used a production system to simulate the problem-solving process of multi-step geometry problems. They found that a high search strategy involved a more complex model which required more resources of working memory than a simpler model. Therefore, computational models indirectly indicate that high problem-solving search does increase cognitive load.

Performance During Acquisition. Instructional time and performance accuracy were both used to indicate the level of cognitive load. Specifically, fewer errors and shorter instruction time indicated a lower cognitive load. Many studies (Owen & Sweller, 1985; Sweller, Chandler, Tierney, & Cooper, 1990; Sweller & Cooper, 1985) supported this assumption. For example, Owen and Sweller (1985) used two kinds of trigonometry problems: one had conventional specific goals and another was goal-free. Goal-free problems request students to find as many variables (e.g., angles) as they can without specifying a goal. Fundamental errors and trigonometric errors were considered in this experiment as indicators of cognitive load. The results revealed that the goal-free condition reduced extraneous load and demonstrated higher levels of performance than the specific-goal condition.

Error Profiles. Error rates also were used to indirectly measure the level of cognitive load because error rates may be positively correlated with the levels of cognitive load. Ayres and Sweller (1990) indicated that in geometry problems, students often make mistakes at the particular steps which require greater resources of working memory, namely, more errors could be made if these steps require a high cognitive load. Ayres (2001) used mathematic problem-solving tasks to indicate that error rates varied at different points, and high error rates were found at the steps which had many variables to be considered, namely, high error rates occurred at the steps with a high intrinsic load.

Time-on-task. This method is based on the relationship between the cognitive load and the time which learners invest in learning (Lee, 2013). A positive relationship was found between cognitive load and time (Wright & Ayton, 1988). Specifically, the decision time increases because the difficulty of learning tasks increases, resulting in a positive relationship between measures of objective task difficulty and decision time. Finally, Chandler and Sweller (1991, 1992) indicated that the instructional time could also be used to measure cognitive load indirectly as mentioned in the *performance during acquisition* section.

It should be noted that cognitive load theory does not extensively focus on indirect measures and they are not used in many research studies. Also indirect measures are not sensitive to the differences among types of cognitive load and so are not widely used and accepted. However, the subjective measure of cognitive load (see next section) is widely used and sensitive to the differences among types of cognitive load.

2.7.2 Subjective Measures of Cognitive Load

Paas (1992) suggested a direct measure of cognitive load which uses subjective ratings of mental effort. The theoretical base of this measure is that learners are able to self-estimate the amount of mental effort invested during learning and testing (Paas, 1992). A definition of mental effort was suggested by Paas, Tuovinen, Tabbers and van Gerven (2003) as the aspect of cognitive load which is actually used to deal with the demands imposed by the task. This measurement method uses a 7-point or 9-point scale. Both scales vary from a very, very low mental effort (number 1) to very, very high mental load (numbers 7 or 9). This measure requires learners to rate their mental effort in the learning and/or test phases by, for example, circling a number in the range from 1 to 9 (when using 9-point scale). This measure is widely used in empirical studies as it is highly reliable and simple to be implemented (Paas, Van Merriënboer, & Adam, 1994).

An alternative to the above subjective rating scale is based on asking learners to rate how difficult or easy they find the task to study. Marcus et al. (1996) demonstrated that subjective measures of difficulty could be sensitive to the different levels of element interactivity of a task. Ayres (2006) used mathematic problems to demonstrate the effectiveness of a subjective measure of difficulty in measuring the variation of intrinsic load. Chen, Epps and Chen (2011) compared four methods of measuring cognitive load: subjective rating, eye tracking, completion time and accuracy. They used 5-level addition problems. Their difficulty increased with increases in level. Specifically, the difficulty was relevant to how many digits and how many numbers to be carried were involved in addition. Level 1 to 3 used one digit for each of the four numbers, which had no number to carry, one number to carry and two numbers to carry respectively. Levels 4 and 5 used two-digit addition with one number to carry produced at different points: only lower value digits belonged to level 4. If both digits were used, they belonged to level 5, during each addition process. Results indicated that subjective rating was the most sensitive measure to the five levels. In addition, eye tracking was as sensitive as completion time which could be used to classify two to three levels.

Subjective ratings of mental effort and difficulty both assume that learners are able to self-estimate the amount of mental effort invested during learning and/or testing phases and the degree of learning difficulty. Van Gog and Paas (2008) suggested that measures of mental effort and measures of difficulty might be two distinct tools but with some relations (this view was also supported by Ayres and Youssef, 2008). Firstly, the measure of difficulty is not always consistent with effort, as some difficult tasks are too demanding and learners cannot actually allocate mental effort to deal with these tasks. Secondly, van Gog and Paas (2008) found that the timing of the measure may also influence the results. For example, some researchers use the measure of mental effort after the acquisition phase and others use it after solving posttest problems. However, in some experiments, there were group differences on performance tests but no statistically significant differences between the two different measures of cognitive load (Cuevas, Fiore, & Oser, 2002; Hummel, Paas, & Koper, 2004; Kester, Kirschner, & Merriënboer, 2005). Galy, Cariou and Mélan (2012) indicated that alertness might be a factor which interfered with subjective-rating measures.

However, some issues concerning subjective ratings have appeared. Firstly, as learners have difficulty to recognize the exact quantity of different types of cognitive load, some studies failed to use this measurement efficiently (Paas, 1992; Xie & Salvendy, 2000). Secondly, Lee, Plass, and Homer (2006) indicated that students who were under 15 years old found it especially difficult to distinguish the meaning of the three types of cognitive load. Thirdly, subjective ratings are sensitive to different types of cognitive load and are convenient, but suffer from a lack of objectivity (Lee, 2013).

2.7.3 Other Measures of Cognitive Load

Secondary Task Method or dual-task methodology, is a traditional way to measure working memory load (Britton & Tesser, 1982; Kerr, 1973; Park & Brünken, 2015). This method requires learners to allocate extra cognitive load to the secondary task which is attached to the primary task. For example, participants are required to solve a mathematics problem as a primary task as well as to give specific responses to a particular sound which is used as a secondary task during problem solving. Even though the secondary task traditionally requires less working memory resources than the primary task, if the primary task requires significant resources of working memory, the performance of the secondary task may be interrupted. Learning experience of problem solving could also be regarded as an example of dual-task process. Sweller (1988) mentioned that two processes were involved in problem solving: solving the problem itself and learning from problem solving process. Therefore, if learners treat solving problems as the primary task and this problem is complex, then less learning from this problem will occur.

Marcus et al. (1996) used diagrams rather than text only to show that this format could reduce element interactivity which is associated with the intrinsic load of the primary task and then improve performance on the secondary task. Chandler and Sweller (1996) used recalling a letter as the secondary task to indicate that superior instruction led to better performance on the secondary task and indeed reduced cognitive load. The research of Brünken, Steinbacher, Plass and Leutner (2002), Brünken, Plass and Leutner (2004) and van Gerven, Pass, van Merriënboer and Schmidt (2006) used dual-task methodology to explain the benefits of dual-modality (audiovisual) presentations. However, as the primary task and the secondary task need to share cognitive resources, and the secondary task usually influences the first task which is relevant to learning, so this method is difficult to practice in a real classroom setting (Lee, 2013).

Physiological Measures of Cognitive Load. Paas, van Merrienboer and Adam (1994) used physiological measures of cognitive load through spectral analysis of heart rate, however, few researchers have used this method. Recently, research about using this kind of measure has appeared again. Based on some research studies, five main indexes were used to indicate a change in cognitive load: pupil size (Kahneman & Beatty, 1966), fMRI (Paas, Ayres, & Pachman, 2008; Whelan, 2007), EEG (Antonenko, Paas, Grabner, & van Gog, 2010), eye tracking (Just & Carpenter, 1976; Mayer, 2010; Park, Korbach, & Brünken, 2015) and language complexity (Khawaja, Chen, & Marcus, 2010). Among those indexes, pupil size has age limitation, both EEG and fMRI are not sensitive to the differences among types of cognitive load. Luckily, the EEG method has an advantage of reflecting various types of cognitive load (Antonenko & Niederhauser, 2010), and some evidence has emerged that eye tracking can also be used to measure fluctuations in cognitive load (Underwood, Jebbett, & Roberts, 2004).

In addition, some research has addressed the relation of brain waves and cognitive load by using EEG. Neubauer, Freudenthaler and Pfurtscheller (1995) used gifted children as participants and found that gifted children released higher Alpha power than average children, which indicated that gifted children used less mental effort. A similar result was obtained by Gerě and Jaušcvec (1999). Antonenko and Niederhauser (2010) discussed alpha, beta and theta waves and their results suggested that instruction with guidance reduced those brain waves, which may be related to extraneous load.

Besides using the EEG to measure a change of cognitive load, a change of pupil size was also investigated to measure the change of amount of cognitive load. Piquado, Isaacowitz and Wingfield (2010) reported some results which were about the relation between pupil size and the amount of cognitive load. They used two materials: digit numbers (4-digit, 6-digit and 8-digit) and sentences with different levels of syntactic complexity, to investigate the change of pupil size between old and young adults. Finally, they indicated that both age groups showed larger pupil sizes when they experienced a larger short-term memory load. What's more, only young adults' pupil size was sensitive to the change of syntactic complexity of sentences. Therefore, the change of pupil size which is a relatively direct measurement may have advantages in finding an objective way to measure cognitive load when individuals experience mental tasks with varying difficulty.

Galy et al. (2012) pointed out a factor which affected the physiological measures of cognitive load. In their research, the higher alertness the subjects had, the larger the heart rate variation would be. Therefore, heart rate might be affected by alertness. However, they also pointed out that compared to subjective-rating measures, the heart rate variability was invalid, and insensitive to subtle mental load variations.

2.7.4 Measures of Different Types of Cognitive Load

With the development of cognitive load theory, many researchers have become interested in measuring different kinds of cognitive load rather than the total cognitive load imposed on working memory (Ayres, 2006; DeLeeuw & Mayer, 2008). As intrinsic and extraneous load add to the total load, therefore, if one type of load is kept constant, the other one can be measured through variations in the total load. For example, Ayres (2006) kept extraneous load constant and measured the differences in intrinsic load. In this experiment, students were asked to solve a set of mathematic problems, and it was assumed that as students had received instructions before and no extra instructions were provided during this task, the extraneous load was constant during this experiment. After completing problems, students were required to rate how easy or difficult they found each computation, and the results indicated that students could identify differences in intrinsic load under the condition that extraneous load was constant. Further, Ayres (2006) suggested that students with more domain-specific knowledge were better at recognizing differences in intrinsic load.

Cierniak, Scheiter and Gerjets (2009, p. 318) used different descriptions of questions in subjective rating scales to measure different types of cognitive load, such as *'How difficult was the learning content for you? How difficult was it for you to learn with the material? How much did you concentrate during learning?'* However, no expected match between learning outcomes and types of cognitive load was found (Gerjets, Scheiter, Opfermann, Hesse, & Eysink, 2009). Psychometric distinctions between different kinds of cognitive load require students to indicate which particular cognitive load they are experiencing, which is quite difficult for learners, especially for novices who have less capability to distinguish between different kinds of cognitive load (Sweller et al., 2011). The alternative way of psychometric measurement of individual cognitive load is to use randomly controlled experiments which vary one kind of cognitive load and then keep other kinds of cognitive load constant.

Therefore, Lee (2013) applied various measurements to measure different types of cognitive load, then discussed the correlations among them and learning outcomes by using language learning materials. The final results indicated that difficulty rating by self-reporting could be an indicator of cognitive load, but brain waves, such as T7 which is relevant to problem solving or T3 which is relevant to process verbal information might be used to partially explain cognitive load.

2.7.5 Static or Dynamic Concepts of Cognitive Load

Based on the previous discussion, subjective ratings of cognitive load are widely used in research within a cognitive load framework as they are very easy to be implemented and sensitive to the different kinds of cognitive load. However, this measurement usually reflects the total amount of cognitive load and regards the cognitive load as a static concept. In this connection, Zheng and Cook (2012) discussed whether cognitive load is a static or a dynamic concept. The measurement of mental effort suggested by Paas and van Merriënboer (1994) fits the static view, but this view is not complete (Xie & Salvendy, 2000). Therefore, Xie and Salvendy (2000) suggested a dynamic view of cognitive load that involves instantaneous load, peak load, accumulated load, average load and overall load. The instantaneous load reflects its dynamic fluctuations during complex learning, which is the dynamic concept; peak load refers to the maximum load which learners experience; average load reflects a mean level or degree of load experienced per unit of time; overall load corresponds to the load learners have experienced during the whole task; accumulated load is the load that has gradually accumulated or built up over a task and is assumed to be the total amount of load experienced at task completion (Xie & Salvendy, 2000). Considering these differences, for example, the physiological measures mentioned above provide online measurements of cognitive load, such as via heart rate, brain activity and pupillometry, and results matched the dynamic concepts indicated by Xie and Salvendy (2000), such as instantaneous load, peak load and accumulated load. Zheng and Cook (2012) challenged the traditional cognitive load measurements and used online measurements to record participants' pupil diameter when they studied complex graphics. Finally, they suggested that offline and online methods measured different aspects of cognitive load that might not be significantly correlated.

2.7.6 New Instrument of Measuring Different Types of Cognitive Load

Leppink, Paas, Van der Vleuten, Van Gog and Van Merriënboer (2013) proposed a new instrument for measuring different types of cognitive load. This method uses a ten-item questionnaire to measure intrinsic load, extraneous load and germane load separately. For intrinsic load, the items are: *the topic/topics involved in this activity was/were very complex; the formulae learned in this activity were very complex; the concepts and definitions involved in this activity were very complex.* For extraneous load, the items are: *during this activity, the* *instructions and explanations were very unclear; in term of learning, the instructions and explanations were very unclear; a lot of unclear language was included in instructions and explanations.* For germane load, the items are: *the activity really enhanced the understanding of topic/topics; the activity really enhanced the knowledge and understanding of this domain; the activity really enhanced the understanding of formulae; the activity really enhanced the understanding of concepts and definitions.* Learners are required to circle a number from 0-10 (0: stands for not at all the case; 10: stands for completely the case) to indicate the degree to which those items reflected the effectiveness of the learning phase. Based on preliminary evidence, this method of measuring different types of cognitive load may provide new insights on convergence between objective and subjective measures (Leppink et al., 2013).

In order to investigate the validity of this new-developed questionnaire, Leppink, Paas, Van Gog, van Der Vleuten and Van Merrienboer (2014) tested this measurement in the field of statistics as well as language. The final results indicated that the assessment of germane cognitive load was not reliable, and the positive correlation between intrinsic load and extraneous load suggested that students had difficult in distinguishing between them. Based on cognitive load theory, if these three types of cognitive load are additive (Sweller et al., 1998), the correlations among them should be around zero, which means that they should be independent on each other, but only the correlation between intrinsic load and germane load was around zero. Therefore, it was still difficult to distinguish different types of cognitive load. In addition, the correlation between task performance and germane load was positive, but still was not statistically significant. Finally, Leppink et al. (2014) suggested that it might be worth testing the effect of wording used in this new measurement.

2.8 Summary of Chapter 2

In this chapter, different kinds of cognitive load were introduced. Intrinsic load reflects the nature of learning material, and is determined by element interactivity. Learning materials with high levels of element interactivity are associated with high intrinsic load, while the materials with low levels of element interactivity are associated with low intrinsic load. Two different ways (pre-training and segmenting tasks) are used by researchers to manage intrinsic load. In cognitive load theory, the concept of understanding associated with intrinsic load is only relevant to high-element-interactivity materials rather than low-element-interactivity materials. Extraneous load is determined by the way learning materials are presented, i.e., their instructional design. An optimal instructional design should reduce extraneous load. Germane load (germane resources) is defined as the actual working memory resources are allocated to accommodate the intrinsic load. According to Kalyuga (2011) and Sweller (2010), this kind of load should not be considered as an independent type of load in cognitive load theory.

Under the framework of cognitive load theory, different kinds of cognitive load can be added to form the total load which is imposed on working memory. As the capacity of working memory is limited, the appropriate management of each type of load is important. Extraneous load which is irrelevant to learning or schema acquisition must be reduced or even eliminated. Intrinsic load, which is relevant to learning, should be appropriately controlled. However, if intrinsic load is very low, a high extraneous load may not overload working memory. Learning will be interfered with if learning materials are characterized by a high intrinsic load with a high extraneous load. This chapter also discussed the issue of measuring of cognitive load. Subjective ratings of cognitive load remains the most widely used measurement method as it is easy to be implemented even in realistic classroom settings and is sensitive to the changes in cognitive load. In order to measure different types of cognitive load individually, the experimental design should control one type of load and vary the other types that can be thus measured based on observed variations in total cognitive load.

In Chapter 3, I will narrow down to begin discussing one of cognitive load effects - the worked example effect. This effect also is one of key concepts of my whole thesis.

Chapter 3 Worked Example Effect

During the mid-1950s to the 1970s, many cognitive psychologists used the paradigm of learning by examples to examine and to describe processes of how a concept forms (Bourne, Goldstein, & Link, 1964; Bruner, Goodnow, & Austin, 1956; Tennyson, Woolley, & Merrill, 1972). Most example-based research at the time aimed at investigating the acquisition of a target concept after requiring learners to view various examples or non-examples. Later, the focus of cognitive psychologists changed to more complex forms of knowledge rather than simple concept learning, such as how persons at different levels of expertise use knowledge to interpret experience and solve problems in specific domains (Atkinson, Derry, Renkl, & Wortham, 2000). The most outstanding and interesting research during the time when problem-solving based learning was very popular was the research about worked-example based learning conducted within the cognitive load theory framework. This work demonstrated that using worked example-problem solving pairs could lead to better performance than problem-based learning (Cooper & Sweller, 1987; Sweller & Cooper, 1985). In cognitive load theory, the worked example effect has been investigated over many years. It is the major effect within this framework that sets up an opposite perspective to constructivist views of instruction.

The use of worked examples is an instructional tool, which provides the professional solution of a problem for a learner to study (Atkinson et al., 2000). There is no specific definition for a worked example, but the typical components of a worked example include a problem statement and associate procedures for solving this problem (Atkinson et al., 2000). An example from algebra is provided (p99, Sweller et al., 2011):

Make a as the subject of the equation, $(a + b) / c = d$.

Solution

$$(a + b) / c = d$$

$$a + b = dc$$

$$a = dc - b$$

A body of research has indicated the positive effects of worked examples on students' learning. Worked examples were successfully used in the domains of algebra (Sweller & Cooper, 1985), statistics (Paas, 1992), geometry (Paas & Van Merriënboer, 1994b; Schwonke, Renkl, Krieg, Wittwer, Aleven & Salden, 2009) and physics (Reisslein, Atkinson, Seeling, & Reisslein, 2006; Van Gog, Kester, & Paas, 2011; Van Gog, Paas, & van Merriënboer, 2006) etc. The later study of Carroll (1994) also suggested that using worked examples could provide particular help for low achieving students or students with learning disabilities in mathematics. Similar findings found in the domain of geometry suggested that when studying 2D and 3D mental rotation, using worked examples could facilitate learning (Pillay, 1994).

3.1 The Worked Example Effect and Human Cognitive Architecture

According to the worked example effect in cognitive load theory, worked examples that provide full guidance to learners on how to solve a problem can result in better performance than a problem solving condition which provides no guidance. The worked example effect is directly relevant to the human cognitive architecture (Sweller et al., 2011) discussed in Chapter 1. According to the borrowing and reorganizing principle, when students study worked examples, the worked out solutions provide the schemas that can be borrowed and be used to solve similar problems. In contrast, students in a problem solving condition need to randomly generate their own solutions and then test their effectiveness, which reflects the randomness as genesis principle.

As randomly generating solution moves would produce more interactive elements which may exceed the capacity of working memory, so this would break the narrow limits of change principle. Consequently, high extraneous load would be imposed on working memory. Based on the key characteristics of working memory, when it deals with novel information, it has a very limited duration and capacity, which indicates that it is counterproductive to borrow many interactive elements of information from the environment. Therefore, worked examples that provide learners with well-structured schemas that can be used to get around working memory limitations by encapsulating many elements of information into a single entity allow learners to deal with the narrow limits of change principle.

After schemas are borrowed from instructors who construct worked examples and are reorganized (reconstructed) accordingly, they will be transferred to long-term memory for storing, according to the information store principle. Finally, if such schemas are successfully stored in long-term memory, they can be retrieved from long-term memory when needed to guide activities required for successful functioning in an external environment, according to the environmental organizing and linking principle. Therefore, based on the characteristics of human cognition, studying worked examples should have advantages over problem solving activities.

3.2 The Worked Example Effect and Element Interactivity

As discussed in Chapter 2, element interactivity is an index to show the difficulty of learning materials. Interactive elements are defined as elements that must be processed simultaneously in working memory as they are logically related (Sweller et al., 2011). The levels of element interactivity are determined by the nature of learning materials as well as the expertise of learners (See Chapter 5).

Some researchers have investigated ways to reduce the element interactivity of learning materials to improve learners' performance by using adapted worked examples (Ayres, 2006; Blayney, Kalyuga, & Sweller, 2010; Gerjets, Scheiter, & Catrambone, 2004; Pollock et al., 2002), such as worked examples presented in isolated-element format. The following sections will explain adapted worked examples in details.

Gerjets et al. (2004) compared the molar presentation of worked out solutions with modular presentation of solutions in worked examples. The "molar way of learning" refers to considering the category of problems with their associated solutions, demonstrating high levels of element interactivity; the "modular way of learning" presents learners with partly independent modules which could be used to solve part of this problem, with each module meaningful in isolation. Therefore, the modular strategy reduces the level of element interactivity by adapting the original worked examples.

Gerjets et al. (2004) used problems that were about the probability of correctly guessing the first three places out of seven runners in a 100m race. In the molar condition, the critical features of this problem were identified and then learners were provided the worked example by using a general formula ($n!$ / $(n-k)!$) that calculated the total number of permutations. The final answer was obtained by directly inserting the relevant numbers into that general formula. In the modular condition, learners were shown the solutions when considering each event (1^{st}, 2^{nd} and 3^{rd}), and then the final answer was obtained by doing a multiplication of the three probabilities. The modular condition which only needed learners to consider one event at a time was found to be superior to the molar condition which needed learners to consider the three conditions simultaneously on similar and transfer problems, involving types of learners (both low and high prior knowledge).

Similarly, Ayres (2006) found that bracket expansion tasks were high in element interactivity and difficult for novices. In Ayres' experiment, tasks like 4(3x-6)-5(7-2x) were used. This kind of task was assumed to be high in element interactivity, as learners needed to consider numbers and mathematical symbols simultaneously. Ayres then used an isolated-elements method to require learners to do one calculation at a time, for example, learners only needed to calculate $4 \times 3x$ in a given time. The isolated-elements way reduced the level of element interactivity and less experienced learners benefited more from this way, whereas, more experienced learners benefited more from the interacting-elements way (full worked example pairs).

Previous research on altering the levels of element interactivity of worked examples demonstrates that worked examples used to test for the worked example effect within the framework of cognitive load theory could be high in element interactivity. Researchers have investigated some ways to reduce levels of element interactivity of worked examples in order to facilitate learning, especially for novices' learning.

Within the framework of cognitive load theory, if materials presented to learners are high in element interactivity associated with a high intrinsic load, it is critical to control the extraneous load in order to avoid overloading working memory. However, if materials are low in element interactivity (i.e., the intrinsic load of materials is low), controlling extraneous load may not be necessary. This defines the element interactivity effect.

The element interactivity effect affects other cognitive load effects including the worked example effect (Sweller et al., 2011). Therefore, whether worked example effect is obtainable may depend on the levels of element interactivity of specific materials. Alternatively speaking, if materials are low in element interactivity for given learners, the use of worked examples may be ineffective.

Therefore, according to the element interactivity effect, if we want to obtain a worked example effect, the presented learning materials should be high in element interactivity (namely, materials with high intrinsic load). If materials are low in element interactivity, learners still have sufficient working memory resources to process the extraneous load imposed by suboptimal instructions, which may reduce the effectiveness of worked examples. The levels of element interactivity also are determined by the expertise of learners, which again affects the effectiveness of worked examples. This issue will be discussed in Chapter 5.

3.3 Research Paradigms Used to Study the Worked Example Effect

Research using worked examples involves two kinds of design patterns: a worked example followed by a similar conventional problem or a conventional problem followed by a worked example. The first pattern is supported by cognitive load theory and allows learners to build their own mental model of the problem solution procedure and then test and practice this solution procedure on that similar conventional problem. Sweller and Cooper (1985) demonstrated the effectiveness of worked example-problem solving pairs (study-solve strategy) consisting of a worked example followed by a similar to-be-solved problem. This sequence originated from the assumption that students would be more motivated to study the worked example provided they know that a similar problem would need to be solved after studying this worked example (Sweller & Cooper, 1985).

An alternative pattern corresponds to the sequence (problem solving-worked example pairs) in which learners first meet some difficulties during initial problem solving and then they might be more motivated to study the following worked example. However, it is questionable whether learners can accurately identify the deficiencies of their performance. For example, Van Gog, Kester and Paas (2011) compared three conditions (worked examples only, example-problem pairs and problem-example pairs) with problem solving only. In this experiment, participants were novices in applying Ohm's law to determine potential problems in electrical circuits. The results indicated that invested mental effort for training tasks was lower in the worked examples only condition and example-problem pairs, compared to the problem-example pairs and problems only conditions. Among all of conditions, the example-problem pairs required the lowest level of mental effort. The performance on test tasks supported the effectiveness of example-problem pairs indicating that higher test performance was obtained in the examples only condition and example-problem pairs.

Reisslein et al. (2006) also compared the effect of example-problem pairs and problem-example pairs using engineering materials. Their results suggested that the interaction of example-problem and problem-example sequences with levels of learner expertise should be considered: novices benefited more from example-problem pairs, while learners with higher levels of expertise benefited more from the problem-example pairs.

Paas (1992) and Paas and Van Merriënboer (1994b) conducted experiments within the domain of basic statistics and geometry. In Paas' experiment, participants were randomly assigned to 3 conditions: conventional problem solving, worked example and partial worked-out example conditions. As the whole experiment was computer-based, if participants who were in the conventional problem solving condition could not correctly solve problem in the required time or the available time elapsed, the computer would automatically provide an example for them to study, namely, the problem solving-worked example sequence, whereas, for the worked example condition, participants could study the provided worked example until they fully understood it and then moved to solve a similar problem, namely, the worked example-problem solving sequence. A similar design was used in the experiments of Paas and Van Merriënboer, demonstrating similar results that the worked example condition (the worked example-problem solving sequence) was more effective than the conventional problem solving condition (the problem solving-worked example sequence). Again, in both studies, the authors mentioned the consideration of the effect of learner's expertise on learning worked examples. Novices may refer to worked example more frequently during problem solving than more experienced learners. Therefore, novices may not be able to effectively diagnose their deficiencies of performance if they are presented with problem-example pairs.

Concerning the sequence of problem solving followed by direct or delayed instruction (such as worked examples), the research of invention activity for future learning (Schwartz & Bransford, 1998; Schwartz, Chase, Oppezzo, & Chin, 2011; Schwartz & Martin, 2004) is designed to prepare students to learn or solve problems in new environment. Productive failure (Clifford, 1984; Kapur, 2008, 2011, 2012; Kapur & Bielaczyc, 2011; Kapur & Rummel, 2012), which indicated that direct instruction may be provided after exploratory study in order to get long-term lasting learning results, supports this sequence, demonstrating that novices may improve their performance before being presented direct instruction.

On the one hand, Schwartz and Martin (2004) compared invention activity with tell and practice activity in the learning of statistics. Students in the invention activity condition attempted to construct relevant statistics formulae first and then were presented a lecture, whereas, in tell and practice condition, the teacher directly instructed students about the core concepts. Results revealed that the invention activity condition outperformed tell and practice condition in the immediate post-test and delayed test. Similar results were obtained by Schwartz and Bransford (1998). They randomly divided participants into three conditions, analyzing activity followed by a lecture, analyzing activity followed by another analyzing activity and summarizing activity followed by a lecture, to learn the concept of schema in Psychology. The analyzing activity followed by a lecture was superior to the other two conditions.

On the other hand, the study of productive failure provided further evidence of the advantage of problem solving followed by direct or delayed instruction (Kapur, 2010, 2011, 2012; Kapur & Bielaczyc, 2011; Westermann & Rummel, 2012). In the experiments, students worked as a group in a generation and exploration phase to study and solve some ill-structured problems first and then presented their results at a later lecture, compared to students who received worked examples or a lecture first. Consistent results were found that collaborative problem solving activities followed by delayed direct instruction from teachers outperformed those in the direct instruction condition including worked examples or a lecture presented by teachers first.

These results directly contradict the results of research comparing a worked example/problem sequence with a problem/worked example sequence. It might be noted that the experiments on invention activities and productive failure consistently violate a basic principle of randomized, controlled experimental designs. They routinely alter multiple variables simultaneously rendering it impossible to determine causality.

A question is whether different possible types of worked examples would produce equally positive effects or whether there should be preferred types of worked examples in specific conditions. The following sub-sections will discuss research studies into the effectiveness of different types of worked examples. The next chapter will also discuss some suboptimal designs of worked examples based on other cognitive load effects.

3.4 Types of Worked Examples with Their Effectiveness

3.4.1 Worked Examples with Sub-goal Structures

Worked examples provide professional solutions to facilitate students' learning. However, if students face a novel problem which requires many changes of steps learned from a previous worked example, but sharing the same goal or structure with that worked example (Catrambone, 1995), students will not be able to solve this novel problem smoothly. Students have a tendency to memorize a set of steps learned from a previous worked example rather than learning meaningful representative structures from that worked example. Catrambone (1998) suggested that if students could form a hierarchical structure that contains different levels of solution procedures, then this kind of structure may facilitate their abilities to solve novel problems. Higher and lower levels of knowledge are connected with each other by casual or other kind of links (Gentner, 1983). When students solve a problem, the higher level of a solution procedure can be broken down to a lower level of procedure and then again broken down to relatively even lower levels. Sub-goals represent a kind of knowledge associated with this hierarchical structure. Dufresne, Gerace, Hardiman and Mestre (1992) and Eylon and Reif (1984) suggested that students who formed a hierarchical goal structure solved novel problems more successfully than students who memorized a step-by-step worked example. Therefore, if we use worked examples with sub-goal structures, it may make students more successful in solving novel problems.

There are at least two ways to explain what a sub-goal is. For example, Newell (1990) and VanLehn (1988) explained a sub-goal as the goal generated by students when they arrive at a dead-end when solving problems. A sub-goal can also be regarded as a task structure which can be taught to a learner (Catrambone & Holyoak, 1990; Dixon, 1987). In this chapter, the sub-goal structure used with worked examples is considered as a meaningful conceptual piece of an overall solution procedure (Catrambone, 1998).

Catrambone (1995) tested worked examples with sub-goal structures within the domain of probability. In his experiment, the problem required students to use the Poisson equation to predict the average number of suitcases owned per lawyer. In order to find the average, students needed to firstly calculate the total probability number. Students were randomly assigned to two conditions: either presented with a worked example including sub-goal which was labeled the step used to calculate the total probability number or presented with a worked example without a sub-goal. Results indicated that students presented with a worked example including a sub-goal mentioned the sub-goal more frequently than students who studied a traditional worked example. Students in the worked example with sub-goal condition (labeled steps used to calculate the total probability number) more successfully predicted the average suitcase owned per lawyer. The students in the traditional worked example condition frequently mentioned that there was not enough information to solve this problem. This result indicated that the labeled steps which were used to calculate the total probability number helped students to learn the structure of the worked example which allowed them to successfully calculate the average number for that problem. Another method to form sub-goals used in Catrambone's experiments was visual isolation. Catrambone (1995) compared a condition in which the worked example integrated the steps for calculating the total probability number together with other steps with the condition of separately circling the steps for calculation of the total probability number as a sub-goal separate from the other steps. Results indicated that the separated sub-goal format was superior to the integrated format for learning.

Catrambone (1998) further tested the above model by using students with different levels of math background and worked examples with different types of label cues. In other words, the effects of an interaction between levels of learner prior knowledge and the types of label cues on transfer were investigated. Final results indicated that using labels (abstract or superficial cues versus no cues) to induce learners to form sub-goals helped them in solving novel problems. Moreover, learners with a strong math background were more likely to form suitable sub-goals from the abstract label cues requiring more mental effort, whereas learners with a weak math background seemed to be more reliant on the semantic content of the label to form a goal which was superficial and more restrictive. The abstract sub-goal was more associated with the structure of a problem rather than the surface features of a problem.

Implications of Worked Examples with Sub-goal Structures. Students tend to memorize a set of steps presented in worked examples and then may have difficulties in successfully solving novel problems. Catrambone (1995, 1998) suggested a kind of worked example that included sub-goal structures. Catrambone (1998) indicated that novel problems in a specific domain might share the same structure with sub-goals similar to those in the problems used in worked examples, but the way to achieve these sub-goals in novel problems could be different. Compared to remembering a series of steps presented in a conventional worked example, studying worked examples with sub-goal structures could reduce the search space when learners solve novel problems and consequently reduce working memory load, as the search space is constrained by each sub-goal.

Studying worked examples with sub-goal structures helps learners to build hierarchical structures or schemas that contain different levels of solution procedures, which facilitates transfer. In addition, using different types of cues and considering different levels of learner expertise also have effects on transfer. Abstract cues could be used for more experienced learners, while superficial cues should be used for less experienced learners. To sum up, when we design a worked example, the relationships between steps should be considered in order to help learners form suitable sub-goals. The worked example should provide some casual or explanatory links between solution steps to help students categorize the sub-goals.

3.4.2 Worked Examples with Self-explanations

Self-explanations were proposed by Chi, Bassok, Lewis, Reimann, and Glaser (1989). This phenomenon demonstrates that learners have a tendency to self-find reasonable explanations for each step which is indicated in a worked example and then learners who try to self-explain reasons for each step have a better performance than those who do not (Chi et al., 1989). Initially, Chi et al. (1989) suggested a model of self-explanations that the learner generated the hidden information missing from a presented worked example. Later, they amended their model by considering self-explanations as a dual process: one process is to generate the hidden information missing from the presented worked example; another process is to complete learners' original schemas (Chi, 2000). Renkl (1997) considered the characteristics of leaners who self-explained and then indicated the categories of learners who were successful in using self-explanations. One kind of learner is principle-based explainers who have a relatively low level of expertise and elaborate on the principle of each step. Another kind of self-explainer is anticipative reasoners who have relatively high levels of expertise and anticipate the next step in an example and then check whether it corresponds to the actual step.

Initially, self-explanations were not incorporated in cognitive load theory research. Later, within the framework of cognitive load theory, Clark, Nguyen, and Sweller (2006) defined self-explanations as a dialogue between students and a worked example which could facilitate the study of this worked example and help students form a schema relevant to this worked example. Within the context of cognitive load theory, self-explanations require students to establish the interactions among elements of a worked example as well as between those elements and learner prior knowledge (Sweller et al., 2011). In addition, from the view of human cognitive architecture, self-explanations may only be suitable for learners with high levels of expertise. For novices who do not have relevant principles stored in the long-term memory, self-explanations may make them randomly generate potential principles related to each step of a worked example, according to the randomness as genesis principle. In order to process those interactive elements generated by self-explanations, learners need to use the resources of working memory. Therefore, a high cognitive load may be imposed on novices' working memory. Learners who have sufficient knowledge stored in long term memory can directly retrieve relevant knowledge (principles) from the knowledge base via the environmental organizing and linking principle, which can reduce working memory load. Based on this view, self-explanations may be more suitable for learners with relatively higher levels of expertise.

Research about combining worked examples with self-explanations has revealed positive effects on learning. Chi et al. (1989) indicated that learners could process a worked example more deeply by self-explaining it and learn more than those who only deal with the surface features of the worked example. A similar study of Chi, Leeuw, Chiu, and LaVancher (1994) demonstrated that studying text with self-explanations could facilitate knowledge acquisition.

Renkl, Atkinson, and Maier (2000) investigated a fading procedure that combined the worked example with problem solving. Firstly, a complete worked example was given;

secondly, a step of this worked example was omitted and then more steps were omitted until the worked example was changed to a conventional problem. Results of their experiment indicated that the fading procedure had positive effects on near-transfer tasks, but no effect on far-transfer tasks (Renkl, Atkinson, Maier, & Staley, 2002). However, when the fading procedure was combined with self-explanation prompts, the final results showed a positive effect on far-transfer tasks (Atkinson, Renkl, & Merrill, 2003). Similar results were obtained by Van Merriënboer, Kester and Paas (2006). They mentioned that teaching complex tasks with self-explanation prompts could improve transfer performance due to balancing intrinsic load and germane resources.

Booth, Lange, Koedinger, and Newton (2013) compared the effect of guided practice with worked examples with the effect of guided practice alone. Participants in the guided practice with worked examples condition were prompted to explain the reason why this step was used. The final results indicated that guided practice with worked example which included prompted self-explanations was superior to guided practice alone in the acquisition of conceptual knowledge and without sacrificing procedural knowledge.

Two kinds of self-explanations were compared by Renkl, Stark, Gruber, and Mandl (1998). In that study, bank apprentices were required to study the mathematics procedures used in finance. They were divided into two groups: one group was required to verbalize their thoughts concurrently (spontaneous self-explanations), while the second group (elicited self-explanations) received instructions on how to generate self-explanations. Final results demonstrated superior performance of the elicitation condition on both near and far transfer tests.

However, the effect of worked examples with self-explanations on learning is not always positive. Mwangi and Sweller (1998) did not find positive effects of worked examples with self-explanations on two-step arithmetic word problems. The explanation for this failure was that working memory load was increased when learners translated mathematics procedures into verbal form (verbal principles). Similarly, Conati and VanLehn (1999, 2000) did not obtain positive results from learning from worked examples by prompting self-examinations in computer-based environment. Hausmann and Chi (2002) also indicated negative results by requiring students to type in their own written self-explanations within a computer-based environment.

Renkl (1997) investigated differences in self-explanations generated by different students. Worked examples of probability problems were given to college students who were asked to verbalize their thoughts about these worked examples simultaneously. The results revealed that there were qualitative differences in self-explanations between successful learners and less successful students. Successful students tended to provide principle-based explanations.

Explanation for the effectiveness of worked examples with self-explanations may be that students could produce a number of inference rules which could be connected to suitable actions associated with specific conditions, and then these rules could be turned to procedural knowledge. Chi (2000) extended this argument indicating that students generated inference rules and repaired their mental model.

Self-explanations also has been applied to other forms of study other than in worked examples. Roy and Chi (2005) summarized the application of self-explanations in multimedia learning and suggested a positive effect of self-explanations in multimedia learning. Rittle-Johnson (2006) compared direct instruction (showing how to solve math problems) with invention (presented problems first and then with feedback after) by using or not using self-explanations. Results indicated an advantage for the self-explanations condition.

Implications of Worked Examples with Self-explanations. Worked examples with self-explanations have been confirmed to be positive for learners' study. The self-explanation process requires learners to deal with the worked example more deeply compared to learners who do not self-explain worked examples, which facilitates knowledge acquisition. However, from the views of cognitive load theory, worked examples with self-explanations may also increase the interactive elements which must be processed in working memory simultaneously. Therefore, the self-explanations technique may be more suitable for learners with relatively higher levels of expertise (Renkl, 1997) and also the effectiveness of using worked example with self-explanation depends on levels of learners' expertise.

3.4.3 Product-oriented and Process-oriented Worked Examples

Product-oriented and process-oriented worked examples are relevant to worked examples with self-explanations. As discussed in the previous sections, worked examples with self-explanations may be more suitable for students with more prior knowledge. Some failures in finding a positive effect of worked examples with self-explanations indicated that novice students might not be able to provide adequate explanations for each step because of a lack of prior knowledge (Chi et al., 1989; Renkl, 1997). Process-oriented examples may solve this problem.

Product-oriented worked examples are typical examples which provide problem solutions only and focus on the results of effective task performance. Process-oriented worked examples are examples which not only present professional solutions, but also provide reasons behind each step, so they provide ways for novices to mimic experts' problem-solving procedures during training. According to the characteristics of these two kinds of worked examples, product-oriented examples provide strategic knowledge ("how" information), while, process-oriented examples provide principled knowledge ("why" information) as well as strategic knowledge. Therefore, students should benefit more from process-oriented worked examples during their initial stage of acquisition of cognitive skills (Van Gog, Paas, & Van Merriënboer, 2004). In addition, studying process-oriented worked examples at the training stage would help learners to construct cognitive schemas, which would produce better transfer performance (Van Gog, Paas, & van Merriënboer, 2008).

According to different constituent skills (recurrent or non-recurrent), Van Gog et al. (2004) suggested a way of how to design a process-oriented worked example. For tasks including only recurrent constituent skills (algorithmic and rule-based behaviour) that have a narrow problem space, it is possible to reach a solution by correctly applying operators. In contrast, tasks including non-recurrent constituent skills have multiple possible solutions. Therefore, it is important to embrace strategic information in worked examples to reduce search.

Van Gog et al. (2008) compared product-oriented worked examples with process-oriented worked examples in the domain of electrical circuits. Participants received two sessional interventions: in the first session, participants were presented with either product-oriented worked examples or process-oriented worked examples; in the second session, participants received either the same format of worked examples (e.g., product-product examples) or different formats of worked examples (e.g., product-process examples). Results suggested that participants who studied the process-oriented examples had higher efficiency on the first transfer test (following the first session) than those who studied product-oriented examples and participants in product-oriented worked examples invested more cognitive effort. For the second transfer test, the process-product condition had a higher efficiency than the process-process condition. However, the process-product condition did not have a significantly higher efficiency than the product-product and product-process conditions. There were no significant results found for mental effort ratings as well.

The effect of process-oriented and product-oriented worked examples was also investigated within the domain of complex cognitive skills. Van Gog et al. (2006) used process-oriented examples in troubleshooting tasks. Participants were randomly assigned to four training conditions: CP (conventional problem solving: no solution and no process information given); PCP (conventional problem solving with process information given); WE (worked example: solution given, no process information) and PWE (worked example with process information given). Results of training phase suggested that product-oriented worked examples (WE) required less time on training tasks; participants in WE and PWE invested less mental effort, but the conditions with process information given required more mental effort during training. The effect of process-oriented and product-oriented worked examples on near and far transfer tests suggested similar results: participants with solutions given were superior to those without solutions, but they spent more time on solving transfer tasks. In terms of mental effort, participants in conditions with process information given invested more mental effort. This research demonstrated that process information added to worked examples (process-oriented worked examples) might increase intrinsic load which could result in participants using more time and investing more mental effort.

