PSYCHOLOGY OF MATHEMATICS LEARNING:

CONSTRUCTING KNOWLEDGE

Endah Retnowati, M.Ed., Ph.D.

NIP. 19801228 200212 2 003

DEPARTMENT OF MATHEMATICS EDUCATION

FACULTY OF MATHEMATICS AND NATURAL SCIENCES

YOGYAKARTA STATE UNIVERSITY

2014

TABLE OF CONTENTS

LIST OF FIGURES

Chapter I   Introduction

Psychology of mathematics learning

Psychology is a science that has been introduced since very long time ago. It lies between biology and culture that specifically explores the causes of thought, act, and behavior. Educational psychology is one of the branches that narrow the science of psychology into the educational field. The scope of educational psychology can be cognitive development, physical development, social and moral development, motivation, intelligence, individual differences, cognitive processes, testing, measurement and assessment as well as classroom teaching. The educational psychology that is specified on cognition is usually called psychology of learning. Scientists in this field attempt to uncover the process of learning and thinking as well as how this affects learners' behaviour.

Mathematics as knowledge or a school subject has its own character that might differ from the others. Basically it has objects that are structurally rigid in patterns and relations. It might become an open thought of connections between mathematical objects too. Its objects are mostly a result of abstractions that are represented in symbols and notations. Therefore, mathematics cognition should be based on these characters in order to construct knowledge and therefore facilitate learning. Accordingly, the psychology of mathematics learning is a specific of psychology science on mathematics cognition.

Meaningful learning

In behavior psychology view, learning is often defined as a permanent change of behavior and hence, learning activity is focused on giving feedback or reward to stimulate change of behavior. However, this definition has been doubted by many cognitive psychologists. In cognitive psychology view, learning is defined as an activity of knowledge construction that changes the structure of previously acquired knowledge. Sweller (1999) defines learning as restructuring knowledge in long term memory where we hold knowledge permanently. It can be argued that eventually knowledge constructed by learners might affect not only their thought but also changes of their acts or behavior.

According to the outcome, learning can be distinguished into two: rote learning and meaningful learning. Through rote learning, a learner might not understand deeply the material since the acquisition is not accompanied by an effort to link among elements of learning material. A learner simply memorizes them. Although memorizing some elements is important to learning, learning is not in-depth. Learners might not fully understand the meaning and the use of the knowledge.

On the other hand, through meaningful learning, a learner puts some efforts to understand the learning material deeply. Understanding is reached by linking

materials to be learned with previously learned knowledge. In addition to this, Mayer (1999) emphasizes that the critical role of learning is to facilitate the transfer of skill, that is the ability of a learner not only to construct knowledge but to apply knowledge into problem solving. Furthermore, it is mentioned that transfer of knowledge occurs when a learner possesses well-structured knowledge in which relevant element of knowledge are closely connected. This could occur when meaningful learning is facilitated.

The above discussion indicates the importance of our knowledge on cognitive architecture and thus meaningful learning occurs. Learning itself indeed involves activities in our memory system and therefore it should be used as the ground on creating effective and efficient learning instructions.

This book specifically covers cognitive psychology perspective on learning mathematics meaningfully. The following chapters include how our cognition works during knowledge construction and how effective mathematics knowledge construction can be facilitated.

Chapter II   Human Cognitive Architecture

Human cognitive architecture refers to the structure of human memory and its functions when processing information, learning and problem solving (Sweller, 2003). There are three major components of memory: sensory memory, working memory and long-term memory. Cognitive activities such as how learners acquire and automate knowledge, solve problems, and develop expertise are reviewed in the second section.

Human Memory

The human memory system is often considered an information processing system. To describe how information is processed and organised in different memory stages, the modal model is historically considered a useful guide.

The modal model

An early version of the modal model was developed by Waugh and Norman (1965), who partitioned human memory into two independent memory systems: primary and secondary memory (see Figure 1). A key feature was that rehearsal in

primary memory enabled stimuli to be encoded into secondary memory. In this

model, primary memory was assumed to have limited capacity and, as a consequence,

most incoming stimuli if unrehearsed were forgotten.
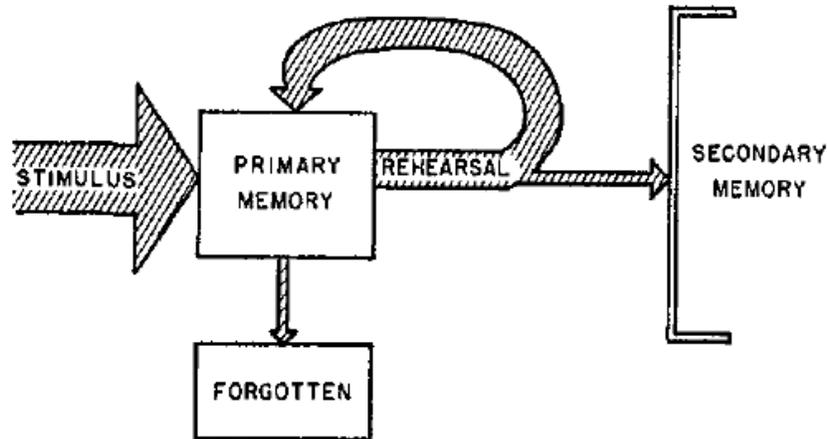
Figure 1. The primary and secondary memory system (Waugh & Norman, 1965, p. 93)

Shiffrin and Atkinson (1969) developed a more complex model, commonly

referred to as the modal model. This model separated the memory system into a short-

term store and long-term store, and included a sensory register which was the first

part of the memory system to receive information from the environment (see Figure