Implications of Product-oriented and Process-oriented Worked Examples. Product-oriented worked examples provide strategic knowledge ("how" information) to learners, whereas, process-oriented worked examples include not only principled knowledge ("why" information), but also strategic knowledge to learners. According to research on self-explanations, students may not provide high-quality explanations for each step in worked examples, therefore, externally provided high-quality explanations (namely, process-oriented worked examples) may improve students' learning and transfer performance. Within the framework of cognitive load theory, process-oriented worked examples may increase germane resources, which could deepen students' conceptual understanding, induce understanding of the reasons behind each step and indicate how experts select a strategy (Van Gog et al., 2004). However, adding process information to examples may increase intrinsic load, which in turn increases the time-on-task and the investment of mental effort.

3.4.4 Worked Examples in Ill- or Less- structured Domains

Most of the previous research on the effectiveness of worked examples focused on using worked examples in well-structured domains, such as algebra (Sweller & Cooper, 1985), statistics (Paas, 1992), geometry (Paas & Van Merriënboer, 1994; Schwonke et al., 2009) and physics (Reisslein et al., 2006; Van Gog et al., 2011; Van Gog et al., 2006). Much less research investigated using worked examples in less-structured or ill-structured domains. Research using worked examples in well-structured domains suggests that learners (especially novices) who study worked examples have better performance than those who solve conventional problems at the initial stage of acquisition of cognitive skills, resulting in the worked example effect. Therefore, whether worked examples in ill- or less- structured domains also can result in worked example effect is another interesting topic to investigate.

Worked examples in a well-structured domain involve specific concepts, procedures, a clear goal and a clear solution path. Usually this kind of worked example results in a unique solution. Compared to worked examples in a well-structured domain, worked examples used in less-structured or ill-structured domains involve less certain concepts and procedures, as well as less clear goals and solution paths. Therefore, this kind of worked example has multiple possibilities and solutions (Jonassen, 1997; Van Merriënboer, 1997). A worked example in ill- or less- structured domains should be different to a worked example used in well-structured domains (Schworm & Renkl, 2007). However, the format of worked examples in ill- or less- structured domains should also share the characteristics of worked examples in well-structured domain, namely, they must provide professional solutions of the problems to students (Kyun, Kalyuga, & Sweller, 2013).

The research in example-based learning in music (Owens & Sweller, 2008), design history (Rourke & Sweller, 2009) and social psychology (Hübner, Nückles, & Renkl, 2010) can be regarded as using worked examples in ill- or less- structured domains. Rourke and Sweller (2009) demonstrated a worked example effect within an ill-structured task domain that involved recognizing design styles used by different designers. This was an ill-structured domain, as students needed to consider and combine many factors which determine the style of a designer. The experiments used participants with or without domain knowledge, and they were randomly assigned to two groups: a worked example group and a problem-solving group. There were three stages in the experiments: the first stage was a lecture about design history; in the second stage, the worked example group studied worked examples and solved paired problems, while the problem-solving group just received problem-solving practice; stage 3 was a test stage. The experiments demonstrated the worked example effect. However, only participants with domain-specific knowledge demonstrated the worked example effect for both near and far transfer tests.

Similar research involving worked example used in less structured or ill-structured domains was conducted in the area of reasoning in law education (Nievelstein, Van Gog, Van Dijck, & Boshuizen, 2013). Worked examples used in this experiment were associated with less-structured tasks. The participants who studied worked examples demonstrated better performance. Interestingly, in this research, with the increased levels of expertise of participants, worked examples did not become redundant.

The language learning area (such as English literature) is another ill-structured domain. Kyun et al. (2013) used Korean university students whose mother language was not English. Students were divided into two groups: one group received conventional essay questions with model answers (worked examples) and followed by similar questions to solve; another group was required to write conventional essays without examples. The results demonstrated a worked example effect in this ill-structured domain. However, the effectiveness of worked examples decreased with increasing levels of learners' expertise.

Owens and Sweller (2008) demonstrated the worked example effect in music learning which belongs to ill-structured domain. In the first experiment, four kinds of examples were provided: the first example explained the relations between note values and a crotchet beat; the second example showed the functions of lower and upper numbers (such as 4/4 time signature); the third and fourth examples provided explanations about the relationship of note values to quaver (eighth note) and minim (half note) beats respectively, with the sequence of materials similarly structured to the first worked example. These worked examples either combined the text with musical notation to reduce split-attention or separated them. The results demonstrated a worked example effect when the text was combined with musical notation. In the second experiment, auditory materials were used as experiment materials, and similar results were obtained.

Implications of Worked Examples in Ill- or Less- structured Domain. There is much less research on using worked examples in ill- or less- structured areas than in well-structured domains. The research discussed above provides some empirical evidence about the effectiveness of worked examples in ill- or less- structured domains. Although, the format of worked examples in ill- or less- structured domains is different from that discussed in well-structured domains, the worked example effect is still robust. Therefore, for well-structured or ill-structured domains, worked examples represent a tool which can facilitate learning.

3.4.5 Correct vs. Incorrect Worked Examples

Previous studies involving worked examples focused mostly on using correct worked examples, and very few studies considered the nature of the solution in the examples (Booth et al., 2013). Ohlsson's (1996) theory of learning from errors indicated that as the initial knowledge used early in learning to solve problems was very general, students usually could not choose the right way to solve problems. In order to make the right choice of a solution, students needed to know the errors they made and how the general knowledge induced them to make this kind of error. Then students would know what additional information should be added to make the solution correct. Similarly, the overlapping waves theory (Siegler, 1998) suggested similar ideas: learners know how to use various strategies (effective and ineffective strategies) in a given situation. With increased expertise, in order to replace the ineffective strategy, learners should understand what strategy is wrong and why it is wrong (Siegler, 2002). Therefore, it is reasonable to believe that using the incorrect worked examples should be effective for students' learning as well.

Some studies have indicated the benefits of using incorrect worked examples to facilitate students' learning (Durkin & Rittle-Johnson, 2012; Rittle- Johnson, 2006; Siegler, 2002; Siegler & Chen, 2008). These studies used a combination of correct and incorrect worked examples. Recent work suggests that if children are asked to self-explain a combination of correct and incorrect worked examples, it will be more effective than using the correct worked examples only (Durkin & Rittle-Johnson, 2012; Rittle- Johnson, 2006; Siegler, 2002; Siegler & Chen, 2008), similar results have been extended to older students (Huang, Liu, & Shiu, 2008) and adults (Curry, 2004).

Booth et al. (2013) investigated the effect of three kinds of worked examples (correct only, incorrect only and correct + incorrect) on the acquisition of conceptual knowledge of algebra. Results indicated that the inclusion of incorrect worked examples could be preferred: firstly, students performed better after explaining incorrect examples; secondly, the combination of correct and incorrect examples made students learn more conceptual knowledge, and finally, by studying incorrect worked examples, students improved the encoding of conceptual features in equations.

Some explanations of the effectiveness of using incorrect examples suggested that using incorrect examples could help students recognize and rectify incorrect procedures and then improve procedural knowledge. This is consistent with the view of Chi, Feltovich, and Glaser (1981), according to which learning accurate and deep features of the problems could help learners build their expertise, and incorrect examples could draw students' attention to particular features of the problems that make the procedures incorrect.

However, the effectiveness of using incorrect worked examples may not be observed for all learners. Große and Renkl (2007) found that incorrect examples were ineffective for relatively less experienced learners. The reason is that relatively novice learners may not be able to locate and identify the errors in examples, as they usually make the same errors when they solve similar types of problems alone. It seems that the research in using incorrect examples should also consider the expertise of learners.

Implications of Correct vs. Incorrect Worked Examples. Previous research used correct examples to demonstrate the worked example effect, and not many studies investigated the nature of solutions presented in worked examples. According to Ohlsson's and Siegler's theories, if students could know which solutions are wrong and why they are wrong, this knowledge may help them to narrow down their search for solutions and find the right strategies. The positive results of using combined correct with incorrect worked examples have supported the effectiveness of using incorrect worked examples. When teachers design their instructions, a combination of correct and incorrect examples may be another effective way to facilitate students' learning. When students learn incorrect examples, they may focus on the key features of the corresponding problems, which would be helpful in improving problem solving. However, the effectiveness of incorrect examples may be influenced by the level of learner expertise. Novice learners may have difficulties in identifying the inappropriate steps in such worked examples, as they tend to make the same mistakes when they solve the similar problems alone.

3.5 Other Issues Raised by Worked Example Research

3.5.1 Group Study or Individual Study

Most research has discussed the superiority of worked examples over problem solving for individual studies (Atkinson et al., 2000; Sweller, 1999), and only a few studies directly compared the worked example with problem solving in group study condition (Kirschner, Paas, & Kirschner, 2009), especially within a cognitive load framework. Kirschner et al. (2009) mentioned two cognitive load consequences of using worked examples in group study : firstly, in group study, group members share the intrinsic load of a task; secondly, the additional transaction cost would occur in group study, as communication between group members needs working memory resources. This transaction cost could be related to both intrinsic and extraneous types of cognitive load.

Direct research of comparing worked examples with problem solving in group study or individual study condition from a cognitive load perspective assumed that problem solving might result in better performance than worked examples in group study if intrinsic load will be shared between group members (Retnowati, Ayres, & Sweller, 2010). However, the experimental results indicated that problem solving combined with collaboration was not supported and worked example effect could possibly be extended to group study condition (Retnowati et al., 2010).

3.5.2 Worked Example and Cognitive Tutors

In cognitive load theory, worked example condition has been usually compared with problem solving condition which does not provide any external supports. Therefore, some researchers (Koedinger & Aleven, 2007; McLaren, Lim, Gagnon, Yaron, & Koedinger, 2006) deemed that the superiority of worked examples might be due to the fact that the control was the unsupported problem-solving group (Mwangi & Sweller, 1998; Zhu & Simon, 1987). Namely, the design of worked example and problem solving conditions could not be regarded as equivalent. Cognitive Tutors provide individual support for learning when students solve problems. Firstly, they select suitable problems; secondly, provide in-time feedback; and lastly, present hints. However, an important limitation of cognitive tutors is that they make students focus on problem-solving skills rather than deeper conceptual understanding (Schwonke et al., 2009). Schwonke et al. (2009) compared a cognitive tutor that included a sequence of worked examples with a traditional cognitive tutor which provided the usual forms of support but without worked examples. Final results demonstrated that students who received worked examples required less time and demonstrated deeper conceptual understanding. Therefore, the worked example effect is robust, even compared to well-supported problem solving conditions used in cognitive tutors (Schwonke et al., 2009).

3.5.3 Worked Example Effect and the Problem Completion Effect

There has been evidence showing that only when learners have difficulty in solving conventional problems, do they study worked examples in depth (Chi et al., 1989). In order to address this issue, Van Merrienboer and Krammer (1987) suggested that in order to make sure learners pay enough attention to the worked example, completion problems should be provided to them.

Completion problems provide explicitly some solution steps and then leave some key steps for learners to complete. For example,

$$2x+10=14$$

$$2x=14-10$$

$$x=?$$

Van Merriënboer (1990) conducted the first extensive research on the effectiveness of completion tasks within the framework of cognitive load theory using a computer programming course. Students were randomly assigned to two groups: the traditional strategy group, in which students were required to design and code new computer programs, and the problem completion group, in which students were required to modify and extend existing computer programs. The problem completion group was superior to the traditional strategy group according to the posttest results. Van Merrienboer and De Croock (1992) did a similar study using computer programming. Their results also indicated that problem completion was better than traditional problem solution.

Paas (1992) compared three conditions - worked example, problem completion and conventional problem solving - to investigate the problem completion effect. Results showed that worked example and problem completion groups were superior to a conventional problem solving group. However, the superiority of problem completion tasks may be limited to far transfer posttest tasks (van Merriënboer, Schuurman, de Crook & Paas, 2002).

3.5.4 New Design Format for Worked Examples

The usual format for using worked examples within the framework of cognitive load theory is worked example-problem solving pairs, which has been tested by a large number of experiments indicating positive effects on learning within a variety of domains. Miller (2010) suggested another format of worked example – a *three-step model of worked example* according to which a worked example contains three steps: 1. Providing a worked-out example; 2. Giving a problem to be discussed or solved by students in a small group; 3. Giving another similar problem to be solved by students alone or in a small group. This method which extends the traditional format of worked example studies was demonstrated to lead to better performance compared to the traditional format (Miller, 2010).

3.6 Summary of Chapter 3

Within the framework of cognitive load theory, the worked example effect has attracted more attention than other cognitive load effects, as it conflicts with the views of Problem-based learning (Sweller et al., 2011). Worked examples provide full guidance to learners, resulting in better results than a problem solving condition which does not provide guidance to learners. The worked example effect usually uses a worked example-problem solving sequence for the worked example group. A body of research has demonstrated that worked example-problem solving pairs have positive effects on learning performance, compared to problem solving only. Although the worked example effect has been obtained in a series of experiments, the conditions for obtaining the effect are also critical. The nature of the learning materials and the expertise of learners (see Chapter 5) may be two major factors influencing the effectiveness of worked examples. The nature of the learning materials was discussed by considering the levels of element interactivity. Based on the element interactivity effect, worked examples of problems used within the framework of cognitive load theory should be high in element interactivity. Another issue discussed in this Chapter concerned the types of worked examples with their effectiveness. From previous research, providing worked example with an explicit sub-goal structure, and using self-explanations can further facilitate learning. However, some concerns about using worked examples with self-explanations suggest that novices may not be able to effectively self-explain (Chi et al., 1989; Renkl, 1997). This problem may be solved by studying process-oriented worked examples which provide professional solutions with the reason for each step. Most research about testing the worked example effect has indicated this effect in well-structured domains with less research using worked examples in ill- or less- structured domains. Research also has used a combination of correct with incorrect worked examples indicating another way to effectively use worked examples.

Some other issues which have been raised when studying worked examples indicated that the worked example effect also could be obtained in group study which reduces the extraneous load rather than individual study only. Researchers also have found that using a problem completion format was effective for learning.

It has been argued that worked examples are only effective if unsupported problem-solving is used as a comparison. Comparing a cognitive tutor using worked examples with a traditional cognitive tutor that did not use examples revealed that the worked example effect was robust. However, Darabi, Nelson, and Palanki (2007) indicated that a worked example was not sufficient to help learners to build their own schema. The structure of a worked example may also influence learning effectiveness (Catrambone, 1994; Catrambone & Holyoak, 1990; Mwangi & Sweller, 1998; Ward & Sweller, 1990; Zhu & Simon, 1987). This issue will continue being discussed in Chapter 4 by considering other cognitive load effects which may affect the design of effective worked examples.

All in all, more experiments are needed to further investigate the effectiveness of worked examples discussed in cognitive load theory. However, based on current empirical results, teaching with worked examples can facilitate students' learning, especially novices' learning.

Chapter 4 Other Cognitive Load Effects that May Influence the Design of Effective Worked Examples

In Chapter 3, different kinds of worked examples and corresponding empirical evidence were discussed. However, not all worked examples were demonstrated to be always effective. A possible reason for some failures to obtain a worked example effect was that the internal structure or presentation formats of the examples precluded effective learning by generating extraneous cognitive load. Within the framework of cognitive load theory, there are a number of other cognitive load effects that are related to such sources of extraneous cognitive load, and therefore need to be considered when designing an effective and efficient worked example. Some researchers indicated that the suggestion of using worked examples as a substitute for problem solving had some controversies (Gabrys, Weiner, & Lesgold, 1993). Therefore, it seems that worked examples are not always the best substitute for problem solving when some factors are taken into consideration (e.g., the design of a worked example or levels of learner expertise - see Chapter 5). The design of a worked example is quite critical, especially the design of its structure (Catrambone, 1994; Catrambone & Holyoak, 1990; Mwangi & Sweller, 1998; Ward & Sweller, 1990; Zhu & Simon, 1987). Therefore, the following sections will discuss this issue by considering some other relevant cognitive load effects which may be related to the design of worked examples and the effectiveness of using worked examples.

4.1 The Split Attention Effect

Some worked examples might impose a heavy burden on students' working memory because of their suboptimal design (Mwangi & Sweller, 1998; Tarmizi & Sweller, 1988; Ward & Sweller, 1990). One such suboptimal instructional design discussed within the framework of cognitive load theory involves situations in which students are required to mentally integrate at least two separate sources of information in order to fully understand the whole task. This type of instructional situation could lead to split attention during learning which may impose high levels of extraneous cognitive load.

According to cognitive load theory, split-attention happens when students are required to split their attention between multiple sources of information that have been separated either spatially or temporally (Sweller et al., 2011). The next sections will firstly discuss split-attention caused by spatially separated information, and then split-attention raised by temporally separated information.

4.1.1 Split-attention Caused by Spatially Separated Information

Geometry learning can be used as an example to explain the split-attention effect in the situation where multiple sources of information are spatially separated. When we present students a geometry example, if the geometric diagram and the full solution are separate, students need to scan these two sources of information. Namely, students need to first hold information from worked out solutions in working memory and then to search the relevant information provided by the associated geometric diagram in order to understand the solution procedure. If the solution has many steps, students have to hold too much information in their working memory simultaneously and repeatedly search for relevant information provided by the geometric diagram, which may impose high levels of extraneous cognitive load and interfere with learning. Therefore, the split-attention effect occurs when learners who are presented with physically integrated materials have better performance than those presented with the corresponding split-source materials which require random search-and-match activities irrelevant to learning (Sweller et al., 2011).

In addition, the characteristics of multiple sources of information may also influence the effect. Those multiple sources of information should be unintelligible and unlearnable in isolation (Sweller et al., 2011). Therefore, the split-attention effect occurs under the condition when two or more sources of information must be processed together in order for learners to understand the whole learning material which is being presented (Sweller et al., 2011).

Chandler and Sweller (1992) obtained the split-attention effect with different kinds of worked examples in Numerical Control programming. One group was presented with a conventional split-source form of the example, while another group received the modified version in which text and picture were physically integrated. The modified group demonstrated significantly better performance than the conventional group. In another experiment, Chandler and Sweller (1992) also showed the existence of a split-attention effect in the traditionally used format for an experimental report- Sample, Materials, Procedures, Results and Discussion- which involves relevant sources of textual information. This format required readers to mentally integrate information from different sections, which naturally increased extraneous load. However, when the physically integrated format was presented to readers, the results indicated that the integrated format was easier to process with higher performance scores than the conventional split-source format.

The split-attention effect is not restricted to the above domains. Ayres and Youssef (2008) found a split-attention effect with economics materials that involved diagrammatic representations. Similar results were found in calculating tax liability (Rose & Wolfe, 2000): a physically integrated format of instruction and a decision aid for calculating tax liability resulted in better performance than a split-source format. The split-attention effect was also found in the domain of second language learning. Lee and Kalyuga (2011) investigated the effectiveness of different configurations for locating pinyin (a transcription system) and

Chinese characters in worked examples used to teach Chinese as a second language. They found that a vertical format placing pinyin directly above the Chinese characters reduced split-attention, while a horizontal arrangement of pinyin and Chinese characters required learners to perform search-and-match processes that increased extraneous cognitive load.

4.1.2 Split-attention Caused by Temporally Separated Information

Another kind of split-attention could be induced when multiple sources of information are presented at separate times rather than separate locations, which is called temporal split-attention. Baggett (1984) compared two versions of an instructional film. In the first version, the narrative was presented before (7, 14, 21s) the corresponding visual information or after (7, 14, 21s); another version presented visual and auditory information simultaneously. The better performance was observed from the simultaneous presentation format, which supported the temporal split-attention effect.

Mayer and Anderson (1991, 1992) provided further evidence about the temporal split-attention effect by comparing a group presented with a narrative before animation with a group presented with an integrated (concurrent) format of narrative and animation. The concurrent presentation resulted in superior performance demonstrating a temporal split-attention effect.

4.1.3 Ways to Manage Split-attention Sources

The most widely used way to avoid split-source presentations is to integrate multiple sources of information. An alternative method is to direct learners’ attention to the right information. Kalyuga, Chandler and Sweller (1999) directed learners’ attention to corresponding parts of an electrical circuit by clicking on the paragraph being studied. Results indicated better performance in comparison with a format in which text is presented below the diagram. Using visual cueing to reduce random search could also improve learning results. Tabbers, Martens and van Merriënboer (2000) used a red colour to indicate the parts of a diagram that corresponded to the written or spoken text. Combining segmentation of long text with signaling techniques is another way. Florax and Ploetzner (2010) segmented long text and labeled each of the smaller segments with the same number used to label the corresponding part of the diagram. They compared five conditions of worked examples: continuous text + unlabeled picture, segmented text + unlabeled picture, continuous text + labeled picture, segmented text + labeled picture and spatially integrated text and picture. The results of this study indicated that there was no difference in learner post-test performance between the worked examples with spatially integrated text and picture and those with segmented text and labeled picture but these two conditions resulted in better performance than the other three conditions. This research not only supported the traditional results discussed in cognitive load theory that (1) spatial proximity was very critical for the split-attention effect; (2) an integrated text and picture could reduce extraneous load and improve learning, but also indicated an alternative way of dealing with split-attention situations (using segmented text with a labeled picture, even when these two sources of information are separated).

Similar results were obtained by Clark and Mayer (2008) and Mayer (2005). As shorter text which includes fewer interactive elements may be low in element interactivity, students could process segmented information without overloading working memory instead of dealing with long, split-source information. Clark and Mayer (2008) and Mayer (2005) also labeled pictures in worked examples, which was regarded as a signaling technique and could be used to avoid the shortcoming of split-source materials. This method reduces searching between multiple sources of information, resulting in reduced extraneous load and improved learning.

Implications of the Split-attention Effect. The split-attention effect is directly relevant to the design of worked example. Based on the discussion above, split-attention presentation format which may involve spatially or temporally split-attention could eliminate any advantages of worked examples. Worked examples involving two or more sources of information should be designed in a way that these sources of information are physically integrated. The characteristics of multiple sources of information also affect the occurrence of the split-attention effect. The split sources of information should be related to each other, but not re-describe each other. Therefore, when designing a worked example involving multiple sources of information, the connections between these sources of information should be considered to decide whether they should be integrated (split-attention effect) or whether some of re-described (redundant) sources of information should be deleted from this worked example demonstrating the redundancy effect which will be discussed in the following section.

4.2 The Redundancy Effect

This effect looks similar to the split-attention effect, which also discusses the relations among multiple sources of information. However, the difference between these two effects is the characteristics of materials used to test these two effects. For the redundancy effect, the multiple sources of information can be understood separately without referring to other information. For example, a piece of information is presented in textual form which can be fully understood, but the same information is also explained in oral form, which is repeated. However, for the split-attention effect, multiple sources of information only can be understood by considering all the information together.

Within the framework of cognitive load theory, any information which is not required for learning is regarded as redundant information (Sweller et al., 2011). Based on human cognitive architecture, such information is likely to be processed by working memory and therefore, increases working memory load. A body of empirical evidence has shown negative effects of redundancy on learning.

The first study on redundancy within the framework of cognitive load theory used learning materials consisting of text and diagrams that could be understood separately (Chandler & Sweller, 1991). With the diagram which essentially re-described the information of the text in the domain of engineering, students who studied the separated text and diagram performed better than those who studied the integrated text with diagram. Assuming a learner who needs to learn a new computer program, the common way is to study the manual which requires the learner to focus on the keyboard and information displayed on the computer screen, which may incur split attention. However, if we use a modified manual which combines text with pictorial information of the keyboard and computer screen, the physical computer may be redundant for learners. Sweller and Chandler (1994) and Chandler and Sweller (1996) compared two versions of instruction: instruction in a modified manual which included text and pictorial information of a keyboard and computer screen and instruction using the same modified manual plus access to a computer. The results of experiments indicated that the modified manual only group was superior to the modified manual plus access to computer group. Subjective ratings of cognitive load also indicated that a lower cognitive load was found with the modified manual only group rather than the modified manual plus access to computer group. Therefore, we could conclude that performance would deteriorate when the same information was presented redundantly in a manual as well as on the computer screen, namely, it mattered with whether the same information was re-presented as redundant information.

Like the split-attention effect, some factors may moderate the redundancy effect. One of these factors is the independence of information sources. Under the split-attention effect, if different sources of information are separated, meaningful learning will be inhibited. However, according to the redundancy effect, if students are presented simultaneously text and the diagram which re-describes the information of the text, learning will be interfered with, as redundancy is triggered. The redundant information requires more working memory resources to process and consequently, fewer working memory resources are left to process information which is relevant to study.

Another factor influencing the redundancy effect is level of element interactivity. If materials are low in element interactivity, it seems that no significant benefits will be obtained if redundant information is deleted. When learning materials are low in element interactivity (with a low intrinsic load), even if the extraneous load is high, the total working memory load could still be within the capacity limits of working memory. However, if students are presented with learning materials that are high in element interactivity (with high intrinsic load), redundant information that would increase extraneous load, and then working memory may be overloaded and learning may be inhibited.

In multimedia learning, the redundancy effect may occur when the same explanations of a picture or animations are presented in both visual (on-screen text) and auditory modalities. In this case, the pacing of presentations might moderate the redundancy effect. System-controlled pacing restricts the time period during which learners need to locate the relevant visual information mentioned by auditory form within the presented text, resulting in increased levels of extraneous load. On the other hand, under learner-controlled pacing, learners can review information at their own pace. Kalyuga, Chandler and Sweller (2004) presented written and auditory versions of the same information to students either with system-controlled pacing or learner-controlled pacing. The two forms of information were presented simultaneously, or were presented sequentially. The results showed that better performance of post-test scores and lower ratings of cognitive load were found testing students who were presented with written and auditory information sequentially under system-controlled pacing. More research further discussed the issue of audio-visual redundancy, Mayer, Heiser, and Lonn (2001) indicated that university learners who studied narrations with animations simultaneously were superior to those who were presented animations with narrations and on-screen text which repeated the content of narrations. Similar results were obtained by Jamet and Le Bohec (2007) by teaching the development of memory models. They compared three groups: one was presented diagrams with spoken information, one group was shown the same diagram and spoken information with written explanations labeled on the diagram sentence by sentence, the other group was shown the same diagram, and spoken information with written explanations labeled on the diagram as a whole text. Results revealed that the written text was redundant, leading to the first group being superior to the other two groups.

Lastly, the length of instructional segments may also be a factor that affects the redundancy effect. Kalyuga et al. (2004) found a redundancy effect when presenting lengthy textual materials. However, Moreno and Mayer (2002) indicated a reverse redundancy effect. They found that when there were not any visual diagrams involved, the same verbal information which was presented simultaneously in written and auditory form could result in better performance than auditory-only form. The analysis of materials used in these two experiments indicated that Moreno and Mayer (2002) used lengthy materials with appropriate breaks that provided time for learners to partially construct or consolidate mental models relevant to each fragment before moving to the next segment. However, the materials used in the experiment of Kalyuga et al. (2004) had no breaks within the text. Therefore, if learners have sufficient time to deal with smaller textual segments, presenting visual and auditory information simultaneously might eliminate the redundancy effect, as no high extraneous load would be imposed on working memory, and could even improve learning.

Implications of Redundancy Effect. This effect implies that worked examples with redundant information will produce negative effects on learning. Interestingly, this effect is counterintuitive (Sweller et al., 1998), as many people believe that the same information repeatedly presented in different formats will be beneficial to students. However, empirical evidence obtained on many occasions has revealed that redundant information may impose a high extraneous load and interfere with learning. If such information is presented in a split-source form, learners can possibly ignore the redundant information. However, if redundant information is presented to learners in a physically integrated form, they have to process the redundant information and unnecessarily allocate additional working memory resources, thus increasing cognitive load. Similar to the split-attention effect, the redundancy effect also involves more than one source of information, but the characteristics of multiple sources of information are different from those causing the split-attention effect. Multiple sources of information causing the redundancy effect should be individually intelligible, namely, learners can understand the whole material by studying individual sources of information, and the other presented information is redundant for learning. Therefore, when designing worked examples, we should not include redundant information, especially in physically integrated format.

4.3 Variability Effect

Unlike extraneous cognitive load which must always be reduced, intrinsic cognitive load should be optimized (Sweller et al., 2011). Within cognitive load theory, researchers have investigated how to structure a series of worked examples to foster transfer of learning. One of the suggested methods is by using varied context examples resulting in the variability effect which increases intrinsic load to enhance transfer (Clark et al., 2006). The variability effect is based on example-based instruction using highly varied worked examples to improve transfer performance compared to the use of low varied worked examples. Two relevant factors should be considered when using varied worked examples: firstly, the interactive elements of the corresponding task could be increased resulting in an increased intrinsic cognitive load; secondly, learners should have sufficient working memory resources to deal with this increased cognitive load. The rationale for using varied context examples is that learners who are exposed to such examples could presumably distinguish the relevant and irrelevant features of worked examples (van Merrienboer & Sweller, 2005), leading to the acquisition of conditionalized schemas that make learners know under what conditions to apply corresponding principles (Clark et al. 2006).

The first experiment conducted to investigate the variability effect within the framework of cognitive load theory compared low and high variability examples describing the application of some geometry principles (Paas & van Merrienboer, 1994). An example of the task used in this study was applying Pythagoras' theorem to calculate the distance between two points. The low-variability examples only changed the values, whereas the high-variability examples changed the values as well as the problem structure. The results indicated that the high-variability worked examples were superior to low-variability worked examples on transfer performance, demonstrating the variability effect. A similar experimental design was used by Quilici and Mayer (1996) to investigate the variability effect within the domain of statistics. Two groups were compared: one group received worked examples that varied surface features of problems, whereas, the second group studied worked examples that varied structural features of problems. Results revealed that students who studied varied structure worked examples were better on sorting task performance based on structure of problem. Therefore, a successful categorization could be improved by presenting students with increased structural variability rather than surface variability.

Another form of variability is associated with contextual interference, which is relevant to the sequence of problems. Low contextual interference is defined as a series of problems which could be solved by using the same set of skills; high contextual interference is related to a sequence of tasks that require different skills but are positioned next to each other (Sweller et al. 2011). Van Merrienboer et al. (2002) used the following format to explain the low and high contextual interference situations: for example, if A, B and C were three kinds of tasks that needed different skills, the sequence of B-B-B, A-A-A, C-C-C belonged to low contextual interference, while the sequence of C-B-A, B-A-C, B-C-A belonged to high contextual interference.

De Croock, van Merriënboer, and Paas (1998) investigated low and high contextual interferences by requiring learners to troubleshoot system failures. The results indicated that performance of students in the high contextual interference group was better than performance of those in the low contextual interference group. However, the mental effort analysis revealed no differences between low and high contextual interference. Van Merriënboer et al. (2002) had suggested that the tasks used by De Crook et al. (1998) were low in element interactivity, so they used more complex tasks relevant to programming skills to compare the effects of low and high contextual interference conditions. It was demonstrated that more time and mental effort were needed for the high contextual interference group than the low contextual interference group, which supported the assumption that high contextual interference tasks would generate more cognitive load (De Crook et al., 1998). The most important result was that the increased intrinsic load due to increased variability produced fewer errors on transfer test, which could be explained by increased germane resources (Sweller et al., 2011).

Implications of Variability Effect. Transfer ability is regarded as an important aspect within the framework of cognitive load theory. Using varied context or high-variability examples is a way to foster transfer skills. Although element interactivity and intrinsic load are increased with the increase of variability, learners are provided chances to experience a greater range of problems, which is helpful in building flexible and well-connected schemas to solve new problems (van Merrienboer & Sweller, 2005). However, two conditions under which the increased variability of worked examples is effective should be addressed. Firstly, the learners should have sufficient working memory resources to deal with an increased intrinsic load. For example, the students with low levels of prior knowledge may not benefit from varied context examples, as they may not have more working memory resources to handle the increased intrinsic load. Secondly, the task itself could be added more interactive elements to increase the intrinsic load. Therefore, based on the variability effect, teachers may design a series of worked examples including low-variability worked examples to teach students first, and then with the increase of learners' expertise, high-variability worked examples could be designed to help students form flexible schemas to distinguish relevant and irrelevant features of worked examples, which could facilitate their transfer skills.

4.4 Summary of Chapter 4

In this Chapter, some cognitive load effects which may affect the design of worked examples were discussed. The split-attention effect indicates that multiple sources of information that cannot be understood in isolation should be physically integrated to reduce extraneous cognitive load. In order to use this effect in the design of worked examples, we should ensure that multiple sources of information do not re-describe each other and so lead to the split-attention effect. It implies that any information that is not required for study is redundant and could impose extra working memory load. Therefore, this kind of information should be deleted to reduce extraneous load. Similar to the split-attention effect, the redundancy effect also involves multiple sources of information, but the difference is that these multiple sources of information are individually intelligible as they re-describe each other. Similar to other instructional theories, cognitive load theory has also considered how to enhance transfer skills by studying worked examples. Varied context examples that increase intrinsic load leading to an increased requirement of germane resources, let learners process worked examples more deeply and form conditionalized schemas to inform them when to use relevant principles, thus enhancing their transfer skills. However, the design and use of varied context examples should take into account the characteristics of learners (they need to have sufficient working memory resources) and the task itself (increased variability will add more interactive elements).

According to these cognitive load effects, if we need to present multiple sources of information to students in a worked example, we should consider whether these multiple sources of information are individually intelligible, and then decide whether they should be physically integrated or deleted. Specifically, if textual information and diagrams are essential and do not re-describe each other, we should design the corresponding worked example by physically integrating these sources of information; otherwise, one of them should be deleted to avoid redundancy. In order to foster transfer skills, high-variability worked examples could be used to provide a greater range of problems to learners to form flexible and well-connected schemas.

Chapter 5 Expertise Reversal Effect in Example-Based Learning

In Chapter 4, cognitive load effects which may affect the effectiveness of worked examples were discussed. As mentioned in Chapter 3 and 4, the characteristics of learners may also influence the effectiveness of worked examples. As an example, the expertise of learners may determine the effectiveness of worked examples. This Chapter 5 discusses this issue in detail.

The expertise reversal effect is regarded as a form of the redundancy effect (see Chapter 4). The information that is beneficial to novices becomes redundant to more experienced learners. Therefore, the expertise reversal effect focuses on the interaction between the characteristics of learners (levels of learner's expertise) and the characteristics of learning tasks, or we can say that this effect is relevant to an interaction between a basic cognitive load effect (redundancy effect) and levels of learner's expertise. Kalyuga and Renkl (2010) mentioned that the expertise reversal effect mainly focused on the role of learner knowledge and discussed the relationship between the expertise of learners and the effectiveness of instruction: effective instructional methods that reduce extraneous load for novices become ineffective or even hinder the learning of knowledgeable learners (Kalyuga, 2007). Two formats of this effect have been investigated: An ordinal interaction (the instruction is effective for novices, but has no effects on more experienced learners) and a disordinal interaction (the instruction is effective for novices, but has negative effects on more experienced learners) (Nievelstein et al., 2013).

5.1 Human Cognitive Architecture and the Expertise Reversal Effect

Knowledge structures (schemas) play an executive role in complex cognitive processes by guiding learners' attention and performance. Learners who do not have relevant knowledge stored in long-term memory need externally explicit instructions as knowledge-based executive function, such as worked examples used in instruction. If external guidance is not provided, novices may have to randomly generate solutions (Randomness as Genesis Principle), which will occupy working memory resources, leaving few resources available for learning. More experienced learners, on the other hand may not need external instructions, as they already have relevant schemas in long-term memory. If we present information which they already know, they need to reconcile the external information with the long-term memory knowledge, which will impose an extraneous load; finally, for intermediate levels of learners, external instructions function when they deal with unfamiliar information, whereas knowledge stored in long-term memory is used when they process familiar information.

Therefore, from the perspective of human cognitive architecture, we can conclude that instructions which are suitable for novices (as external instructions compensate their incomplete knowledge base) may be ineffective or even detrimental to more experienced learners (as presenting redundant information will generate an extraneous load).

5.2 Empirical Evidence for the Expertise Reversal Effect

As the expertise reversal effect considers the factor of learner's expertise, it should be discussed in relation to different knowledge levels of participants. Research demonstrating the expertise reversal effect can be divided into two categories: longitudinal studies, which use the same participants trained from novices to experts after several training sessions; or cross-sectional studies, which chooses participants who differ in levels of expertise.

5.2.1 Expertise Reversal Effect Studies in Technical Domains

The expertise reversal effect was initially investigated in a series of longitudinal studies by intensively training groups of technical apprentices from novices to experts in the domain of engineering (Kalyuga, Chandler, & Sweller, 1998, 2000, 2001). Experiments using electrical wiring diagrams as instructional materials revealed an expertise reversal effect by employing split-attention situations (Kalyuga et al., 1998). In these experiments, the format of sections of text integrated with diagrams was compared with a diagrams-text separated format. Results indicated that the integrated format was superior to the diagrams-text separated format for novices, but after a period of training, the effectiveness of the integrated format deceased compared to the increasing effectiveness of the diagrams-text separated format. Subjective ratings of cognitive load further supported the hypothesis that diagrams alone were easier to be processed by more knowledgeable learners, whereas, the integrated format was more suitable for novices who needed external textual instructions to understand the presented diagrams. The results were in line with the expertise reversal effect. With the increase of learner's expertise, textual information which had been beneficial to novices became redundant for more knowledgeable learners.

The following experiments of Kalyuga et al. (2000, 2001) and Kalyuga, Chandler, Tuovinen, and Sweller (2001) provided more evidence of the expertise reversal effect. In mechanical engineering, novices benefited more if narrated explanations used to explain how to use specific diagrams were presented together with relevant animated diagrams. However, integrating narrated explanations with animated diagrams interfered with study after novices had received a series of intensive training sessions which developed their expertise in the relevant domain. Finally, for these more knowledgeable students, diagrams alone were superior to the diagrams with narrations format.

In another study, worked examples that were used to instruct learners in how to program industrial equipment were superior to conventional problem solving initially. However, this result was reversed after learners received intensive training (Kalyuga et al., 2001). Another experiment of Kalyuga et al. (2001) compared a worked example group with a problem solving group using instructions in writing Boolean switching equations for relay circuits. At the start of the experiment, learners did not have sufficient knowledge. However, after learners received more extensive training, the problem solving group outperformed the worked example group as guidance was redundant for more knowledgeable learners, indicating the expertise reversal effect.

Kalyuga et al. (2001) obtained a full expertise reversal effect when they compared worked examples with exploratory instructions in writing switching equations for relay circuits. The results demonstrated that worked examples had been superior to exploratory instruction at the beginning, but with training, the situation was reversed as an increase in learners’ expertise, demonstrating that exploratory instruction was better than worked examples for more knowledgeable learners.

Brunstein, Betts, and Anderson (2009) used algebra-like questions in the Cognitive Tutor and found that with sufficient practice, minimal guidance was better than full guidance; with less practice, the situation was reversed. This result was in line with the expertise reversal effect. For more experienced learners, more external explanation will be redundant, resulting in an extra cognitive load.

5.2.2 Cross-sectional Studies

Some evidence about the expertise reversal effect comes from a body of cross-sectional studies. The next sections will discuss this issue within different domains.

Expertise Reversal Effect in Science Instruction. Kalyuga and Sweller (2004) investigated the expertise reversal effect in studying coordinate geometry. Based on pre-test scores, participants were divided into two groups: relatively more experienced learners and relatively less experienced learners, and were randomly assigned to a worked example group or a problem solving group. A post-test indicated an interaction of instructional formats and learner's expertise. Less knowledgeable learners benefited more from the worked example format with the opposite result found for more experienced learners.

By using scientific texts and pictures, Seufert (2003) provided a partial expertise reversal effect by investigating the effect of direct support (providing specific guidance), indirect support (using questions which did not include specific guidance) and no help on students' performance. There were no differences among these three groups for more experienced learners, but for less experienced learners, both direct and non-direct support groups were significantly better than the no-help group, furthermore, the direct-support group was superior to the non-direct support group.

In the domain of biology (McNamara, Kintsch, Songer, & Kintsch, 1996), students were presented with learning materials which were highly coherent. According to results, this kind of learning materials was more suitable for less-knowledgeable students rather than for more experienced students, as such materials were redundant for them (more experienced learners already have relevant knowledge in their long-term memory). The interaction between instructional formats and learner's expertise revealed an expertise reversal effect.

Expertise Reversal Effect in the Domain of Social Science Subjects. The expertise reversal effect also has been found with ill- or less-structured domains. Kyun et al. (2013) used participants with different levels of prior knowledge studying English literature. The knowledge level of participants in three experiments gradually decreased. The general experiment results found that significant advantages of worked examples on post-tests gradually appeared as the knowledge level of learners decreased. There were no differences for retention, near or far transfer tests for the most experienced learners in Experiment 1; in

Experiment 2 (relatively less knowledgeable learners), the worked example group was superior to the problem solving group on the retention test only; in the last experiment, the participants were the least knowledgeable, and results revealed that the worked example group was significantly better on the retention test and marginally superior on the near and far transfer tests. The results were in line with the expertise reversal effect.

Another experiment in the area of English literature also demonstrated an expertise reversal effect. Oksa, Kalyuga, and Chandler (2010) compared two instructional formats used in studying Shakespearean plays. One group received the material which combined Modern English explanations with Shakespeare's original old English line by line; another group was presented with the materials which put the Modern English explanations as footnotes or notes. The participants, who were less knowledgeable about Shakespearean plays, demonstrated better performance with an integrated format, whereas, for the participants who were Shakespearean experts, the separated format was better.

However, some studies in other less-structured domains did not find expertise reversal effect. Nievelstein et al. (2013) compared novices and experienced law students on reasoning legal cases. Experiment results indicated that for more knowledgeable learners, worked examples did not have any negative effects, namely, the expertise reversal effect was not found for more knowledgeable learners by using worked examples. The explanation may be that less-structured tasks need a longer time to master in order to form relevant schemas. Similarly, in art education, Rourke and Sweller (2009) compared studying worked example with problem solving by asking different levels of university students to identify a designer's style. In their experiments, the first year and the second year university students were used as participants with the former ones considered as less knowledgeable in visual literacy skills and the history of design. Results consistently demonstrated the benefits of studying worked examples for both the first and the second year university students, demonstrating the worked example effect. There was no expertise reversal effect obtained even though the level of expertise was increased between experiments. It was suggested that the second year students did not have critical domain-specific knowledge which could be used to solve relevant identify problems, so worked examples were not redundant for these students. More research needs to be done by testing less-structured problems to further investigate the expertise reversal effect in such domains.

5.3 Other Issues Raised by Expertise Reversal Effect Research

5.3.1 The Expertise Reversal Effect and the Isolated Elements Effect

The isolated elements effect indicates that initially presenting a set of isolated elements of information rather than whole complexes of interactive elements in instructional materials may reduce excessive intrinsic load. A disadvantage of this strategy is that the information is in the form of isolated elements and students cannot learn the relationships between the isolated elements initially. However, these isolated elements could help learners form partial schemas at the outset and then form the whole schema at the following phase after receiving instructions in the relationships among those isolated elements (Pollock et al., 2002). This isolated-element strategy also complies with the expertise reversal effect.