2). The short-term store controlled the flow of information from the sensory register

into the short-term store and eventually into the long-term store. The control process

involved activities such as analysing incoming stimuli, activating rehearsal, searching

and retrieving strategies, and altering, modifying, encoding and transferring

mechanisms between the different memory systems (Shiffrin & Atkinson, 1969).
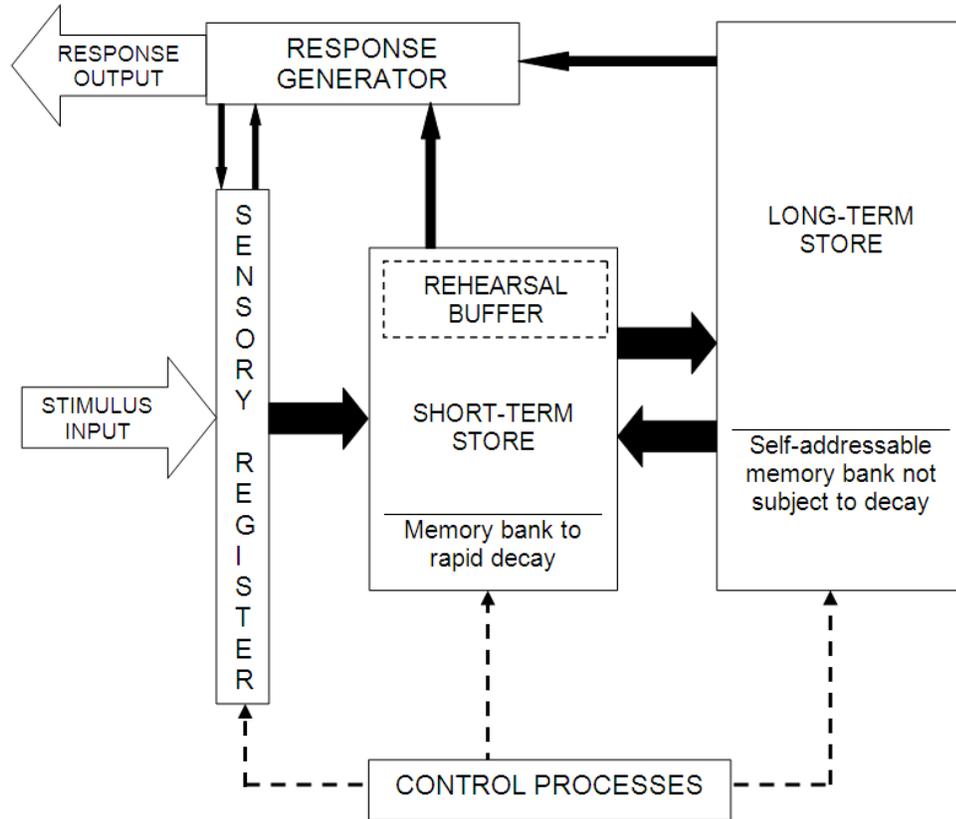
Figure 2. The flow chart of the memory system, adapted from Shiffrin and Atkinson (1969, p. 180). Solid lines describe path of information transfer, broken lines illustrate paths of control activities.

R.C. Atkinson and Shiffrin (1971) later revised their modal model as a unified system by placing the control process inside the short-term store (see Figure 3). It was assumed that the short-term store was central to the cognitive system, playing the major role in information processing, such as rehearsing, encoding, and retrieving strategies.
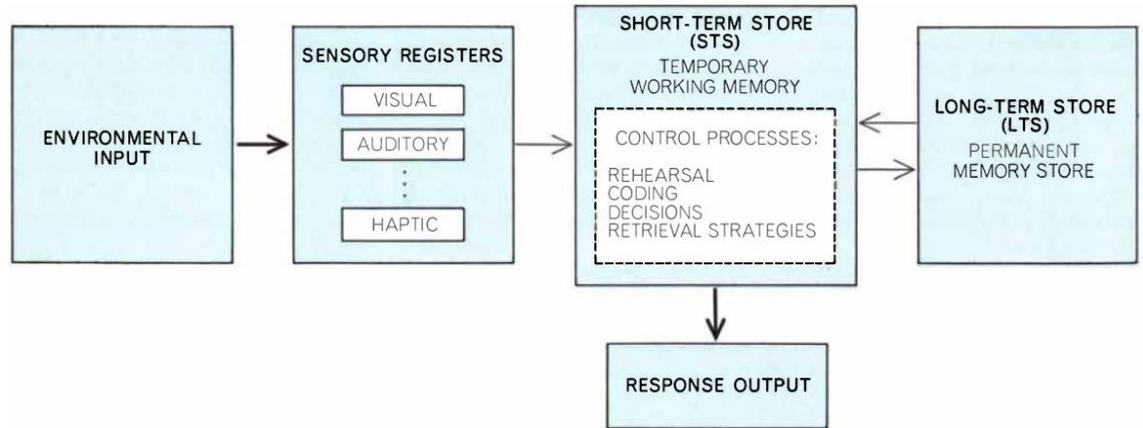
Figure 3. The information flow (R.C. Atkinson & Shiffrin, 1971, p. 82)

These early models provided a general idea of how memory works and how information flows between the different memory components. Other researchers (e.g., Baddeley, 1992, 2002; Ericsson & Kintsch, 1995) have built on these and developed more sophisticated models that more closely match the various cognitive processes that occur when humans learn and solve problems. Nevertheless, many memory models recognise that there are three major components to human memory, which are described in more detail in the following sections.

Sensory memory

Sensory memory is assumed to be the first memory system that holds stimuli received from the environment (R. C. Atkinson & Shiffrin, 1968). We have five

sensory registers (e.g., sight, hearing, smell, taste, touch) that detect stimuli from the environment. However, most of the research into sensory memory has focused on visual and auditory stimuli. By allocating attention or cognitive resources, sensory memory perceives incoming stimuli (R. C. Atkinson & Shiffrin, 1968). Prior knowledge stored in long term memory allows sensory memory to recognise patterns of stimuli (pattern recognition) which, once recognised, is assigned meaning in the short-term store (Cowan, 2005). To illustrate, when we read these letters, because we have prior knowledge about scripts, sensory memory through the visual sense will recognise these letters as scripts. We also know that scripts have meaning and we must pay attention to them. Therefore sensory memory encodes and transfers the stimuli (scripts) into working memory to construct their meaning.

A major feature of sensory memory is that it has a severely limited capacity and duration (Darwin, Turvey, & Crowder, 1972; Sperling, 1960; von Wright, 1972).The visual and auditory sensory registers have been extensively investigated. Sperling (1960), as well as von Wright (1972), demonstrated that the visual register (iconic store) is able to hold only a limited number of visual stimuli (e.g., a set of letters, numbers or icons) for a fraction of a second, and once the cue (sight) has disappeared, the number of stimuli retained by the visual sensory register rapidly decreases over time. Darwin et al. (1972) found that the auditory register (echo store) is also only able to hold a limited number of auditory stimuli (e.g., spoken letters) for a few

seconds (i.e., less than 3 seconds) and then the number decreases as soon as the cue (echo) disappears.