Blayney et al. (2010) compared two instructional formats: one used an isolated-interactive elements format; another one used a fully interactive elements format. This experiment was in the domain of accountancy with university students as participants. The results revealed an expertise reversal effect by considering different formats of interactive elements (isolated or not). Novices benefited more from an isolated-interactive elements format, compared to more knowledgeable learners who demonstrated superior results after studying with a fully interactive elements format. In this experiment, if more knowledgeable learners were presented with an isolated-interactive elements format, they needed to integrate those isolated simple elements with their knowledge base, which required extra working memory resources, interfering with learning.

Therefore, an instructional sequence from an isolated elements format to fully interactive elements format is beneficial for less-knowledgeable learners. More knowledgeable learners do not need to process information in an isolated form, as they can fully use their long-term memory knowledge to deal with a fully interactive elements format.

5.3.2 The Expertise Reversal Effect and the Variability Effect

Studying highly variable worked examples rather than worked examples with similar characteristics is more beneficial, demonstrating the variability effect (see Chapter 4). The effect of learner's expertise on grouping problems by structural variability and surface variability has been tested.

Scheiter and Gerjets (2007) investigated the effect of presenting algebra word problems based on surface features or via structural features by using students with different levels of expertise. Considering surface features, problems were either motion problems or finance problems; in terms of structural features, problems in a particular category required the same kind of algebraic equation to obtain a solution but different equations for problems in a different category. Scheiter and Gerjets (2007) divided students into two groups: students who were less-knowledgeable about sorting problems via structural features and those who were more knowledgeable. The results of the experiment indicated that less-knowledgeable students showed superior performance when presented algebra word problems based on surface features, as they needed to learn how to structurally group problems, whereas, more knowledgeable students were better if problems were grouped based on structural features. More-knowledgeable learners did not need to learn how to group problems by structural features. Being taught how to distinguish problems by structural features was redundant for them. This research was in line with the expertise reversal effect.

5.3.3 Pre-training with the Expertise Reversal Effect

Using a worked example or isolated-element format could be regarded as a kind of pre-training. The expertise reversal effect was obtained with both these instructional methods. Research has also indicated an interaction between other forms of pre-training and levels of expertise.

Clarke, Ayres, and Sweller (2005) compared a sequential format with a concurrent format for using a spreadsheet to learn mathematics with students at different levels of expertise. Students in the sequential format received training about using spreadsheets first and then applied this knowledge to learn mathematical concepts; students in the concurrent format received instructions about how to use a spreadsheet and how to apply the spreadsheet to learn mathematical concepts simultaneously. Results showed that students with less knowledge about using a spreadsheet learned mathematics more effectively with the sequential format, whereas, students who were familiar with the use of a spreadsheet benefited more from the integrated format.

Research discussed in Chapter 3 also indicated the relationship between pre-training and the expertise reversal effect by using different kinds of worked examples. Van Gog et al. (2008) investigated the effectiveness of product-oriented worked examples which provided full problem solution steps only and process-oriented worked examples which showed full problem solution steps as well as relevant reasons for the steps, in the area of electrical circuits troubleshooting. Participants were randomly assigned to four conditions: product-product, product-process, process-process, process-product sequences. In the first phase (pre-training), there were no differences between conditions, but for the measures of cognitive load, process-oriented examples produced higher efficiency than product-oriented examples initially. After the second training session, when learners became more experienced, the information provided by process-oriented examples was redundant for them, so product-oriented examples were superior to process-oriented examples, demonstrating an expertise reversal effect.

5.3.4 The Expertise Reversal Effect and Aptitude-treatment Interactions

Aptitude-Treatment Interactions (ATIs) suggest that different student's aptitudes may affect the results of different treatments, which may result in different learning outcomes (Cronbach, 1967; Cronbach & Snow, 1977; Mayer, Stiehl, & Greeno, 1975; Shute & Gluck, 1996; Snow & Lohman, 1984). In ATIs, the aptitudes can be defined as knowledge, skills, learning styles or personality characteristics (Sweller et al., 2011). The expertise reversal effect demonstrates the interaction between prior knowledge or achievement and the effectiveness of instructional treatment. Therefore, the expertise reversal effect may be relevant to the Aptitude-Treatment Interactions. Tobias (1976, 1987, 1989) indicated a consistent pattern of results indicating that high-experienced learners required low levels of instructional support, whereas low-experienced learners needed high levels of instructional support, which was in line with the expertise reversal effect.

5.4 Adaptive Learning Environments

According to the expertise reversal effect, guidance which is suitable for novices may have no effect or have negative effects for more experienced learners. Therefore, on the one hand, if instructors could provide guidance which could dynamically change with the change of learner's expertise, then learners could be taught appropriately. Namely, instructors should tailor the learning environment by considering learners' levels of expertise and making instruction suitable for learning at each stage. On the other hand, when to change the content of instruction and how to change it are also important issues relevant to the adaptive learning environments.

5.4.1 Rapid Assessment Evaluating Levels of Expertise

In order to provide suitable instructions with a change of learner's expertise, we need a tool to dynamically assess learners' levels of expertise. This tool should be rapid, so instructors could obtain in-time feedback and then use it to tailor the format of learning tasks used in instruction according to changing learner expertise.

Schemas stored in long-term memory determine the characteristics of working memory and the information working memory can process currently (discussed in Chapter 1). Therefore, if a tool could assess the changes of knowledge stored in long-term memory, then we could use it to assess levels of learners' expertise. At the initial stage of study, learners have not formed mature and complete schemas which could be used to guide their study. If they are required to solve a conventional problem, they have to randomly search associated solutions, which will impose an extraneous load. Therefore, providing a worked example is a way to compensate the lack of knowledge in long-term memory, which could reduce the random search processes. After a period of study, as learners' expertise increases, the schemas of learners' gradually become mature and complete. If instructors still provide full worked examples to learners, the presented guidance which has been stored in learners' long-term memory will become redundant and impose an extraneous load. The presence or absence of an organized knowledge base in long-term memory is the critical factor to distinguish novices and experts (De Groot, 1965). Therefore, the rapid assessment should be a direct indicator of the change of knowledge stored in long term memory.

Kalyuga and Sweller (2004) suggested a rapid assessment method, which asked learners to indicate their first step towards solution of each problem within a limited time. This assessment is named as “first-step diagnostic assessment”. Under their assumptions, experienced learners could provide the first step which is more advanced and could skip more intermediate steps, whereas novices may only provide some single, random steps. Kalyuga and Sweller (2004) commented on the first-step diagnostic assessment method as a way which was more rapid and direct for diagnosing levels of learner’s expertise compared to traditional tests which require learners to provide full solutions. This assessment has been tested for validity in many well-structured domains, such as algebra, coordinate geometry and arithmetic word problems.

A more general format of this assessment was first used in the research concerning sentence comprehension (Kalyuga, 2006), which is an ill-structured domain. This more general format does not require learners to provide the first step of each problem within a limited time, but it requires the presentation of each sentence within a limited time with some possible explanations provided immediately afterwards, and then participants were required to rapidly answer “Right”, “Wrong” or “Don’t know” for each possible explanation based on that sentence. The difficulty of the presented sentence was increased gradually. This alternative rapid assessment showed high correlation with traditional knowledge tests (Kalyuga, 2008).

Kalyuga (2006) also introduced a rapid assessment method emphasizing types of schemas used in a specific domain. Five identified schemas associated with solving arithmetic word problems were considered (Marshall, 1993, 1995). If a student can provide a correct solution which is associated with a corresponding schema, which means that this student has already formed that kind of schema. In this experiment, Kalyuga (2006) only used four identified schemas (Change, Group, Vary and Restate schemas) to construct arithmetic test problems. A change schema refers to a cognitive structure which allows students view the situation in which the value of a variable changes permanently over time as a familiar situation; a group schema is a cognitive structure which allows students to identify as a familiar situation in which small values are combined together to become a big value; a vary schema relates to a situation which indicates a permanent relationship between the amounts of two different things, in which if one thing goes up or down, another thing changes in a fixed way; a restate schema refers to a situation in which there is a known relationship between two things, but a restatement indicates different values of those two things from their initial stages.

Based on those four schemas used to solve arithmetic word problems, Kalyuga (2006) designed 20 problems, in which four problems used each schema respectively, the other 16 problems involved the combinations of any two schemas (e.g., change-change, change-vary, vary-restate etc.). Results suggested that the rapid test required significantly less time than a traditional test and a high correlation was obtained between this rapid test and traditional test scores. Therefore, Kalyuga (2006) indicated that this rapid assessment method could potentially be used in adaptive learning environments, but it should be replicated in other domains in the future.

5.4.2 Fading Strategy for Tailoring Learning Environments

After knowing when to tailor instruction, next we need to know how to make changes in the format of instruction, such as making changes to the format of worked examples. One way is to change the full worked examples to completion tasks, which is relevant to fading procedures used to alter a full worked example (Van Merriënboer, 1990; Van Merriënboer, Kirschner, & Kester, 2003).This fading strategy is based on the assumption that learners will still have sufficient working memory resources to process increasing problem-solving demands with an increase of expertise.

The completion tasks include a problem statement, partially worked-out solutions and some solution steps left for learners to complete. Compared to traditional worked example-problem solving pairs, a fading strategy can provide dynamic instructions which may fit the changes in learner's expertise. Traditional worked examples are fixed to provide full guidance to learners, ignoring the dynamic aspect of learner's expertise. By using a fading strategy, the transition from a full worked example to a conventional problem may avoid providing redundant information to relatively more experienced learners.

The general format of fading procedures is that a full worked example is provided initially to learners, followed by a completion task which omits a step of that full worked example, and then the similar completion task which requires learners to complete two omitted steps and so on. Two issues should be considered when using a fading strategy: the direction of fading and the pace of fading. The next sections will discuss these two issues separately.

For the first issue, two fading strategies have been investigated within cognitive load theory: backward fading and forward fading. Backward fading provides a full worked example and then for the second task, the last step of that full worked example is omitted; for the next task, the last two steps are omitted and so on. The forward fading strategy is similar to the backward fading strategy. The only difference is that the steps are omitted from the first step of that full worked example to the first two steps and so on.

Renkl et al. (2002) compared these two strategies (backward and forward fading) with traditional worked example-problem solving pairs. For both classroom and lab-based experiments, the fading strategy was superior to traditional worked example-problem solving pairs on near transfer tests. In addition, the backward fading was beneficial on far-transfer tests as well.

Renkl and Atkinson (2001) investigated the fading strategy with a self-explanation procedure. University students were required to provide probability rules when they studied how to calculate probability. Backward fading procedures with or without self-explanation prompts were compared. The results indicated that the fading procedure with self-explanation prompts was superior for both near and far transfer tests. Similarly, Atkinson et al. (2003) investigated the effects of combining self-explanation with a fading strategy within a computer-based environment or real-classroom setting. The results extended the results obtained by Renkl and Atkinson (2001). They indicated that (1) this combined strategy did not sacrifice learning time; (2) it fostered learner performance on not only near-transfer tasks, but also far-transfer tasks with different levels of learners (university students as well as high school students). However, this combined instruction may only be suitable for well-structured tasks, such as mathematics and physics.

The research results relevant to combining the fading procedure with self-explanation supported the cognitive apprenticeship approach (Collins, Brown, & Newman, 1989). The cognitive apprenticeship approach indicated that learners should solve problems with scaffolding provided by instructors or mentors. With increasing levels of learner expertise, worked examples are changed to scaffolded problem solving and then to conventional problem solving, which is consistent with the cognitive apprenticeship approach. This combined instruction is also in line with the reflection strategy suggested by the cognitive apprenticeship approach, according to which learners are encouraged to reflect on self-explanations (Atkinson et al., 2003).

Some researchers suggested that how to fade the procedures (backward or forward) and which step should be faded depends on the structure of the task and the content of the to-be-presented learning materials (Renkl, Atkinson, & Große, 2004). Therefore, it did not matter whether to fade the procedures in a backward or forward direction.

The pace of fading is another issue which should be mentioned when we use fading strategy. Reisslein, Sullivan, and Reisslein (2007) compared three paces of fading: immediate, fast and slow fading. For the immediate fading group, learners solved problems directly after instruction; for the fast fading group, learners received a full worked example and then steps were omitted gradually (an omitted step for the second task, then two omitted steps for the next task and so on) based on the backward fading procedure; for the slow fading group, every two worked examples, a step was omitted. The results of this experiment indicated an interaction between pace of fading and levels of learner's expertise. More knowledgeable learners benefited more from immediate and fast fading conditions, compared to less experienced learners who learned more from the slow pacing condition. Therefore, well-guided and slow-paced instructional procedures that reduce extraneous load are more suitable for novices.

Schwonke, Renkl, Salden, and Aleven (2011) also considered different fading ratios. The ratio of worked example steps and to-be-solved steps varied from one to-be-solved step to four to-be-solved steps. Two kinds of knowledge were tested: conceptual and procedural knowledge. All participants were assessed by a pre-test which involved three principles (interior angle, major-minor arc and exterior angle theorems) initially. The general results of the pre-test showed that participants found the exterior angle theorem was the hardest to use and the major-minor arc theorem was the easiest to use. Therefore, the exterior angle theorem was defined as a difficult principle and the major-minor arc was an easy principle to learn and use in this experiment. Post-test problems requiring the use of each principle were presented to test procedural and conceptual knowledge. Based on the post-test, the results indicated that the smallest ratio of worked example and problem solving (namely, problem solving only) had an advantage for procedural knowledge only. For difficult mathematics problems which were based on difficult principles, the bigger the ratio (slow fading) was, the more benefits of procedural and conceptual knowledge learners would obtain. The results concerning conceptual knowledge in this experiment were in line with the expertise reversal effect: with the development of learners' conceptual knowledge, the fading sequences of a worked example were better than a full worked example (Schwonke et al., 2011).

Implications of Fading Strategy. A fading strategy follows from the expertise reversal effect. The expertise reversal effect indicates that instruction which is beneficial for novices may have no effect or have a negative effect for relatively more experienced learners. The additional instructional guidance may become redundant for learners with higher levels of expertise, but beneficial for novices who do not have a sufficient knowledge base. In order to make instruction adaptive to the changing learner's expertise, we need to diagnose learner's levels of expertise in real time, so a rapid assessment method which requires the learner to indicate the first solution step for a problem has been suggested. By using this assessment approach to obtain in-time feedback on actual level of learner's expertise, instruction can be tailored to individual learner. A fading strategy is a method to provide in-time instructions based on the current levels of learner's expertise. In this strategy, backward and forward fading procedures were discussed. A body of research demonstrated that backward fading was superior to forward fading. From a cognitive load perspective, backward fading may require less cognitive load compared to forward fading which omits the first and perhaps most critical step for learners (Ayres & Sweller, 1990). Researchers also discussed how quickly to fade solution procedures. They suggested that immediate and fast fading were more suitable for experienced learners, while, slow fading was beneficial for novices.

In order to use a fading strategy properly, we also need to consider two points: firstly, the tasks selected should be high in element interactivity (based on the element interactivity effect); secondly, each faded task is only suitable for learners whose level of expertise is appropriate for this stage's task.

5.5 Summary of Chapter 5

This Chapter focused on the expertise reversal effect, which influences the effectiveness of worked examples by taking into account levels of learner's expertise. This effect considers the interaction between characteristics of learners and characteristics of learning materials. Novices who do not have mature schemas in their long-term memory need learning materials which provide external instructions, such as studying worked examples, while for more experienced learners who have already equipped with relevant schemas in their long-term memory, external instructions may become redundant and impose an extraneous load. Therefore, a learner-adapted environment is critical for facilitating a learner's study schedule.

In this Chapter, the expertise reversal effect was discussed based on longitudinal as well as cross-sectional studies. Empirical evidences for the expertise reversal effect relevant to effective usage of worked examples have been discussed in different subject domains (Engineering, science and social science). No matter which kind of domain, the results consistently showed that with an increase of expertise, a full worked example should be gradually faded to a full problem solving, which avoids redundancy. However, presenting a full worked example is more beneficial at the initial stage of learning.

In order to tailor the teaching environment based on learner’s expertise, a fading strategy is commonly used. It shows how to transit from a full worked example to a conventional problem with the changes in learner’s expertise. Two kinds of fading strategy were discussed: forward-fading and backward-fading strategies. Based on research results, backward-fading may be more beneficial.

The expertise reversal effect is based on the central role of the knowledge base in human cognition. Therefore, we may assume that this cognitive load effect may have influence on other cognitive load effects, for example, on the worked example effect discussed in Chapter 3. Worked examples are effective for learning, but their effectiveness should depend on the level of learner’s expertise. With an increase in learner’s expertise, a full worked example which provides full guidance should be replaced with instructional methods with reduced guidance, as experienced learners have sufficient working memory resources to deal with the increasing demands of problem solving.

From Chapter 3 to Chapter 5, I have discussed the worked example effect and other research studies that are relevant to the worked example effect. The last chapter of my literature review will touch another effect which is another key concept of my thesis – the generation effect.

Chapter 6 The Generation Effect

This Chapter will discuss the generation effect. As I indicated before, the generation effect should be obtained with materials low in element interactivity, therefore, this Chapter will mainly focus on materials used to test the generation effect that are indeed low in element interactivity. Relevant generation effect research studies will be reviewed first, and then some other research studies associated with generation activity will also be discussed.

Researchers have been interested in the generation effect since Bobrow and Bower (1969) demonstrated that recall of a noun word pair was better if people generated their own sentence linking the two words of a noun word pair rather than being simply read a similar linking sentence externally provided. This effect (namely, the generation effect), indicating that items which are generated based on a given stimulus and an encoding rule are better remembered than the same items which are simply read by subjects (Slamecka & Graf, 1978), has been found in various tests, such as cued recall, recognition as well as free recall tests. McElroy and Slamecka (1982) demonstrated another version of the generation effect according to which active participation in the learning process (generation) produced better retention than passive observation (presentation/reading).

The traditional paradigm used in investigating the generation effect is word pairs, which includes a stimulus with a rule and the first letter of the target word as the cue. However, some other paradigms are also used to obtain the generation effect. Anderson, Goldberg, and Hidde (1971) applied incomplete sentences as contexts in which the to-be-generated target was a highly probable completion, such as "The doctor looked at the time on his (*watch*)". The group who filled in the blanks learned more than the group who just read the whole sentence. Glisky and Rabinowitz (1985) applied single words with missing letters, like ALC-H-L, in their experiments as testing materials. According to the results of those experiments, no matter what kind of paradigm was applied, the generation effect was found in each situation. From a cognitive load perspective, all these types of materials are characterized by relatively low levels of elements interactivity (including sentence completion tasks, assuming the participants are sufficiently proficient readers). The levels of element interactivity for specific studies will be mentioned throughout the following sections concerning the empirical evidence of generation effect.

6.1 Generating and Reading

Since the generation effect was described by Slamecka and Graf (1978), a comparison between reading and generating has been widely discussed. Begg and Snider (1987) compared generating and reading from the perspective of knowledge processing. They indicated that reading and generating were both encoding processes. The available knowledge of an intact word was enough to identify the word when reading, but the knowledge used to identify an incomplete word when generating would be sufficient only if the details that were provided by the available information are correctly processed. Therefore, the difference between generating and reading was that generating needed deeper processing of knowledge to reach a criterion of identifiability about that incomplete word than reading, demonstrating that the generating process requires more cognitive effort. In addition, according to the "lexical activation hypothesis" suggested by McElroy and Slamecka (1982), generating an item requires more activation of an item's attributes of semantic memory than the same item which is read. Similarly, Nairne, Pusen, and Widner (1985) suggested that generating activated more associations of the target which pre-existed in semantic memory than reading, and then these activated associations could be used as retrieval routes to finally retrieve target items. Differences between generating and reading can also be found based on the mechanism of the recognition process. According to Mandler's proposal (1980), recognition is based on two types of reconstruction of a to-be-remembered item: (a) an item is reconstructed by finding a retrieval route; (b) an item is reconstructed by re-experiencing surface features associated with the item. Based on these two types of reconstruction, the first reconstruction shows that generating may be superior to reading, as the generating process involves retrieving and requires cognitive effort during study, whereas, the second type of reconstruction intends to show that reading may be better for memory retention than generating, as participants are externally presented the to-be-remembered items, which provides chance to reconstruct items through re-experiencing their surface features. Glisky et al. (1985) mentioned that the difference which appears between reading and generating might lie in the processes involved in contacting the lexical unit: for reading, this process might be automatic and effortless, whereas generating a word involved an interaction between the structural information of the word fragment and knowledge system. All in all, since Slamecka and Graf (1978) established the generation effect, the generation process has been mostly proved as a process which is superior to the reading process.

6.2 The Generation Effect with Simple Materials

6.2.1 The Generation Effect with Semantic and Meaningless Materials

Slamecka and Graf (1978) used regular word pairs (such as using the opposite rule to generate a word when presented *hot-c*__ for the generation group; reading *hot-cold* for the reading group) with five rules (associate, category, opposite, synonym and rhyme) as cues in various memory tests (free recall, cued recall and recognition tests). Both the learning and testing materials were low in element interactivity. The authors found that differences between generation and read conditions were highly significant under all the tests. Following this research, a body of similar research has also indicated the generation effect with words which are semantically meaningful. For example, similar results were found by Glisky and

Rabinowitz (1985) when they used a single word fragment (such as ALC-H-L as the fragments for ALCOHOL) as a stimulus as well as a cue under the measurement of recognition test, which was another experiment using low element interactivity materials. This research further indicated that the generation effect was not always dependent on the relations between separate cues and separate targets (such as opposite: *hot-c__*); the single word fragments themselves (such as ALC-H-L) could provide cues and be the targets. They also suggested that only the items which were generated during the encoding stage were better recalled during the test stage.

Similarly, Slamecka and Fevreiski (1983) tested single words (low in element interactivity) in two formats based on how many letters of a targeted item were provided: high information (e.g., *hot-c_ _d*) or low information (e.g., *hot-c__*) format, and compared these two generating conditions with reading conditions. The low information format was more difficult to be successfully generated than the high information format. Therefore, they suggested that the accuracy of generation during the encoding stage might affect the performance on memory tests. Interestingly, the free-recall (low in element interactivity) performance of participants in the two generating versions (no matter whether they successfully generated or failed to generate) were better than performance of participants in the reading condition. It was suggested that generation not only required access to the semantic characteristics of the word, but also needed to establish the surface features of the word. The failed generation process could fail to reveal the surface characteristics of the word, but could still be better than reading because of the processing of semantic features of the word during the encoding stage.

By using very incomplete words or moderately incomplete words as testing materials (low in element interactivity), Snodgrass and Kinjo (1998) found the same effect size of the generation effect for these two kinds of material in a free recall test compared to recall of complete words.

In addition to the generation effect obtained with semantically meaningful words, McFarland, Frey, and Rhodes (1980) obtained the generation effect with sentences. They asked participants to decide whether an experimenter-generated word was suitable for a specific context or asked participants to generate a word suitable for a specific context. As participants were sufficiently proficient readers to effortlessly infer the context and only needed to deal with a word at a time, therefore, the materials were low in element interactivity. At the test stage, the participants were required to choose the words used during the encoding stage, therefore the test tasks were also low in element interactivity. Results indicated that self-generation was superior to experimenter-generated condition.

However, McElroy and Slamecka (1982) found that there was no generation effect in recognition and recall tests when the experimental materials were non-words which were designed based on the formal letter transposition rule (the first three letters of the stimulus are in backward order and then are put after the consonant which serves as the first letter of targeted word), for example, *preet-terp*. The material used in this experiment was low in element interactivity again. Similar results were found when materials were meaningless letter bigrams (e.g., E C), nonunitised 2-digit numbers (e.g., 2, 8) and unfamiliar compounds (e.g., cheese-ketchup) (all were low in element interactivity). The generation effect was not found for any of them (Gardiner & Hampton, 1985).

Thus, it seems that the generation effect may only be reliable for materials which have real meaning or are semantically meaningful. In all the cases of the generation effect obtained with semantic materials, the materials and tests were low in element interactivity.

6.2.2 The Generation Effect with Pictures

There are not many research studies that have investigated the generation effect by using pictures as experimental materials. The first research that used pictures to test the generation effect was reported by Peynircioğlu (1989). The report indicated that participants who drew pictures (such as a glass two-thirds full) had better memory of the pictures than those who just copied or looked at the pictures. Other research (Pring, Freestone, & Katan, 1990) that tested the generation effect by using raised shape pictures compared the performance of sighted and blind children. In this experiment, all sighted children wore a blindfold in order to be as blind as real blind children. In the generation condition, a cue was given (e.g., Chair) and a raised picture which described a close semantic associate was provided. Children in this group were required to generate the name of the picture (e.g., Table). In another condition, the picture and what this picture described were given, and children were required to feel the picture and then to repeat the name of the picture. Results revealed that the generation group was superior for the group which was told to repeat the name of the picture. Interestingly, the generation effect was only found with sighted children, and the reverse generation effect was found with blind children, as blind children might only deal with sensory processing rather than conceptual processing (Pring, 1988). Therefore, Kinjo and Snodgrass (2000) suggested that the main factor of the generation effect which was obtained with pictures might be the contribution of deeper sensory processing, which did not support the semantic explanation of the generation effect used with words. However, they also suggested that the generation effect using pictures was obtained because of the combined effects of sensory and semantic activation as well as additional cognitive operations. For the experiments mentioned in this section, we also can conclude that the to-be-remembered pictures were all low in element interactivity as the participants were familiar with the content of those pictures and the final tests were simple memorization tests which were low in element interactivity as well.

6.2.3 The Generation Effect in the Domain of Arithmetic

Testing the generation effect in the domain of arithmetic provides additional empirical evidence. Previous research studies, such as when and how to use a calculator to improve students' arithmetic ability (Pyke & LeFevre, 2011; Rittle-Johnson & Kmicikewycz, 2008) or research on the generation effect with multiplication problems (McNamara & Healy, 2000), have demonstrated the generation effect in arithmetic.

McNamara (1995) investigated the generation effect by comparing using a calculator with generating the answers mentally using simple multiplication problems (e.g., 2x6=12). The materials were low in element interactivity for the intended learners. Grade-2 students who had not received formal instruction in school on multiplication were divided between the generation and reading conditions, but they knew the answers to the simple multiplication problems presented (such as 1x1, 2x4). The training problems were simple multiplication problems with an ascending order of multipliers (the multipliers ranged from 2 to 7 and products ranged from 12 to 30), which indicated that the materials used were low in element interactivity. After several blocks of training (generation or reading), students were tested with 12 training problems, 10 reversals of training problems with a descending order of multipliers and products and another 12 easy problems (products were less than or equal to 10), students were required to spend only a limited time on any one problem of this test. Therefore, the test can be regarded as a cued recall test which was low in element interactivity. The study demonstrated that, in terms of accuracy, generation training was effective for students with low prior knowledge, but ineffective for students with relatively higher prior knowledge who could correctly answer some multiplication problems, demonstrating that the generation effect might be dependent on the knowledge base of students. Thus, a calculator should be used only after students had acquired a basic knowledge of arithmetic.

However, McNamara (1995) mentioned that the procedure (multiple trials-test) used to test the generation effect in this experiment might be different from the traditional procedure (single trial-test) used to test the generation effect, therefore, the generation effect obtained in this experiment might be called as a "generation advantage".

Pyke and LeFevre (2011) used another kind of material (such as G+4=K) which was low in element interactivity for adults in this experiment. Adults needed to count 4 letters down from G and then got the answer K. After the learning phase, participants were asked to solve problems used in the learning phase and four new problems as well as to find out whether they recalled the answers or counted. They obtained similar results that showed that students who generated the answers of arithmetic problems or students who generated the answers first and then were given feedback from a calculator learned better than students who read answers to those arithmetic problems directly from the calculator. They further pointed out that a calculator should be used after students had acquired a basic knowledge of arithmetic.

McNamara and Healy (2000) again investigated the generation effect by using multiplication problems. They used simple (such as 6x80=480) as well as difficult (such as 16x8=128) multiplication problems with undergraduate students as participants in their experiments, so the materials were low in element interactivity for these participants. After training, the participants were required to write down all of the answers they had seen during the training phase. Results indicated a generation effect, but they did not support the effort theory of the generation effect (Griffith, 1976; McFarland et al., 1980), as for difficult multiplication problems, they found lower accuracy in the generation group, compared to the higher accuracy obtained with simple multiplication problems.

Rittle-Johnson and Kmicikewycz (2008) used arithmetic materials consisting of 3 or 4 times 11, 12 or 13, such as 4x11. Students were assigned to the generation group (generating answers first and then checking them with a calculator) and the reading group (reading answers directly from a calculator). The groups were tested with 12 problems (6 training problems and 6 new problems), which might include problems high in element interactivity for Grade 3 students, such as 4x13. The results indicated that firstly, with an increased level of expertise, benefits of the generation effect decreased, which supported the result of McNamara (1995); secondly, the generation effect might be extended to relatively new problems resulting in enhanced transfer ability. However earlier, McNamara and Healy (2000) had suggested an opposite view. They indicated that the first factor for generation was to engage in the connection of stimulus and target rather than to generate or to produce the target. It seems that the generation effect may only be obtained for items which have been stored in our long-term memory.

Overall, materials used in the experiments mentioned in this section were also low in element interactivity, with the exception of the experiment of Rittle-Johnson and Kmicikewycz's, and the tests were mostly memorization tasks which again were low in element interactivity.

6.2.4 The Generation Effect and Contextual Memory

Previous research concerning the generation effect was mainly about item memory. Marsh, Edelman, and Bower (2001) described another kind of generation effect which was about context memory. In their research, the main hypothesis was that the generation advantage could be extended to test the memory of the context where generation happened. Results confirmed their hypothesis. They found that firstly, subjects were better at remembering the place where they generated the words than remembering the place where they read these words, no matter how difficult the generation tasks were; secondly, subjects who generated items were better at remembering the color or font of the generated items as well as at recognizing the context (e.g., left or right computer) in which they generated. They explained these results with the following reasons: generation involved elaborate processing which might combine many features in memory including all aspects of an event which were relevant to the generating stage. Although this experiment used different materials to test the generation effect, the nature of materials (e.g., the color, the font etc.) was easy for participants, namely, the materials were low in element interactivity and the test used in this experiment was also low in element interactivity.

6.3 The Generation Effect in Classroom Settings

The generation effect research described above was conducted in lab settings. However, some researchers transferred their focus from the lab to natural settings to investigate whether the generation effect could be found in such settings as well. Foos, Mora, and Tkacz (1994) found the generation effect by using university students as participants in real settings and indicated that the previous failures in finding the generation effect in such settings were due to the fact that researchers did not separate the targeted items and non-targeted items used in the experiments. In their experiment, five-page text (such as "the work of being a bee") was used. Participants were required to either generate the outline, the questions and the questions with answers of the text or study an experimenter-generated outline, questions and questions with answers. The test items were divided into targeted and non-targeted items for multiple choice or fill-in-the-blank tests. Results indicated that students who generated their own materials demonstrated better performance at those targeted self-generated materials on the multiple-choice and fill-in-blank tests. However, the performance of the generation group on non-targeted items was not better than the performance of the study group. Therefore, Foos et al. (1994) suggested that the generation effect could be obtained in natural settings, but only for targeted items that had been generated during the study phase.

The materials used by Foos et al. (1994) might be high in element interactivity, as the content of reading was a relatively lengthy story for university students. However, the tests used in the generation effect experiments were memorization tests that were low in element interactivity, as participants who generated materials only performed better on targeted items. The generation effect was found only with participants who successfully recalled what they generated previously.

As indicated above, Rittle-Johnson and Kmicikewycz (2008) also found the generation effect in natural settings. They extended the results of Foos et al. (1994), as the results suggested that the generation effect could be found not only with items targeted during the learning stage, but also with related but relatively new problems (such as 3x8) if students had low levels of expertise. However, with an increase of expertise, the advantage of generating answers decreased.

6.4 Immediate vs. Delayed Testing

According to Table 1 that summarizes the results of experiments testing the generation effect, most studies (35 out of 38) successfully demonstrated the effect after a delay or distraction. A smaller proportion of studies (19 out of 26) demonstrated the effect using immediate tests. While the effect seems to be obtainable on both immediate and delayed tests, it may be more likely using delayed rather than immediate tests as can be seen from Table 1, though that conclusion is far from certain. Furthermore, even if it is more likely, it is unclear why. One explanation that has been offered is that the process of generation itself can improve memory and that the effect of generation is more durable on memory (Schweickert, McDaniel, & Riegler, 1994).

Table 1. *Research Studies of the Generation Effect*

Experiment	Immediate	Delayed
Begg, Vinski, Frankovich, and Holgate (1991)	√	√
Burns (1992)	×	
Burns (1996)		√
Burns, Curti, and Lavin (1993)	×	√
Buyer and Dominowski (1989)		√
Carroll and Nelson (1993)		×
Chechile and Soraci (1999)	√	
Crutcher and Healy (1989)		√
de Winstanley and Ligon Bjork (1997)		√
DeWinstanley and Bjork (2004)	√	√
Dick, Kean, and Sands (1989)	√	
Donaldson and Bass (1980)	√	
Fiedler, Lachnit, Fay, and Krug (1992)	√	
Flory and Pring (1995)		√
Gardiner (1988)		√
Gardiner (1989)	√	
Gardiner and Arthurs (1982)		√
Gardiner, Dawson, and Sutton (1989)		√

Gardiner, Gregg, and Hampton (1988)		√
Gardiner and Hampton (1985)		√
Gardiner and Hampton (1988)		√
Gardiner and Rowley (1984)		√
Ghatala (1983)	√	
Glisky & Rabinowitz (1985)		√
Graf (1981)		√
Greenwald and Johnson (1989)	√	√
Grosofsky, Payne, and Campbell (1994)	√	√
Hertel (1989)		√
Java (1994)	√	
Johnson, Raye, Foley, and Foley (1981)	√	√
Johnson, Schmitt, and Pietrukowicz (1989)		√
Kinoshita (1989)		√
Liu and Lee (1990)	√	
Lutz, Briggs, and Cain (2003)		×
MacLeod and Daniels (2000)		√
McDaniel, Waddill, and Einstein (1988)	×	
McElroy (1987)	√	
McElroy and Slamecka (1982)	×	
McFarland et al. (1980)	√	√
McFarland, Warren, and Crockard (1985)		√
McNamara and Healy (1995)		√
McNamara & Healy (2000)		√

Mulligan and Duke (2002)	×	
Nairne et al. (1985)		√
Nicolas (1996)		√
Olofsson and Nilsson (1992)		√
Payne, Neely, and Burns (1986)	×	
Rabinowitz and Craik (1986)	√	√
Reardon, Durso, Foley, and McGahan (1987)	√	
Schmidt (1992)		√
Schmidt and Cherry (1989)		×
Schweickert et al. (1994)		√
Slamecka & Fevreiski (1983)	√	√
Soloway (1986)	√	
Steffens & Erdfelder (1998)	×	

Note. "√" refers to an experiment that obtained the generation effect; "×" refers to an experiment that did not obtain the generation effect.

6.5 Explanations of the Generation Effect

This section discusses some conditions which influence obtaining the generation effect. Firstly, the cognitive mechanisms of generation effect and then some factors influencing the generation effect.

Liu and Lee (1990) proposed two stages involved in generation effect: firstly, recalling initial features; secondly, locating retrieval routes. In their explanation, the subjects who successfully recalled a target word, firstly recalled the initial features of the target, which was similar to the tip-of-the-tongue phenomenon (if we come up with some features of a target word initially, then accumulating more features of the target word will let us reveal the whole target). Finally, they found associated retrieval routes leading from the initial features to the entire to-be-generated target. Therefore, if there are only a few initial features and few associated retrieval routes that are distinctly different, then a generation advantage may be obtained. This model was supported by Slamecka and Fevreiski (1983), based on their study obtaining the generation effect for words rather than non-words. They reasoned that it was plausible to regard generation as a process which had different stages: starting to process appropriate semantic features and then revealing surface features of targets. For words, the number of initial features that are semantic is larger than that of non-words, which may provide more chances to reach the second stage for words. According to their results, unsuccessful generation is an incomplete generation process in which only semantic features are processed, without reaching the stage of revealing surface features of the targets.

With more empirical evidence of the generation effect obtained, various factors which potentially influence the effect have been revealed. Categorizing those factors, two main types of explanation of the generation effect are usually mentioned: (1) the lexical activation hypothesis with semantic memory; (2) the suggestion that the generation process itself, not semantic memory, enhances memory.

Semantic memory is defined as a person's general knowledge by McElroy and Slamecka (1982). An explicit explanation of the generation effect is that this effect occurs only if semantic memory is involved in the encoding process. More specifically, an interpretation based on lexical activation was suggested. It means that the generated task leads to an enhanced activation of semantic features which consequently increases the likelihood of gaining access to the memory trace. There are two versions of the Lexical Activation Hypothesis. The strong version of this hypothesis claims that "the generational product must be a word in one's vocabulary" (Slamecka & Fevreiski, 1983, p.161); the weak version of this hypothesis indicates that numbers as well as words which are represented in the subjective lexicon can be regarded as a generation product.

Donaldson and Bass (1980) proposed that subjects spontaneously perform a "semantic adequacy check" which underlines the memorial advantage for generated items. Graf (1980) used meaningful sentences and anomalous sentences, but only found a generation effect with meaningful sentences. Similarly, McElroy and Slamecka (1982) used words and non-words with different tests, processes and various encoding rules, but the generation effect only was obtained with words rather than non-words. Therefore, they supported the assumption that the lexical status of stimulus materials had an influence on whether generated items would be remembered better than read items. Again, Gardiner and Hampton (1985) found a generation effect for integrated word pairs rather than unfamiliar word pairs. Interestingly, Slamecka and Fevreiski (1983) found that the generation effect occurred even when the self-generation attempt failed. The result may remind us of the distinction between the surface feature and the semantic features of an item. Self-generation of surface features is not essential to the generation effect, as the item that failed to be generated was recalled equally as well as items that had been generated successfully, which indicated that even though a self-generation attempt failed, the semantic features of items were still processed. Based on the experiments of Slamecka and Fevreiski (1983), the surface and semantic attributes should be regarded separately for recognition and for free recall, as the surface attribute is the key factor for recognition, while the free recall relies more on the semantic attribute.

Some researchers suggested that a number of findings contradicted the Lexical Activation Hypothesis. For example, Gardiner and Rowley (1984) found a generation effect with numbers as the stimuli; Gardiner and Hampton (1985) reported a generation effect for meaningful bigrams (e.g., US), unitized two-digital numbers (e.g., 28). Those stimuli were much like non-words. However, according to the definition of semantic memory by McElroy and Slamecka (1982), those stimuli do pre-exist in semantic memory. Therefore, it is reasonable to find the generation effect with those stimuli. Based on the experiments that used noun compounds, unitized two-digital numbers and letter bigrams, Gardiner and Hampton (1985) added a condition under which semantic memory was involved in explaining the generation effect. They suggested that the involvement of semantic memory was critical to the generation effect only in that the items to be recalled must form some integrated functional units.

Relational theory appears from the work of Donaldson and Bass (1980), which suggests that generation strengthens the relationship between stimulus and response. Considering this theory, the involvement of semantic memory in obtaining the generation effect is also addressed. Graf (1980) used cued recall and word-pair recognition tests, which are sensitive to the relationship between stimulus and response, to test the generation effect. The effect was found only for meaningful sentences, but not for the anomalous sentences. Therefore, Graf (1980) suggested that generating increased the integration relationships among words which were within a sentence and this integration would be easier if the sentence was meaningful. In sum, some advantages of generated targets appear as they are encoded in relation to their contextual cues which are relevant to semantic memory.

This perspective supports the assumption that generation superiority is due to intrinsic differences between the generating and reading tasks. According to this perspective, it is the generation process itself, not semantic processing, that acts to enhance memorability of generated materials.

Cognitive effort is a possible explanation of the generation effect that considers the generation process itself rather than semantic memory. This view emphasizes that generating involves deeper processing of an item's meaning than reading. The deeper processing produces better retention results at the test stage. Therefore, more cognitive effort (Tyler, Hertel, McCallum, & Ellis, 1979) is required when generating than reading. Griffith (1976) and McFarland et al. (1980) suggested that the generation effect was attributable to greater amounts of cognitive effort required by generation tasks. If the goal of processing in the generation stage is to make stimuli cognitively identifiable, reaching that goal will be more effortful (Tyler et al., 1979). Therefore, generating processes require more precise discrimination (Jacoby, Craik, & Begg, 1979) as well as elaboration of stimuli (Craik & Tulving, 1975) than reading processes.

A number of memory theorists (Jacoby, 1983; Kolers, 1979; Tulving, 1979) have suggested that compatibility between encoding and retrieval processes is a critical determinant of remembering. This view was tested by using word fragments (such as AL_OHO_) without specific rules (Glisky & Rabinowitz, 1985). They let participants re-generate a word during a recognition test before making their recognition decision. The results indicated that if operations used in the study phase were the same as operations used in the retrieval phase, a significant advantage of generation would be found. Specifically, some participants received the generate-same material (generating the same missing letters of a word at the study and retrieval stages) and the others received the generate-different material (generating different letters of a word at the study and retrieval stages). They found significant differences between these two conditions. The generate-the-same materials produced more benefits than the generate-different materials. This study demonstrated that the generation effect occurred when the same operations used in the study stage were retained at the retrieval stage. Therefore, encoding/retrieval compatibility belonging to the generation process itself affected the generation effect. Of course, this encoding specificity effect for generating at retrieval was found only when the specific encoding operations were reinstated.

Lutz et al. (2003) investigated the generation effect by using non-words and sentences. In their research, the non-words were readable, but no generation effect was found with this kind of material. For sentences, the generation effect was found for clichés which were similar to familiar items whose retrieval routes had been stored in long-term memory. Therefore, Lutz et al. (2003) indicated that the generation effect would be found if the representation of items had been stored in memory. Based on transfer-appropriate-processing, if the cues used during the study phase are matched with those used during the test phase, a positive generation effect would be obtained. Otherwise, a negative generation effect would be found. Generally speaking, the compatibility of retrieval routes during the study and test stages might be another factor to affect the generation effect by considering the generation process itself.

Other factors which may affect obtaining a generation effect are: visual familiarity, format, item-specific features and within/between list design. The following sections will discuss them in details.

Jacoby (1983) showed that generating items was disadvantageous if subjects experienced visual familiarity with items which were read. Johns and Swanson (1988) compared two groups using non-words as stimuli. One group was given feedback (namely, representing words to participants) at the end of the study period, another group was not. The study found that the generation effect was not obtained in the no-feedback group, while it was found in the feedback group. This finding supported the factor of visual familiarity. In terms of the standard no-feedback paradigm, subjects had opportunities to experience intact correct words in the reading condition, but never had a chance to experience intact correct words after generating when they were in the generation condition. Therefore, it might be too difficult to remember non-words that have never been seen (especially the correct forms of them) after generation, which might have caused no generation advantage observed in the study. However, McElroy and Slamecka (1982) obtained different results when considering the factor of visual familiarity in their research with non-words. In order to make sure that generated responses would also be visually exposed to participants, all items in this experiment were presented twice. The authors did not find a generation effect with those items, however, the effect was found with meaningful sentences which were also provided to subjects as feedback after generating. Therefore, McElroy and Slamecka (1982) suggested that it was not the exposure condition (visual familiarity) that led to the absence of a generation effect with non-words.