The limitation of sensory memory has specific consequences, in that sensory memory is seldom able to perceive every incoming stimulus or transfer every stimulus into working memory. Unattended stimuli in sensory memory are quickly forgotten (R. C. Atkinson & Shiffrin, 1968, 1969; Waugh & Norman, 1965). As the environment changes rapidly, any memory stored is soon replaced by new stimuli.

In summary, sensory memory involves perception, pattern recognition and the assignment of meaning using prior knowledge stored in long-term memory. Its limited capacity and duration dictate that, at any given time, only a limited amount of stimuli can be perceived by a learner. If assignment of meaning is made, then stimuli are transferred into working memory.

Working memory

Baddeley and Hitch (1974) introduced the term working memory to replace the terms primary memory or short-term memory used by R. C. Atkinson and Shiffrin (1968, 1971). It was argued that the function of this memory component was not simply to provide a temporary store for information, but also to conduct complex cognitive activities. Working memory is also related to consciousness, because when we are actively processing or storing information in working memory, we are aware

of such processes (Baddeley, 2007; Sweller, 2003). However, the cognitive processes conducted in working memory can be both controlled and automatic (Baars & Franklin, 2003). Processing the meaning of new text, for an example, can be a controlled process because we may be searching for meaning and conscious of doing this, whereas while reading the actual words in the text, it will be more automatic because we are familiar with the actual words.

  Working memory: limited capacity. Like sensory memory, working memory has a limited capacity. Under most circumstance, working memory is only able to hold about five to nine chunks (seven plus or minus two) of information simultaneously (Miller, 1956). Miller's findings were revisited more recently by Cowan (2000), who revealed that working memory can only hold less than four chunks of information simultaneously during more complex cognitive activities, such as evaluating, contrasting, or combining new and old knowledge. Ericsson and Kintsch (1995) asserted that these limitations apply for novel material only. When working memory deals with well-learned material stored in long-term memory, its limitation reduces. Other researchers, such as Oberauer and Hein (2012), have argued that learners can only pay attention to a single chunk of information at a given time; meanwhile, Sweller et al. (2011) have suggested that the relative capacity of working memory depends entirely on the complexity of the cognitive process performed. Learning new skills or solving high complexity problems in novel situations

considerably reduces the capacity of working memory to deal with the amount of information presented.

Working memory: limited processing duration. Not only is working memory limited in capacity, but it is also limited in duration. Peterson and Peterson (1959) found that information stored in working memory starts to disappear within a few seconds and is completely lost after 20 seconds because of interference, decay and replacement by subsequent information. Rehearsal can maintain the information in working memory and overcome its loss. But if information is not rehearsed, it will be quickly forgotten (Waugh & Norman, 1965).

Baddeley and Hitch's working memory model. In contrast to the unitary system of working memory proposed by Atkinson and Shiffrin, Baddeley and Hitch (1974) developed a more precise model of working memory. This model sought to explain how received visual or auditory information was processed. As can be seen in Figure 4, the model consisted of three components (Baddeley, 1992).
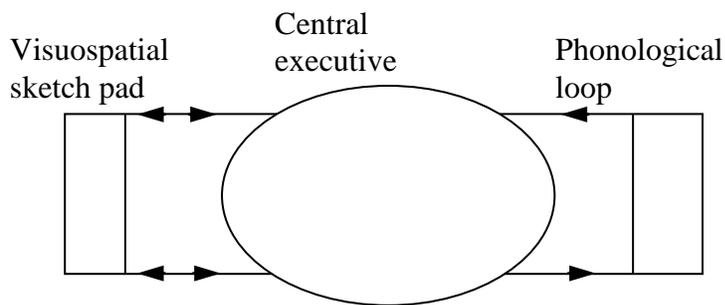
Figure 4. The simplified model of working memory system, adapted from Baddeley (1992, p.557)

There is a *central executive*, which is assumed to be the central controller that guides how stimuli entering working memory are processed, as well as organising the other two components: the *visuospatial sketch pad* and the *phonological loop* (Baddeley, 1992, 1996; Baddeley & Hitch, 1974). These two components are also known as the slave systems of the central executive.

The two subsidiary (slave) components are distinguishable by the different types of information they store and process (Baddeley & Hitch, 1974, Baddeley, 2000). The visuospatial sketch pad stores and maintains visual imageries, such as shapes and colours. It is also able to perform spatial comparisons or movements of visual representations and mentally rotate images. The phonological loop is the verbal analog to the visuospatial sketch pad. Its function is to retain verbal and acoustic information. It is assumed to have a temporal phonological store that is able to hold the information, which will decay after few seconds unless the information is refreshed through articulatory rehearsal. Baddeley

(1996, 2007) asserted that the phonological loop has evolved during language comprehension through the process of rehearsal and response production.