Kirsner (1973) showed impaired performance if the case of items was altered from the study stage to the test stage. Similarly, Hock, Throckmorton, Webb, and Rosenthal (1981) also tested words and non-words in either the same case or different cases during the study stage. They found that recognition of both words and non-words was impaired in the change-case condition. Other research that considered the changed format was conducted by Nairne et al. (1985). They did not find the generation effect with non-words. Considering the format change between study and test stages, read items were seen in both the experimenter’s typewritten format and the subject’s own handwriting during the study stage, while the generated items were seen only in the subject’s handwriting. However, during the test stage, the format resembled their appearance in the experimenter’s typewritten format. Johns and Swanson (1988) used different formats of non-words during both the study and test stages. In this study, half of the items were in the lowercase and the other half were in the uppercase, while in the test, all of the test items were in the lowercase. Their results replicated those of Kirsner’s (1973) and showed that the impairment in recognition performance which was associated with changing an item’s format was larger for non-words than for words. Therefore, the change of formats may also affect the generation effect.

Specific features of responses are another factor that has been investigated. Glisky and Rabinowitz (1985) found the generation effect by using single words without cues or context, which suggested that the features of items themselves affected the generation effect, as subjects processed the items themselves only during the generation stage. Hirshman and Bjork (1988) found a similar result showing that only the item-specific factor influenced the generation effect in a free recall test. However, they indicated that a single factor might not be enough to explain the generation effect, which suggested that in order to explain the generation effect, multiple factors might need to be considered. Therefore, two sub-theories were formed according to the features of items.

Single-factor theories assume that the generation effect is caused by a single factor. Two kinds of a single factor are proposed by researchers to explain the generation effect: firstly, the generation effect enhances the processing of shared features of stimulus and response; secondly, the generation effect enhances the processing of features of responses only (Burns, 1990). A theory based on the first factor is a stimulus-response relational theory. A theory based on the second factor is a response-oriented theory, which again can be divided into two possible approaches: one is based on a response-specific factor, which indicates that the generation effect only improves the features of responses that are not shared with another response or stimulus; another approach is based on a response-relational factor, which suggests that the generation effect enhances the features shared among responses. Begg, Duft, Lalonde, Melnick, and Sanvito (1989) supported response-oriented factor theory as an explanation of the generation effect. Provided the generation effect enhances stimulus–response processing, the advantage of the generation effect should be found with the stimulus as well. However, their results did not support this hypothesis. The opposite view was expressed in the research conducted by Greenwald and Johnson (1989). They found a small but significant generation effect with the stimulus used in unexpected free-recall, cued-recall and recognition tests. This study therefore supported the suggestion that the generation effect enhanced stimulus-response processing.

Multiple-factor theories assume that the generation effect should be explained by multiple factors rather than a single factor. According to Burns (1990), a multiple-factor theory requires that the stimulus-response relation and response-oriented factors both should be addressed. Similarly, Hirshman and Bjork (1988) found the generation effect with cued recall (which is sensitive to stimulus-response processing as well as response-oriented processing) and free recall tests (which are sensitive to response-oriented processing), thus supporting the multiple-factor theory.

Another version of multiple-factor theory was proposed by McDaniel and Waddill (1990). The three factors suggested in their research were: the information provided by the letters of each word, the information paired with each word, and the information from the whole list. The second factor included two situations: an explicit cue accompanying words (e.g., *HOT-C_*: antonym) or an implicit cue which provides only a relevant hint of the word (e.g., *TALL-S_*). The results suggested that when the targets themselves and the relationships among initial words and targeted words were obvious to subjects, they would consider the information which came from the whole list, and then a generation effect could be found using a free recall test.

Finally, how to present testing materials is another factor that may affect obtaining the generation effect. Slamecka and Katsaiti (1987) compared three types of design: presenting material including items for reading only, presenting material with items for generating only and presenting material which mixed read and generated items. Two main findings were reported: firstly, the recall level was increased for items used to be generated in a mixed list; secondly, the generation effect was obtained by using a mixed list rather than a pure read list or generated list. Slamecka and Katsaiti (1987) explained the results based on the selective displaced rehearsal hypothesis. This hypothesis assumed that participants in a mixed list condition spent more time on generated items than on read items, which led to an increased level of performance with generated items. However, this hypothesis may only be effective for free recall tests.

6.6 Other Studies Related to Generation Activity

Apart from studies concerning paired associates., some other research studies may be also relevant to generation activity, such as learner-generated drawing and ICAP theory. This section will give more details.

Many research studies have investigated the effectiveness of using self-drawing diagrams, and most of them indicated that self-drawing diagrams during learning and problem solving could lead to deeper comprehension with students more engaged in learning (Leopold, Doerner, Leutner, & Dutke, 2015; Leopold & Leutner, 2012; Van Meter, 2001). Leopold and Leutner (2012) compared instructions to generate diagrams with main idea selection and summarization in two experiments. The results were consistently positive for the self-drawing strategy which was found to deepen learning and comprehension, whereas, negative for the other two text-based strategies. Again, Leopold et al. (2015) investigated connections between text and pictures by assigning participants into four groups: picture group, text only group, integration group (who wrote main concepts right by the corresponding components of the pictures), separation group (who wrote down main concepts besides the corresponding pictures). Transfer and comprehension scores were better for the text-picture group rather than the text-only group. Van Meter (2001) compared a drawing strategy with a reading condition, and similar results were found. Although the drawing strategy required more time during learning, participants engaged more compared to the reading condition.

But some research studies also pointed out some negative results when drawing a diagram or picture (Hegarty & Kozhevnikov, 1999; Snowman & Cunningham, 1975). Hegarty and Kozhevnikov (1999) compared a schema-base representation which showed spatial relations, with a picture-based representation which showed the visual appearance of objects in this problem. The results suggested that a schematic representation was positively related with success compared to a pictorial representation which was negatively related with success. Similarly, Snowman and Cunningham (1975) suggested that a no drawing effect was found by requiring participants to recall factual information.

Chi and Wylie (2014) suggested the I (interactive) C (constructive) A (active) P (passive) framework, which assumed that the performance of learners would increase when students were more engaged with the learning materials. This theory suggests another view of learning. Cognitive load theory suggests using worked examples to show students how to solve problems, while, ICAP emphasizes that students should actively engaged in learning. ICAP has been tested in different learning environments and indicates that as students transfer from *passive* to *interactive* forms of learning, the change of knowledge processes leads to better learning outcome.

6.7 Summary of Chapter 6

In this chapter, the generation effect was discussed according to its theoretical perspective and a body of empirical evidence. Slamecka and Graf (1978) described this interesting and robust memory phenomenon with various conditions and with different memory tests (e.g., recall tests and recognition tests). The traditional paradigm used in the generation effect uses word pairs. In addition to the use of word pairs as materials, other materials have also been used to investigate the generation effect, such as multiplication problems, sentences, and pictures. However, not all of these experimental materials are suitable for obtaining the generation effect. Many research studies have indicated that the materials used in the generation effect may need to be semantically meaningful. McElroy and Slamecka (1982), Lutz et al. (2003) found no generation effect for non-words, even if these non-words were readable. Similar results were found when the materials were meaningless letter bigrams (e.g., E C), nonunitised 2-digit numbers (e.g., 2, 8) and unfamiliar compounds (e.g., cheese ketchup) (Gardiner & Hampton, 1985). Although different kinds of materials have been used to test the generation effect, a common characteristic of those materials (and the corresponding memory tests) is that they are low in element interactivity.

Most results in the generation effect research have been obtained in lab settings, but some researchers also focused on natural settings. Foos et al. (1994) found the generation effect in real settings, but only with the items targeted during the encoding stage. Rittle-Johnson and Kmicikewycz (2008) also demonstrated the generation effect in natural settings.

Researchers have identified various factors that may affect the generation effect. One of the suggested explanations is semantic memory which is based on real word materials. Other factors used to explain the generation effect include visual familiarity and cognitive effort etc. With the experimental evidence of the generation effect, some general theories for explaining the effect have appeared. Burns (1990) proposed a single-factor theory, involving stimulus-response relational theory and response-specific theory. However, some researchers indicated that a single factor could not explain the generation effect very well. Therefore, the multiple-factor theory was considered. The multiple-factor theory combines those factors used in the single-factor theories. It suggests that in order to explain the generation effect, we need to consider the shared features between stimulus and response as well as the unique features of the response.

Apart from considering the generation effect tested with simple materials, some other studies may be also relevant to generation activity. In this Chapter, learner-generated drawing and ICAP theory were reviewed to discuss another side of generation.

From the traditional generation effect to more general generation activity, it seems that learning outcomes may be also highly associated with students' engagement. Cognitive load theory suggests to use designed worked examples to show students how to solve problems, while, ICAP theory and learner-generated drawing studies may have indicated that considering the limited capacity of working memory, if students actively engage (generate) in learning, students' learning outcomes still could be improved.

According to the review in the previous Chapters, the generation effect may require simple, low element interactivity materials, while, the worked example effect may require high element interactivity materials. Therefore, the connection between the two effects may be the concept of element interactivity. The levels of learner expertise may be another factor influencing the relation of these two effects. With an increase of learner's expertise, materials high in element interactivity will be changed to low element interactivity materials which may be suitable for generation. Chapter 7 will discuss those hypotheses in details.

Chapter 7 Research Question and Hypotheses

7.1 Research Question

Previous chapters have discussed the worked example and generation effects in details. According to Cognitive Load Theory, worked examples, which provide full guidance to learners on how to solve a problem, can result in better performance than a problem solving condition which has no guidance, resulting in the worked example effect. In contrast, the generation effect occurs when learners generate responses rather than being shown them and results in better performance than a presentation condition which provides an answer to a question.

Therefore, based on the description of the worked example and generation effects, we can find an obvious contradiction between them: for the worked example effect, providing full guidance is more efficient than providing no guidance; for generation effect, the situation is reversed. This contradiction is exactly the research question which is going to be investigated in this thesis by conducting a series of experiments.

The next sections are going to raise hypotheses of my research and briefly introduce the experiments.

7.2 Element Interactivity and Worked Example Effect

As discussed in Chapter 2, element interactivity is the basic concept for determining types of cognitive load. Interactive elements are defined as elements that must be processed simultaneously in working memory as they are logically related (Sweller et al., 2011). An element which should be processed in working memory can be a symbol, concept or procedure, and it is characteristically a schema. Before a schema forms, its components must be processed in working memory as individual elements, but after the acquisition of the schema, these individual elements can be incorporated into this more complex structure to be an entity processed in working memory. Element interactivity levels can be determined by estimating the number of interacting elements in learning materials (Sweller & Chandler, 1994; Tindall-Ford et al., 1997).

There are a variety of documented conditions under which the worked example effect, like all cognitive load effects, will not be obtained (Sweller et al., 2011). One of those conditions concerns levels of element interactivity. The worked example effect should only be obtained if element interactivity is high resulting in a high intrinsic cognitive load. If intrinsic cognitive load is high, extraneous cognitive load should be controlled by using worked examples instead of problem solving. Therefore, a comparison of worked examples with problem solving results in the worked example effect only under high element interactivity conditions.

However, if materials are low in element interactivity with the intrinsic cognitive load of materials low, instructional procedures associated with cognitive load theory such as procedures based on the worked example effect, no longer apply. Controlling extraneous cognitive load is unnecessary when the intrinsic cognitive load of the materials used is low because the total cognitive load may not exceed working memory limits. The possible occurrence of cognitive load effects such as the worked example effect under high but not low element interactivity conditions provides an instance of the element interactivity effect.

Element interactivity does not just depend on the characteristics of the materials. It also depends on the knowledge base of learners. High element interactivity material for novices may be low element interactivity material for more knowledgeable learners. Accordingly, as levels of expertise increase, we might expect the worked example effect to reduce and eventually reverse. Precisely this effect has been obtained (Kalyuga et al., 2001) providing an example of the expertise reversal effect (Kalyuga et al., 2003). An interpretation of this finding that accords with the present theoretical framework is that increased expertise reduces element interactivity and with reduced element interactivity, a result similar to the generation effect appears.

7.3 Element Interactivity and Generation Effect

In marked contrast to the worked example effect, the generation effect occurs when learners who generate responses themselves perform better than those who study presented answers to questions (Slamecka & Fevreiski, 1983; Slamecka & Graf, 1978). This phenomenon is robust for various situations and different kinds of memory tests such as cued, free recall and recognition tests. The traditional format in studies of the generation effect is to use word pairs that include a stimulus as the cue and the first letter of the target word, with a rule indicating how the response is to be generated, for example, *COLD-H__ (OPPOSITE)*. Irrespective of specific formats, the generation effect has been found in many studies with generation conditions resulting in better memory traces than presentation conditions.

The majority of the studies have used highly familiar, low element interactivity materials that are unlikely to have imposed a heavy working memory load. As examples, Slamecka and Graf (1978) used regular word pairs with five rules (associate, category, opposite, synonym and rhyme) as cues to obtain the generation effect using free recall, cued recall and recognition tests. Each paired associate can be learned without reference to any other pair and so is low in element interactivity. McFarland et al. (1980) obtained the generation effect with sentences. They asked participants to fill in a missing word to generate meaningful sentences. Schemas for sentences ensure that element interactivity is low. McNamara and Healy (2000) used arithmetic problem solving tasks with one group of undergraduate learners required to calculate the answers to multiplication problems whereas another group was presented the answers. Participants then were required to recall the answers to the problems. The group that generated the answers was better able to remember them than the group presented the answers. In this experiment, McNamara and Healy (2000) used problem solving tasks but the post-test consisted of a simple memory task requiring recall of low element interactivity material. Most, possibly all of the literature on the generation effect used material that imposed a low working memory load due to low element interactivity.

7.4 Hypotheses of a Series of Experiments in this Thesis

As discussed in the Chapter 6, there are several hypotheses to account for the generation effect. None of them can explain (or are intended to explain) the worked example effect, the results of which are contrary to those expected according to the generation effect. Indeed, all of these hypotheses are directly contradicted by the worked example effect. Solving problems is inferior to studying worked examples but requires more effort than studying worked examples (Paas, 1992; Paas & Van Merriënboer, 1994b), contradicting the cognitive effort hypothesis; takes longer than studying worked examples (Cooper & Sweller, 1987; Sweller & Cooper, 1985), contradicting any hypothesis based on generation taking more time than presentation; and requires less disparity between learning and test conditions than studying worked examples, contradicting the transfer appropriate processing hypothesis.

The hypotheses used to explain the generation effect cannot be used to explain the worked example effect. Of course, the explanation of the worked example effect, that the guidance provided to assist learners in solving problems reduces working memory load thus facilitating learning is equally inapplicable to the generation effect or, indeed, the reverse worked example effect (Kalyuga et al., 2001). Those contrary results lead to different hypotheses. Those hypotheses suggest that the categories of material used to demonstrate the worked example and generation effects are different, potentially resolving the contradiction.

The contradiction between the worked example and generation effects may be resolved by considering element interactivity as a critical factor. The worked example effect (i.e., the superiority of high levels of guidance) may occur for high element interactivity materials whereas the generation effect (i.e., the superiority of low levels of guidance) may be applicable for low element interactivity materials. Accordingly, it was hypothesized that the specific type of effect could be predicted by considering the levels of element interactivity of the corresponding materials.

The specific hypotheses are as follows:

Hypothesis 1: A dis-ordinal interaction of levels of guidance and element interactivity will be obtained. High levels of guidance may be superior to low levels of guidance using materials high in element interactivity resulting in the worked example effect, whereas low guidance may be superior to high guidance using materials low in element interactivity resulting in the generation effect. As indicated above, complex material may need explicit guidance to assist learners to understand the material. Simpler material may not require explicit guidance.

Hypothesis 2: At higher levels of learner expertise, the interaction of guidance and element interactivity may disappear because higher levels of expertise should reduce element interactivity. High guidance may become inferior to low guidance with materials that are high in element interactivity for more expert learners (i.e., the worked example effect will disappear), whereas low guidance may remain superior to high guidance with low element interactivity materials demonstrating the generation effect. Thus, for learners with high levels of prior knowledge, the generation effect was hypothesized to be obtained for all materials with the interaction of guidance and element interactivity disappearing.

7.5 Overview of Experiments

In this thesis, in order to resolve the contradiction between the generation effect and the worked example effect and test our hypotheses, five experiments were included to investigate the interaction of levels of guidance and levels of element interactivity by using students with different levels of expertise. All of the experiments applied a mixed factorial design (See Figure 7.1 The general experimental design) with repeated measures on the variable of element interactivity.

Guidance / Element interactivity	Low	High
High	Problem Solving Group	Worked-example Group
Low	Generation Group	Presentation Group

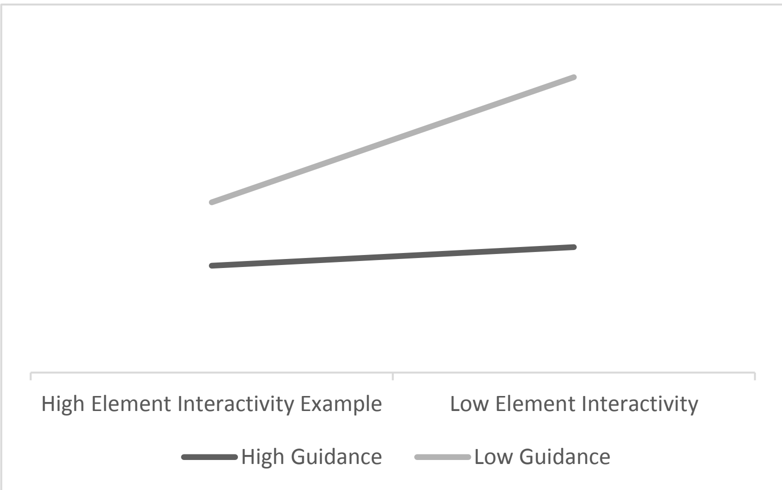


Figure 7.1. The General Experimental Design

Experiment 1, 2 and 3 tested students' (grade 4, 6 and 7) geometry knowledge. Each experiment had two phases: the generation effect phase (35 minutes) and the worked example effect phase (35 minutes). The second phase might or might not have a delay after the first phase. Experiment 1, 2 and 3 shared similar experimental procedures (Figure 7.2 and 7.3). Specifically, in the first phase, four steps were involved: participants were randomly assigned; participants studied geometry formulae; participants generated or were presented geometry formulae and finally the free-recall test. Similar steps were used for the second phase: participants were randomly re-assigned; participants reviewed geometry formulae; participants self-solved problems or studied worked examples and finally were presented the similar transfer test.

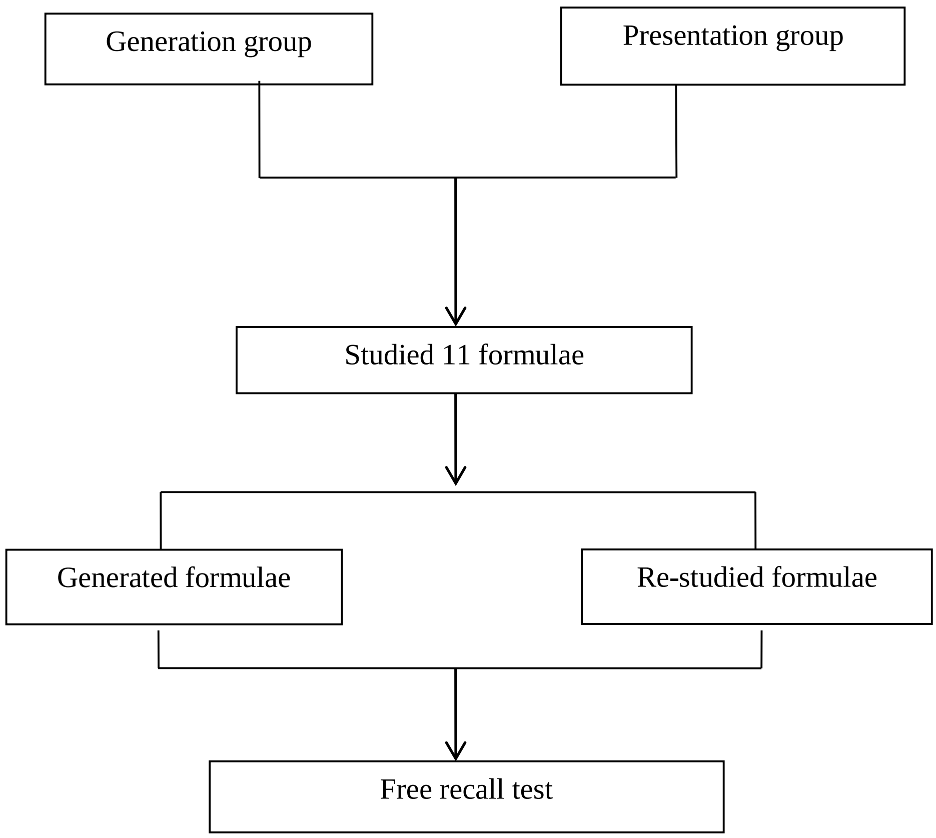


Figure 7.2. Phase 1: Procedure Used to Test for the Generation Effect

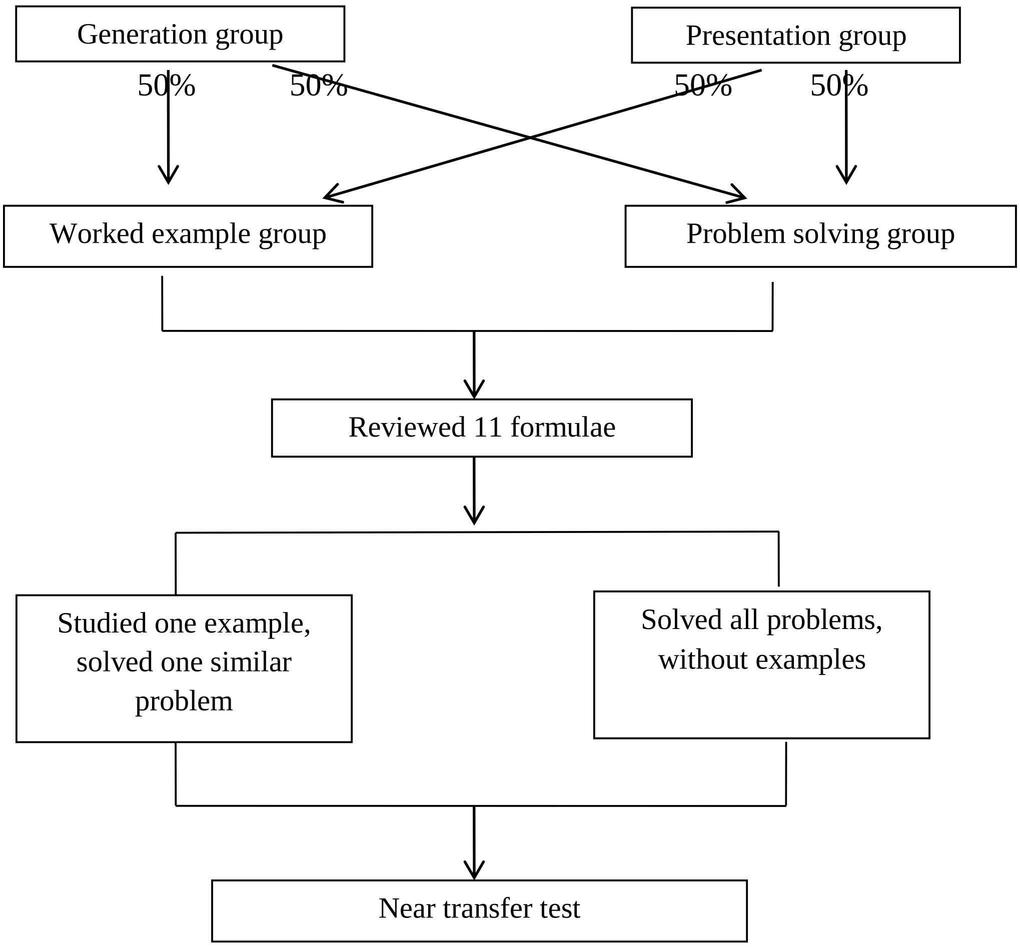


Figure 7.3. Phase 2: Procedure Used to Test for the Worked Example Effect

The test materials were similar for Experiments 1, 2 and 3. However, Experiment 2 and 3 slightly changed the materials used to test for the worked example effect and the similar transfer test. In Experiment 1, the materials used to test for the worked example effect involved two kinds of geometry questions: calculating the area of composite shapes and calculating the area of shaded parts. In order to be more suitable for participants and to facilitate the marking procedure (i.e., marking the same types of problems rather than different types of problems), in Experiments 2 and 3, only calculating the area of composite shapes was used.

Experiment 4 and 5 used older students (grade 10 and 11) to continue investigating the interaction of levels of guidance and levels of element interactivity to test our hypotheses in the domain of trigonometry which is more difficult. The experimental design and procedures were as same as the experimental design and procedures used in the previous three experiments. The test materials used in Experiment 4 and 5 were changed to test trigonometry knowledge and immediate and delayed tests were both used. Participants were required to memorize 10 trigonometry formulae during the first phase of the experiment and then to study how to apply trigonometry formulae to simplify trigonometry expressions in the second phase of the experiment.

7.6 Overview of the Participants of the Experiments

As discussed in Chapter 5, the effectiveness of using worked examples may be influenced by the levels of learner's expertise, so different levels of learners were chosen in the five experiments.

In Experiment 1, grade 4 students who had not previously studied algebra were used. They only studied some of the geometry formulae employed in this experiment, so they were regarded as novices; in Experiment 2, grade 6 students were used. They had previously studied all those formulae as well as how to calculate the area of a composite shape and the area of the shaded parts of diagrams. However, relevant schemas were not completely formed, compared to grade 7 students who were the participants in Experiment 3. Therefore, the level of learners' expertise gradually increased in the first three experiments.

Experiment 4 and 5 used the same experimental design. Grade 10 students who participated in Experiment 4 had not previously studied trigonometry, whereas, Experiment 5 employed grade 11 students who had studied trigonometry and its applications, including how to simplify trigonometry expressions. Therefore, grade 10 students were novices and grade 11 students were regarded as relatively expert.

PART 2: EMPIRICAL STUDIES

Chapter 8 Experiment 1

Five 2 (guidance: low vs. high) x 2 (element interactivity: low vs. high) mixed factorial designs were used from Chapter 8 to Chapter 12 to investigate hypotheses indicated in Chapter 7. The second variable was a within-subject variable.

The purpose of Experiment 1 was to investigate the hypothesis of a dis-ordinal interaction between levels of guidance and levels of element interactivity using Grade 4, primary school learners studying geometry topics that were either high or low in element interactivity for these students. As indicated above, according to the element interactivity effect, if learning materials are low in element interactivity, then other cognitive load effects, such as the worked example effect, are unlikely to be obtained. Therefore, in this experiment, high-element interactivity materials were used to test for the worked example effect by comparing studying worked examples (high guidance) with problem solving (low guidance). Low-element interactivity materials were used to test for the generation effect by presenting learners with answers to memory formulae (high guidance) or having them generate answers themselves (low guidance).

8.1 Method

8.1.1 Participants

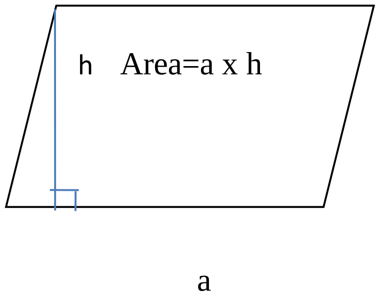
The participants were 41 Grade 4 students from a primary school in Chengdu, China. They were approximately 10 years old. They were randomly assigned to either the generation or presentation group in the first phase of the experiment and then half of the students from the generation group were randomly assigned to the problem solving group and the other half to the worked example group in the second phase. Similarly, half of the students from the presentation group were allocated randomly into either the worked example or problem solving group. Four students in the generation group and four in the presentation group did not complete the entire procedure. These eight students were eliminated from the data analyses, leaving 33 students. In class, all students only had studied the area and perimeter formulae of squares and rectangles and so were regarded as novices with respect to other formulae used in this study to test for the generation effect. Similarly, none of the students had been taught to solve the problems used to test for the worked example effect.

8.1.2 Materials

To test for the generation effect, 11 geometry formulae were chosen from textbooks used in primary and secondary schools. There were three surface area formulae, four perimeter formulae and four area formulae (see Appendix 1). In order to allow the Grade 4 students who had not as yet studied algebra understand the formulae, the presentation avoided algebraic expressions, such as “ab” which were replaced by a × b.

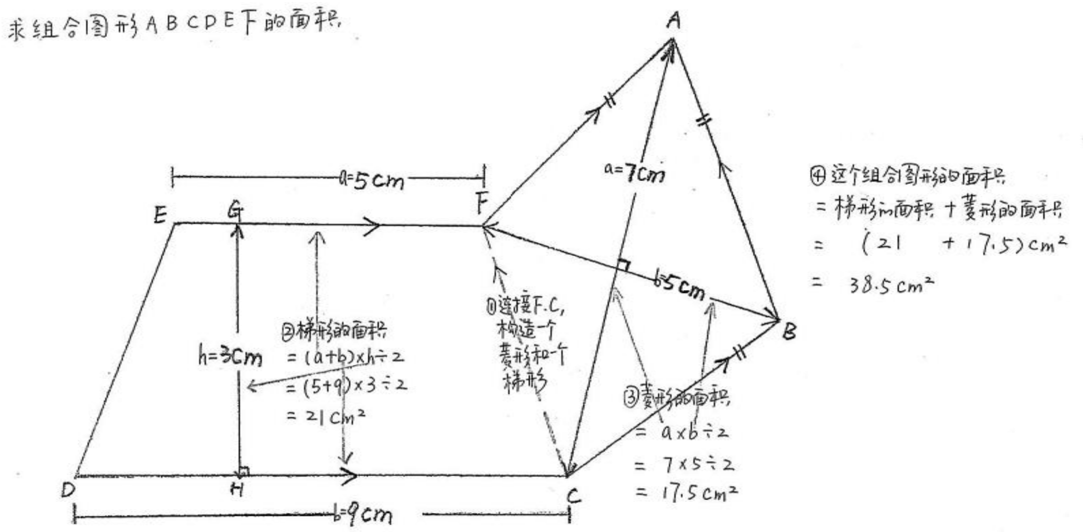
Levels of element interactivity were estimated for all materials using the method illustrated by the following examples. Assume students are asked to remember the formula used to calculate the area of a parallelogram (see an example in Appendix 1 below), a task used to test for the generation effect (see Appendix 1). Based on the concept of element interactivity, Grade 4 students, needed to memorize the five elements of the equation, $Area = a \times h$. While there are five elements, they do not interact and neither do they need to be connected to the diagram. Each can be memorized separately and if one is forgotten, it does not affect any of the others. For example, if a pro-numeral is forgotten and replaced by a different symbol, it has no effect on any of the other symbols. Therefore, this material required learners to deal with only one element at a time and so the interacting element count is one.

Parallelogram:



An Example from Appendix 1

In contrast, assume students are asked to calculate the area of a composite shape (see an example from Appendix 3), a task used to test for the worked example effect (see Appendix 3). Grade 4 students who had not learned previously how to calculate the area of composite shapes needed, firstly, to identify the four equal length lines, including the missing line, FC, that form a rhombus. They then needed to identify the four lines that constitute a trapezium, again including the missing line, FC. Next, they had to determine that they needed to calculate the areas of both requiring another two elements. To actually calculate the two areas, students needed to know the meaning of *a* and *b* in the rhombus and the multiplication relationship between them, as well as the meaning of *a*, *b*, and *h* in the trapezium and addition, multiplication, and division by two, which together involved nine elements. Finally, adding together those two separate area values involved another element. Therefore, in total, it can be estimated that about 20 interacting elements were involved in this task rendering it very high in element interactivity for Grade 4 students.



An Example from Appendix 3

Students were given three booklets printed on A3 paper. The first booklet contained 11 basic geometry formulae (see Appendix 1) and was common to both groups. All of the perimeter formulae were presented first followed by all of the area formulae and then all of the surface area formulae. The second booklet differed for the generation and presentation groups. For the presentation group, the content was identical to the first booklet, whereas for the generation group, only the names of each of 11 formulae and their relevant geometric shapes (or only the names of each of 11 formulae) were included as students needed to generate each formula by themselves (see Appendix 2). The third booklet (common to both groups) was blank to allow students to write out their answers in the free recall test. They were required to write out as many formulae as they had studied from the first and second booklets.

To test for the worked example effect, students again were given three booklets. The first booklet was identical to the first booklet (including 11 basic geometry formulae) used to test for the generation effect (see Appendix 1). Its function was revision (review) of the previously learned information. The second booklet differed for the worked example (see Appendix 3) and problem solving groups (see Appendix 4). The booklet for the worked example group contained two worked examples each followed by a similar problem for students to solve. Students in the problem solving group were required to solve the same four problems by themselves with no worked examples provided. The third booklet contained five test problems for students to solve (see Appendix 5). The first three problems required learners to calculate the area of a composite shape and the other two required students to calculate the area of shaded sections of the diagrams. All booklets had a cover page identical to the one for the generation effect.

8.1.3 Procedure

The generation effect phase of the experiment lasted for one class period of 35 minutes. Prior to studying the first booklet, students were re-seated according to the group into which they were randomly placed (see **Participants** section above) (7 minutes).

The *study stage* (10 minutes). After being re-seated, students began studying the first booklet. They could make notes in this booklet if they needed to. After 10 minutes, all students handed in this booklet.

The *generation or presentation stage* (10 minutes). The experimenter distributed the second booklet to students in the generation and presentation groups separately. Students in the generation group were required to generate all of the formulae they had studied in the first booklet, whereas students in the presentation group were required to again study those formulae. No one could hand in this booklet before 10 minutes had elapsed. Any students who completed their task in less than 10 minutes were told to review the material again. After 10 minutes, all students handed in this booklet.

The *free recall test stage* (8 minutes). The test required students to write out as many of the formulae that they had studied in the first and second booklets. Students could only hand in their test booklet after eight minutes had elapsed. Therefore, if students finished early, they were required to review their answers. When scoring the test, a correct formula was awarded one mark. Therefore, the maximum score in the free recall test was 11. Each student's score out of 11 was converted into a percentage score for analysis providing the scores testing knowledge of the low-element interactivity material.

The worked example effect phase of the experiment also lasted for one class period of 35 minutes that occurred four hours later on the same day after the generation effect phase. Prior to the experiment, the students already had been randomly chosen from the generation and presentation groups to form the worked example and problem solving groups (see **Participants** section above). Students were re-seated according to the group to which they had been allocated (7 minutes).

The *study stage* (10 minutes): The procedure for this stage was identical to that used for the equivalent stage in the generation effect phase of the experiment.

The *worked-example or problem-solving stage* (10 minutes). The general procedure was identical to that used in the generation effect phase. Students in the worked example group were required to first study the worked example of Problem 1 indicating how to calculate the area of a composite shape and then to solve a similar problem (Problem 2). A similar procedure was followed for Problem 3 (a worked example) and Problem 4 (a similar problem of Problem 3 that students had to solve themselves rather than study as a worked example). Students in the problem solving group were required to solve the same four problems (Problems 1 - 4) used in the worked example group by themselves, with none of the problems presented as worked examples.

The *similar transfer test stage* (8 minutes). Again, the general procedure was identical to that used in the generation effect phase. The test required students to solve five problems (see Appendix 5). Students could obtain a maximum of four marks on each of the first three problems (one mark for constructing the line to divide the composite shape into two basic geometry shapes; one for correctly using the area formula of one of the basic geometry shapes; one for correctly using the area formula of the other basic geometry shape; one for adding the two areas). The maximum score for both of the last two problems was also four (one for calculating the area of the whole shape; one for correctly using the area formula of one of the basic geometry shapes; one for correctly using the area formula of the other basic geometry shape; one for subtracting the area of the non-shaded parts from the total area). Each student's total score out of 20 (five problems each with a maximum score of four) was converted to a percentage score for analysis. The internal reliability of this test using Cronbach's α was .72 after deleting the 3rd test question to increase the reliability of the test. These scores provided the dependent variable testing for knowledge of the high-element interactivity material.

8.2 Results and Discussion

Means and standard deviations of percentage test score results may be found in Table 2. These results were analyzed using a 2 (levels of guidance) x 2 (levels of element interactivity) ANOVA with repeated measures on the element interactivity factor. All means, standard deviations presented in Table 2 and the statistical analyses following the table were based on the four test questions remaining after eliminating Question 3 but it should be noted that the patterns of significance were identical to those obtained using all five test questions. Cohen's *d* was used through this experiment to indicate effect size following the calculation of a t-test value.

Table 2. *Mean (SD) Percentage Correct Test Score Results After Eliminating Q3 for Experiment 1*

Guidance	Low element interactivity	High element interactivity
High (N=16)	37.0 (11.38)	31.6 (24.20)
Low (N=17)	50.2 (13.16)	18.7 (13.07)

Note. The maximum score for Low element interactivity test was 11; the maximum score for High element interactivity test was 16.

Based on the data of Table 2, the main effect of guidance was not significant, $F(1, 31) = .002$, MSe = 241.94, $p = .964$, $\eta_p^2 = 0$. The main effect of element interactivity was significant, $F(1, 31) = 19.85$, MSe = 281.86, $p < .001$, Wilks' Lambda = .610, $\eta_p^2 = .390$. Low element interactivity material percentage correct test scores were higher than the high element interactivity test scores. The interaction between guidance and element interactivity was of primary interest in this experiment and was significant, $F(1, 31) = 9.98$, MSe = 281.86, $p = .004$, Wilks' Lambda = .756, $\eta_p^2 = .244$.

Following the significant interaction, simple effects tests were conducted. For the low element interactivity material testing for the generation effect, the effect of guidance was significant, $t(31) = -3.08$, $SE_{diff} = 4.30$, $p = .002$ (1-tailed), Cohen's $d = .96$. The mean percentage correct scores indicated that low guidance was superior to high guidance demonstrating a generation effect.

For the high element interactivity material testing for the worked example effect, the effect of guidance also was significant, $t(31) = 1.92$, $SE_{diff} = 6.71$, $p = .03$ (1- tailed), Cohen's $d = .64$. The mean percentage correct scores indicated that high guidance was superior to low guidance demonstrating a worked example effect.

Note that a one-tailed test was used as there were clear directional hypotheses: for low element interactivity materials, low guidance should be superior to high guidance, indicating a generation effect, while, for high element interactivity materials, high guidance should be superior to low guidance, indicating a worked example effect.

In Experiment 1, it was hypothesized that an interaction of guidance and element interactivity would be obtained. High guidance was predicted to be superior to low guidance using materials high in element interactivity, while low guidance was predicted to be superior to high guidance with materials low in element interactivity. The results of Experiment 1 confirmed this hypothesis with a dis-ordinal interaction of guidance and element interactivity obtained. The simple effect tests indicated that students who generated formulae during a study stage better memorized those formulae than students presented the formulae, in line with the generation effect. For materials high in element interactivity, students who studied worked example-problem pairs were better at solving test problems than students who only solved problems by themselves during the study stage, in line with the worked example effect.

Obtaining the worked example effect may be explained by the human cognitive architecture outlined in Chapter 1. Based on the borrowing and reorganizing principle, novices under worked example conditions can imitate professional solutions which reduces extraneous load, so more working memory load is available to deal with intrinsic load, leading to better performance than problem solving. Solving a problem requires novices to randomly search solutions using the Randomness as Genesis Principle, imposing a high extraneous load which hinders learning.

The generation effect is not a cognitive load theory effect because it can be obtained using information that does not impose a heavy cognitive load. It can be explained by other factors discussed in Chapter 6. Firstly, during generation, participants might invest more cognitive effort than a presentation group, which may enhance memory traces, as indicated by Tyler, Hertel, McCallum and Ellis (1979). Secondly, the generation process might enhance the features of responses (Burns, 1990), therefore, participants in the generation group may more successfully recall the whole formulae (responses) than the presentation group and so performed better in the free recall test which is sensitive to response-oriented processing (Hirshman & Bjork, 1988).

Chapter 9 Experiment 2

Experiment 2 again tested for an interaction between levels of guidance and levels of element interactivity using older, more knowledgeable learners using similar materials to those of Experiment 1. It was hypothesized that the interaction should be reduced or eliminated using these students who had a reduced requirement for worked examples. Specifically, for low element interactivity materials, the generation effect should be obtained again with more knowledgeable students. However, for high element interactivity materials with learners at higher level of expertise, the worked example effect should be reduced, eliminated or even reversed due to the expertise reversal effect.

The general materials were similar to those used in Experiment 1. Low element interactivity materials were used to test for the generation effect by presenting learners with answers to memory formulae (high guidance) or having them generate answers themselves (low guidance), whereas high-element interactivity materials were used to test for the worked example effect by comparing studying worked examples (high guidance) with problem solving (low guidance). The same two phases of Experiment 1 were used in Experiment 2. Students firstly were presented the low element interactivity materials to test for the generation effect followed by the high element interactivity materials to test for the worked example effect.

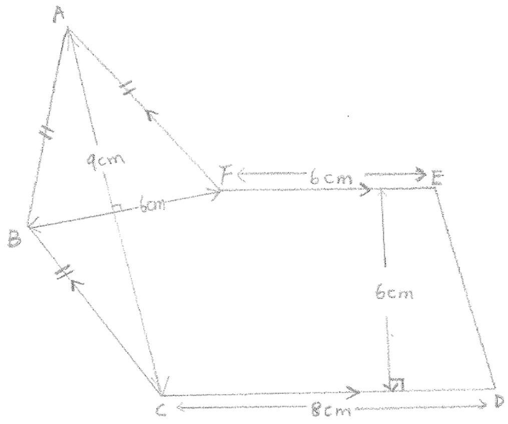
9.1 Method

9.1.1 Participants

The participants were 50 Grade 6 students, from a primary school in Chengdu, China. They were approximately 12 years old. The method of assigning students to the generation or presentation group followed by the worked example or problem solving group was identical to Experiment 1. Three students did not complete the entire procedure of the first phase of Experiment 2 and were eliminated from the data analysis, leaving 47 students. In previous classes, all students had studied most of the area and perimeter formulae used in this study to test for the generation effect and had been taught how to solve the problems used to test for the worked example effect. Therefore, Grade 6 students had been provided with more tuition prior to this experiment than the Grade 4 students tested in Experiment 1.