Building on his initial model, Baddeley (2000) later revised the model by adding an *episodic buffer* (see Figure 5).
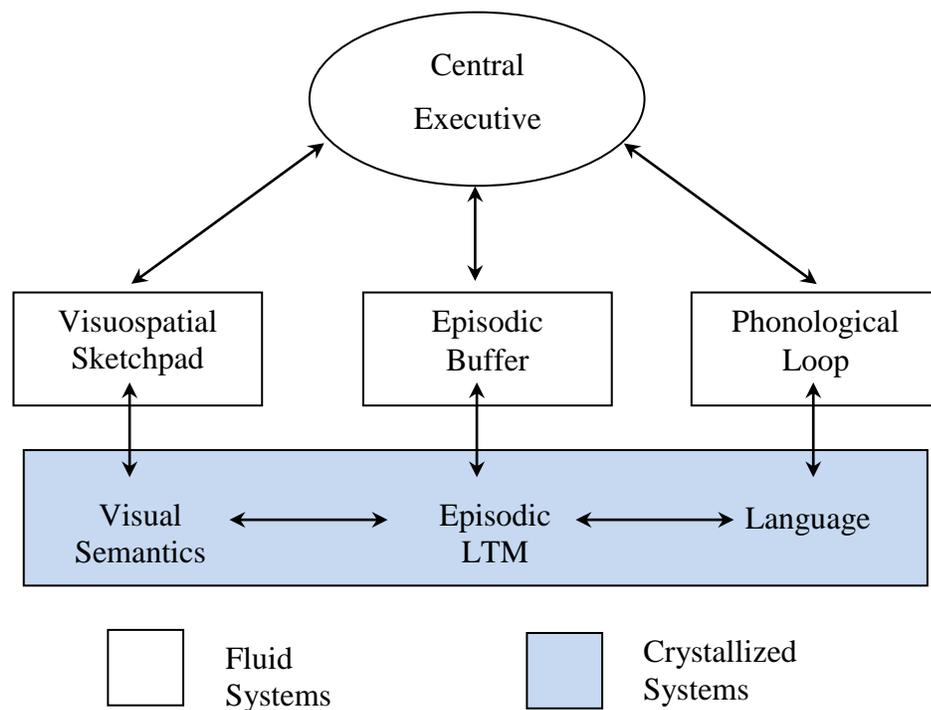
Figure 5. The revised model of the working memory system proposed by Baddeley (2000, p. 421).

The unshaded areas represent fluid systems that have the capacity for attention and temporary storage and are unchanged by learning. The shaded areas represent 'crystallized' cognitive systems that are capable of holding information or knowledge.

The episodic buffer is assumed to have a limited capacity when accumulating information from the slave systems. Specifically, the episodic buffer was proposed as a subsystem that integrates information from both slave systems into unified chunks forming coherent episodes. Baddeley further explained that the integrating process involves an interface between *episodic long-term memory* and the stimuli stored in the slave systems (Baddeley, 2000, 2001). Episodic knowledge (i.e., the recollection of individual events) in long term memory is retrieved by the central executive into the episodic buffer to assist the integration of information in order to represent the information in terms of space and time (Baddeley, 2001).

Moreover, Baddeley (1996) contended that the central executive plays the major role in working memory because it has a function equivalent to a supervisory attention system that plays an important role in managing attention resources for the subsidiary systems. Specifically, the central executive has control over the episodic buffer enabling it to integrate new knowledge from the visuaospatial sketchpad and the phonological loop, with information stored in long term memory (Baddeley, 2000, 2001).

It has been highlighted that sensory memory and working memory are restricted in capacity. In contrast, long-term memory has an extremely large capacity, and exerts considerable influence over working memory. In the following section long-term memory and how knowledge is organised are discussed.

Long term memory

Long term memory is considered to be a permanent store of information accumulated and transformed over human lives (R. C. Atkinson & Shiffrin, 1968; Sweller et al., 2011). It provides a permanent storage repository in our cognitive architecture and has a virtually unlimited capacity. It is argued that long-term memory is central to human cognition, since the purpose of learning is to change the knowledge structures contained within long-term memory (see Sweller et al., 2011).

Initial evidence that long term memory has an unlimited storage capacity came from De Groot (1978) in his classic study on chess. The study was completed between 1938 and 1943, and was originally designed to analyse how chess players think when making their moves. The participants consisted of well-known chess grandmasters, masters, champions and less-skilled players, who were asked to think aloud when deciding what moves to make. De Groot predicted that grandmasters would use sophisticated strategies and tactics, but found little difference between the methods employed by chess players with different levels of expertise.

De Groot also asked participants to look at specific chess configurations on a board for a short time interval (10 to 15 seconds), and then asked them to reconstruct the configurations, after the board was withdrawn from sight. De Groot found that grandmasters were able to efficiently grasp the configuration in less time (about five seconds), more accurately reconstruct it, and could then indicate the best associated next move, compared to the less-skilled players. He further found that grandmasters

could visualise chess configurations as meaningful structures, which enabled them to understand the underlying concepts involved. They possessed a large knowledge base of chess positions and appropriate moves that gave them a distinct advantage over less experienced chess players. De Groot suggested that the vast amount of chess structures remembered by grandmasters were a result of many years of experience. De Groot's study greatly influenced the study of human cognition because it illustrated, for the first time, that expertise in chess was a result of knowledge stored in long-term memory, rather than superior on-the-spot problem solving and planning.

Chase and Simon (1973) replicated de Groot's finding that master chess players were superior to novel players when asked to memorise real chess configurations. However, when shown chessboards where the pieces were randomly placed, experts had no better recall than novices. This last result demonstrated that chess experts do not have a greater working memory capacity than novices as such. When trying to remember random configurations, all players were constrained by Miller's restricted capacity findings. It was only when the chess pieces created a meaningful pattern that experts had better recall as a result of a superior long-term memory.

Further research by Simon and Gilmartin (1973) using a computer program to simulate the chess position reconstruction process, suggested that master players have acquired hundreds of thousands of chess configurations in their long term memory which enables them to create meaningful chunks for chess positions. These chunks

not only allow superior short-term recall of chess games, but also enable them to recognise a game situation, and the best way to proceed in the game.

Following research into chess, studies in other domains, such as in electronics (Egan & Schwartz, 1979), physics (Chi, Feltovich, & Glaser, 1981; Chi, Glaser, & Rees, 1982), and mathematics (Sweller & Cooper, 1985), have provided more understanding on how expert learners have acquired unlimited amounts of knowledge. Consistently it has been shown that knowledge stored in long term memory is the major difference between experts and novices in any given domain (see Ericsson, Charness, Feltovich, & Hoffman, 2006). Not only do experts have more knowledge than novices, but also they possess knowledge that is more connected and sophisticated (see Section 1.3 for more detail). The following section discusses schema theory, which provides a well-researched explanation of how knowledge is constructed and structured in long term memory, and how it becomes automated.

Schema Theory

Schema theory is used to explain how knowledge is stored in long term memory (Sweller et al., 1998). A schema is an internal mental representation in human cognition, and its definition emerged from Bartlett's study of thinking processes as well as Piaget's theory of cognitive development (Mayer, 1977).