9.1.2 Materials

To test for the generation effect in this experiment, the same materials that had been used to test for the generation effect in Experiment 1 were used, except that all formulae were in algebraic form. However, the materials used to test for the worked example effect in Experiment 1 were changed in Experiment 2 to better fit Grade 6 students and facilitate the marking procedure (see Appendix 6: materials used for the worked example group; see Appendix 7: materials used for the problem solving group). In this experiment, the problems used in the second booklet that divided students into worked example and problem solving groups retained the first two problems that were concerned with calculating the area of a composite shape used in Experiment 1. The last two problems were changed to the similar problems (see an example in Appendix 6) used in the test phase of Experiment 1 (Questions 1, 2 and 3). Therefore, the test questions testing for the worked example effect used in the third booklet also had to be changed in this experiment. Only the first two test questions used in the test of Experiment 1 were used in the Experiment 2 test, with all of the test questions requiring a calculation of the area of a composite shape. Therefore, this test was different from the corresponding test used in Experiment 1 (see Appendix 8).



An Example from Appendix 6

When counting the number of interacting elements to evaluate the effective level of element interactivity in instructional materials of this experiment, the relatively higher level of learner expertise was taken into account. For example, for the area of a parallelogram formula in the material for testing the generation effect (used above in Experiment 1 to demonstrate the procedure), it was considered that Grade 6 students had acquired the relevant knowledge, so the relevant schema which combined single elements of this formula was already available in their long-term memory. Therefore, they did not need to consider the meaning of *l* and *h* separately, and the relation among *l*, *h* and the area formula. They could just use the stored schema to deal with this memorization task. Therefore, the number of interacting elements for Grade 6 students should be one (the relevant schema as a whole entity to be processed in working memory).

A similar reduction of the number of interacting elements applied to materials used for testing the worked example effect. For example, in the case of the task used above in Experiment 1 to illustrate the procedure, when calculating the area of a given composite shape, Grade 6 students are likely to have already acquired relevant knowledge allowing them to perceive and calculate the areas of a rhombus and a trapezium as single units to be processed in working memory. They are unlikely to need to consider the basic lines which form a rhombus or trapezium, as these individual elements have already been integrated into previously acquired schemas which can be regarded as a single entity when calculating the area of a rhombus or trapezium. But they are unlikely to be ready to retrieve the formulae for the areas of a rhombus or a trapezium together as a single entity nor are they likely to readily recognize a rhombus or trapezium within a complex shape consisting of a combination of both geometric shapes. In order to solve a question concerning a diagram consisting of a rhombus and trapezium combined, the more complex schema required is unlikely to have been acquired by these Grade 6 students. When faced with a figure consisting of a rhombus and a trapezium, we can estimate that they must recognize the rhombus within the complex figure (1st element), recognize the trapezium (2nd elements) and realize that the total area of the shape (3rd element) requires the addition (4th element) of both areas (5th and 6th elements).

9.1.3 Procedure

The general procedures were identical to Experiment 1.

9.1.4 Scoring

The scoring system for the generation effect was identical to Experiment 1. For the worked example effect, only the scoring system used to calculate the area of a composite shape was used. The internal reliability of the test for the worked example effect using Cronbach's α was .78, after eliminating the 1st test question to increase the reliability of the test.

9.2 Results and Discussion

Means and standard deviations of percentage test score results may be found in Table 3 and Table 4. These results were analysed using a 2 (levels of guidance) x 2 (levels of element interactivity) ANOVA with repeated measures on the element interactivity factor. All means and standard deviations presented in Table 3 and the statistical analyses following this table were based on the four test questions remaining after eliminating Question 1 and it should be noted that the patterns of significance were slightly changed compared to those obtained using all five test questions. All means and standard deviations presented in Table 4 and the description of statistical analyses following this table were based on the five test questions. Cohen's *d* was used through this experiment to indicate the effect size.

Table 3. *Mean (SD) Percentage Correct Test Score Results After Eliminating Q1 for Experiment 2*

Guidance	Low element interactivity	High element interactivity
High (N=22)	71.4 (20.15)	91.5 (10.83)
Low (N=25)	80.9 (13.82)	84.0 (18.67)

Note. The maximum score for Low element interactivity test was 11; the maximum score for High element interactivity test was 16.

Based on the data of Table 3, the main effect of guidance was not significant, F(1, 45) = .099, MSe = 255.89, p = .755, η_{p^2} = .002. The main effect of element interactivity was significant, F(1, 45) = 11.40, MSe = 276.22, p = .002. Wilks' Lambda = .798, η_{p^2} = .202. High element interactivity material percentage test scores were higher than the low element interactivity test scores. The interaction effect of levels of guidance and levels of element interactivity was of primary interest in this experiment and also was significant, F(1, 45) = 6.15, p = .017, Wilks' Lambda = .880, η_{p^2} = .120.

Following the significant interaction result, simple effects tests were conducted. For the low-element interactivity materials testing for the generation effect, the effect of guidance was significant, t(46) = -2.02, SE_{diff} = 4.89, p = .025 (1-tailed), Cohen's d = .57. The mean percentage correct scores indicated that low guidance was superior to high guidance, indicating the generation effect was obtained.

For the high element interactivity material testing for the worked example effect, the effect of guidance was also significant, t(46) = 1.90, SE_{diff} = 4.85, p = .03 (1-tailed), Cohen's d = .54. The mean percentage correct scores indicated that high guidance was superior to low guidance, indicating that the worked example effect was obtained. But the effect was smaller than the one obtained in Experiment 1, where the effect size was d = .64.

Table 4. *Mean (SD) Percentage Correct Test Score Results of All Five Questions for Experiment 2*

Guidance	Low element interactivity	High element interactivity
High (N=22)	71.4 (20.15)	90.9 (10.08)
Low (N=25)	80.9 (13.82)	85.6 (16.09)

Note. The maximum score for Low element interactivity test was 11; the maximum score for High element interactivity test was 20.

Based on the data of Table 4, the main effect of guidance was not significant, F(1, 45) = .466, MSe = 226.56, p = .498, η_{p^2}= .010. The main effect of element interactivity was significant, $F(1, 45) = 13.71$, $MSe = 250.39$, $p = .001$. Wilks' Lambda = .766, $\eta_{p^2} = .234$. The mean percentage correct scores indicated that high element interactivity scores were higher than low element interactivity scores. Finally, The interaction effect of guidance and element interactivity was of primary interest in this experiment and also was significant, $F(1, 45) = 5.16$, $p = .028$. Wilks' Lambda = .766, $\eta_{p^2} = .103$.

Following the significant interaction result, simple effects tests were conducted. For the low-element interactivity materials testing for the generation effect, the effect of guidance was significant, $t(46) = -2.02$, $SE_{diff} = -9.88$, $p = .025$ (1-tailed), Cohen's $d = .57$. The mean percentage correct scores indicated that low guidance was superior to the high guidance, indicating the generation effect was obtained.

For the high element interactivity material testing for the worked example effect, the effect of guidance was marginally significant indicating a possible trend, $t(46) = 1.60$, $SE_{diff} = 6.68$, $p = .06$ (1-tailed), Cohen's $d = .46$. The mean percentage correct scores indicated that high guidance was marginally ($p = .06$) superior to the low guidance, indicating a possible worked example effect, smaller than the one obtained in Experiment 1, where the effect size was $d = .68$. Note that a one-tailed test was used as I had a clear directional hypothesis.

It was hypothesized in Experiment 2 that the interaction of guidance and element interactivity should be reduced or eliminated using relatively more knowledgeable students who had a reduced requirement for worked examples. The generation effect should be obtainable with these students but with their higher levels of expertise, the worked example effect should be reduced, eliminated or even reversed due to the expertise reversal effect. This experiment provided some evidence supporting this hypothesis, although the evidence was not strong.

The same factors as those mentioned previously in Experiment 1 could be used to explain the obtained generation effect, but for the weaker worked example effect, the results require further discussion. According to the expertise reversal effect discussed in Chapter 5, with the increase of learners' expertise, high guidance may become redundant for these learners causing higher levels of extraneous load, but the worked example effect was still obtained (or a clear trend towards the effect was observed based on the marginally significant difference obtained when all five test questions were used) although with reduced effect size compared to Experiment 1. It indicated that participants in Experiment 2 were indeed more knowledgeable than those in Experiment 1, because the worked example effect was weaker (the effect size dropped from .64 to .54 for four questions and from .68 to .48 based on five questions), suggesting that they had a reduced requirement for external guidance (e.g., worked examples). However, they might not have as yet acquired complete schemas to be able to readily retrieve the required solutions as the worked examples were not fully redundant for these participants, therefore, they still needed some external guidance. The following experiment attempted to further reduce this need by using even more knowledgeable learners.

Chapter 10 Experiment 3

Experiment 3 again tested for an interaction between guidance and element interactivity with older, more expert learners using the same materials as in Experiment 2. It was again hypothesized that the interaction should be reduced or eliminated using students with higher levels of prior knowledge who had a reduced requirement for worked examples. It was expected that the generation effect should be easily obtainable with these more knowledgeable students but the worked example effect should be reduced, eliminated or even reversed due to the expertise reversal effect.

The general procedure in Experiment 3 was similar to that used in Experiments 1 and 2, including the same two phases.

10.1 Method

10.1.1 Participants

The participants were 38 Grade 7 students, from a secondary school in Chengdu, China. They were approximately 13 years old. They were randomly assigned to groups using the procedure of Experiments 1 and 2. Two students in the first phase of this experiment did not complete the entire procedure. These two students were eliminated from the data analysis, leaving 36 students. In class, all the participating students had previously studied the area and perimeter formulae used in this study to test for the generation effect. Similarly, all students had been taught to solve the problems used to test for the worked example effect approximately a year prior to the experiment. Therefore, Grade 7 students were regarded as relatively more experienced with respect to the formulae as well as the problems used in Experiment 2.

10.1.2 Materials

The experimental materials used in Experiment 2 were used again in Experiment 3.

When counting the number of interacting elements to evaluate the effective level of element interactivity in instructional materials of this experiment, the relatively higher level of learner expertise was also taken into account. For example, for the area of a parallelogram formula in the material for testing the generation effect (used above in Experiment 1 to demonstrate the procedure), it was considered that Grade 7 students who were more knowledgeable than Grade 6 students used in Experiment 2, could just use the stored schema to deal with this memorization task. Therefore, the number of interacting elements for Grade 7 students should be one (the relevant schema as a whole entity to be processed in working memory).

A similar reduction of the number of interacting elements applied to the material used for testing the worked example effect. For example, in the case of the task used above in Experiment 1 to illustrate the procedure, Grade 7 students are likely to have already acquired the relevant schemas for perceiving and calculating the areas of a rhombus and a trapezium, and could retrieve these schemas from their long-term memory as single units to be processed in working memory, thus reducing the number of interacting elements to two. Combining the values of these two areas using the schema for composite shapes as an entity results in the total number of interacting elements for this problem for Grade 7 students to be three, which is a low level of element interactivity.

10.1.3 Procedure

The general procedure was identical to that used in Experiments 1 and 2.

10.1.4 Scoring

The scoring procedure for the generation effect and the worked example effect was identical to that used in Experiment 2. The internal reliability of the test for the worked example effect using Cronbach's α was .90, after eliminating the 1st test question to increase the reliability of the test.

10.2 Results and Discussion

Means and standard deviations of percentage correct test score results may be found in Table 5. These results were analyzed using a 2 (levels of guidance) x 2 (levels of element interactivity) ANOVA with repeated measures on the element interactivity factor. All means and standard deviations presented in Table 5, as well as the following statistical analyses following the table were based on the four test questions remaining after eliminating Question 1 but it should be noted that the patterns of significance were identical to those obtained using all five test questions.

Table 5. *Mean (SD) Percentage Correct Test Score Results After Eliminating Q1 for Experiment 3*

Guidance	Low element interactivity	High element interactivity
High (N=19)	64.4 (19.24)	64.1 (35.95)
Low (N=17)	77.2 (9.61)	77.9 (24.91)

Note. The maximum score for Low element interactivity test was 11; the maximum score for High element interactivity test was 16.

Based on the data of Table 5, the main effect of guidance was significant, $F(1, 34) = 5.24$, $MSe = 605.98$, $p = .028$, $\eta_{p^2} = .134$. The main effect of element interactivity was not significant, $F(1, 34) = .001$, $MSe = 610.02$, $p = .971$. Wilks' Lambda = 1.00, $\eta_{p^2} = .000$. The interaction of guidance and element interactivity was of primary interest in this experiment, and it was not significant, $F(1, 34) < 1$, $p = .933$. Wilks' Lambda = 1.00, $\eta_{p^2} = 0$.

It was hypothesized that when using older, more knowledgeable students in Experiment 3, the interaction of guidance and element interactivity should be further reduced compared to the previous experiments or eliminated because the worked example effect could be eliminated or reversed with increases in expertise. Results of this experiment supported this hypothesis with no interaction of guidance and element interactivity obtained. Increased guidance had a similar negative effect on both higher and lower element interactivity materials. In other words, in contrast to Experiments 1 and 2, the generation effect was obtained for both lower and higher element interactivity materials with no sign of the worked example effect for the high element interactivity material.

The results of the above three experiments taken together represent an expertise reversal effect. With the gradual increase in the level of learner expertise, external guidance gradually became unnecessary for learners. High levels of guidance (e.g., worked examples) were redundant for the most knowledgeable participants, presumably leading to increased levels of extraneous load and relatively decreased performance compared to problem solving group. As for the generation effect for low element interactivity materials, the results of all three experiments confirmed our hypotheses and the explanations mentioned in the conclusion to Experiment 1 could be applied to all three cases of the observed generation effect.

Chapter 11 Experiment 4

The purpose of Experiment 4 and Experiment 5 was to test whether similar results to those obtained in the previous experiments generalized to more complex material studied by older, more sophisticated learners. Grade 10, secondary school learners studying trigonometry topics that were either high or low in element interactivity were used.

11.1 Method

11.1.1 Participants

The participants were 50 Grade 10 students from a secondary school in Chengdu, China. They were approximately 16 years old. Six students' data were not complete, so they were excluded for final analysis, leaving 44 students' performance for data analysis. They were randomly assigned to either the generation or presentation group in the first phase of the experiment and then to either the worked example or problem solving group in the second phase using the same method as in the previous three experiments. No students had studied trigonometry in class and so all were regarded as novices.

11.1.2 Materials

To test for the generation effect, ten trigonometry formulae were chosen from textbooks used in secondary school. There were five sine formulae and five cosine formulae (see Appendix 9).

Levels of element interactivity were estimated for all materials using the method illustrated in the previous experiments. As an example, students were asked to remember the formula, $\sin(A+B) = \sin A\cos B + \cos A\sin B$, a task used to test for the generation effect. Based on the concept of element interactivity, as Grade 10 students did not have relevant knowledge of the formulae, elements of the equation had to be memorized. For a simple memory test (as opposed to a test requiring students to use the formula), each of the elements is independent of all other elements. Whether a particular element is remembered or forgotten should not affect memory of any of the other elements. Accordingly, the element interactivity count is one because the elements do not interact.

In contrast, assuming students are asked to simplify a trigonometry formula, such as

$$
\begin{aligned}
&\sqrt{2}\sin(\alpha+\frac{\pi}{4})\\
&=\sqrt{2}\left(\sin\alpha\cos\frac{\pi}{4}+\cos\alpha\sin\frac{\pi}{4}\right)\\
&=\sqrt{2}\left(\frac{\sqrt{2}}{2}\sin\alpha+\frac{\sqrt{2}}{2}\cos\alpha\right)\\
&=\sin\alpha+\cos\alpha
\end{aligned}
$$

a task used to test for the worked example effect. Grade 10 students who had not learned previously how to simplify trigonometry expressions needed to remember the single symbols of the expression of the first line (e.g., “sin”, “+”), totaling 11 elements, including multiplication that is implied without a specific symbol. Next, in order to obtain the expression of the second line, students had to recall the relevant formula (sin (A+B) = sinAcosB + cosAsinB), totaling 16 elements, and apply it to the first line resulting in a formula including 19 elements. Then they needed to calculate the value of sine and cosine indicated in the third line, totaling eight elements. Finally, they needed to open the final brackets to give the last line, involving 17 elements. Therefore, in total, it can be estimated that about 71 interacting elements were involved in this task, which is a very complex (high in element interactivity) task for Grade 10 students, compared to memorizing the trigonometry formula considered above. Because they interact, if any of these elements were forgotten or incorrect, the other elements could not be processed. Of course, previous knowledge acquired when learning the formulae will reduce the number of interacting elements considerably.

Students were given three A4 papers when testing for the generation effect. The first paper contained ten basic trigonometry formulae and was common to both groups (see Appendix 9). The second paper differed for the generation and presentation groups. For the presentation group, the content was identical to the first paper, whereas for the generation group, only the names of each of the ten formulae were included as students needed to generate each formula by themselves (see Appendix 10). The last paper (common to both groups) was blank to allow students to write out their answers in the free recall test. They were required to write out as many of the formulae that they remembered from studying the first and second papers.

To test for the worked example effect, students again were given three A4 papers. The first paper was identical to the first paper used to test for the generation effect. Its function was revision of the previously learned information. The second paper differed for the worked example and problem solving groups. The paper for the worked example group contained two worked examples each followed by a similar problem for students to solve (see Appendix 11). Students in the problem-solving group were required to solve the same four problems by themselves with no worked examples provided (see Appendix 12). The last paper contained five test problems for students to solve (see Appendix 13).

11.1.3 Procedure

The general procedure for testing the generation and the worked example effects was identical to the procedure used in the previous experiments. When scoring the test used to test for the generation effect, a correct formula was awarded one mark. Therefore, the maximum score in the free recall test was 10. Each student's score out of 10 was converted into a percentage score for analysis providing the scores testing knowledge of the low-element interactivity material. In the second phase, students could obtain a maximum of three marks on each of the problems, one for correctly using the formula, one for correctly calculating the sine value, and one for correctly calculating the cosine value, to test for the worked example effect. Each student's total score out of 15 (five problems each with a maximum score of three) was converted to a percentage score for analysis. The internal reliability of this test using Cronbach's α was .76 after deleting the 4th test question, as around 75% students were able to fully solve this problem. These scores provided the dependent variable measuring knowledge of the high element interactivity material.

A week after the immediate test, 36 students (two students' data were excluded as they were not complete, leaving 34 students for final analysis) of the same class took the delayed tests for the worked example and generation effects. The content of the delayed tests was exactly the same as the immediate tests. Using Cronbach's α, the internal reliability of the test for the worked example effect was .75, again after deleting the 4th test question, with around 65% of students able to fully solve this problem.

11.2 Results and Discussion

Means and standard deviations of percentage test score results may be found in Table 6 (immediate test) and Table 7 (delayed test). These results were analyzed using a 2 (levels of guidance) x 2 (levels of element interactivity) ANOVA with repeated measures on the element interactivity factor. All means, standard deviations and statistical analyses were based on the four test questions remaining after eliminating Question 4. It should be noted that the patterns of significance were slightly different from those obtained using all five test questions in that the simple effects test for the worked example effect was only marginally significant. All other patterns of significance using all five test questions were identical to those reported above.

Table 6. *Mean (SD) Percentage Correct Test Score Results of Immediate Test After Eliminating Q4 for Experiment 4*

Guidance	Low element interactivity	High element interactivity
High (N=23)	70.4 (33.77)	22.1 (10.54)
Low (N=21)	71.4 (25.93)	11.1 (12.43)

Note. The maximum score for Low element interactivity test was 10; the maximum score for High element interactivity test was 12.

Based on the data of Table 6, the main effect of guidance was not significant, $F(1, 42) = .97$, MSe = 566.332, $p = .330$, $\eta_{p^2} = .023$. The main effect of element interactivity was significant, $F(1, 42) = 134.09$, MSe = 483.34, $p < .001$, Wilks' Lambda = .239, $\eta_{p^2} = .761$. The low element interactivity material percentage correct test scores were higher than the high element interactivity test scores. The interaction between guidance and element interactivity was of primary interest in this experiment, but was not significant, $F(1, 42) = 1.64$, MSe = 483.34, $p = .208$, Wilks' Lambda = .963, $\eta_{p^2} = .037$.

Table 7. *Mean (SD) Percentage Correct Test Score Results of Delayed Test After Eliminating Q4 for Experiment 4*

Guidance	Low element interactivity	High element interactivity
High (N=17)	44.7 (26.72)	19.6 (11.00)
Low (N=17)	65.9 (19.70)	11.8 (12.17)

Note. The maximum score for Low element interactivity test was 10; the maximum score for High element interactivity test was 12.

Based on the data of Table 7, the main effect of guidance was not significant, $F(1, 32) = 1.88$, MSe = 402.88, $p = .180$, $\eta_p^2 = .055$. The main effect of element interactivity was significant, $F(1, 32) = 94.31$, MSe = 282.78, $p < .001$, Wilks' Lambda = .253, $\eta_p^2 = .747$. The low element interactivity material percentage correct test scores were higher than the high element interactivity test scores. The interaction between guidance and element interactivity was significant, $F(1, 32) = 12.66$, MSe = 282.78, $p = .001$, Wilks' Lambda = .717, $\eta_p^2 = .283$.

Following the significant interaction, simple effects tests were conducted. For the low element interactivity material testing for the generation effect, the effect of guidance was significant, $t(32) = 2.63$, $SE_{diff} = 8.05$, $p = .007$ (1-tailed), $d = 0.83$. The mean percentage correct scores indicated that low guidance was superior to high guidance demonstrating a generation effect.

For the high element interactivity material testing for the worked example effect, the effect of guidance also was significant, $t(32) = -1.97$, $SE_{diff} = 3.98$, $p = .03$ (1- tailed), $d = 0.65$. The mean percentage correct scores indicated that high guidance was superior to low guidance demonstrating a worked example effect.

In Experiment 4, it was hypothesized that an interaction of guidance and element interactivity would be obtained. High guidance was predicted to be superior to low guidance using materials high in element interactivity, whereas low guidance was predicted to be superior to high guidance with materials low in element interactivity. The results of the immediate test of Experiment 4 showed no statistically significant interaction and therefore, the worked example and generation effects were not obtained in the immediate test. Interestingly, the results of the delayed test of Experiment 4 confirmed the hypothesis with an interaction of guidance and element interactivity obtained. The simple effect tests indicated that students who generated formulae during a study stage a week previously better memorized those formulae than students presented the formulae in the previous week, in line with the generation effect. For materials high in element interactivity, students who studied worked example-problem solving pairs were better at solving the delayed test problems than students who only solved problems by themselves during the study stage a week previously, in line with the worked example effect.

According to the review of relevant studies of the generation effect (See Table 1.), we assumed that the generation effect might be more likely to be obtained on a delayed test, which was confirmed in this experiment. The results might also support that the hypothesis that the process of generation itself could improve memory and that the effect of generation was more durable on memory (Schweickert et al., 1994).

Additionally, the worked example effect also was obtained on the delayed but not the immediate test. This finding will need more investigation, as the worked example effect is usually tested on an immediate test. However, in terms of obtaining the worked example effect itself with this group of participants, human cognitive architecture can again be used to explain the results. For novices, studying worked examples reduces extraneous load, leading to better performance than problem solving as more working memory resources are needed to solve problems than to process instructor provided solutions.

Chapter 12 Experiment 5

Experiment 5 again tested for an interaction between guidance and element interactivity with older, more expert learners using the same materials as in Experiment 4. It was hypothesized that the interaction should be reduced or eliminated using more expert students who had a reduced requirement for worked examples. The generation effect should be obtainable with more knowledgeable students but with increased expertise, the worked example effect should be reduced, eliminated or even reversed due to the expertise reversal effect. In effect, all of the materials should be low in element interactivity for these relatively knowledgeable students. The general procedure was similar to that used in previous experiments. The same two phases of the previous experiments also were used in Experiment 5.

12.1 Method

12.1.1 Participants

The participants were 52 Grade 11 students, from a secondary school in Chengdu, China. They were approximately 17 years old. Two students' data were excluded as they were incomplete, leaving 50 students. They were randomly assigned to groups using the procedure of the previous experiments. In class, all students previously had studied the trigonometry formulae used in this study to test for the generation effect. Similarly, all students had been taught how to simplify trigonometry expressions used to test for the worked example effect in their previous semester. Therefore, the Grade 11 students were regarded as relatively expert with respect to the formulae as well as the problems used in Experiment 5.

12.1.2 Materials

The test materials were similar to those used in Experiment 4 with the forth question used to test the worked example effect slightly changed to .

When counting the number of interacting elements to evaluate the effective level of element interactivity in instructional materials of this experiment, the relatively higher level of learner expertise was taken into account. Grade 11 students only needed to retrieve the schema which is a single element when testing for the generation effect, so the element interactivity count is one.

A similar reduction of the number of interacting elements applied to materials used for testing the worked example effect. For example, in the case of the task used above in Experiment 4 to illustrate the procedure, Grade 11 students are likely to have already acquired the relevant schemas for simplifying trigonometry expressions. They can transfer those schemas from their long-term memory as single units to be processed in working memory, thus reducing the number of interacting elements to one.

12.1.3 Procedure

The general procedure was identical to Experiment 4. A week after the immediate test, 50 students from the same class (four students of this class were absent, leaving 46 students for final analysis) were tested again with the same tests used for testing the generation and worked example effects a week earlier.

12.1.4 Scoring

The scoring procedure for the generation effect and the worked example effect was identical to that used in Experiment 4. The internal reliability of the immediate test for the worked example effect using Cronbach's α was .49, after eliminating the 4th test question, as around 80% of the students could fully solve the test. The content of the delayed tests was the same as the immediate tests. The internal reliability of the delayed test for the worked example effect using Cronbach's α was .29 after deleting the 4th test question as over 80% of the students could solve this problem. These low α levels require an explanation. The low consistency of the four test questions used may be due to the relative expertise of participants. In Experiment 5, learners were more knowledgeable than in Experiment 4. That knowledge allowed many to fully solve many of the problems. Where a full solution was not obtained, learners tended to obtain a score of 0 rather than 1 or 2, resulting in scores on each problem having a wide range. Many learners obtained a score of 3 on one problem but a score of 0 on another, a common occurrence on problem solving tasks. The consequence is low consistency using Cronbach's α.

12.2 Results and Discussion

Means and standard deviations of percentage correct test score results may be found in Table 8 (immediate test) and Table 9 (delayed test). These results were analyzed using a 2 (levels of guidance) x 2 (levels of element interactivity) ANOVA with repeated measures on the element interactivity factor. All means, standard deviations and statistical analyses following the tables were based on the four test questions remaining after eliminating Question 4 and it should be noted that the patterns of significance were changed compared to those obtained using all five test questions, with none of the significant results due to guidance indicated below when using four test questions obtained when using five test questions.

Table 8. *Mean (SD) Percentage Correct Test Score Results of Immediate Test After Eliminating Q4 for Experiment 5*

Guidance	Low element interactivity	High element interactivity
High (N=25)	87.6 (16.15)	76.7 (28.05)
Low (N=25)	97.2 (8.91)	82.7 (16.12)

Note. The maximum score for Low element interactivity test was 10; the maximum score for High element interactivity test was 12.

Based on the data of Table 8, the main effect of guidance was significant, F(1, 48) = 4.676, MSe = 325.289, p = .036, η_{p^2} = .089. The main effect of element interactivity was significant, F(1, 48) = 11.009, MSe = 368.206, p = .002. Wilks' Lambda = .813, η_{p^2} = .187. The low element interactivity material percentage correct test scores were higher than the high element interactivity test scores. The interaction of guidance and element interactivity was of primary interest in this experiment but was not significant, F(1, 48) = .220, MSe = 368.206, p =.641. Wilks' Lambda = .995, η_{p^2} = .005.

Table 9. *Mean (SD) Percentage Correct Test Score Results of Delayed Test After Eliminating Q4 for Experiment 5*

Guidance	Low element interactivity	High element interactivity
High (N=24)	88.3 (22.20)	75.7 (22.78)
Low (N=22)	95.5 (8.00)	83.7 (16.96)

Note. The maximum score for Low element interactivity test was 10; the maximum score for High element interactivity test was 12.

Based on the data of Table 9, the main effect of guidance was significant, F(1, 44) = 4.404, MSe = 298.653, p = .042, η_{p^2} =.091. The main effect of element interactivity was significant, F(1, 44) = 8.573, MSe = 397.942, p = .005. Wilks' Lambda = .837, η_{p^2} = .163. The low element interactivity material percentage correct test scores were higher than the high element interactivity test scores. The interaction of guidance and element interactivity was of primary interest in this experiment but was not significant, F(1, 44) = .012, MSe = 397.942, p = .915. Wilks' Lambda = 1.000, η_{p^2} = 0.

It was hypothesized that when using older, more knowledgeable students in Experiment 5, the interaction of guidance and element interactivity should be reduced compared to the Experiment 4 or eliminated. The worked example effect was predicted to be eliminated or reversed with increases in expertise thus reducing or eliminating the interaction. The results of this experiment supported this hypothesis with no interaction of guidance and element interactivity obtained. Instead, increased guidance had a negative effect with materials used to test for both the worked example and generation effects. In other words, in contrast to

Experiment 4, the generation effect was obtained on immediate and delayed tests for both sets of material with no sign of the worked example effect.

The results replicated the pattern of Experiment 3 with the generation effect obtained for both low and high element interactivity materials. In addition, Experiments 4 and 5 reflected an expertise reversal effect again. More knowledgeable learners who already had acquired relevant schemas were used in Experiment 5, therefore high guidance (e.g., worked examples) was redundant for them causing a high extraneous load.

PART 3: GENERAL DISCUSSION

Chapter 13 General Discussion

In this dissertation, five experiments were conducted to investigate the relations between the worked example effect and the generation effect by resolving an obvious contradiction between them. Worked examples which provide full guidance are superior to problem solving which has no guidance, demonstrating the worked example effect, whereas, generation which requires learners to self-generate answers is superior to the presentation which externally provides intact answers, demonstrating the generation effect. This contradiction was assumed to be resolved by considering materials with different levels of element interactivity. The worked example effect may occur using materials high in element interactivity, while, simpler materials are suitable for the generation effect (Hypothesis 1).

However, with an increase of learner expertise, the interaction of guidance and element interactivity should be reduced and then eliminated, as the worked example effect should be reversed giving the generation effect with the increase of learner expertise, demonstrating the expertise reversal effect (Hypothesis 2). These two main hypotheses were tested in the five experiments. Experiments 1 to 3 tested the two main hypotheses using an immediate test in the domain of geometry, whereas, Experiments 4 and 5 were designed to test the two hypotheses using trigonometry on both immediate and delayed tests. The next sections will summarize the results of the five experiments, discuss the results of five experiments accordingly and then will point out some educational implications of this study and its potential limitations.

13.1 Results and Discussion of Experiments 1, 2 and 3

Experiment 1 provided support for Hypothesis 1 by demonstrating a dis-ordinal interaction of guidance and element interactivity with Grade 4 students who were regarded as novices for the learning materials used. Specifically, for materials high in element interactivity, high guidance in the form of worked examples was superior to low guidance in the form of problems to solve, demonstrating the worked example effect. High guidance permits the borrowing and reorganizing principle to come into play because instructional information is provided to learners, allowing them to “borrow” information from instructors whereas low guidance requires the randomness as genesis principle to be used because learners must generate responses randomly if relevant information is not available to them. Borrowing information from others should reduce cognitive load compared to generating the information oneself. For materials low in element interactivity, low guidance in the form of learners generating formulae was superior to high guidance in the form of learners being presented the formulae indicating the generation effect. Reducing cognitive load by using the borrowing and reorganizing principle is unnecessary for low element interactivity material because the cognitive load is already low and indeed, based on the generation effect, attempts to reduce cognitive load are likely to be counterproductive.

The second hypothesis was that as levels of learner expertise increase, actual levels of element interactivity for these learners should decrease and so the interaction between element interactivity and guidance obtained in Experiment 1 should reduce or disappear to be replaced by superior performance by both the low guidance, generation and the problem solving groups over their high guidance controls represented by the presentation and worked examples groups. According to the environmental organizing and linking principle, with increased expertise, interacting elements should be incorporated into schemas held in long-term memory and so should no longer impose a heavy working memory load. If so, the worked example effect should no longer be obtainable.

In Experiment 2, similar materials were used to test Grade 6 students who had more knowledge in the domain of geometry. The interaction of guidance and element interactivity should be reduced or eliminated using students who had a reduced requirement for worked examples. The generation effect should be obtainable with more knowledgeable students but with increased expertise, the worked example effect should be reduced, eliminated or even reversed due to expertise reversal effects. There was some evidence for this hypothesis. Both the generation and worked example effects again were obtained although with reduced effect sizes compared to Experiment 1.

Experiment 3 again used similar materials to those presented to Grade 4 students in Experiment 1 and Grade 6 students in Experiment 2 but this time presented to Grade 7 students. For Grade 7 learners, the materials constituted revision or review because they had studied the topic a year previously. As assumed by the second hypothesis, the interaction of guidance and element interactivity disappeared with the increase in levels of expertise because the worked example effect reversed with the increase in expertise. The generation effect was still robust and was found for both sets of materials. In other words, low guidance was superior to high guidance for both sets of materials. Therefore, the second hypothesis was confirmed.

A comparison of the worked example effect in Experiments 1 and 2 and its reversal in Experiment 3 provides a clear example of the expertise reversal effect. In Experiment 1, the worked example group was superior to the problem solving group, demonstrating the worked example effect. That effect was reduced when using more knowledgeable students in Experiment 2 and totally reversed in Experiment 3 by using the most experienced students. The worked examples were redundant for the more knowledgeable learners and so instead of obtaining a worked examples effect, we obtained a generation effect. With increases in expertise, most cognitive load effects first disappear and then reverse. In the case of the worked example effect, studying worked examples is superior to solving problems when testing novices but this difference disappears and then reverses with increases in expertise in the domain (Kalyuga, Chandler, Tuovinen, & Sweller, 2001). The contrasting results of Experiments 1, 2 and 3 may be due entirely to the expertise reversal effect.

Indeed, it may be plausible to suggest that the interaction between levels of guidance and element interactivity found in Experiment 1 is itself a form of the expertise reversal effect. According to cognitive load theory, changes in expertise result in changes in element interactivity as interacting elements are subsumed into knowledge held in long-term memory resulting in changes in effective working memory capacity limits. Material that is high in element interactivity for novice learners should be lower in element interactivity for relatively more knowledgeable learners. Instead of having to deal with large numbers of interacting elements via the narrow limits of change principle, many elements can be dealt with simultaneously using the environmental organizing and linking principle. In Experiment 1, the low element interactivity material that yielded the generation effect consisted of information that learners could easily learn. They had sufficient knowledge to be able to acquire the information readily by generating it rather than having it presented. They did not have sufficient information to easily generate the problem solutions of the high element interactivity information. When using participants who did have sufficient information to generate solutions readily in Experiment 3, the generation effect was obtained for both sets of material.

Since its inception, cognitive load theory has been applied largely, though not entirely, to novices for whom the material they were required to learn in a given area was complex and difficult due to the working memory load that was imposed. The theory was never intended to apply to information that was difficult for reasons other than a heavy working memory load. Having to learn a large number of elements that do not interact provides an example of an area that can be difficult for students for reasons other than a heavy working memory load.

Once knowledge held in long-term memory renders information simple rather than complex, cognitive load theory has used the redundancy effect (Chandler & Sweller, 1991) to explain why information should be generated rather than presented (Kalyuga, Ayres, Chandler, & Sweller, 2003). Unnecessarily processing redundant information may increase cognitive load. It is possible that unnecessarily reading presented information may be more cognitively demanding than generating that information oneself when the information is highly familiar. Nevertheless, any one or a combination of the reasons discussed in the literature review chapters to this dissertation may provide suitable explanations of the generation effect.

13.2 Results and Discussion of Experiment 4 – 5

In Experiment 4 and 5, the materials were changed from geometry to the more difficult topics of trigonometry but using older, more knowledgeable learners. In Experiment 4, it was again hypothesized that an interaction of guidance and element interactivity would be obtained. High guidance was predicted to be superior to low guidance using materials high in element interactivity, whereas low guidance was predicted to be superior to high guidance with materials low in element interactivity. In addition, as both immediate and delayed tests were used, based on the previous literature, it was also assumed that the worked example effect would be obtained on an immediate test, while, the generation effect was more likely to be found on a delayed test. The results of the immediate test of Experiment 4 showed no statistically significant interaction and therefore, both the worked example and generation effects were not obtained in the immediate test. Interestingly, the results of the delayed test of Experiment 4 confirmed the hypothesis with an interaction of guidance and element interactivity obtained. The simple effect tests indicated that students who generated formulae during a study stage a week earlier better memorized those formulae than students presented the formulae, in line with the generation effect. For materials high in element interactivity, students who studied worked example-problem solving pairs were better at solving the delayed test problems than students who only solved problems by themselves during the study stage a week earlier, in line with the worked example effect. The results also supported the hypothesis that the generation effect was indeed more likely to be obtained on a delayed test with no sign of the worked example effect on an immediate test.

In Experiment 5, it was assumed that when using older, more knowledgeable students, the interaction of guidance and element interactivity should be reduced or eliminated compared to Experiment 4. The worked example effect was predicted to be eliminated or reversed with increases in expertise thus reducing or eliminating the interaction. Results of this experiment supported this hypothesis with no interaction of guidance and element interactivity obtained. Instead, increased guidance had a negative effect on both the materials used to test for the worked example and generation effects. In other words, in contrast to Experiment 4, the generation effect was obtained on both immediate and delayed tests for both sets of material with no sign of the worked example effect.

The failure to obtain the generation or the worked example effects on the immediate tests but the indication of both effects on the delayed tests in Experiment 4 requires further discussion. With respect to the generation effect, as can be seen from Table 1 which summarized studies of the generation effect in Chapter 6, the effect is somewhat more likely using delayed than immediate tests although many studies have obtained the effect on immediate tests. In Experiment 4, I failed to obtain the effect on the immediate test but obtained it on the delayed test, in conformity with the majority of studies.

Unlike the generation effect, the worked example effect has rarely been tested using delayed tests. It can readily be obtained using immediate tests (Sweller, et al. 2011). The current results suggest that subsequent work on the worked example effect should include delayed as well as immediate tests.

Based on the current results, any explanation of an increased likelihood of the generation effect on delayed as opposed to immediate tests also would need to explain the increased likelihood of the worked example effect on delayed tests. There is no obvious cognitive explanation. While the explanation of Schweickert et al. (1994) that generation results in stronger, more persistent memory traces is intuitively plausible, it cannot explain why generating problem solutions results in a decrease in delayed performance compared to studying those solutions. In other words, whether element interactivity is high or low affects whether the generation effect or a reverse generation effect is obtained but does not alter whether the two effects are more or less likely to be obtained on immediate or delayed tests. Perhaps any instructional technique that improves learning is more likely to leave a durable trace compared to a technique that results in less learning. Further research clearly is required.

13.3 Educational Implications

The results obtained in this series of experiments have clear educational implications. They suggest that when dealing with complex material that learners may have difficulty understanding, high levels of guidance are likely to result in enhanced performance over lower levels of guidance. In contrast, when dealing with simple material that is easy for students to understand either because there are few interacting elements or because previously high element interactivity material has been learned and incorporated into knowledge held in long-term memory, learners should practice generating responses rather than being shown them. Most curricula include both high and low element interactivity materials. Based on the current study, learners should be encouraged to generate responses when dealing with low element interactivity material but should have complex, high element interactivity concepts and procedures explicitly demonstrated. Considering the results obtained from the delayed test, contrary to some claims about the ineffectiveness of explicit instruction for longer-term learning, demonstrating worked examples on how to deal with complex materials to novice learners may have a durable, positive effect.

13.4 Limitations of This Study and Future Work

A major limitation of the current study is that whereas the concept of cognitive load was used to hypothesize and explain the findings, no independent measure of cognitive load was used. Cognitive load usually is assessed using subjective ratings of mental effort or task difficulty (Paas & van Merrienboer, 1993). It can be difficult to measure using young students tested under relatively standard, ecologically valid classroom conditions. Lee (2013) indicated that students younger than 15 years of age might not be suitable participants for using subjective ratings of cognitive load. As the participants used in the first three experiments were 10-13 years in age, subjective rating of cognitive load may not be appropriate. For experiments 4 and 5, as my main purpose was to replicate the obtained results of the previous three experiments, subjective ratings of cognitive load were again not used. In future studies, it may be useful to obtain independent measures of cognitive load.

From the perspective of cognitive load theory, some other cognitive load effects that are relevant to the worked example effect or the expertise reversal effect may be also considered, such as the guidance fading effect which gradually fades solutions with the increase of learner expertise or the isolated elements effect, discussed below. Guidance fading provides learners with partial worked examples with key steps to be completed by the learner. It has been suggested that this alternative format of regular worked examples may encourage students to fully study presented worked examples (Van Merrienboer & Krammer, 1987). Completion problems and guidance fading have elements of both worked examples and generation for different parts of a problem. It may be of interest to test whether worked example effects and generation effects are obtained on different sections of the same problem depending on whether a solution is presented or needs to be generated for a particular section, along with interactions with levels of element interactivity. For example, if the problem is low in element interactivity itself, such as a one-step problem, generation may be superior to worked examples, whereas, a different result may be obtained if the problem is high in element interactivity, such as make *a* the subject of the equation, $[(a+b)/c]-e=d$, with the full solution being:

$$[(a+b)/c]-e = d$$

$$[(a+b)/c] = d+e$$

$$(a+b) = (d+e)c$$

$$a = (d+e)c-b$$

$$a = dc+ec-b$$

If we omit steps backwards, such as the last step, we may require students to generate that step, which only requires them to open the bracket. With more steps omitted, worked examples may be presented in order to provide support for the increased number of interactive elements involved; if we omit steps forwards, worked examples may be effective as it may be more difficult to generate key steps at the beginning of a problem (Sweller et al., 2011).

Therefore, the relevant hypotheses may be that firstly, if we omit one step backwards, the generation effect may be found, as the last one step is relatively easy to figure out; secondly, if we omit steps forwards, the worked example effect may be found, as it may be more difficult to generate some key steps without external support initially; thirdly, the worked example effect may be obtained if we omit more than one step, as more interactive elements are involved. However, with an increase of learner expertise, the generation effect may be obtained for all parts of a problem.

The isolated element effect presents learners (especially novices) with some simple but logically connected concepts to learn initially in order to reduce the levels of element interactivity. The advantage of using isolated element material is that learners only need to consider the interactive elements within smaller rather than larger units, which may reduce levels of element interactivity (Sweller et al., 2011). Isolated elements reduce the intrinsic load of worked examples, with high element interactivity materials being artificially converted to materials low in element interactivity. Therefore, we may investigate relations between isolated forms of worked examples and generation as learners are presented with relatively low element interactivity materials to learn in the isolated format. We may test the hypothesis that requiring learners to generate responses to isolated elements while presented with worked examples for integrated elements is superior to presenting responses to isolated elements while requiring learners to generate responses to complete problems.

As an example, a 2 (format: isolated elements vs. integrated elements) x 2 (guidance: worked example vs. problem solving) experimental design may be used. The experiment would include two phases: in the first phase, novice students will be randomly assigned to the four groups: a) an isolated elements group with the problems presented as worked examples, in which students are instructed how to solve 103 x 97 by using the formula x^2-y^2 = (x-y)(x+y). The first worked example could be, 103 = 100 + 3, the second one, 97 = 100 - 3, the third one, (100 + 3)(100 - 3) = 100^2 - 3^2 and the last one is 100^2 - 3^2 = 10000 - 9 = 9991. The four worked examples would be presented in sequence; (b) an isolated elements group with the problems presented as problems to be solved, we can design a set of problems to ask students to generate solutions such as, 103 = 100 + ?; 97 = 100 - ?; (100 + ?)(100 - ?) = 100^2 - $?^2$; 100^2 - $?^2$ = 10000 - $?^2$. These questions also will be presented sequentially; (c) an integrated elements group with the problems presented as worked examples. For this condition, the whole worked example will be presented: $103 \times 97 = (100 + 3)(100 - 3) = 100^2 - 3^2 = 10000 - 9 = 9991$. For this case, the all the interactive elements are presented simultaneously, which will impose a higher intrinsic load than the sequential presentation; (d) an integrated elements group with the problems presented as problems to be solved. In this group, students are required to solve 103 x 97 by using the formula $x^2-y^2 = (x-y)(x+y)$ themselves without any worked examples presented. After the first phase, with learner expertise relatively increased, an integrated element format will be used for all four groups. For the two groups presented worked examples in the first phase, worked examples will again be used while for the two groups presented problems, the integrated element format will be presented as problem solving only. Subsequent experiments could test the effects of switching from worked examples to problems and vice-versa between the two phases.

The hypothesis for the test phase following the two learning phases is that an interaction between the format and guidance may be obtained. Specifically, the isolated elements group with the problems presented as problems to be solved in the first phase will be superior to the isolated elements group with the problems presented as worked examples, whereas, the integrated elements group with the problems presented as worked examples will outperform the integrated elements group with the problems presented as problems to be solved.

It is helpful to manipulate cognitive load by using different factors, like a variation of learners` expertise or a variation of learning material in the future research. It may be also necessary to test for cognitive load and thereby providing evidence for its successful operation.

The other limitations may be as follows: firstly, the delayed tests were as the same as immediate tests for testing the delayed generation and worked example effects in experiment

4, so the delayed effects may be caused by the practice effect; secondly, for the delayed tests, the small sample size of experiment 4 may be a limitation, as is the case for all reported experiments with small sample sizes. Therefore, as mentioned above, the relationship between the generation and the worked example effects on the delayed test requires more research.

13.5 Conclusions

The present study was established to initially resolve a contradiction between the worked example and the generation effects. The results of five reported experiments have revealed some evidence about how to resolve this contradiction. The three main conclusions are:

Firstly, the worked example effect occurs using materials which were complex for students, while simple materials can be used to obtain the generation effect. Therefore, the contradiction between the worked example and the generation effects may be resolved by considering different types of materials.

Secondly, the relations between the worked example and the generation effects might also depend on the factor of learner expertise. For novices, high guidance (e.g., worked examples) was needed when they were presented with complex problems, but with the increase of learner expertise, high guidance (e.g., worked examples) was redundant for them and the low guidance group outperformed the high guidance group. Therefore, a worked example effect was converted to a generation effect providing an example of the expertise reversal effect.

Last but not the least, the above relations can be extended to a delayed rather than an immediate test. This conclusion is interesting, as most studies within the framework of cognitive load theory tested the worked example effect directly after the study phase with rare testing of this effect on a delayed test. However, the generation effect was found more often on a delayed test, and this result was obtained in Experiment 4. Therefore, the worked examples and generation effects may both have durable, positive effects on learning under appropriate conditions.

References

Anderson, R. C., Goldberg, S. R., & Hidde, J. L. (1971). Meaningful processing of sentences. *Journal of Educational Psychology*, *62*, 395-399. doi: 10.1037/h0031625

Antonenko, P., & Niederhauser, D. S. (2010). The influence of leads on cognitive load and learning in a hypertext environment. *Computers in Human Behavior*, *26*, 140-150. doi: 10.1016/j.chb.2009.10.014

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review*, *22*, 425-438. doi: 10.1007/s10648-010-9130-y

Atkinson, Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*, 181-214. doi: 10.2307/1170661

Atkinson, 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

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, *2*, 89-195. doi: 10.1016/S0079-7421(08)60422-3

Ausubel, D. P. (1964). Some psychological and educational limitations of learning by discovery. *The Arithmetic Teacher*, *1*, 290-302.

Ayres, P. (2001). Systematic mathematical errors and cognitive load. *Contemporary Educational Psychology*, *26*, 227-248. doi: 10.1006/ceps.2000.1051

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. (2013). Can the isolated-elements strategy be improved by targeting points of high cognitive load for additional practice? *Learning and Instruction*, *23*, 115-124. doi: 10.1016/j.learninstruc.2012.08.002

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

Ayres, P., & Youssef, A. (2008). Investigating the influence of transitory information and motivation during instructional animations. *Paper presented at the Proceedings of the 8th international conference on International conference for the learning sciences-* Volume 1.

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

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

Baggett, P. (1984). Role of temporal overlap of visual and auditory material in forming dual media associations. *Journal of Educational Psychology*, *76*, 408-417. doi: 10.1037/0022-0663.76.3.408

Bandura, A. (1986). *Social foundations of thought and action*: Englewood Cliffs, NJ: Prentice Hall.

Barrows, H. S. (1980). *Problem-based learning: An approach to medical education*: Springer Publishing Company.

Bartlett, F. C. (1932). *Remembering: An experimental and social study*. Cambridge: Cambridge University.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*, 610-632. doi: 10.1016/0749-596X(89)90016-8

Begg, I., & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 553-563. doi: 10.1037/0278-7393.13.4.553

Begg, I., Vinski, E., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory & Cognition*, *19*, 487-497. doi:10.3758/BF03199571

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*, 277-287. doi: 10.1007/s11251-009-9105-x

Bobrow, S. A., & Bower, G. H. (1969). Comprehension and recall of sentences. *Journal of Experimental Psychology*, *80*, 455-461. doi: 10.1037/h0027461

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, *25*, 24-34. doi: 10.1016/j.learninstruc.2012.11.002

Bourne, L. E., Goldstein, S., & Link, W. E. (1964). Concept learning as a function of availability of previously presented information. *Journal of Experimental Psychology*, *67*, 439-448. doi: 10.1037/h0043205

Britton, B. K., & Tesser, A. (1982). Effects of prior knowledge on use of cognitive capacity in three complex cognitive tasks. *Journal of Verbal Learning and Verbal Behavior*, *21*, 421-436. doi: 10.1016/S0022-5371(82)90709-5

Bruner, J. S. (1961). The act of discovery. *Harvard educational review*.

Bruner, J. S., Goodnow, J., & Austin, G. (1956). A Study of Thinking. *New York: John Wiley & Sons, Inc*, 14, 330.

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., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology*, *49*, 109-119. doi: 10.1027//1618-3169.49.2.109

Brunstein, A., Betts, S., & Anderson, J. R. (2009). Practice enables successful learning under minimal guidance. *Journal of Educational Psychology*, *101*, 790-802. doi: 10.1037/a0016656

Burns, D. J. (1990). The generation effect: A test between single-and multifactor theories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 1060-1067. doi: 10.1037/0278-7393.16.6.1060

Burns, D. J. (1992). The consequences of generation. *Journal of Memory and Language*, *31*, 615-633. doi: 10.1016/0749-596X(92)90031-R

Burns, D. J. (1996). The item-order distinction and the generation effect: The importance of order information in long-term memory. *The American Journal of Psychology*, *109*, 567-580. doi: 10.2307/1423395

Burns, D. J., Curti, E. T., & Lavin, J. C. (1993). The effects of generation on item and order retention in immediate and delayed recall. *Memory & Cognition*, *21*, 846-852. doi:10.3758/BF03202752

Buyer, L. S., & Dominowski, R. L. (1989). Retention of solutions: It is better to give than to receive. *The American Journal of Psychology*, 353-363. doi: 10.2307/1423055

Carroll, M., & Nelson, T. O. (1993). Effect of overlearning on the feeling of knowing is more detectable in within-subject than in between-subject designs. *The American Journal of Psychology*, 227-235. doi: 10.2307/1423169

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

Catrambone, R. (1994). Improving examples to improve transfer to novel problems. *Memory & Cognition*, *22*, 606-615. doi: 10.3758/BF03198399

Catrambone, R. (1995). Aiding subgoal learning: Effects on transfer. *Journal of Educational Psychology*, *87*, 5-17. doi: 10.1037/0022-0663.87.1.5

Catrambone, R. (1996). Generalizing solution procedures learned from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1020-1031. doi: 10.1037/0278-7393.22.4.1020

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General*, *127*, 355-376. doi: 10.1037/0096-3445.127.4.355

Catrambone, R., & Holyoak, K. J. (1990). Learning subgoals and methods for solving probability problems. *Memory & Cognition*, *18*, 593-603. doi:10.3758/BF03197102

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*, 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*, 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. (1973). Perception in chess. *Cognitive Psychology*, *4,* 55-81. doi**:** 10.1016/0010-0285(73)90004-2

Chechile, R. A., & Soraci, S. A. (1999). Evidence for a multiple-process account of the generation effect. *Memory*, *7*, 483-508. doi: 10.1080/741944921

Chen, S., Epps, J., & Chen, F. (2011). A comparison of four methods for cognitive load measurement. *Paper presented at the Proceedings of the 23rd Australian Computer-Human Interaction Conference*.

Chi, M. T. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. *Advances in Instructional Psychology, 5*, 161-238.

Chi, M. T., 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:10.1207/s15516709cog1302_1

Chi, M. T., 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., Glaser, R., & Rees, E. (1982). *Expertise in problem solving*. Sternberg RJ, Advances in The Psychology of Human Intelligence Vol. 1, 1982: Erlbaum, Hillsdale, NJ.

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

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*, 219-243. doi: 10.1080/00461520.2014.965823

Choi, H. H., van Merriënboer, J. J., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load. *Educational Psychology Review, 26*, 225-244. doi: 10.1007/s10648-014-9262-6

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*, 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*: Pfeiffer. San Francisco.

Clark, R. C., & Mayer, R. E. (2008). Learning by viewing versus learning by doing: Evidence-based guidelines for principled learning environments. *Performance Improvement*, *47*, 5-13. doi: 10.1002/pfi.20028

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*, 15-24. doi: 10.1007/BF02504794

Clifford, M. M. (1984). Thoughts on a theory of constructive failure. *Educational Psychologist*, *19*, 108-120. doi:10.1080/00461528409529286

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser, 18*, 32-42.

Conati, C., & VanLehn, K. (1999). *Teaching meta-cognitive skills: implementation and evaluation of a tutoring system to guide self-explanation while learning from examples*: Artificial Intelligence in Education, Amsterdam: IOS Press.

Conati, C., & VanLehn, K. (2000). Further results from the evaluation of an intelligent computer tutor to coach self-explanation. *Paper presented at the Intelligent Tutoring Systems*.

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

Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences*, *24,* 154-176.

Craig, R. C. (1956). Directed versus independent discovery of established relations. *Journal of Educational Psychology*, *47,* 223-234. doi: 10.1037/h0046768

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268-294. doi: 10.1037/0096-3445.104.3.268

Cronbach, L. J. (1967). How can instruction be adapted to individual differences. *Learning and Individual Differences*, 23-39.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods*: A handbook for research on interactions: Irvington.

Crutcher, R. J., & Healy, A. F. (1989). Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 669-675. doi: 10.1037/0278-7393.15.4.669

Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: Use of diagrams in computer-based training environments. *Instructional Science, 30*, 433-464. doi: 10.1023/A:1020516301541

Curry, L. A. (2004). The effects of self-explanations of correct and incorrect solutions on algebra problem-solving performance. *Paper presented at the Proceedings of the 26th annual conference of the cognitive science society*.

Darabi, A. A., Nelson, D. W., & Palanki, S. (2007). Acquisition of troubleshooting skills in a computer simulation: Worked example vs. conventional problem solving instructional strategies. *Computers in Human Behavior, 23*, 1809-1819. doi: 10.1016/j.chb.2005.11.001

De Croock, M. B., van Merriënboer, J. J., & Paas, F. G. (1998). High versus low contextual interference in simulation-based training of troubleshooting skills: Effects on transfer performance and invested mental effort. *Computers in Human Behavior*, *14*, 249-267. doi:10.1016/S0747-5632(98)00005-3

De Groot, A. (1965). *Thought andchoice in chess*. The Hague: Mouton.

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

De Winstanley, P. A., & Bjork, E. L. (1997). Processing instructions and the generation effect: A test of the multifactor transfer-appropriate processing theory. *Memory*, *5*, 401-422. doi: 10.1080/741941392

De Winstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition*, *32*, 945-955. doi:10.3758/BF03196872

Dick, M. B., Kean, M. L., & Sands, D. (1989). Memory for internally generated words in Alzheimer-type dementia: Breakdown in encoding and semantic memory. *Brain and Cognition, 9*, 88-108. doi**:** 10.1016/0278-2626(89)90046-8

Dixon, P. (1987). The processing of organizational and component step information in written directions. *Journal of Memory and Language, 26*, 24-35. doi: 10.1016/0749-596X(87)90060-X

Doherty-Sneddon, G., & Phelps, F. (2005). Gaze aversion: a response to cognitive or social difficulty? *Memory & Cognition, 33*, 727-733. doi:10.3758/BF03195338

Donaldson, W., & Bass, M. (1980). Relational information and memory for problem solutions. *Journal of Verbal Learning and Verbal Behavior, 19*, 26-35. doi: 10.1016/S0022-5371(80)90488-0

Dufresne, R. J., Gerace, W. J., Hardiman, P. T., & Mestre, J. P. (1992). Constraining novices to perform expertlike problem analyses: Effects on schema acquisition. *The Journal of the Learning Sciences, 2*, 307-331. doi:10.1207/s15327809jls0203_3

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*, 206-214. doi: 10.1016/j.learninstruc.2011.11.001

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition, 7*, 149-158. doi:10.3758/BF03197595

Erez, A., & Isen, A. M. (2002). The influence of positive affect on the components of expectancy motivation. *Journal of Applied Psychology, 87*, 1055-1067. doi: 10.1037//0021-9010.87.6.1055

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

Eylon, B. S., & Reif, F. (1984). Effects of knowledge organization on task performance. *Cognition and Instruction, 1*, 5-44. doi:10.1207/s1532690xci0101_2

Fiedler, K., Lachnit, H., Fay, D., & Krug, C. (1992). Mobilization of cognitive resources and the generation effect. *The Quarterly Journal of Experimental Psychology, 45*, 149-171. doi:10.1080/14640749208401320

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*, 216-224. doi: 10.1016/j.learninstruc.2009.02.021

Flory, P., & Pring, L. (1995). The effects of data-driven and conceptually driven generation of study items on direct and indirect measures of memory. *The Quarterly Journal of Experimental Psychology, 48*, 153-165. doi:10.1080/14640749508401382

Foos, P. W., Mora, J. J., & Tkacz, S. (1994). Student study techniques and the generation effect. *Journal of Educational Psychology, 86*, 567-576. doi: 10.1037/0022-0663.86.4.567

Gabrys, G., Weiner, A., & Lesgold, A. (1993). *Learning by problem solving in a coached apprenticeship system*. In M. Rabinowitz (Ed.), Cognitive science foundations of instruction (pp. 119-147). Hillsdale, NJ: Lawrence Erlbaum Associates.

Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology, 83*, 269-275. doi:10.1016/j.ijpsycho.2011.09.023

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16*, 309-313. doi:10.3758/BF03197041

Gardiner, J. M. (1989). A generation effect in memory without awareness. *British Journal of Psychology, 80*, 163-168. doi: 10.1111/j.2044-8295.1989.tb02310.x

Gardiner, J. M., & Arthurs, F. S. (1982). Encoding context and the generating effect in multitrial free-recall learning. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 36*, 527-531. doi: 10.1037/h0080651

Gardiner, J. M., Dawson, A. J., & Sutton, E. A. (1989). Specificity and generality of enhanced priming effects for self-generated study items. *The American Journal of Psychology*, 295-305. doi: 10.2307/1423051

Gardiner, J. M., Gregg, V. H., & Hampton, J. A. (1988). Word frequency and generation effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 687-693. doi: 10.1037/0278-7393.14.4.687

Gardiner, J. M., & Hampton, J. A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 732-741. doi: 10.1037/0278-7393.11.1-4.732

Gardiner, J. M., & Hampton, J. A. (1988). Item-specific processing and the generation effect: Support for a distinctiveness account. *The American Journal of Psychology*, *101*, 495-504. doi: 10.2307/1423227

Gardiner, J. M., & Rowley, J. M. (1984). A generation effect with numbers rather than words. *Memory & Cognition, 12*, 443-445. doi:10.3758/BF03198305

Geary, D. C. (2007). *Educating the evolved mind: Conceptual foundations for an evolutionary educational psychology*: IAP.

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

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science, 7*, 155-170. doi: 10.1016/S0364-0213(83)80009-3

Gerě, I., & Jaušcvec, N. (1999). Multimedia: Differences in cognitive processes observed with EEG. *Educational Technology Research and Development, 47*, 5-14. doi: 10.1007/BF02299630

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. (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*, 360-370. doi: 10.1016/j.chb.2008.12.015

Ghatala, E. S. (1983). When does internal generation facilitate memory for sentences? *The American Journal of Psychology, 96*, 75-83. doi: 10.2307/1422210

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306-355. doi:10.1016/0010-0285(80)90013-4

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

Glisky, E. L., & Rabinowitz, J. C. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 193-205. doi: 10.1037/0278-7393.11.2.193

Graf, P. (1980). Two consequences of generating: Increased inter-and intraword organization of sentences. *Journal of Verbal Learning and Verbal Behavior, 19*, 316-327. doi: 10.1016/S0022-5371(80)90248-0

Graf, P. (1981). Reading and generating normal and transformed sentences. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 35,* 293-308. doi: 10.1037/h0081193

Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography. *Experimental Brain Research, 112,* 103-111. doi: 10.1007/BF00227183

Greenwald, A. G., & Johnson, M. M. (1989). The generation effect extended: Memory enhancement for generation cues. *Memory & Cognition, 17,* 673-681. doi: 10.3758/BF03202628

Griffith, D. (1976). The attentional demands of mnemonic control processes. *Memory & Cognition, 4*, 103-108. doi:10.3758/BF03213261

Grosofsky, A., Payne, D. G., & Campbell, K. D. (1994). Does the generation effect depend upon selective displaced rehearsal? *The American Journal of Psychology*, *107*, 53-68. doi: 10.2307/1423289

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction, 17*, 612-634. doi: 10.1016/j.learninstruc.2007.09.008

Hausmann, R. G., & Chi, M. H. (2002). Can a computer interface support self-explaining. *Cognitive Technology*, *7,* 4-14.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*, 684-689. doi:10.1037/0022-0663.91.4.684

Hertel, P. T. (1989). The generation effect: A reflection of cognitive effort? *Bulletin of the Psychonomic Society*, *27*, 541-544. doi: 10.3758/BF03334663

Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 484-494. doi: 10.1037/0278-7393.14.3.484

Hock, H. S., Throckmorton, B., Webb, E., & Rosenthal, A. (1981). The effect of phonemic processing on the retention of graphemic representations for words and nonwords. *Memory & Cognition, 9*, 461-471. doi:10.3758/BF03202340

Huang, T. H., Liu, Y. C., & Shiu, C.Y. (2008). Construction of an online learning system for decimal numbers through the use of cognitive conflict strategy. *Computers & Education, 50*, 61-76. doi: 10.1016/j.compedu.2006.03.007

Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction, 20*, 18-29. doi: 10.1016/j.learninstruc.2008.12.001

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

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*, 2526-2528. doi: 10.1126/science.286.5449.2526

Jablonka, E., Lamb, M., & Zeligowski, A. (2014). *Evolution in four dimensions : Genetic, epigenetic, behavioral, and symbolic variation in the history of life* (Rev ed., Life and mind). Cambridge, Massachusetts: MIT Press.

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22*, 485-508. doi: 10.1016/S0022-5371(83)90301-8

Jacoby, L. L., Craik, F. I., & Begg, I. (1979). Effects of decision difficulty on recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 18*, 585-600. doi: 10.1016/S0022-5371(79)90324-4

Jamet, E., & Le Bohec, O. (2007). The effect of redundant text in multimedia instruction. *Contemporary Educational Psychology, 32,* 588-598. doi: 10.1016/j.cedpsych.2006.07.001

Java, R. I. (1994). States of awareness following word stem completion. *European Journal of Cognitive Psychology, 6*, 77-92. doi:10.1080/09541449408520135

Jeffries, R., Turner, A. T., Polson, P. G., & Atwood, M. (1981). *The processes involved in software design*. Cognitive Skills and their Acquisition, 254-284.

Johns, E. E., & Swanson, L. G. (1988). The generation effect with nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 180-190. doi: 10.1037/0278-7393.14.1.180

Johnson, M. K., Raye, C. L., Foley, H. J., & Foley, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *The American Journal of Psychology*, *94*, 37-64. doi: 10.2307/1422342

Johnson, M. M., Schmitt, F. A., & Pietrukowicz, M. (1989). The memory advantages of the generation effect: Age and process differences. *Journal of gerontology*, *44*, 91-94. doi: 10.1093/geronj/44.3.P91

Jonassen, D. H. (1997). Instructional design models for well-structured and III-structured problem-solving learning outcomes. *Educational Technology Research and Development, 45*, 65-94. doi: 10.1007/BF02299613

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8,* 441-480. doi: 10.1016/0010-0285(76)90015-3

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583-1585. doi: 10.1126/science.154.3756.1583

Kalyuga, S. (2006). Assessment of learners' organised knowledge structures in adaptive learning environments. *Applied Cognitive Psychology, 20*, 333-342. doi: 10.1002/acp.1249

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. (2008). When less is more in cognitive diagnosis: A rapid online method for diagnosing learner task-specific expertise. *Journal of Educational Psychology, 100,* 603-612. doi: 10.1037/0022-0663.100.3.603

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

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

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13,* 351-371. doi: 10.1002/acp.1773

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*, 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*, 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: The Journal of the Human Factors and Ergonomics Society, 46*, 567-581. doi: 10.1518/hfes.46.3.567.50405

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., & Sweller, J. (2004). Measuring Knowledge to Optimize Cognitive Load Factors During Instruction. *Journal of Educational Psychology*, *96*, 558-568. doi: 10.1037/0022-0663.96.3.558

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*, 379-424. doi: 10.1080/07370000802212669

Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science, 38*, 523-550. doi: 10.1007/s11251-009-9093-x

Kapur, M. (2011). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science, 39*, 561-579. doi: 10.1007/s11251-010-9144-3

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*, 651-672. doi: 10.1007/s11251-012-9209-6

Kapur, M., & Bielaczyc, K. (2011). Classroom-based experiments in productive failure. *Paper presented at the Proceedings of the 33rd annual conference of the cognitive science society* (pp 2812-2817).

Kapur, M., & Rummel, N. (2012). Productive failure in learning from generation and invention activities. *Instructional Science*, *40*, 645-650. doi: 10.1007/s11251-012-9235-4

Kazemi, F., & Ghoraishi, M. (2012). Comparison of Problem-based Learning Approach and traditional teaching on attitude, misconceptions and mathematics performance of University Students. *Procedia-Social and Behavioral Sciences, 46*, 3852-3856. doi:10.1016/j.sbspro.2012.06.159

Keislar, E. R., & Shulman, L. M. (1966). Learning by Discovery: A Critical Appraisal: *Proceedings of a Conference on Learning by Discovery: Rand McNally*.

Kerr, B. (1973). Processing demands during mental operations. *Memory & Cognition, 1*, 401-412. doi: 10.3758/BF03208899

Kester, L., Kirschner, P. A., & Merriënboer, J. J. (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

Khawaja, M. A., F. Chen, and N. Marcus (2010). “Using Language Complexity to Measure Cognitive Load for Adaptive Interaction Design.” *Proceedings of International Conference on Intelligent User Interfaces (IUI 2010)*, pp. 333–336. Hong Kong, China: ACM.

Kinjo, H., & Snodgrass, J. G. (2000). Does the generation effect occur for pictures? *The American Journal of Psychology*, *113*, 95-121. doi: 10.2307/1423462

Kinoshita, S. (1989). Generation enhances semantic processing? The role of distinctiveness in the generation effect. *Memory & Cognition*, *17*, 563-571. doi: 10.3758/BF03197079

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. *Computers in Human Behavior*, *25,* 306-314. doi:10.1016/j.chb.2008.12.008

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*, 75-86. doi: 10.1207/s15326985ep4102_1

Kirsner, K. (1973). An analysis of the visual component in recognition memory for verbal stimuli. *Memory & Cognition, 1*, 449-453. doi: 10.3758/BF03208907

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction effects of direct instruction and discovery learning. *Psychological Science, 15*, 661-667. doi: 10.1111/j.0956-7976.2004.00737.x

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*, 239-264. doi: 10.1007/s10648-007-9049-0

Kolers, P. A. (1979). Reading and knowing. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 33*(2), 106-117. http://dx.doi.org/10.1037/h0081704

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

Kyun, S., Kalyuga, S., & Sweller, J. (2013). The Effect of Worked Examples When Learning to Write Essays in English Literature. *The Journal of Experimental Education, 81,* 385-408. doi: 10.1080/00220973.2012.727884

Lai, P., & Tang, K. C. C. (1999). Constraints affecting the implementation of Problem-Based Learning (PBL) strategy in university courses. *In J. Marsh (Ed.), Implementing problem based learning: Proceedings of the First Asia Pacific Conference on Problem Based Learning,* pp. 49-54.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*, 1335-1342.

Lee, H. (2013). Measuring cognitive load with electroencephalography and self-report: focus on the effect of English-medium learning for Korean students. *Educational Psychology*, *34*, 838-848. doi: 10.1080/01443410.2013.860217

Lee, C. H., & Kalyuga, S. (2011). Effectiveness of different pinyin presentation formats in learning Chinese characters: A cognitive load perspective. *Language Learning*, *61*, 1099-1118. doi: 10.1111/j.1467-9922.2011.00666.x

Lee, H., Plass, J. L., & Homer, B. D. (2006). Optimizing cognitive load for learning from computer-based science simulations. *Journal of Educational Psychology, 98*, 902-913. http://dx.doi.org/10.1037/0022-0663.98.4.902

Leopold, C., Doerner, M., Leutner, D., & Dutke, S. (2015). Effects of strategy instructions on learning from text and pictures. *Instructional Science, 43*, 345-364. doi: 10.1007/s11251-014-9336-3

Leopold, C., & Leutner, D. (2012). Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learning and Instruction*, *22*, 16-26. doi: 10.1016/j.learninstruc.2011.05.005

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (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., & Van Merrienboer, J. J. (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

Liu, I. M., & Lee, Y. S. (1990). Memorial consequences of generating words and non-words. *The Quarterly Journal of Experimental Psychology, 42*, 255-278. doi: 10.1080/14640749008401221

Lutz, J., Briggs, A., & Cain, K. (2003). An examination of the value of the generation effect for learning new material. *The Journal of General Psychology, 130*, 171-188. doi: 10.1080/00221300309601283

MacLeod, C. M., & Daniels, K. A. (2000). Direct versus indirect tests of memory: Directed forgetting meets the generation effect. *Psychonomic Bulletin & Review, 7*, 354-359. doi: 10.3758/BF03212993

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*, 252-271. http://dx.doi.org/10.1037/0033-295X.87.3.252

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

Marsh, E. J., Edelman, G., & Bower, G. H. (2001). Demonstrations of a generation effect in context memory. *Memory & Cognition, 29*, 798-805. doi: 10.3758/BF03196409

Marshall, S. P. (1993). Assessing schema knowledge. *Test theory for a new generation of tests*, 155-180.

Marshall, S. P. (1995). *Schemas in problem solving*: Cambridge University Press.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*, 14-19. http://dx.doi.org/10.1037/0003-066X.59.1.14

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*: Cambridge University Press.

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction, 20,*167-171. doi: 10.1016/j.learninstruc.2009.02.012

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83,* 484-490. doi: 10.1037/0022-0663.83.4.484

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84,* 444-452. http://dx.doi.org/10.1037/0022-0663.84.4.444

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93,* 187-198. http://dx.doi.org/10.1037/0022-0663.93.1.187

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

Mayer, R. E., Stiehl, C. C., & Greeno, J. G. (1975). Acquisition of understanding and skill in relation to subjects' preparation and meaningfulness of instruction. *Journal of Educational Psychology, 67,* 331-350. http://dx.doi.org/10.1037/h0076619

McDaniel, M. A., & Waddill, P. J. (1990). Generation effects for context words: Implications for item-specific and multifactor theories. *Journal of Memory and Language, 29*, 201-211. doi:10.1016/0749-596X(90)90072-8

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language, 27,* 521-536. doi:10.1016/0749-596X(88)90023-X

McElroy, L. A. (1987). The generation effect with homographs: Evidence for postgeneration processing. *Memory & Cognition, 15*, 148-153. doi: 10.3758/BF03197026

McElroy, L. A., & Slamecka, N. J. (1982). Memorial consequences of generating nonwords: Implications for semantic-memory interpretations of the generation effect. *Journal of Verbal Learning and Verbal Behavior, 21*, 249-259. doi:10.1016/S0022-5371(82)90593-X

McFarland, C. E., Frey, T. J., & Rhodes, D. D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning and Verbal Behavior, 19*, 210-225. doi:10.1016/S0022-5371(80)90182-6

McFarland, C. E., Warren, L. R., & Crockard, J. (1985). Memory for self-generated stimuli in young and old adults. *Journal of Gerontology, 40*, 205-207. doi: 10.1093/geronj/40.2.205

McLaren, B. M., Lim, S. J., Gagnon, F., Yaron, D., & Koedinger, K. R. (2006). Studying the effects of personalized language and worked examples in the context of a web-based intelligent tutor. *Paper presented at the Intelligent tutoring systems*.

McNamara, D. S. (1995). Effects of prior knowledge on the generation advantage: Calculators versus calculation to learn simple multiplication. *Journal of Educational Psychology, 87*, 307-318. http://dx.doi.org/10.1037/0022-0663.87.2.307

McNamara, D. S., & Healy, A. F. (1995). A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems. *Journal of Memory and Language, 34*, 399-416. doi:10.1006/jmla.1995.1018

McNamara, D. S., & Healy, A. F. (2000). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory and Language, 43*, 652-679. doi:10.1006/jmla.2000.2720

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43. doi: 10.1207/s1532690xci1401_1

Miller, D. (2010). Using a three-step method in a calculus class: Extending the worked example. *College Teaching, 58*, 99-104. doi: 10.1080/87567550903521249

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

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*, 156-163. http://dx.doi.org/10.1037/0022-0663.94.1.156

Mulligan, N. W., & Duke, M. D. (2002). Positive and negative generation effects, hypermnesia, and total recall time. *Memory & Cognition, 30*, 1044-1053. doi: 10.3758/BF03194322

Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction, 16*, 173-199. doi: 10.1207/s1532690xci1602_2

Nairne, J. S., Pusen, C., & Widner, R. L. (1985). Representation in the mental lexicon: Implications for theories of the generation effect. *Memory & Cognition, 13*, 183-191. doi: 10.3758/BF03197011

Neubauer, A., Freudenthaler, H. H., & Pfurtscheller, G. (1995). Intelligence and spatiotemporal patterns of event-related desynchronization (ERD). *Intelligence, 20*, 249-266. doi:10.1016/0160-2896(95)90010-1

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University.

Newell, A., & Simon, H. A. (1972). *Human problem solving*: Prentice-Hall Englewood Cliffs, NJ.

Nicolas, S. (1996). The generation effect in a word-stem completion task: The influence of conceptual processes. *European Journal of Cognitive Psychology, 8*, 405-424. doi: 10.1080/713752536

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

Ohlsson, S. (1996). Learning from performance errors. *Psychological Review, 103*, 241-262. http://dx.doi.org/10.1037/0033-295X.103.2.241

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

Olofsson, U., & Nilsson, L. G. (1992). The generation effect in primed word-fragment completion reexamined. *Psychological Research, 54*, 103-109. doi: 10.1007/BF00937138

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

Owens, P., & Sweller, J. (2008). Cognitive load theory and music instruction. *Educational Psychology, 28*, 29-45. doi: 10.1080/01443410701369146

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

Paas, F., Ayres, P., & Pachman, M. (2008). *Assessment of cognitive load in multimedia learning*. Recent Innovations in Educational Technology That Facilitate Student Learning, Information Age Publishing Inc., Charlotte, NC, 11-35.

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. (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 Merrienboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors, 35*, 737-743. doi: 10.1177/001872089303500412

Paas, F., & Van Merriënboer, J. J. (1994a). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*, 351-371. doi: 10.1007/BF02213420

Paas, F., & Van Merriënboer, J. J. (1994b). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*, 122-133. http://dx.doi.org/10.1037/0022-0663.86.1.122

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

Park, B., & Brünken, R. (2015). The Rhythm Method: A New Method for Measuring Cognitive Load—An Experimental Dual-Task Study. *Applied Cognitive Psychology, 29,* 232-243. doi: 10.1002/acp.3100

Park, B., Korbach, A., & Brünken, R. (2015). Do learner characteristics moderate the Seductive-Details-Effect? A Cognitive-Load-Study using eye-tracking. *Journal of Educational Technology & Society, 18,* 24-36. URL: http://www.jstor.org/stable/jeductechsoci.18.4.24

Payne, D. G., Neely, J. H., & Burns, D. J. (1986). The generation effect: Further tests of the lexical activation hypothesis. *Memory & Cognition, 14,* 246-252. doi: 10.3758/BF03197700

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

Peynircioğlu, Z. F. (1989). The generation effect with pictures and nonsense figures. *Acta Psychologica, 70*, 153-160. doi:10.1016/0001-6918(89)90018-8

Piaget, J. (1928). *The judgement and reasoning in children*. London: Routledge and Kegan.

Pillay, H. K. (1994). Cognitive load and mental rotation: structuring orthographic projection for learning and problem solving. *Instructional Science, 22*, 91-113. doi: 10.1007/BF00892159

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology, 47*, 560-569. doi: 10.1111/j.1469-8986.2009.00947.x

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

Polya, G. (1957). *How to Solve It: a new aspect of mathematical method*, ed: London: Penguin.

Portin, P. (2002). Historical development of the concept of the gene. *Journal of Medicine and Philosophy, 27*, 257-286. doi: 10.1076/jmep.27.3.257.2980

Pring, L. (1988). The 'reverse generation effect': A comparison of memory performance between blind and sighted children. *British Journal of Psychology, 79*, 387-400. doi: 10.1111/j.2044-8295.1988.tb02297.x

Pring, L., Freestone, S., & Katan, S. (1990). Recalling pictures and words: Reversing the generation effect. *Current Psychology, 9*, 35-45. doi: 10.1007/BF02686766

Pyke, A. A., & LeFevre, J.-A. (2011). Calculator use need not undermine direct-access ability: The roles of retrieval, calculation, and calculator use in the acquisition of arithmetic facts. *Journal of Educational Psychology, 103*, 607-616. http://dx.doi.org/10.1037/a0023291

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

Rabinowitz, J. C., & Craik, F. I. (1986). Specific enhancement effects associated with word generation. *Journal of Memory and Language, 25*, 226-237. doi:10.1016/0749-596X(86)90031-8

Reardon, R., Durso, F. T., Foley, M. A., & McGahan, J. R. (1987). Expertise and the generation effect. *Social Cognition, 5*, 336-348. doi: 10.1521/soco.1987.5.4.336

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*, 92-103. doi:10.1016/j.learninstruc.2006.02.008

Reisslein, J., Sullivan, H., & Reisslein, M. (2007). Learner achievement and attitudes under different paces of transitioning to independent problem solving. *Journal of Engineering Education, 96*, 45-56. doi: 10.1002/j.2168-9830.2007.tb00914.x

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., 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. (2000). From studying examples to solving problems: Fading worked-out solution steps helps learning. In L. Gleitman, & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 393-398). Mahwah, NJ: Lawrence Erlbaum Associations, http://dx.doi.org/10.1.1.23.6816.

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*, 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

Renkl, A., & Atkinson, R. (2001). The effects of gradually increasing problem-solving demands in cognitive skill acquisition. *EARLI 2001*.

Retnowati, E., Ayres, P., & Sweller, J. (2010). Worked example effects in individual and group work settings. *Educational Psychology, 30*, 349-367. doi: 10.1080/01443411003659960

Rittle-Johnson, B., & Kmicikewycz, A. O. (2008). When generating answers benefits arithmetic skill: The importance of prior knowledge. *Journal of Experimental Child Psychology, 101,* 75-81. doi:10.1016/j.jecp.2008.03.001

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77*, 1-15. doi: 10.1111/j.1467-8624.2006.00852.x

Rose, J. M., & Wolfe, C. J. (2000). The effects of system design alternatives on the acquisition of tax knowledge from a computerized tax decision aid. *Accounting, Organizations and Society, 25,* 285-306. doi:10.1016/S0361-3682(99)00048-3

Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognise designers' styles. *Learning and Instruction, 19*, 185-199. doi:10.1016/j.learninstruc.2008.03.006

Roy, M., & Chi, M. T. (2005). The self-explanation principle in multimedia learning. *The Cambridge handbook of multimedia learning*, 271-286.

Scheiter, K., & Gerjets, P. (2007). *Making your own order: Order effects in system-and user-controlled settings for learning and problem solving.* In order to learn: How the sequence of topics influences learning, 195-212.

Schmidt, S. R. (1992). Evaluating the role of distinctiveness in the generation effect. *The Quarterly Journal of Experimental Psychology, 44*, 237-260. doi: 10.1080/02724989243000028

Schmidt, S. R., & Cherry, K. (1989). The negative generation effect: Delineation of a phenomenon. *Memory & Cognition, 17*, 359-369. doi: 10.3758/BF03198475

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

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

Scholey, A. B., Moss, M. C., Neave, N., & Wesnes, K. (1999). Cognitive performance, hyperoxia, and heart rate following oxygen administration in healthy young adults. *Physiology & Behavior*, *67*, 783-789. doi:10.1016/S0031-9384(99)00183-3

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

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*, 759-775. http://dx.doi.org/10.1037/a0025140

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22*, 129-184. doi: 10.1207/s1532690xci2202_1

Schweickert, R., McDaniel, M. A., & Riegler, G. (1994). Effects of generation on immediate memory span and delayed unexpected free recall. *The Quarterly Journal of Experimental Psychology, 47*, 781-804. doi: 10.1080/14640749408401137

Schwerdt, G., & Wuppermann, A. C. (2011). Sage on the Stage: Is Lecturing Really All that Bad? *Education Next, 11*, 62-67.

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*, 258-266. doi:10.1016/j.chb.2008.12.011

Schwonke, R., Renkl, A., Salden, R., & Aleven, V. (2011). Effects of different ratios of worked solution steps and problem solving opportunities on cognitive load and learning outcomes. *Computers in Human Behavior, 27*, 58-62. doi:10.1016/j.chb.2010.03.037

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology, 99,* 285-296. http://dx.doi.org/10.1037/0022-0663.99.2.285

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction, 13*, 227-237. doi:10.1016/S0959-4752(02)00022-1

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*, 127-190. http://dx.doi.org/10.1037/0033-295X.84.2.127

Shute, V. J., & Gluck, K. A. (1996). Individual differences in patterns of spontaneous online tool use. *The Journal of the Learning Sciences, 5*, 329-355. doi: 10.1207/s15327809jls0504_2

Siegler, R. S. (1998). *Emerging minds: The process of change in children's thinking.* Oxford University Press.

Siegler, R. S. (2002). *Microgenetic studies of self-explanation.* Microdevelopment: Transition processes in development and learning, 31-58.

Siegler, R. S., & Chen, Z. (2008). Differentiation and integration: Guiding principles for analyzing cognitive change. *Developmental Science, 11,* 433-448. doi: 10.1111/j.1467-7687.2008.00689.x

Slamecka, N. J., & Fevreiski, J. (1983). The generation effect when generation fails. *Journal of Verbal Learning and Verbal Behavior, 22,* 153-163. doi:10.1016/S0022-5371(83)90112-3

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human learning and Memory, 4*, 592-604. http://dx.doi.org/10.1037/0278-7393.4.6.592

Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language*, *26*, 589-607. doi:10.1016/0749-596X(87)90104-5

Snodgrass, J. G., & Kinjo, H. (1998). On the generality of the perceptual closure effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 645-658. http://dx.doi.org/10.1037/0278-7393.24.3.645

Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology, 76*, 347-376. http://dx.doi.org/10.1037/0022-0663.76.3.347

Snowman, J., & Cunningham, D. J. (1975). A comparison of pictorial and written adjunct aids in learning from text. *Journal of Educational Psychology, 67*, 307-311. URL: http://dx.doi.org/10.1037/h0076934

Soloway, R. M. (1986). No generation effect without semantic activation. *Bulletin of the Psychonomic Society*, *24*, 261-262. doi: 10.3758/BF03330134

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*, 1-29. doi: 10.1037/h0093759

Steffens, M. C., & Erdfelder, M. C. S. E. (1998). Determ inants of Positive and Negative Generation Effects in Free Recall. *The Quarterly Journal of Experimental Psychology: Section A, 51*, 705-733. doi: 10.1080/713755794

Stotz, K., & Griffiths, P. (2004). Genes: Philosophical analyses put to the test. *History and Philosophy of The Life Sciences*, 5-28.

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

Sweller, J. (1999). *Instructional Design in Technical Areas*. Australian Education Review, No. 43: ERIC.

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

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*, New York: Springer.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 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*, 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.

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

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

Sweller, J. (2003). Evolution of human cognitive architecture. *Psychology of Learning and Motivation*, 43, 216-266. doi: 10.1016/S0079-7421(03)01015-6

Sweller, J. (2010). *What human cognitive architecture tells us about constructivism*. In S. Tobias & T. M. Duffy (Eds.), Constructivist instruction: Success or failure (pp. 127-143). New York, NJ: Routledge.

Tabbers, H., Martens, R., & Merriënboer, V. (2000). Multimedia instructions and cognitive load theory: Split-attention and modality effects. *Paper presented at the National Convention of the Association for Educational Communications and Technology*, Long Beach, CA.

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

Tennyson, R. D., Woolley, F. R., & Merrill, M. D. (1972). Exemplar and nonexampler variables which produce correct concept classification behavior and specified classification errors. *Journal of Educational Psychology*, *63,* 144-152. doi: 10.1037/h0032368

Tettamanti, M., Moro, A., Messa, C., Moresco, R. M., Rizzo, G., Carpinelli, A., . . . Perani, D. (2005). Basal ganglia and language: phonology modulates dopaminergic release. *Neuroreport*, *16*, 397-401. doi: 10.1097/00001756-200503150-00018

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

Tobias, S. (1976). Achievement treatment interactions. *Review of Educational Research*, 61-74.