Schema construction

Bartlett's study. Bartlett (1932) introduced the term schema as "an active organised setting" (p. 209) based on the accumulation of past experiences. In his groundbreaking study, Bartlett analysed the quality of information recall using several methods and materials, such as facial images, novel folk stories, picture writings, descriptive and argumentative prose passages.

Concerning the highly influential folk story, *The war of the ghosts* (p. 65), he observed that subjects could not replicate the material literally (as it is). He found that there were tendencies to transform the material into a general nature, or the rationalisation of unfamiliar characters by changing them into more familiar characters. The British students in the study were unable to make sense of some aspects of the culturally different North-American folk story and therefore interpreted it in terms of their own culture. Hence it was concluded that human memory attempts to make meaning of perceived information by connecting it to something that has been acquired previously. Bartlett named this knowledge structure as a schema. The British students re-constructed the story in terms of their own culture-specific knowledge. Having perceived the information in a certain fashion, schematic knowledge highly influenced their responses.

Piaget's theory. While Bartlett's study shows that prior-schemas influence the construction of new knowledge, Piaget theorised how schemas (schemata) are constructed and re-constructed. According to Wadsworth (1978), Piaget defined a

schema as a structure of knowledge that acts as an active internal reorganisation of information. Piaget asserted that schemas must be constructed purposefully by linking actions on objects with one's own experiences of similar things, and therefore schema construction is an individual unique action (Wadsworth, 1978).

Piaget's original studies investigated children's cognitive development using a clinical observation method (Mayer, 1977). Although the use of this method was criticised, Mayer argued that Piaget's theory has been very influential in the area of human cognition. A key aspect of the theory assumes that schemas evolve because humans have always learned from the environment in order to survive (Mayer, 1977).

The construction of new schemas occurs in two ways: assimilation and accommodation (Mayer, 1977; Wadsworth, 1978). *Assimilation* occurs when new information, which is similar but not identical to the existing schematic knowledge, is integrated or incorporated into an existing schema. To illustrate this, students learn the concepts of square and rhombus by associating them with a known schema for a rectangle, which is a plane shape with two sets of parallel sides.
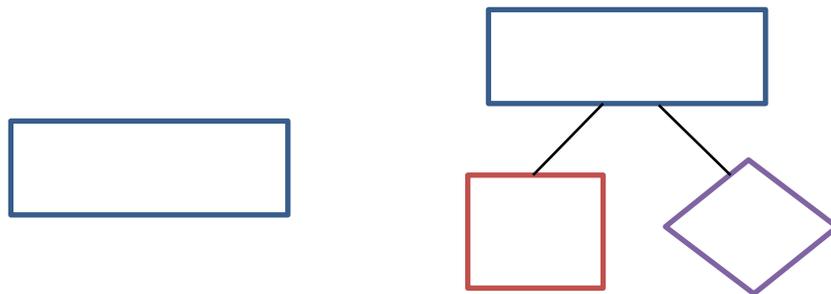
Old schema ------------------- new schemata as a result of assimilation

Figure 6. Assimilation of new schemata

Accommodation occurs when changes in the existing schema are made in order to fit the new information. When a student is learning about the new concept of a square by looking at its similarities and differences with the existing concept of a rectangle, the central concept associated with a rectangle has to be restructured to accommodate a quadrilateral with the same length of sides.
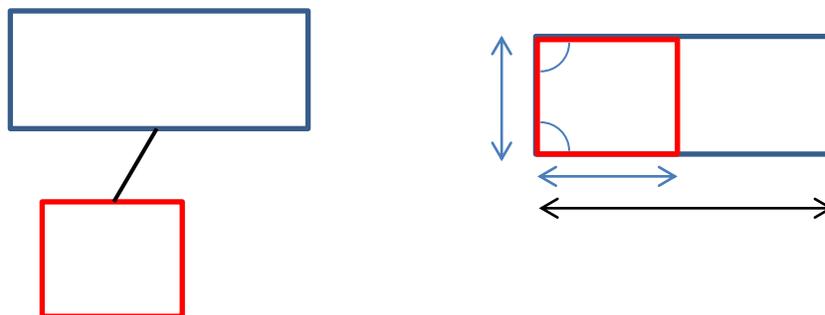
Old schemata as a result of assimilation-- new schemata as a result of accommodation

Figure 7. Accommodation of schemata

However, under most circumstances, assimilating new information and accommodating an old schema to fit a new schema causes *cognitive disequilibrium*. Cognitive disequilibrium leads to a search process of balancing cognitive structures which is called *equilibration*. An active equilibration process allows learners to continuously assimilate new information and accommodate old schemas until a better sense of understanding is obtained. Equilibration maintains the integration of new schemas and also the alteration of the older schema into a more organised representation (Mayer, 1977; Wadsworth, 1978).

A key function of schemas is that they organise information stored in long-term memory (Sweller et al., 2011; Sweller et al., 1998). A number of elements of information can be incorporated into a single schema. For example, four elements, floor, wall, roof and space, can be incorporated into a schema based on the concept of room. The schema has a meaningful interrelated structure linking the properties of the four elements and how they fit together and interact. Furthermore, the element floor can consist of sub-elements such as foundation, base, flat, tile, carpet, rectangular shape, ground, and so forth. This example illustrates how schemas consist of connected information, using various levels of information. Schemas can be grouped together to form broader categories. For example, room can be considered a sub-category of a house schema.

Central to cognitive load theory is the use of schemas to help reducing working memory load (Sweller et al., 2011; Sweller et al., 1998). Because schemas

consist of connected elements, information can be retrieved from long term memory as chunked information. For example, consider the English alphabet. When very familiar with it, students can use the alphabet as a single piece of information, rather than 26 individual elements. Consequently, if asked to remember the alphabet and complete other tasks it may be easy to complete these tasks, as only one other piece of information has to be remembered simultaneously, in addition to completing the tasks. The same argument applies when constructing new schemas; if relevant chunked information in the form of schemas can be accessed from long term memory, more available resources are available in working memory to process the information required to acquire new knowledge.