Tobias, S. (1987). Mandatory text review and interaction with student characteristics. *Journal of Educational Psychology, 79*, 154-161. doi: 10.1037/0022-0663.79.2.154

Tobias, S. (1989). Another look at research on the adaptation of instruction to students characteristics. *Educational Psychologist*, *24*, 213-227. doi:10.1207/s15326985ep2403_1

Tuddenham, R. D. (1966). Jean Piaget and the world of the child. *American Psychologist*, *21*, 207-217. doi: 10.1037/h0023304

Tulving, E. (1979). Relation between encoding specificity and levels of processing. *Levels of Processing in Human Memory*, 405-428.

Tyler, S. W., Hertel, P. T., McCallum, M. C., & Ellis, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human learning and Memory*, *5*, 607-617. doi: 10.1037/0278-7393.5.6.607

Uline, C., & Tschannen-Moran, M. (2008). The walls speak: the interplay of quality facilities, school climate, and student achievement. *Journal of Educational Administration*, *46*, 55-73. doi: 10.1108/09578230810849817

Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *Quarterly Journal of Experimental Psychology Section A*, *57*, 165-182. doi: 10.1080/02724980343000189

Van Gerven, P. W., Paas, F., van Merriënboer, J. J., & Schmidt, H. G. (2006). Modality and variability as factors in training the elderly. *Applied Cognitive Psychology*, *20*, 311-320. doi: 10.1002/acp.1247

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*, 212-218. doi: 10.1016/j.cedpsych.2010.10.004

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

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

Van Gog, T., Paas, F., & van Merriënboer, J. J. (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 Merriënboer, J. J. (1990). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of Educational Computing Research*, *6*, 265-285. doi: 10.2190/4NK5-17L7-TWQV-1EHL

Van Merriënboer, J. J. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*: Educational Technology.

van Merriënboer, J. J., Jelsma, O., & Paas, F. G. (1992). Training for reflective expertise: A four-component instructional design model for complex cognitive skills. *Educational Technology Research and Development*, *40*, 23-43. doi: 10.1007/BF02297047

Van Merriënboer, J. J., 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., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist*, *38*, 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*, 147-177. doi: 10.1007/s10648-005-3951-0

Van Merriënboer, J. J., Schuurman, J., De Croock, M., & Paas, F. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance and training efficiency. *Learning and Instruction*, *12*, 11-37. doi: 10.1016/S0959-4752(01)00020-2

Van Merrienboer, J. J., & De Croock, M. B. (1992). Strategies for computer-based programming instruction: Program completion vs. program generation. *Journal of Educational Computing Research, 8*, 365-394. doi: 10.2190/MJDX-9PP4-KFMT-09PM

Van Merrienboer, J. J., & Krammer, H. P. (1987). Instructional strategies and tactics for the design of introductory computer programming courses in high school. *Instructional Science*, *16*, 251-285. doi: 10.1007/BF00120253

Van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of Educational Psychology, 93*, 129-140. URL: http://dx.doi.org/10.1037/0022-0663.93.1.129

VanLehn, K. (1988). *Problem solving and cognitive skill acquisition*: DTIC Document.

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

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*: Oxford University Press.

Westermann, K., & Rummel, N. (2012). Delaying instruction: evidence from a study in a university relearning setting. *Instructional Science*, *40*, 673-689. doi: 10.1007/s11251-012-9207-8

Whelan, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review*, *2*, 1-12. doi: 10.1016/j.edurev.2006.11.001

Wright, G., & Ayton, P. (1988). Decision time, subjective probability, and task difficulty. *Memory & Cognition*, *16*, 176-185. doi: 10.3758/BF03213487

Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single-and multi-task environments. *Work and Stress*, *14*, 74-99. doi:10.1080/026783700417249

Zheng, R., & Cook, A. (2012). Solving complex problems: A convergent approach to cognitive load measurement. *British Journal of Educational Technology*, *43*, 233-246. doi: 10.1111/j.1467-8535.2010.01169.x

Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction*, *4*, 137-166. doi:10.1207/s1532690xci0403_1

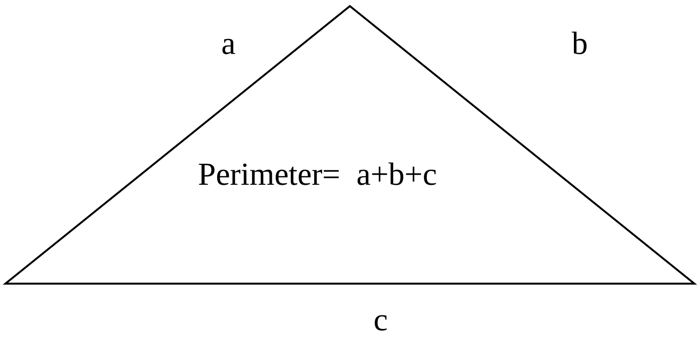
APPENDIX

Appendix 1. Geometry Formulae Used to Test the Generation Effect

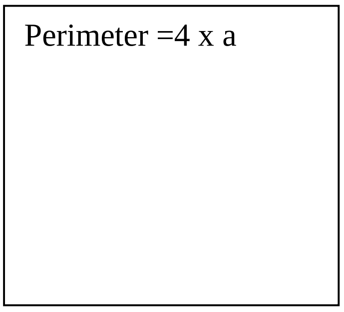
Please study the following formulae carefully.

Perimeter (circumference) formulae:

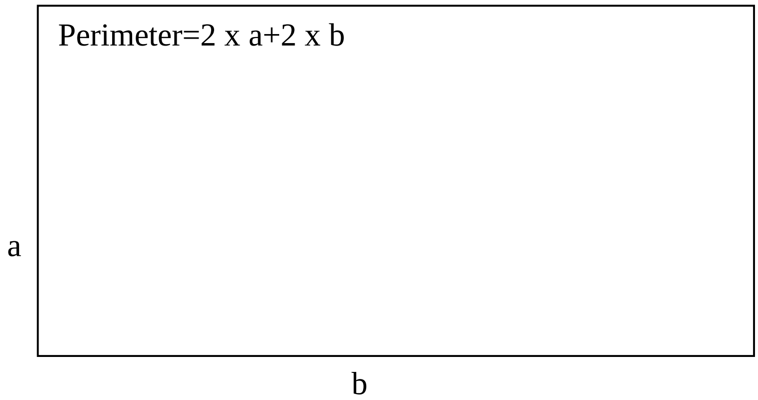
1 Triangle:



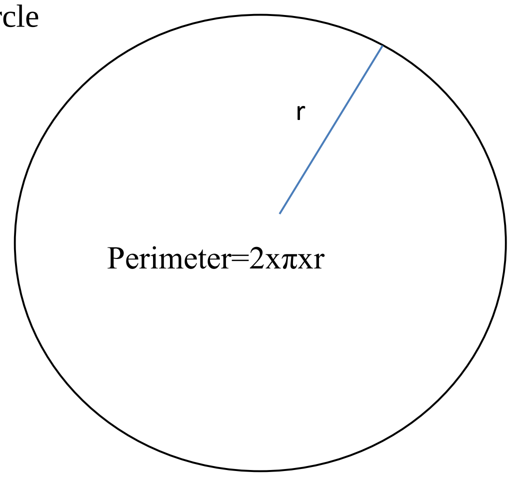
2 Square:



3 Rectangle:

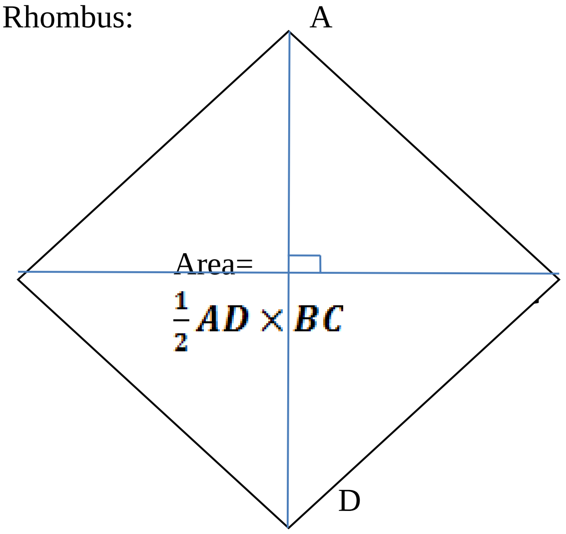


4 Circle

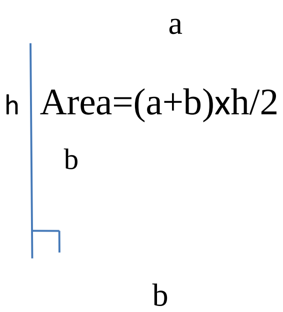


Area formulae:

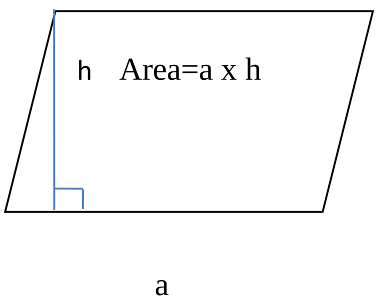
1 Rhombus:



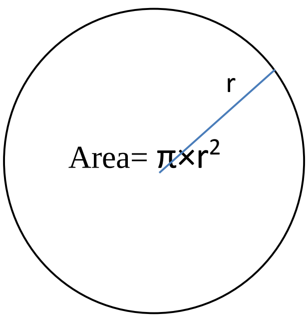
2 Trapezium:



3 Parallelogram:

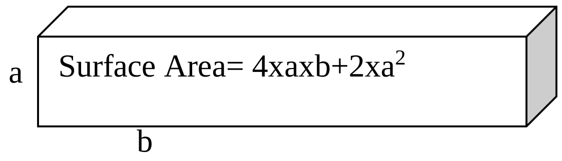


4 Circle:

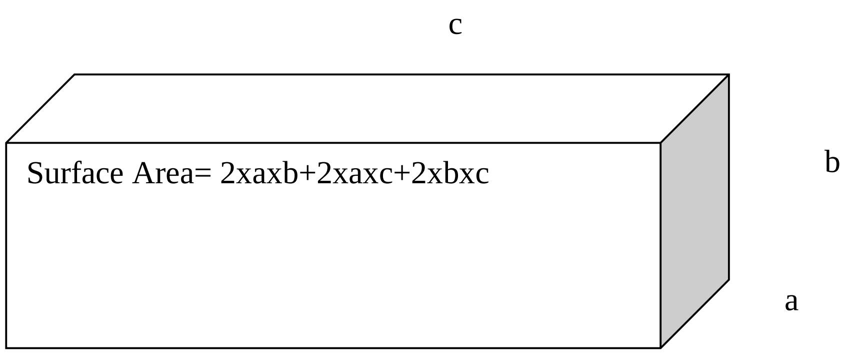


Surface area:

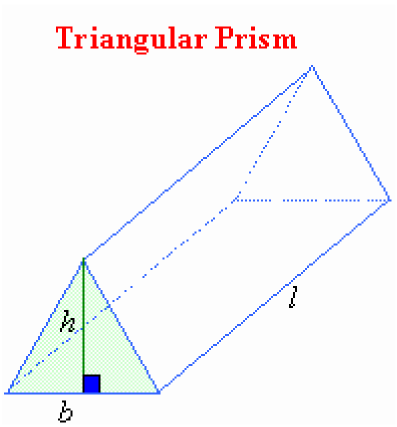
1 Rectangular prism (end with square):



2 Rectangular prism (end with rectangle):



3

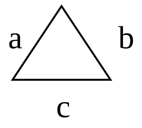


Surface Area= bxh+3xbxl

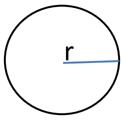
Appendix 2. Booklet Used For Generation Group

Give the formulae of the given shapes:

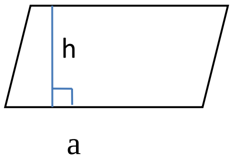
1. perimeter of triangle:



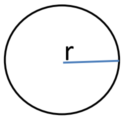
2. circumference (perimeter) of circle:



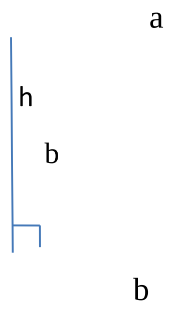
3. Area of Parallelogram:



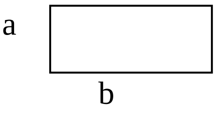
4. Area of Circle:



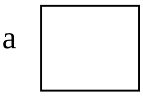
5. Area of Trapezium:



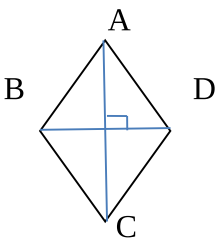
6. perimeter of Rectangle:



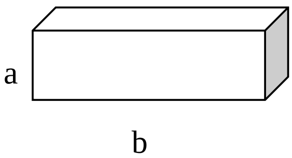
7. perimeter of square:



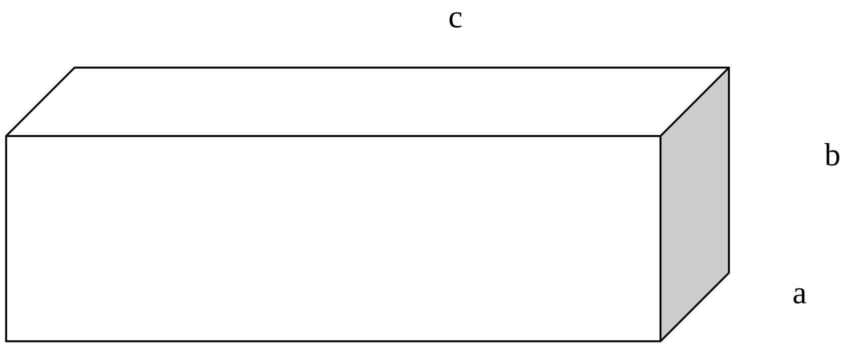
8. Area of Rhombus:



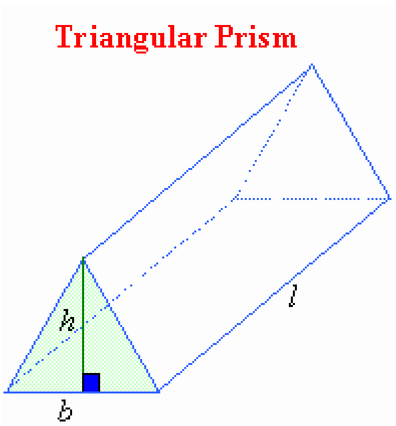
9. surface area of Rectangular prism (end with square):



10. surface area of Rectangular prism (end with rectangle):



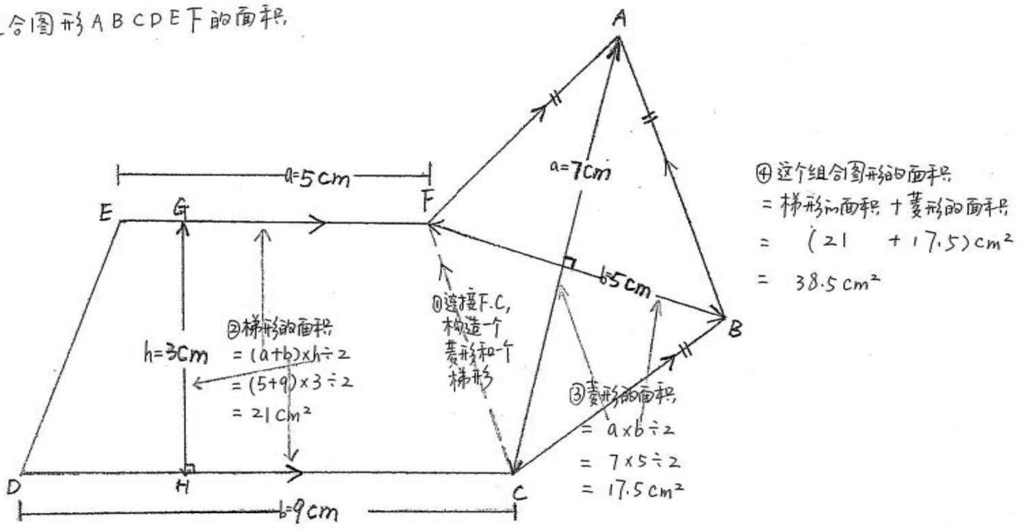
11. surface area of triangular prism:



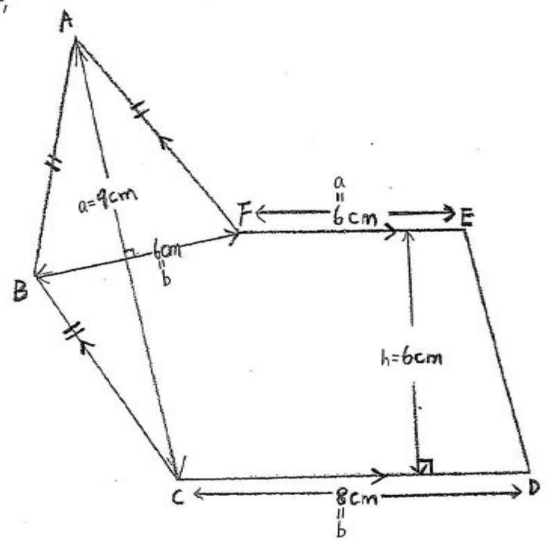
Appendix 3. Booklet Used For Worked Example Group

Please study how to calculate the area of composite shape and then solve a similar problem.

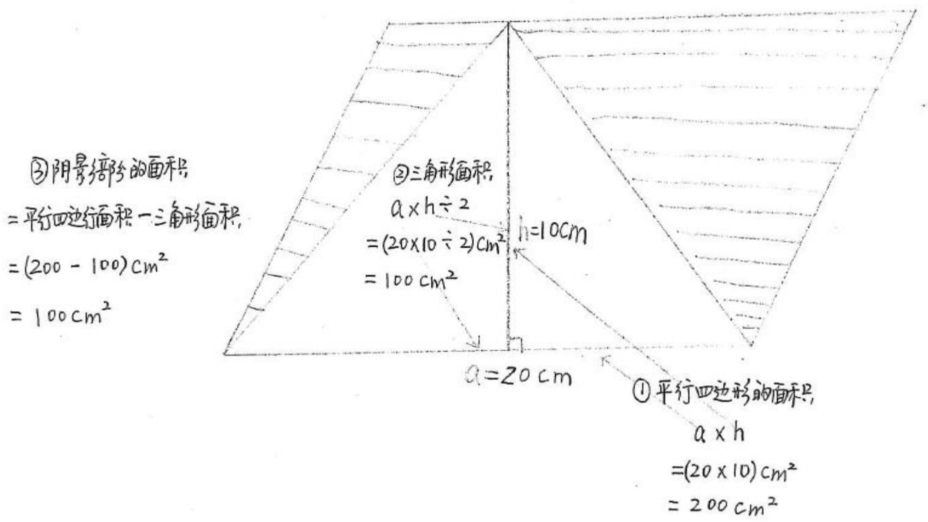
求组合图形ABCDEF的面积

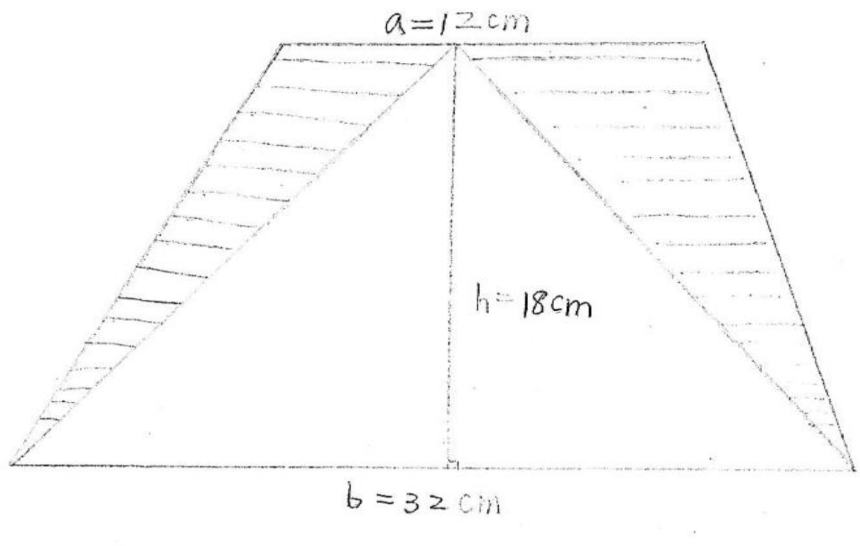


求组合图形ABCDEF的面积



Please study how to calculate the area of shading parts and then solve a similar problem.

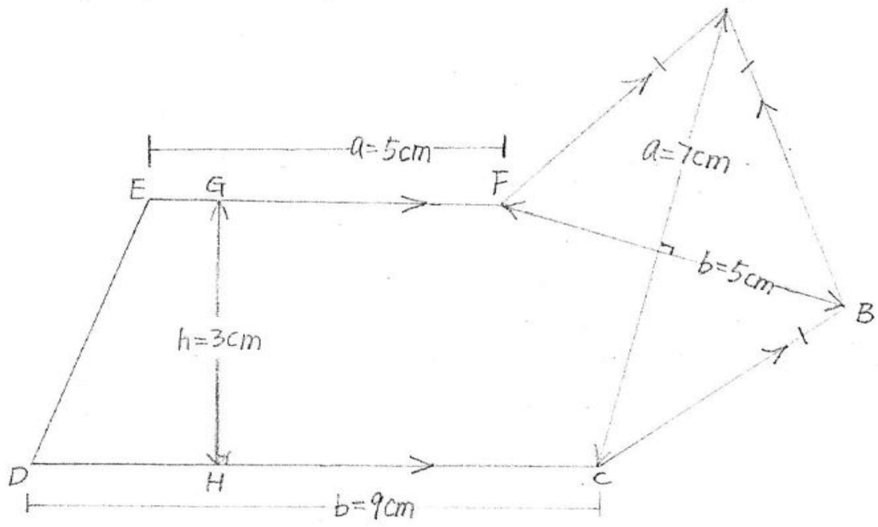


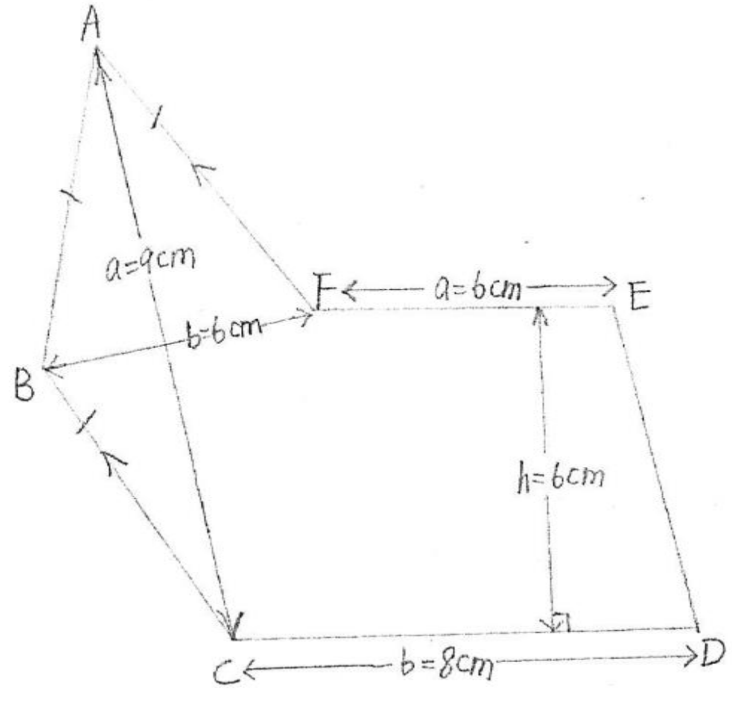


Appendix 4. Booklet Used to Problem Solving Group

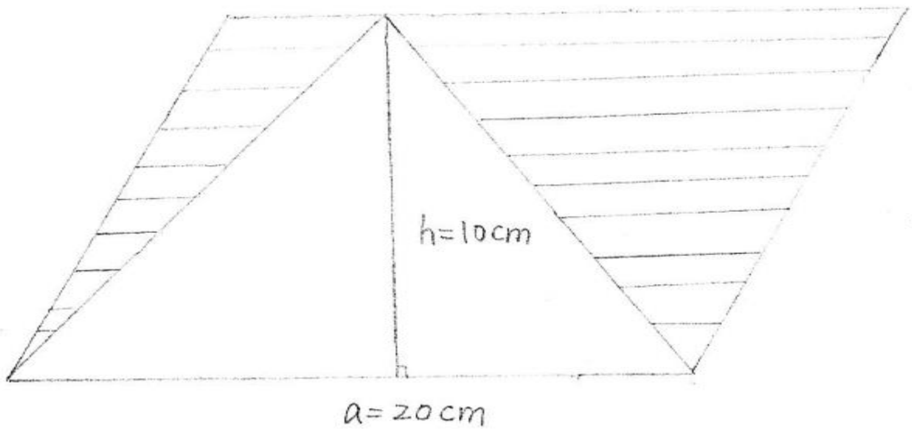
Please calculate the area of composite shapes.

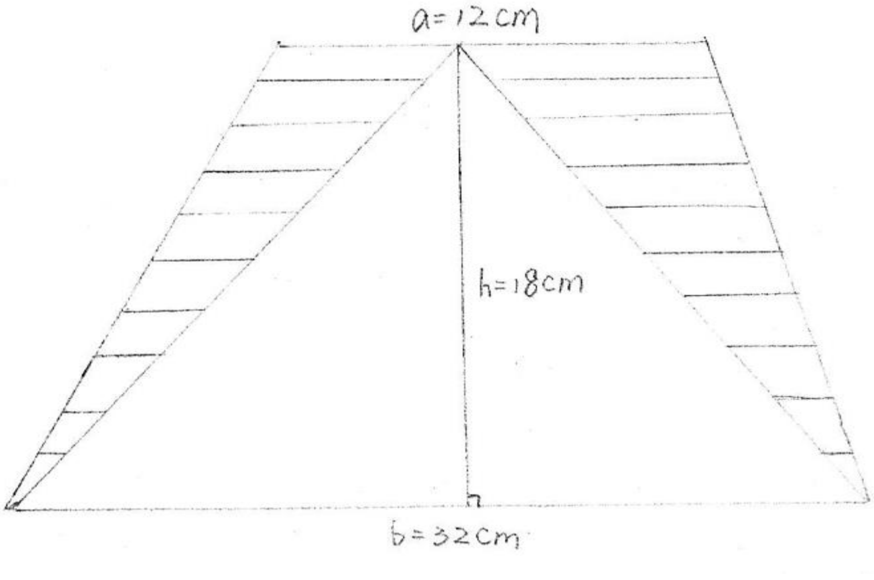
求组合图形面积（ABCDEF）





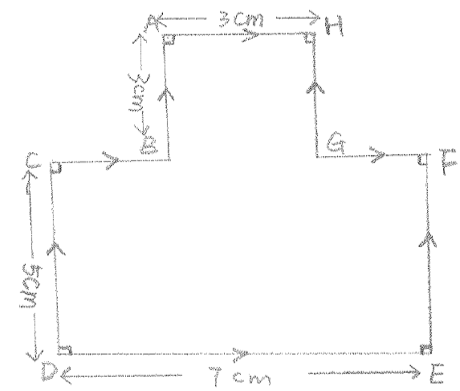
Please calculate the area of shading parts.

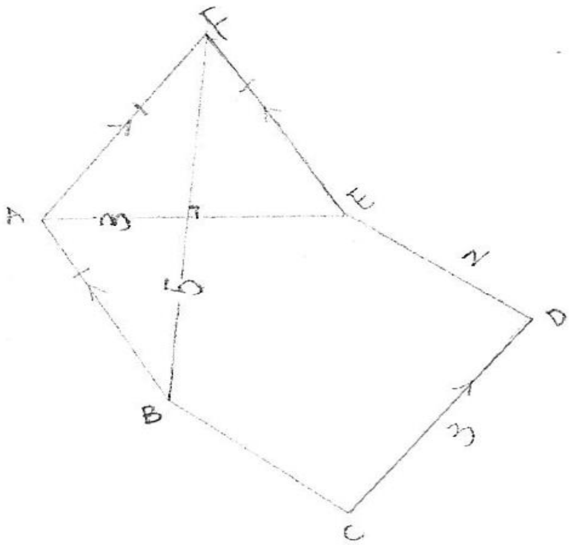


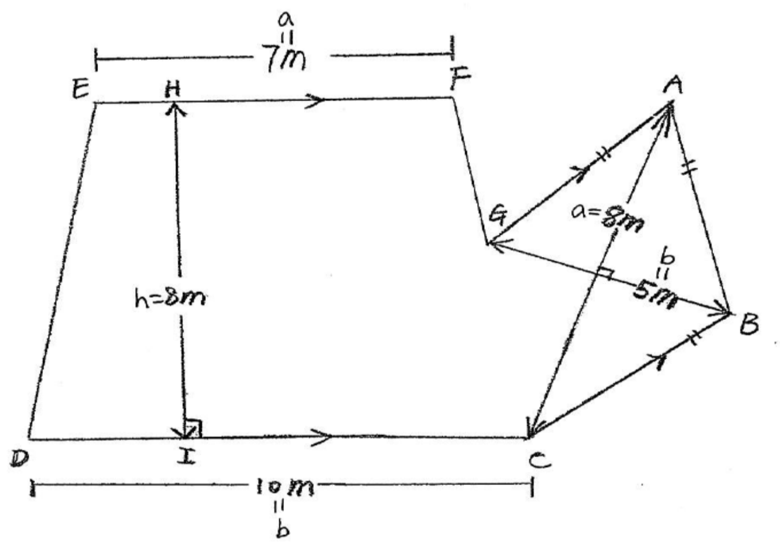


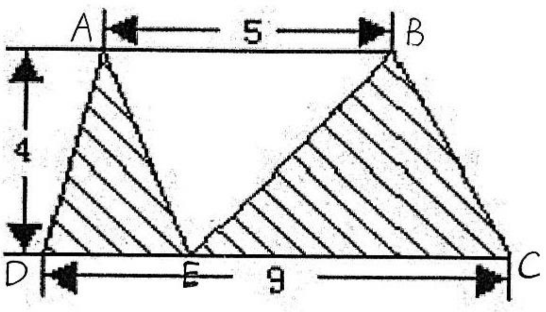
Appendix 5. Testing Materials Used to Test the Worked Example Effect

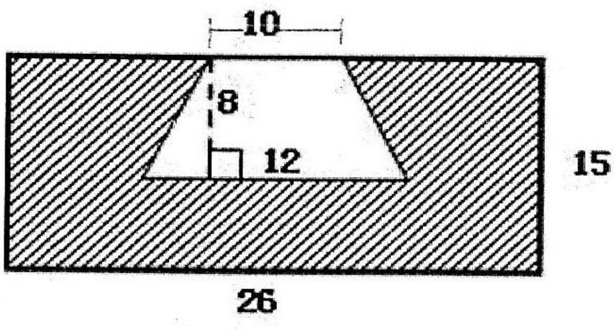
Please calculate the area of following shapes.





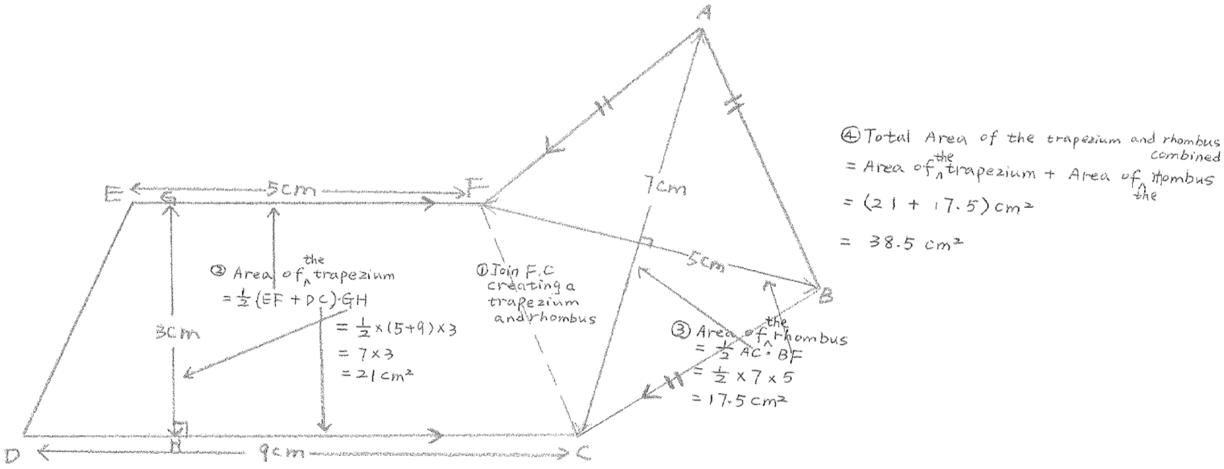

a
7m
E
H
F
A
G
a=8m
h=8m
b
5m
B
D
I
C
10m
b

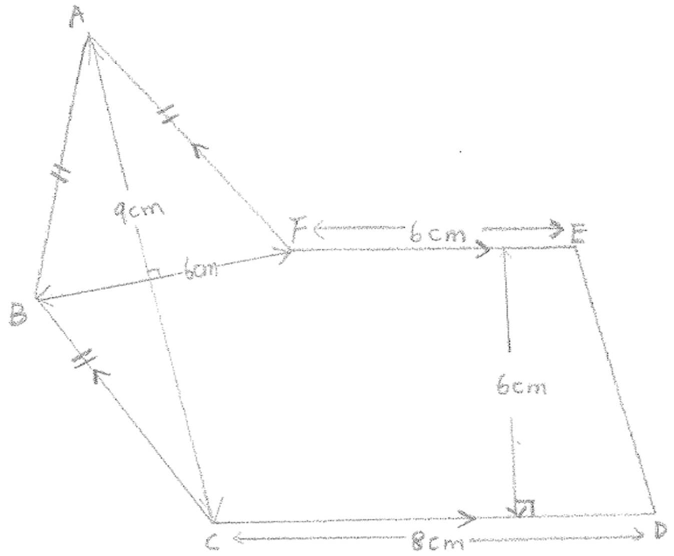

A
5
B
4
D
E
9
C

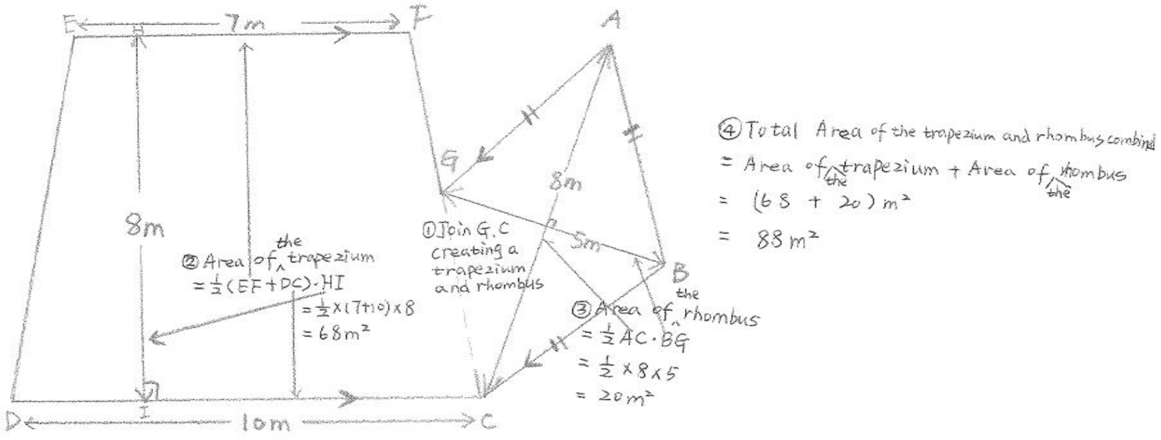

10
8
12
15
26

Appendix 6. Booklet Used for Worked Example Group (Experiment 2 and 3)

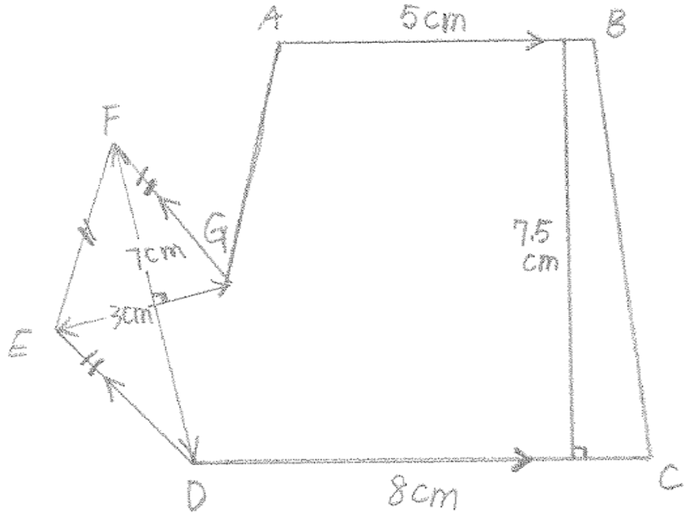
Please study an example and then solve a similar problem







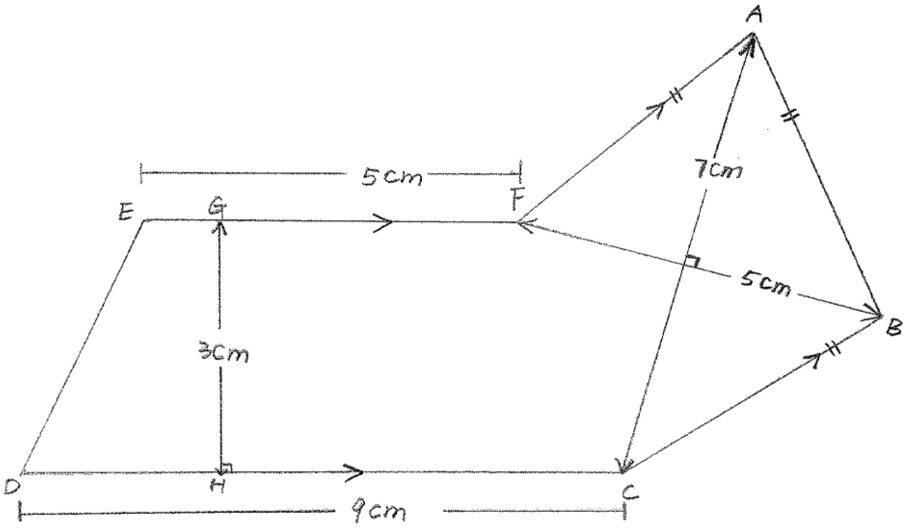
E
7m
F
A
G
8m
5m
B
8m
① Join G, C creating a trapezium and rhombus
② Area of the trapezium
= ½ (EF + DC)·HI
= ½ × (7+10) × 8
= 68m²
③ Area of the rhombus
= ½ AC·BG
= ½ × 8 × 5
= 20m²
D
I
10m
C
④ Total Area of the trapezium and rhombus combined
= Area of the trapezium + Area of the rhombus
= (68 + 20) m²
= 88 m²

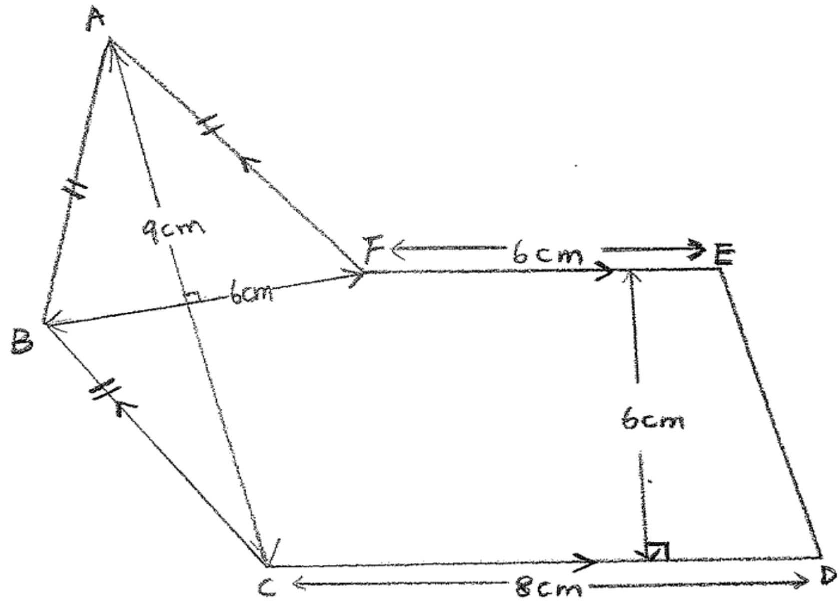


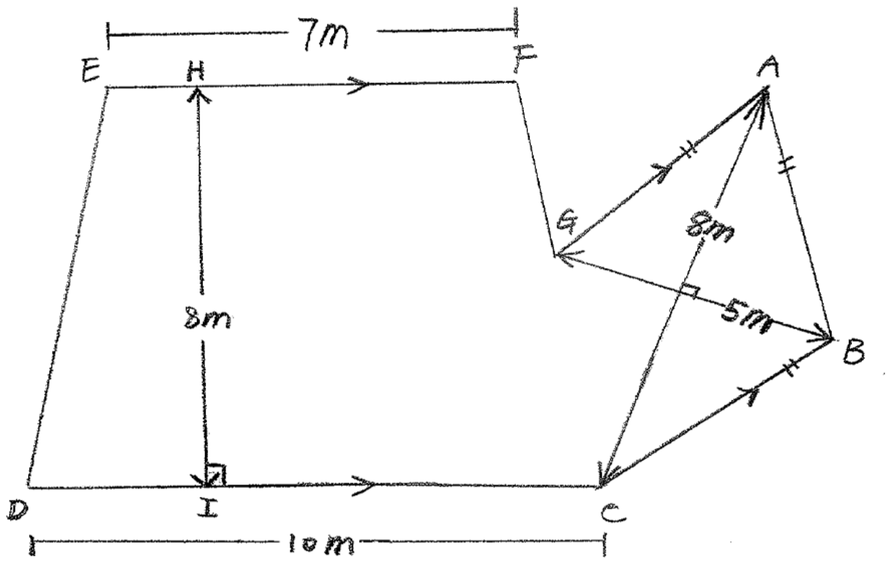
A
5cm
B
F
G
7cm
3cm
E
7.5 cm
D
8cm
C

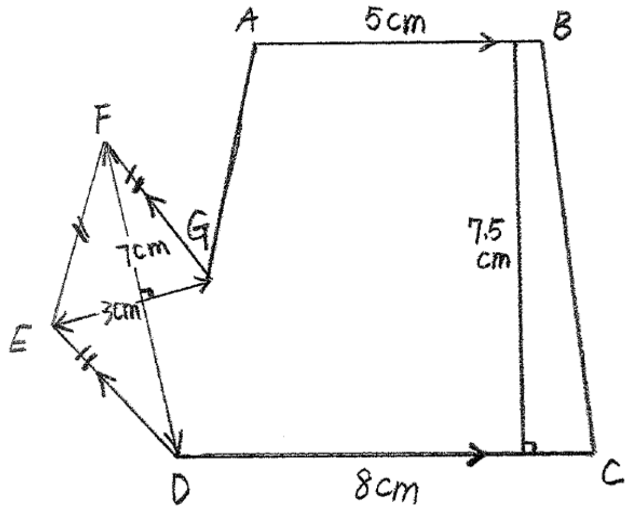
Appendix 7. Booklet Used for Problem Solving Group (Experiment 2 and 3)