Schema automation

Schematic knowledge can help overcome the limitations of a very limited human working memory as described above. Another advantage occurs when schematic knowledge is automated. Schema automation occurs when a schema can be activated effortlessly (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Being able to access schematic knowledge with little effort again has little impact on working memory load, and therefore allows more cognitive resources to be used for schema construction (Sweller et al., 2011; Sweller et al., 1998). It is found that schema automation occurs gradually and only with extensive use of the schema (Cooper & Sweller, 1987; Ericsson, Krampe, & Tesch-Romer, 1993).

The following example illustrates this point. The first time we learn handwriting, we have to consciously pay attention to how to hold a pencil when writing basic text like the alphabet. Focusing on the actual writing can interfere with learning the alphabet. However, after a period of practice, we are able to write more complicated passages without paying much attention to how we hold the pencil, because handwriting skills have become automated. Similarly, when we learn to read, we need to consciously recognise letters and pronounce every syllable. However after many years of reading, we can read familiar text without effort as reading has become automated. Hence working memory resources can be directed to understanding more complex text. In mathematics, the first time we learn simple addition, we have to consciously pay attention on numbers or maybe their representations. However, after some times of practice, we are able to do mental addition and hence we can do multiplication without paying too much attention on the simple addition anymore.

Highly automated schemas are very desirable, but research has shown that extensive practice is required for knowledge and skills to become automated (see Ericsson et al., 2006). Recall, de Groot (1968) found that grandmasters spent many years of playing chess before they were able to efficiently recall chess configurations given a short exposure. Similarly, experts in physics had spent years building knowledge in their domain, enabling them to excel in physics problem solving (Chi et al., 1981; Chi et al., 1982).

Summary

Early attempts to model our cognitive structures were based on the modal model (R. C. Atkinson & Shiffrin, 1968; Waugh & Norman, 1965). This model included three distinct memory systems: sensory memory, working memory and long term memory. Sensory memory is the bridge between our virtual system and the environment. Working memory is where conscious thought processes the meaning of information perceived and transferred by sensory memory. The Baddeley working memory model (Baddeley, 1992, 2000; Baddeley & Hitch, 1974), expanded upon the early models by arguing that working memory itself consists of a central executive and subsidiary components: the visuospatial sketch pad, the episodic buffer and the phonological loop. Each sub-system has its own function controlled by the central executive. Both sensory and working memory have severe limitations in capacity and duration, while long term memory can store unlimited amounts of knowledge. To explain how long term memory stores and uses knowledge, schema theory (Bartlett, 1932; De Groot, 1978) has been highly influential. A schema is a knowledge structure that combines elements of information into the category in which it will be used (Sweller et al., 1998).

Schemas, stored in long term memory, are retrieved into working memory, and used to process new information and subsequently help develop new schematic knowledge. Schema construction is managed by working memory, and because of its limited capacity, automated schemas are very helpful and necessary (Sweller et al.,

1998). Schema automation allows cognitive processes to occur effortlessly in

working memory with little demands on cognitive load, thus making more cognitive

resources available for more complex cognitive tasks, such as problem solving.

Chapter III   Experts and Novices' Cognition

The previous chapter covered human cognition and specifically highlighted how the modal model can describe how knowledge is acquired, how schema theory can explain the way knowledge is stored and organised in long term memory, as well as how schema automation can reduce cognitive load in working memory. It has been pointed out that the study of human cognition was a major impetus in the study of expert cognition (Feltovich, Pritetula, & Ericsson, 2006).

Much research in the field of expertise has been conducted by two general approaches (Chi, 2006): (1) studying how a highly distinctive expert performs in their domain of expertise, or (2) comparing experts to novices. Chi concluded that studies on expertise have not only identified ways in which experts excel, but also ways in which they do not. Building on empirical evidence, generalisable characteristics of experts have been developed (Feltovich et al., 2006; Glaser & Chi, 1988), and these are discussed in the following section. In recent developments, the study of expert cognition and characteristics of expertise have influenced educational goals, particularly in instructional designs. It is argued that the aim of classroom instruction has shifted from a focus on behavioural changes to the development of expertise (Amirault & Branson, 2006; Feltovich et al., 2006).

Generalisable characteristics of expertise

Feltovich et al. (2006) pointed out that the study of expertise has progressed since the early work of Chi and colleagues on expertise characteristics (Glaser & Chi, 1988), covering much wider contexts, many of which are reported in the handbook entitled *Expertise and Expert Performance* edited by Ericsson et al. (2006). Feltovich et al. (2006, pp. 47-60) identified nine characteristics of expertise, which are briefly summarised below.

1. Expertise is limited in its scope and elite performance does not transfer. Feltovich et al. found that there was strong evidence in support of the finding that experts excel in their own domain. In addition when experts have reached an elite level, it is rare for an individual to have a second domain area of expertise. High levels of proficiency in one domain do not transfer to high levels of proficiency in a second domain, even when there may be similarities between domains.

2. Knowledge and content matter are important to expertise. This characteristic elaborates the first characteristic above. Feltovich et al. found that experts rely heavily on their domain specific skills and knowledge to produce a superior performance in various tasks in their domain of expertise. It was found that their specific expertise influenced basic cognitive abilities, such as reasoning and encoding. Experts also acquire rich knowledge in a specific domain and continuously develop their level of expertise. As a consequence, their increasing level of expertise gradually turns them into unique individuals.

3. Expertise involves larger and more integrated cognitive units. Feltovich et al. also asserted that there was strong evidence in support of the notion that experts are able to perceive larger amounts of information in working memory because they have acquired superiority in chunking large amounts of information with their increased experience. They become experts because they have superior encoding and storage skills that allow them to build well organised knowledge in long term memory. This superiority is acquired through extensive practice.

4. Expertise involves functional and abstracted representations of presented information. This means that experts are able to see and represent information at a deeper level using specific principles or rules associated with the problem. Feltovich et al. added that experts can develop more complex (functional and abstracted) representations of information by immediate and integrated access to knowledge relevant to task demands because they have acquired more organised retrieval skills.

5. Expertise involves automated basic strokes. In addition to their effective encoding, storage and retrieval skills, experts have highly automated skills. Feltovich et al. pointed out that this automation results from consistent practice on tasks specified in the domain of expertise over a very long period. It was also found that automation plays a great role not only in accomplishing more complex skills but also in controlling the usability of available knowledge.

6. Expertise involves selective access of relevant information. Feltovich et al. found that experts are better able to pay attention to relevant information by using discriminating cues. They are able to utilise functional and abstracted models to categorise information and transfer their knowledge of past events to new ones. Moreover, they are able to recognise the particular information in a task and adequately use their knowledge to perform that task.

7. Expertise involves reflection. Experts have a good understanding of their own cognitive processes, indicating metacognitive skills. Feltovich et al. noted that research consistently showed that they not only have the capability of planning a solution process, but are able to modify and adjust their plans during the problem solving process. Expert monitoring behaviour has three functions. *Firstly*, it provides efficient and rapid reactions to situation changes. Experts are able to back-track or start again when their reasoning needs to be modified. *Secondly*, experts can simultaneously improve and refine their skills. *Thirdly*, experts can adjust their planning to meet the demands of novel situations. When experts fail they can explain why such procedures were inapplicable, whereas novices cannot.

8. Expertise is an adaptation. Experts have the ability to adapt to cognitive restrictions, such as limitations posed by attention resources and working memory when novel or simultaneous information is present, and an impaired access to long term memory (e.g., forgetting an important aspect). This adaptation consequently encourages them to generate effective applications to task demands.

9. Simple experience is not sufficient for the development of expertise.

Supported by the study of deliberate practice by Ericsson and colleagues (e.g.,

Ericsson et al., 1993), Feltovich et al. concluded that to acquire the above

characteristics of expertise, a learner needs to practice consciously in a working

environment that is designed to achieve performance superiority, and perform that

practice over a substantial period of time. As Ericsson (e.g., Ericsson et al., 1993), has

observed, in most domains it takes 10 years of deliberate practice to achieve expert

status.

Difference in problem solving strategies between experts and novices

As described above, much is known about the characteristics of experts. Of

particular importance to this book, and cognitive load theory in general, is the

difference between experts and novices when solving problems. This section

discusses problem solving and the strategies used by experts and novices.

Problem solving.

Problem solving is an activity to find a solution to a given problem that cannot

be solved immediately (Kantowski, 1977). In other words, no automatic solution is

available to the problem solver. Sweller (1999, p. 3) gives the following example:

"Suppose five days after the day before yesterday is Friday. What day of the week is

tomorrow?" This specific problem seems to be familiar to us, yet the solution may

not be available immediately. According to the prior-knowledge of the problem

solver, solutions may vary from algebraic equations to trial-and-error.

According to Kantowski (1977) problem solving consists of a set of activities

(the process) and actual solutions (the product). Specifically, Anderson (1993)

summarised that problem solving creates a problem space consisting of a number of

states dependent upon the rules of the problem. Problem solving attempts are made to

find connections between the facts or rules within the problem space as well as to

create a path between the given state and the goal state.

Problem solving is common in our everyday life, and so it is essential to be

able to solve problems. As argued by Schmidt, Loyens, Van Gog, and Paas (2007)

problem solving is an important process in learning since it facilitates reasoning and

the ability to explain observable facts and occurrences. However, to solve problems

effectively depends on the level of expertise of the problem solver, since relevant

schemas in the domain are essential to recognise and solve problems. As was

discussed above, experts possess well-developed automated schemas, which enable

them to categorise a problem based on its deep structure and solve it effectively (Chi

et al., 1982, Feltovich et al., 2006). In contrast, novices or less-knowledgeable

learners in a domain do not have sufficient schemas and categorise problems

according to a more superficial structure, resulting in inefficient solutions or no

solution at all.

Problem solving skills

Kantowski (1977) stated that problem solving has two aspects: (1) the *process*: a set of activities, and (2) the *product*: the actual solution. It is noted that an *ill-defined problem* has multiple acceptable products and many possible ways for reaching them, while a *well-defined problem* has only one possible product and one agreed process for reaching it (Brunning et al., 2004).

Brunning et al. (2004) assumed that successful problem solvers engage five component skills: (1) identifying the problem, (2) representing the problem, (3) selecting an appropriate strategy, (4) implementing the strategy, and (5) evaluating solutions. These skills are heavily constrained by domain specific knowledge (secondary knowledge) and general problem solving strategies (biologically primary knowledge). Identifying the problem can be the first challenging part in solving a problem, since it requires creativity, persistence and willingness to think carefully about the problem over sufficient time. The degree to which the problem solver acquires domain specific prior knowledge can determine successful problem finding since prior knowledge facilitates perception and elaboration of new information. For instance, mathematical experts may be less able to identify a medical problem because they may not have sufficient prior knowledge that allows them to identify medical problems, and on the other hand, a doctor can identify medical problems but may not be able to solve some mathematical problems as mathematicians do.

After problem finding, problem solvers may need to represent the problem externally as the amount of information needed to deal with complex problems is constrained by working memory load and so too difficult to solve mentally. In order to find a strategy path, problem solvers should represent the more important components of the problem space: goal state (what we want to accomplish), initial state (what is the given information), operators (objects or concepts that can be used to reach the goal) and constraints on operators (rules or procedures to be used by the operator). It has been argued that the size of a problem space depends on the way the problem is understood or the level of expertise (Bransford, Brown, & Cocking, 2005). More knowledgeable problem solvers tend to categorise the problem space based on principles or solution strategies that are relevant to solve the problem, because they have sufficient knowledge and experience of it. However, less knowledgeable problem solvers rely on the surface structure such as the objects that appear in the problem. Evidence for this hypothesis was shown in the first experiment of Sweller and Cooper (1985) that investigated the algebraic problem representation skills of three different age levels using *Einstellung* and memory tests. *Einstellung* is described as an occurrence of inappropriate use of a previously acquired schema because a problem is incorrectly perceived as belonging to a familiar category that requires the use of that particular schema (Sweller & Cooper, 1985). The results suggested that the more experienced students had the better cognitive representation, indicated by a superior memory of actual algebraic equations and an increased

resistance to *Einstellung* effects in operating on equations, than the less experienced students. As a consequence, providing more practice on a particular problem type as well as on analysing different problems by less experienced problem solvers can improve their ability in categorising the problem space.