Please solve the following four problems



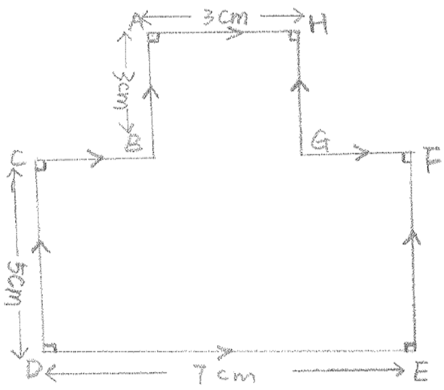


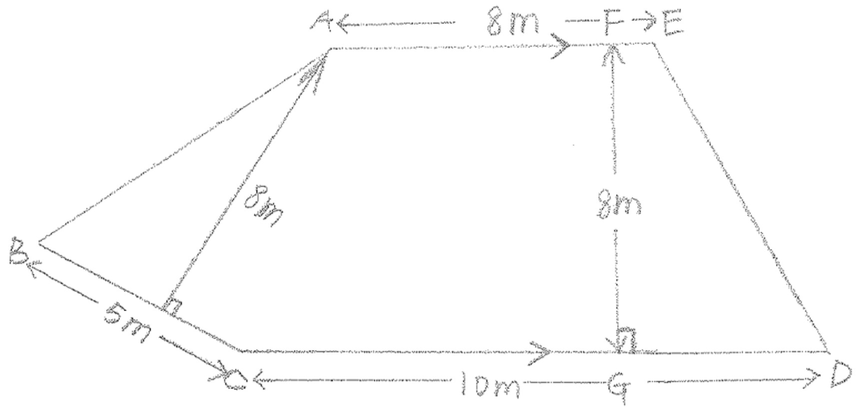

7m
E
H
F
A
G
8m
8m
5m
B
D
I
C
10m

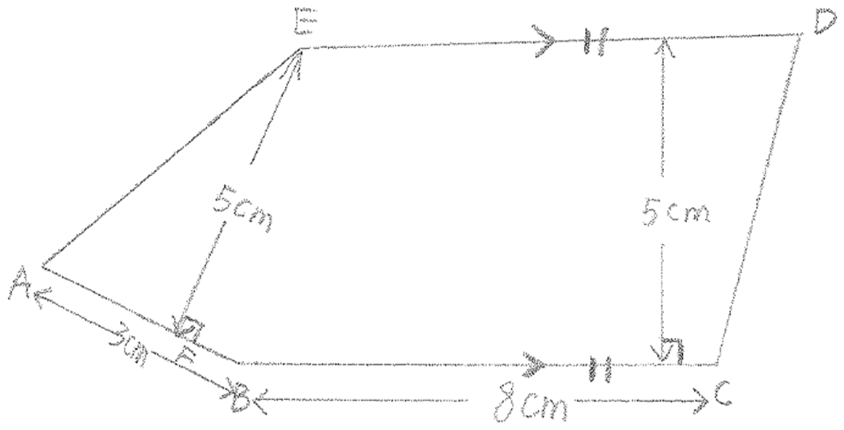

A
5cm
B
F
G
7cm
7.5
cm
E
3cm
D
8cm
C

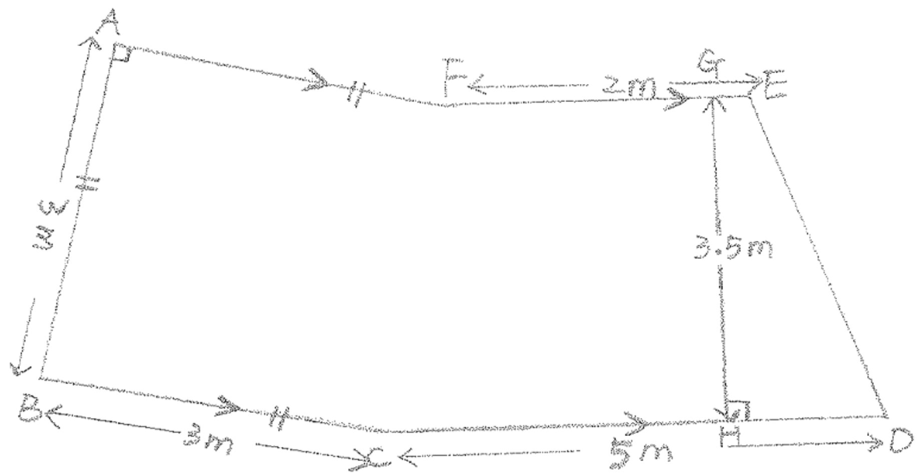
Appendix 8. Test Materials Used to Test the Worked Example Effect (Experiment 2 and 3)

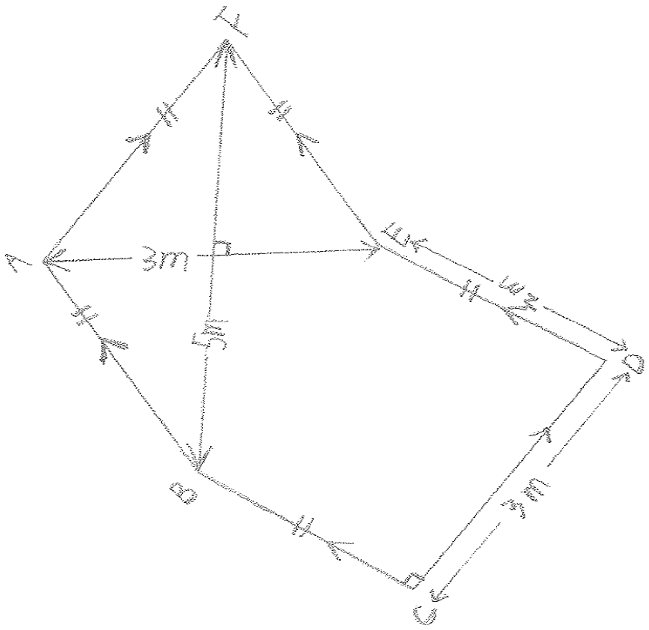
Please solve the following five problems






E
D
5cm
5 cm
A
3cm
F
B
8 cm
C


A
F
2m
G
E
3m
3.5m
B
3m
C
5m
H
D


F
A
3m
5m
E
2m
D
3m
B
C

Appendix 9. Trigonometry Formulae Used to Test for the Generation Effect

Please study the following formulae carefully.

$\sin(A+B)=\sin A\cos B+\cos A\sin B$

$\sin(A-B)=\sin A\cos B-\cos A\sin B$

$\cos(A+B)=\cos A\cos B-\sin A\sin B$

$\cos(A-B)=\cos A\cos B+\sin A\sin B$

$\sin(-A)=-\sin A$

$\cos(-A)=\cos A$

$\sin(\pi+A)=-\sin A$

$\cos(\pi+A)=-\cos A$

$\sin(\pi-A)=\sin A$

$\cos(\pi-A)=-\cos A$

Appendix 10. Booklet Used for Generation Group

Please finish the following formulae.

$\sin(A+B)=$

$\sin(A-B)=$

$\cos(A+B)=$

$\cos(A-B)=$

$\sin(-A)=$

$\cos(-A)=$

$\sin(\pi+A)=$

$\cos(\pi+A)=$

$\sin(\pi-A)=$

$\cos(\pi-A)=$

Appendix 11. Booklet used for Worked Example Group

Please study an example and then solve a similar problem.

Example 1:

$$\sqrt{2}\sin(\alpha+\frac{\pi}{4})$$
$$=\sqrt{2}(\sin\alpha\cos\frac{\pi}{4}+\cos\alpha\sin\frac{\pi}{4})$$
$$=\sqrt{2}(\frac{\sqrt{2}}{2}\sin\alpha+\frac{\sqrt{2}}{2}\cos\alpha)$$
$$=\sin\alpha+\cos\alpha$$

Solve the following problem:

$$\sqrt{2}\sin(\alpha-\frac{\pi}{4})$$

Example 2:

$$\cos\alpha+\sqrt{3}\sin\alpha$$
$$=2(\frac{1}{2}\cos\alpha+\frac{\sqrt{3}}{2}\sin\alpha)$$
$$=2(\sin\frac{\pi}{6}\cos\alpha+\cos\frac{\pi}{6}\sin\alpha)$$
$$=2\sin(\alpha+\frac{\pi}{6})$$

Solve the following problem:

$$\sin\alpha+\sqrt{3}\cos\alpha$$

Appendix 12. Booklet Used for Problem Solving Group

Please simplify the following expressions.

$$\sqrt{2}\sin(\alpha+\frac{\pi}{4})$$

$$\sqrt{2}\sin(\alpha-\frac{\pi}{4})$$

$$\cos\alpha+\sqrt{3}\sin\alpha$$

$$\sin\alpha+\sqrt{3}\cos\alpha$$

Appendix 13. Test Materials Used to Test the Worked Example Effect

Please simplify the following expressions.

$$\sin(\alpha + \frac{7\pi}{4})$$

$$\sqrt{2}\cos(\alpha - \frac{3\pi}{4})$$

$$2\sqrt{3}\cos(2\alpha - \frac{\pi}{6})$$

$$\sin(\alpha - \beta) + 2\cos\alpha \sin\alpha$$

$$(\cos\frac{\pi}{12} - \sin\frac{\pi}{12})(\cos\frac{\pi}{12} + \sin\frac{\pi}{12})$$

Appendix 14. List of Papers

Chen, O., Kalyuga, S., & Sweller, J. (2015). The Worked Example Effect, the Generation Effect, and Element Interactivity. *Journal of Educational Psychology, 107,* 689-704. http://dx.doi.org/10.1037/edu0000018

Chen, O., Kalyuga, S., & Sweller, J. (Under Review). “Relations between the worked example and generation effects on immediate and delayed tests” *Learning and Instruction*.

Chen, O., Kalyuga, S., & Sweller, J. (Resubmitted). “When instructional guidance is needed” *Instructional Science*.

Chen, O., Kalyuga, S., & Sweller, J. (In Press). The Expertise Reversal Effect is a Variant of the More General Element Interactivity Effect. *Educational Psychology Review*. Doi: 10.1007/s10648-016-9359-1

Chen, O. (2015). “THE EFFECT OF PENTOMINO ON THE SPATIAL ABILITY” *Proceeding of International Conference On Research, Implementation And Education Of Mathematics And Sciences 2015* (ICRIEMS 2015)

Educ Psychol Rev
DOI 10.1007/s10648-016-9359-1

REVIEW ARTICLE

The Expertise Reversal Effect is a Variant of the More General Element Interactivity Effect

Ouhao Chen[1] · Slava Kalyuga[1] · John Sweller[1]

Abstract Within the framework of cognitive load theory, the element interactivity and the expertise reversal effects usually are not treated as closely related effects. We argue that the two effects may be intertwined with the expertise reversal effect constituting a particular example of the element interactivity effect. Specifically, the element interactivity effect relies on changes in element interactivity due to changes in the type of material being learned, while the expertise reversal effect also relies on changes in relative levels of element interactivity but in this case, due to changes in relative levels of expertise. If so, both effects rely on equivalent changes in element interactivity with the changes induced by different factors. Empirical evidence is used to support this contention.

Keywords Cognitive load theory · Element interactivity · Expertise · Worked example effect · Generation effect

Within cognitive load theory, the element interactivity and expertise reversal effects are regarded as distinct cognitive load effects. However, empirical evidence obtained recently (Chen et al. 2015), along with previous evidence (Blayney et al. 2010; Kalyuga et al. 2001b; Leahy and Sweller 2005), can be interpreted as indicating that the expertise reversal effect may be a variant of the more general element interactivity effect. In this paper, we review the two effects and suggest possible relations between them.

✉ John Sweller
j.sweller@unsw.edu.au

Ouhao Chen
chenouhao_lucky@126.com

Slava Kalyuga
s.kalyuga@unsw.edu.au

[1] School of Education, University of New South Wales, Sydney 2052, Australia

Published online: 29 January 2016

The Expertise Reversal Effect

The expertise reversal effect focuses on the interaction between levels of learners' expertise and the instructional procedures used. Consider two instructional procedures, one of which results in superior test performance compared to the other when instructing novices. Under the expertise reversal effect, with increases in levels of expertise, the difference between the two procedures first decreases, then is eliminated, and may finally reverse (Kalyuga 2007; Kalyuga and Renkl 2010). Based on these changes in the relative effectiveness of instruction, two formats of this effect can be categorized: an ordinal interaction in which one of two instructional procedures is effective for novices, but is less effective or has no effects when testing more experienced learners, and a disordinal interaction where one instructional procedure is effective for novices with the relative effectiveness reversed for more experienced learners (Nievelstein et al. 2013). Which form occurs depends on the relative levels of expertise of the learners. If the differences in expertise are small, test performance may not include a crossover point resulting in an ordinal interaction. Larger differences in expertise are more likely to include a crossover point resulting in a disordinal interaction.

Evidence for the Expertise Reversal Effect The expertise reversal effect was initially investigated in a series of longitudinal studies by intensively training groups of technical apprentices from novices to experts in the domain of engineering (Kalyuga et al. 1998, 2000, 2001b). In one set of experiments (Kalyuga et al. 1998), text integrated with diagrams was compared with a diagrams alone condition, testing for the redundancy effect. Results indicated that the diagrams and text condition was superior to the diagrams alone condition for novices, but after a period of training, the effectiveness of the diagrams and text condition deceased compared to the increasing effectiveness of the diagrams alone condition. Subjective ratings of cognitive load further supported the hypothesis that diagrams alone were more easily processed by more knowledgeable learners, whereas the diagrams and text condition was more suitable for novices who needed additional textual instructions to understand the presented diagrams. With an increase in learner's expertise, textual information that had been beneficial for novices became redundant for more knowledgeable learners.

Subsequent experiments by Kalyuga et al. (2000, 2001b) and Kalyuga et al. (2001a) provided more data concerning the expertise reversal effect. Kalyuga et al. (2000), using mechanical engineering materials, found novices benefited more if narrated explanations used to explain how to use specific diagrams were presented together with relevant animated diagrams, as opposed to a diagram only condition. However, integrating narrated explanations with animated diagrams interfered with learning after novices had received a series of intensive training sessions which developed their expertise in the relevant domain. For these more knowledgeable students, diagrams alone were superior to the diagrams with narrations format. Kalyuga et al. (2001b) obtained a full expertise reversal effect when they compared worked examples with instructions to explore in writing switching equations for relay circuits. The results demonstrated that worked examples initially were superior to instructions to explore, but after additional training, the advantage was reversed. For more knowledgeable learners, instructions to explore resulted in superior learning than studying worked examples.

In mathematics curriculum areas, Kalyuga and Sweller (2004) investigated the expertise reversal effect in studying coordinate geometry. Participants were assigned to a worked example group or a problem-solving group. A posttest indicated an interaction of instructional formats and learner expertise. Less knowledgeable learners benefited more from the worked example format with the opposite result found for more knowledgeable learners. In other mathematics areas, similar results were found. Brunstein et al. (2009) observed an expertise reversal effect in algebra learning. They found that for students given considerable practice, a low level of guidance was superior to explicit guidance, whereas, for novices who had less practice, high guidance led to better test results than minimal guidance. Similarly, in the domain of statistics, Leppink et al. (2012) assigned students with different levels of expertise to four groups: reading only; answering open-ended questions; answering open-ended questions in which the answer had to include supporting arguments; and studying worked examples that included the type of arguments that students in the previous group were required to generate. Results again confirmed the expertise reversal effect. Specifically, students with low expertise learned more from worked examples, whereas, high-expertise students learned more from answering open-ended questions with supporting arguments. Rey and Buchwald (2011) also observed an expertise reversal effect when asking students to learn the gradient descent (a mathematical optimization algorithm). Students whose expertise was increased by practice during the experiment had higher test scores if they did not receive additional text explaining a relevant animation, whereas students with a low level of knowledge benefited more from the provision of additional text.

An expertise reversal effect has also been demonstrated in the area of English literature. Oksa et al. (2010) compared two instructional formats used in studying Shakespearean plays. One group received material that combined modern English explanations with Shakespeare's original old English line by line, while another group had the modern English explanatory materials presented as footnotes. Participants, who were less knowledgeable about Shakespearean plays, demonstrated better performance with an integrated format, whereas, for the participants who were Shakespearean experts, the separated format was better.

Nückles et al. (2010) found the expertise reversal effect in learning journal writing skills. Students were divided into a group with prompts and a group without prompts in writing journal entries. During the first semester, students with prompts provided more writing strategies and outperformed students without prompts, but at the end of the semester, as the levels of learner expertise increased, the advantage reversed in line with the expertise reversal effect.

Van Gog et al. (2008) demonstrated an expertise reversal effect by comparing product-oriented worked examples and process-oriented worked examples. The first type of worked example only indicates the procedure to solve a problem, whereas the latter includes not only the procedure but also the reasons for each step (Van Gog et al. 2004). Students were divided into product-product, product-process, process-product, and process-process conditions. Results indicated no initial differences between the conditions, but after two sessions of practice, the process-product group was superior to the process-process group because with an increase of expertise, explanations became redundant resulting in an expertise reversal effect.

The expertise reversal effect also has been found in a computer-based learning environment. Rey and Fischer (2013) tested the effect with a computer program teaching statistical data analysis and induced expertise experimentally by providing some extra examples and illustrations in addition to textual explanations during the experiment. Students were randomly assigned to four groups: experts with textual explanations, experts without textual explanations, novices with textual explanations, and novices without textual explanations. Results replicated the expertise reversal effect. Students with a low level of expertise benefited more from the provision of textual explanations compared to the more expert students who performed better without additional textual explanations.

Johnson et al. (2015) investigated the effects of both visual signaling and of the visual presence of an animated pedagogical agent by comparing the performance of four groups: visual signaling with the animated pedagogical agent present; visual signaling without the animated pedagogical agent present; no visual signaling with the animated pedagogical agent present; and no visual signaling without the animated pedagogical agent present. Students were divided into low or high levels of prior knowledge. The results indicated that students with a high level of knowledge performed better without the animated pedagogical agent present, whereas the opposite result was observed for students with a low level of knowledge.

In summary, work on the expertise reversal effect indicates that in a large variety of curriculum areas, novice students benefit from the presentation of additional information and guidance. With increasing levels of expertise, additional information becomes redundant resulting in a reduction or reversal of the advantage. None of these studies explicitly linked the expertise reversal effect with element interactivity.

Element Interactivity

Element interactivity is a basic concept of cognitive load theory. It can be used to determine categories of cognitive load as well as constituting an effect in its own right. Interactive elements are defined as elements that must be processed simultaneously in working memory as they are logically related (Sweller et al. 2011). An element can be a symbol, a concept, or a procedure that must be learned.

Considered from a broad perspective, the concept of element interactivity provides a practically usable approximation for describing the complexity of information involved in learning, especially when the acquisition of domain-specific knowledge in long-term memory is the goal of instruction. As is the case for any theoretical abstraction, ideally, this description should include details of relevant processes and operations, as well as the timescale on which they occur. Of course, some of these details may be difficult to precisely describe and quantify. For example, processes such as making inferences to construct mental representations, integrating them with prior knowledge, or blocking irrelevant information are likely to consume working memory resources but may be difficult to describe in terms of clearly defined interacting elements of information (Kalyuga 2015). However, the elements associated with most cognitive processes can be described and the concept of element interactivity is effective in assessing levels of cognitive load imposed by specific learning tasks on specific categories of learners. Element interactivity levels can be determined by estimating the number of interacting elements in learning materials (Sweller 1994; Sweller and Chandler 1994; Tindall-Ford et al. 1997). That number will depend on both the nature of the material being processed and the levels of expertise of the learner as discussed in the next section.

Element Interactivity and Intrinsic Cognitive Load Intrinsic load is determined by levels of element connectedness that determine the nature of information, and by learners' knowledge (Van Merriënboer et al. 2006). With respect to element connectedness, instructional materials can be divided into high or low element interactivity materials. For example, students learning the symbols of the periodic table in Chemistry can study each symbol individually with no reference to other symbols. Students learning the symbol for hydrogen, *H*, can do so independently of learning the symbol for copper, *Cu,* without considering any relations between them. Such material has a low degree of element interactivity and a low intrinsic cognitive load.

In contrast, a simple algebra equation such as, $x - 3 = 2$, solve for x, is relatively high in element interactivity. In order to understand and solve this problem, students must consider not only the individual symbols, but also the relations among them. All must be processed simultaneously in working memory. If they are considered in isolation, the problem cannot be understood and solved. Therefore, relatively more interactive elements will need to be processed simultaneously in working memory increasing intrinsic cognitive load compared to low element interactivity material that allows fewer elements to be processed simultaneously.

As well of the structure of information, the expertise of learners also affects intrinsic load. Experienced learners who have acquired relevant schemas for the above problem can treat the entire equation and the problem solution as a single element in working memory, thus reducing the intrinsic load. Element interactivity is a combination of the characteristics of the material to be learned and the knowledge base of the learner. It cannot be determined merely by reference to the characteristics of the information alone. When estimating the level of element interactivity, elements that have been combined into a single, higher order element by relatively more knowledgeable learners enable them to reduce working memory load and so need to be taken into account.

Element interactivity is not equivalent to task difficulty because as indicated above, not all elements interact. A task that requires many elements to be learned will be difficult but because not all of the elements may interact, element interactivity may be low. Learning the chemical symbols of the periodic table or the vocabulary of a second language may be very difficult tasks because there are many elements that need to be learned, but element interactivity is low. Each element can be learned independently of every other element. The task is difficult but working memory load and intrinsic cognitive load is low due to low element interactivity.

Element Interactivity and Understanding Element interactivity also can be used to define "understanding." Information will be fully understood if all interactive elements can be processed in working memory simultaneously (Sweller et al. 2011). Nevertheless, the term understanding tends to not be used when dealing with low element interactive information. If someone deals with information low in element interactivity, such as "Cu" stands for "copper", we would not refer to them understanding or failing to understand the relation. If we fail to recall this relation, we will attribute the failure to forgetting or having no prior knowledge rather than failing to understand. Therefore, understanding is only used for materials high in element interactivity.

The distinction between learning by understanding and learning by rote is also related to element interactivity. Learning by understanding increases the number of interactive elements that must be processed in working memory simultaneously. However, if a large number of interactive elements cannot be handled simultaneously, learning by rote reduces the number of interacting elements albeit at the expense of understanding. Of course, learning by understanding is the ultimate goal of instruction.

Element Interactivity and Extraneous Cognitive Load Extraneous cognitive load is imposed by inappropriate instructional procedures. It must be reduced or eliminated (Kalyuga 2011) to provide more working memory resources to deal with intrinsic load, which enhances learning. Extraneous load also is determined by element interactivity (Sweller 2010). It occurs under conditions where element interactivity can be reduced without altering what is learned.

For example, if instructional procedures require learners to study worked examples, they will need to process fewer elements simultaneously in working memory than if instruction requires learners to solve the equivalent problems.

The Element Interactivity Effect This effect indicates that any cognitive load effects, such as the worked example effect, may not be obtained if the intrinsic load is very low (Sweller et al. 2011). The addition of intrinsic and extraneous load determines the total load imposed on working memory. If the intrinsic load is low, a high extraneous load may not matter as the total cognitive load may still be within the capacity of working memory. However, if intrinsic load is high with a high extraneous load imposed by suboptimal instruction, working memory may be overloaded. Total cognitive load needs to be reduced by reducing extraneous load. Under these circumstances, cognitive load effects can be obtained by reducing extraneous load.

A body of evidence has demonstrated the element interactivity effect. Sweller and Chandler (1994) and Chandler and Sweller (1996) tested for the split-attention and redundancy effects using computers and computer manuals with students learning computer applications. They found both effects using high element interactivity material, but the effects disappeared using low element interactivity material. Rey (2011) also found that the split-attention effect was eliminated for low element interactivity information. Similarly, Tindall-Ford et al. (1997) obtained the modality effect according to which learners who were presented instructions on how to read wiring diagrams and tables in spoken form performed better than students presented the same information in written form, using high but not low element interactivity materials. Leahy and Sweller (2005) tested students learning to read a bus timetable and obtained an imagination effect that occurs when learners asked to imagine procedures learn more than do learners asked to study the same procedures. The effect only was obtained using high rather than low element interactivity material.

Marcus et al. (1996) investigated the relation between levels of element interactivity and understanding by comparing identical textual and diagrammatic information when students learned the effects of connecting resistors in series or in parallel. Textual information required learners to process multiple, interacting elements, while diagrammatic information allowed students to use previously acquired knowledge to treat the multiple elements as a single, schematic element. The results revealed that information presented in diagrammatic form reduced element interactivity and cognitive load.

Expertise and the Element Interactivity Effect Because levels of element interactivity depend not only on the nature of the information being processed but also on the expertise of the learner, learner expertise will also affect the element interactivity effect. For given information, higher levels of expertise reduce the level of element interactivity, whereas lower levels of expertise increase the level of element interactivity. Since levels of element interactivity are affected by levels of expertise, we can expect that the occurrence of the element interactivity effect also will be affected by levels of expertise. As is the case with all cognitive load effects, high element interactivity is a necessary condition. The element interactivity effect itself requires high element interactivity. If element interactivity is low due to high levels of expertise, the effect will not be obtained.

The suggestion that expertise alters element interactivity and provides the machinery underlying the expertise reversal effect is the central thesis of this paper. There is considerable empirical evidence for the suggested effects of expertise on element interactivity leading to the expertise reversal effect. That evidence is discussed below in the "Empirical Evidence for the Hypothesis" subsection.

Human Cognitive Architecture and the Reciprocity of Complexity and Expertise

The reason for the equivalent effects of decreases in complexity and increases in expertise can be found in the cognitive architecture that underlies cognitive load theory. Human cognitive architecture (Sweller et al. 2011) can be used to indicate how novel information is acquired and the differences in the manner in which familiar and unfamiliar information is processed (Sweller 2015).

Human Cognitive Architecture

The Borrowing and Reorganizing Principle Almost all of the knowledge we acquire is borrowed from other people via listening, reading, and imitating before being reorganized when combined with previously acquired information.

Randomness as Genesis Principle Borrowed information initially must be created. It is created by a random generation and test process during problem solving.

Narrow Limits of Change Principle Novel information is initially processed by a limited capacity, limited duration working memory.

The Information Store Principle Long-term memory has a large, effectively unlimited capacity to store information transferred from working memory.

Environmental Organizing and Linking Principle Information in long-term memory does not become active until it has been triggered by cues from the environment that induce working memory to choose which knowledge set to use. The specific knowledge set held in long-term memory can be used to govern complex behavior that is suitable for that environment. Unlimited amounts of information can be transferred from long-term to working memory.

Reciprocity Between Levels of Element Interactivity and Expertise

This cognitive architecture explains the reciprocity between levels of element interactivity and expertise. Based on the environmental organizing and linking principle, knowledge held in long-term memory leads to learners' expertise and determines how they perceive and organize information. Novices do not have relevant knowledge stored in their long-term memory (the information store principle). They are likely to perceive novel information as a collection of discrete, interacting elements that can easily overwhelm limited working memory resources. They have not developed knowledge structures used to integrate individual elements, so a task that is presented may contain high levels of element interactivity leading to a high intrinsic load. In addition, if external guidance is not provided, novices may have to randomly generate solutions (randomness as genesis principle) to solve problems, which will cause a high extraneous load, leaving few resources available for learning (narrow limits of change principle).

More knowledgeable learners use their knowledge to integrate individual elements presented by the same task into fewer elements, reducing the levels of element interactivity. When that knowledge is transferred by experts to working memory using the environmental organizing and linking principle, there may be little pressure on working memory resources. Novices who lack relevant knowledge cannot affect such an action. In this manner, levels of expertise have a reciprocal influence on the levels of element interactivity. For given information, low levels of expertise with respect to that information increase the level of element interactivity, whereas high levels of expertise decrease the level of element interactivity. In turn, these changes in element interactivity have instructional consequences.

Relations Between the Element Interactivity and the Expertise Reversal Effects

As discussed above, the element interactivity effect suggests that every cognitive load effect relies on materials that are high in element interactivity. The expertise reversal effect suggests that instruction that is suitable for novices may not be suitable for more knowledgeable learners. If high levels of expertise reduce the levels of element interactivity rendering most cognitive load effects unobtainable, whereas low levels of expertise increase the level of element interactivity, facilitating cognitive load effects, then the expertise reversal effect may be regarded as an example of the more general element interactivity effect.

A specific example can be used to clarify the relation. Consider the expertise reversal affect as it applies to the worked example effect. We know, based on the worked example effect, that novices are more likely to benefit from studying worked examples rather than solving problems. We also know that with increasing expertise, the worked example effect decreases in magnitude, then disappears and finally reverses with problem solving being superior to studying worked examples.

Consider this expertise reversal effect from an element interactivity perspective. For novices, searching for suitable problem moves using the randomness as genesis principle, determining whether a particular move is suitable with respect to the problem goal, remembering which moves have been previously chosen, both possibly successful moves for later use and unsuccessful moves to ensure they are not chosen again, requires the processing of a large number of interacting elements. Working memory tends to be overwhelmed and learning may be inhibited. Far fewer interacting elements need to be processed when studying worked examples by using the borrowing and reorganizing principle. Learning is facilitated resulting in the worked example effect when compared to problem solving.

Now consider more expert learners presented either problems to be solved or worked examples to study. When solving problems, the learner already is likely to have acquired knowledge indicating which moves need to be made for that particular problem. Practicing those moves may be needed but determining which moves to make is relatively straightforward and can be accomplished merely by referring to information held in long-term memory via the environmental organizing and linking principle. Moves are generated by knowledge rather than the random generate and test process of novices. There may be only a single element (or schema) that needs to be retrieved from long-term memory to generate the problem solution. In contrast, if studying a worked example, more expert learners must compare their known problem solution with the redundant solution presented. The consequence is an increase in element interactivity due to redundancy rather than the decrease we find with novices resulting in a reverse worked example effect with problem solving being superior to studying worked examples. That reverse worked example effect is an example of the redundancy effect.

Based on the above argument, comparing problem solving with studying worked examples causes a reverse result depending on whether novices or more expert learners are used. That result is the basis of the expertise reversal effect, but on the current analysis, leads to the conclusion that the expertise reversal effect is caused entirely by changes in element interactivity. In other words, the expertise reversal effect may merely be an example or variant of the element interactivity effect.

Empirical Evidence for the Hypothesis

There are a number of research studies that were designed to simultaneously investigate the expertise reversal and the element interactivity effects within a cognitive load theory framework. These studies may be used to reveal the hypothesized relation between the two effects.

Kalyuga et al. (2001b) looked at the worked example effect. For high element interactivity material, they found, when testing novices, that studying worked examples was superior to problem solving but that with increased expertise, problem solving was superior to worked examples, providing an example of an expertise reversal effect. In contrast, no significant differences were found with materials that were low in element interactivity. In other words, the worked example effect was obtained with high but not low element interactivity material. That worked example effect could not only be eliminated by using different information that was low in element interactivity, but could also be eliminated by increased expertise that had a similar effect to decreased complexity.

Leahy and Sweller (2005) looked at the imagination effect that occurs when learners asked to imagine procedures or concepts learn more than learners who study the information instead. They found the effect using more but not less knowledgeable students. The less knowledgeable students were not able to imagine the procedures and so needed to study the information. This expertise reversal effect only was obtained using high, not low, element interactivity material. Again, element interactivity could be altered either by altering the information or altering the expertise of the learners. The level of element interactivity was influenced by the level of expertise.

Blayney et al. (2010) studied the isolated elements effect and its interaction with levels of expertise. The isolated elements effect occurs when learners presented with very complex information that normally requires them to process more interacting elements than can be handled by working memory, learn more if the information first is presented in isolated form such that relations between interacting elements are omitted. In a subsequent phase, the information is presented in integrated form emphasizing the interactions between elements. The effect occurs when isolated followed by interacting elements phases results in better performance than multiple presentations of the interacting form only. Students first can learn the isolated elements followed by the interactions between the previously learned elements, without overloading working memory in either phase. In contrast, if the full interacting set of elements is presented initially, working memory is likely to be overloaded resulting in decreased learning.

Blayney et al. (2010) found that in accountancy training, less knowledgeable learners benefited more when presented with isolated elements of information, in accord with the isolated elements effect, but more knowledgeable learners benefited more from interactive elements of information. For less knowledgeable learners who demonstrated a standard, isolated elements effect, we can assume that they required the presentation of isolated elements first in order to be able to process excessive amounts of information in working memory, as indicated above. In the case of more knowledgeable learners, element interactivity and intrinsic cognitive load is reduced due to the environmental organizing and linking principle. Since element interactivity is low for these students, reducing it further by unnecessarily presenting isolated elements will inhibit rather than facilitate further learning. In this manner, the expertise reversal effect that was obtained is really a variant of the element interactivity effect.

The failure to find an isolated elements effect using more knowledgeable learners is no different to the failure to find any other cognitive load effect using low element interactivity information (e.g., Sweller and Chandler 1994; Tindall-Ford et al. 1997). High element interactivity information is essential for any cognitive load effect to manifest itself. Increases in expertise reduce element interactivity and low element interactivity eliminates cognitive load effects. If so, it is the reduction in element interactivity with increases in expertise that underlies the expertise reversal effect.

Blayney et al. (2010) manipulated element interactivity by altering the manner in which the same information was presented to more and less knowledgeable learners. Chen et al. (2015), rather than altering the way in which the same information was presented to learners with different levels of expertise, altered what students at different levels of expertise had to learn. Some of the information was low in element interactivity while other information was high. In addition, rather than investigating the isolated elements effect, Chen et al. (2015) investigated the worked example and generation effects.

The worked example and generation effects are interesting because they ostensibly appear to be contradictory. As indicated above, the worked example effect occurs when learners provided with high levels of guidance in the form of worked examples perform better on subsequent test problems than learners presented with the same material as problems to be solved (Cooper and Sweller 1987; Paas 1992; Paas and Van Merriënboer 1994; Renkl 2014; Sweller and Cooper 1985). Requiring learners to solve a problem provides much lower levels of guidance than studying worked examples.

In contrast to the worked example effect, the generation effect occurs when learners are asked to generate responses rather than being provided with the correct responses. This effect has been investigated by various research studies using different types of testing materials. The most commonly used format is paired associates, such as *hot – c_* (opposite) (Slamecka and Graf 1978). Other research studies used single-word fragments (Glisky and Rabinowitz 1985) in which learners had to generate missing letters to complete word fragments, such as ALC-H-L as the fragments for ALCOHOL; incomplete sentences as contexts (Anderson et al. 1971) requiring learners to generate the last word of an incomplete sentence such as “The doctor looked at the time on his (*watch*)”; and algebra materials (McNamara 1995), such as $2 \times 4 = 8$, in which students needed to generate the answer 8 or read the whole formula. Contrary to the worked example effect according to which explicitly providing problem solutions (providing high guidance) benefits learners more than asking them to solve problems (a low guidance condition), the generation effect demonstrates that generating answers in order to memorize information (a low guidance condition) is more effective than providing answers explicitly (high guidance).

Chen et al. (2015) designed experiments to directly investigate the relations between levels of guidance and element interactivity. They hypothesized that the worked example effect required high element interactivity information while the generation effect required low element interactivity information. Two experiments were conducted in the domain of geometry. Learning simple, low element interactivity geometry formulae were used to test for the generation effect. In contrast, learning to solve geometry problems using those formulae, a high element interactivity task, was used to test for the worked example effect. Participants in Experiment 1 were novices while those in Experiment 2 were more knowledgeable learners. The same topic areas were used in both experiments.

The results indicated that when novices were tested in Experiment 1, the worked example effect was obtained for the high element interactivity information, whereas the generation effect was obtained for the low element interactivity information. In Experiment 2 using more knowledgeable learners, a generation effect for learning formulae or reversed worked example effect for learning problem solutions, was obtained for both sets of information. Generating answers or solution procedures rather than studying provided answers or procedures was superior irrespective whether learners were learning the formulae or learning to use the formulae in problems.

These results support the suggestion that the expertise reversal effect depends on changes in levels of element interactivity. In the first experiment, the worked example effect was obtained using high element interactivity information, while the generation effect was obtained using low element interactivity information. In the second experiment, increased expertise rendered all of the information low in element interactivity and a reversed worked example effect and a generation effect were obtained for all information. It also might be noted that using the high element interactivity problem solving information across both experiments yielded an expertise reversal effect. A worked example effect was obtained using low expertise learners in Experiment 1 while a reverse worked example effects was obtained using higher expertise learners in Experiment 2, with the same content material being taught in both experiments. These results provide evidence that the expertise reversal effect is caused by changing levels of element interactivity due to changes in expertise.

Conclusions

In this paper, we have suggested that there are both theoretical and empirical reasons for assuming that the expertise reversal effect is a variant of the element interactivity effect. From a theoretical perspective, it was pointed out that increases in expertise have long been assumed to result in decreases in element interactivity. Element interactivity associated with intrinsic cognitive load only can be varied by changing the task or changing levels of expertise. Based on human cognitive architecture, a primary manifestation of expertise is the ability to treat multiple elements as a single element in working memory thus transforming our ability to function in a variety of environments. With increasing expertise, high element interactivity information is transformed into low element interactivity information, leading directly to the expertise reversal effect. Instructional procedures designed to reduce working memory load for novices under a high element interactivity environment no longer can reduce working memory load in the already low element interactivity environment of more expert learners. The result is the elimination or reversal of usual cognitive load effects. Empirical evidence for this suggestion comes from data indicating that changes in expertise result in changes in element interactivity, ultimately generating the expertise reversal effect.

It should be noted that a similar argument was presented by Wulf and Shea (2002) in the area of motor learning. They suggested that results obtained from simple motor tasks may not generalize to complex tasks. They also suggested that results using simple and complex tasks may be more similar from data obtained after more practice on complex tasks due to increases in expertise. These suggestions from motor learning bear a considerable similarity to the current suggestions based on cognition.

Cognitive load theory and cognitive load effects are intended to have direct instructional implications, and the current work is no exception. Element interactivity is a central concept of cognitive load theory and all cognitive load effects rely on differences in element interactivity between instructional conditions (Sweller 2010). By analyzing element interactivity between instructional conditions, we can predict which instructional procedures are likely to be effective. That analysis simultaneously must take into consideration both the nature of the information learners are processing and the knowledge levels of the learners. Such an analysis of element interactivity leads to the expertise reversal effect and can provide us with guidelines for effective instructional design.

Acknowledgments We thank the China Scholarship Council which provided a research grant (No. 201308110473) to Ouhao Chen.

References

Anderson, R. C., Goldberg, S. R., & Hidde, J. L. (1971). Meaningful processing of sentences. *Journal of Educational Psychology, 62*, 395–399.

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*, 277–287.

Brunstein, A., Betts, S., & Anderson, J. R. (2009). Practice enables successful learning under minimal guidance. *Journal of Educational Psychology, 101*, 790–802.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*, 151–170.

Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology, 107*, 689–704.

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

Glisky, E. L., & Rabinowitz, J. C. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 11*, 193–205.

Johnson, A., Ozogul, G., & Reisslein, M. (2015). Supporting multimedia learning with visual signalling and animated pedagogical agent: moderating effects of prior knowledge. *Journal of Computer Assisted Learning, 31*, 97–115.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539.

Kalyuga, S. (2011). Cognitive load theory: how many types of load does it really need? *Educational Psychology Review, 23*, 1–19.

Kalyuga, S. (2015). *Instructional guidance: a cognitive load perspective*. Charlotte: Information Age Publishing.

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: introduction to the special issue. *Instructional Science, 38*, 209–215.

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96*, 558–568.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 40*, 1–17.
Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*, 126–136.
Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001a). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579–588.
Kalyuga, S., Chandler, P., & Sweller, J. (2001b). Learner experience and efficiency of instructional guidance. *Educational Psychology, 21*, 5–23.
Leahy, W., & Sweller, J. (2005). Interactions among the imagination, expertise reversal, and element interactivity effects. *Journal of Experimental Psychology: Applied, 11*, 266–276.
Leppink, J., Broers, N. J., Imbos, T., van der Vleuten, C. P., & Berger, M. P. (2012). Self-explanation in the domain of statistics: an expertise reversal effect. *Higher Education, 63*, 771–785.
Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 88*, 49–63.
McNamara, D. S. (1995). Effects of prior knowledge on the generation advantage: calculators versus calculation to learn simple multiplication. *Journal of Educational Psychology, 87*, 307–318.
Nievelstein, F., Van Gog, T., Van Dijck, G., & Boshuizen, H. P. (2013). The worked example and expertise reversal effect in less structured tasks: learning to reason about legal cases. *Contemporary Educational Psychology, 38*, 118–125.
Nückles, M., Hübner, S., Dümer, S., & Renkl, A. (2010). Expertise reversal effects in writing-to-learn. *Instructional Science, 38*, 237–258.
Oksa, A., Kalyuga, S., & Chandler, P. (2010). Expertise reversal effect in using explanatory notes for readers of Shakespearean text. *Instructional Science, 38*, 217–236.
Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. *Journal of Educational Psychology, 84*, 429–434.
Paas, F., & Van Merriënboer, J. J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of Educational Psychology, 86*, 122–133.
Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*, 1–37.
Rey, G. D. (2011). Interactive elements for dynamically linked multiple representations in computer simulations. *Applied Cognitive Psychology, 25*, 12–19.
Rey, G. D., & Buchwald, F. (2011). The expertise reversal effect: cognitive load and motivational explanations. *Journal of Experimental Psychology: Applied, 17*, 33–48.
Rey, G. D., & Fischer, A. (2013). The expertise reversal effect concerning instructional explanations. *Instructional Science, 41*, 407–429.
Slamecka, N. J., & Graf, P. (1978). The generation effect: delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 592–602.
Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312.
Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*, 123–138.
Sweller, J. (2015). In academe, what is learned, and how is it learned? *Current Directions in Psychological Science, 24*, 190–194.
Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185–233.
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.
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory.* New York: Springer.
Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*, 257–287.
Van Gog, T., Paas, F., & Van Merriënboer, J. J. (2004). Process-oriented worked examples: improving transfer performance through enhanced understanding. *Instructional Science, 32*, 83–98.
Van Gog, T., Paas, F., & van Merriënboer, J. J. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction, 18*, 211–222.
Van Merriënboer, J. J., 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.
Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin and Review, 9*, 185–211.