Thirdly, problem solving requires selecting an appropriate strategy: that can be a highly structured strategy, namely, an algorithm, or a general problem solving strategy (Geary, 2007), which is broad knowledge that is not connected to a specific domain but generally needed for completing problem solving tasks, for example vocabulary to express ideas, general search information skills or metacognitive knowledge to carry out problem solving activities (Brunning et al, 2004). The problem of finding the volume of a geometrical shape which can be solved using the volume formula is an example of the use of an algorithm based strategy. Expert problem solvers in the domain, not surprisingly, are able to retrieve or to select the appropriate algorithm because of their proficient schematic knowledge or their large experience of planning strategies. However, using an algorithm based strategy is impossible for novice problem solvers because either the algorithm does not exist in their long term memory or they lack expertise in using it. Subsequently, novice problem solvers will use a general problem solving strategy that is traditionally called a heuristic or "*rule of thumb*", trial and error, or means ends analysis. Brunning et al (2004) indicated that people who deal with a very unfamiliar problem may not have

sufficient information or experience to derive a strategic solution plan. They might use trial and error at the start and then, after reaching some preliminary conclusion to the problem, turn to a more efficient strategy. This strategy may work to obtain a solution but such a strategy is considered the least efficient method of problem solving because it does not direct the problem solvers' attention to acquire practical schemas in their long term memory (Sweller, 1999).

Schoenfeld (1980) defined a heuristic strategy as: "a general suggestion or technique which helps problem-solvers to understand or to solve a problem" (p. 795). Heuristic strategies include strategies used by expert problem solvers that are stated as short explanations or clues. Heuristic strategies use working backwards, or a looking back strategy to search for a solution (Kantowski, 1977). Schoenfeld (1980) provided an example of the use of a heuristic strategy in mathematics as follows.

"To solve a complicated problem, it often helps to examine and solve a simpler analogous problem. Then exploit your solution."

Problem 5: Let a, b, and c be positive real numbers. Show that not all three of the terms $a(1-b)$, $b(1-c)$, and $c(1-a)$ can exceed ¼.

(Schoenfeld, 1980, p. 795)

A heuristic method applied to the above examples shows that the problems can be solved by examining and applying a simpler analogous problem to the given

problem. Arguably, the heuristic method may be a source of extraneous cognitive load. In problem 5, an analogous two-variable formation may be used and then expanded to three variables to exploit the answer. This problem seems difficult to solve, indeed, the author stated that an easier way to solve this problem type using analogous problems had not been found. Nonetheless, the author argued strongly that using a heuristic strategy with an example of how that strategy works and training to work with it will facilitate problem solvers, without explicitly explaining how detailed the heuristic strategy is.

The heuristic method imposes a heavy working memory load because, rather than facilitating schema acquisition and automation, a heuristic strategy suggests learners create sub goals or analogous problems that will result in slower learning and the hazard of misconception.

It has been indicated that a heuristic strategy can be applied differently to different problems and to do so, one needs to retrieve other schemas in order to find an analogous problem and then apply means ends analysis or a trial and error approach. Prior knowledge possessed by problem solvers is the reason for this. In addition, it can also be argued that a heuristic method does not always guarantee a solution and even makes problem solving more difficult. Notwithstanding the fact that a heuristic method may provide a stepping stone, it is obvious that a heuristic

strategy is inefficient for learning novel problem solving because it imposes a high cognitive load.

The fourth problem solving skill, implementing the strategy, largely depends on the result of identifying and representing the problem, and selecting the appropriate solution strategy. Notably, expert and novice problem solvers have a clear difference in their implementing a solution strategy since they have a contrasting schema structure. The schematic structure of this knowledge in long term memory largely determines how expert problem solvers derive and implement a solution strategy (Sweller, 1999; Sweller et al., 1998). Expert problem solvers posses a well developed declarative knowledge base about how a problem is structured, procedural knowledge about how to perform a problem solution and conditional knowledge about when and why to use declarative and procedural knowledge. This knowledge is developed by deliberate practice (Erricson, 2003) and so the more experience gained, the better the problem solving strategy. In contrast, novice problem solvers posses either less prior knowledge required to identify and represent the problem or less experience in selecting a strategy to solve domain specific problems. Furthermore, Sweller (1999) and Sweller, et al. (1998) pointed out that less knowledgeable problem solvers coordinate the problem solution phase poorly, consider single solutions based on a noticeable problem space and reach conclusions that may be less transferable to another problems.

The fifth problem solving skill is evaluating the solution both in terms of the process and the product of problem solving (Brunning et al, 2004). Evaluation of the solution allows us to reflect more deeply on the process of problem solving and so understand the application of a specific strategy. Expert problem solvers are more likely to consider more solutions and carefully evaluate solutions before discarding them, unlike novice problem solvers (Brunning et al, 2004). Pawley, Ayres, Cooper & Sweller (2005) investigated the effect of checking instructions in translating a word problem into algebraic equations. In the experiment, Pawley, et al. explicitly instructed students to check whether the equation formed has the same meaning as the given word problem. The results suggested that checking instructions was beneficial for lower knowledge students but not for higher knowledge students. It was found that higher knowledge students were capable of completing the problem better without explicitly instructed to check. Similar to Brunning et al. (2004), Pawley et al. (2005) argued that checking instructions may be a redundant activity for more knowledgeable students because they already posses evaluation skills as part of their problem solving approach.

The problem solving strategies of experts and novices.

Experts are able to understand and categorise problems efficiently and use a forward moving strategy to solve them (Ayres & Sweller, 1990). For example, an expert mathematician can solve an arithmetic word problem by creating equations to

represent the problem mathematically, and then generate the appropriate equations to solve the unknown variables required. They have acquired sufficient schemas permitting them to effectively select the associated steps to move forward from the problem statement towards the problem goal.

Novices or less-knowledgeable learners without schematic knowledge will more likely use a general problem solving strategy, such as means-ends analysis (Ayres & Sweller, 1990; Sweller, 1988). Using means ends analysis, problem solvers try to reduce the distance between the given information and the problem goal by creating sub-sub goals and then examining them individually to find the solution. The inefficiency of means ends analysis has been confirmed by several studies (Ayres & Sweller, 1990; Sweller, 1999; Sweller & Cooper, 1985).

Ayres and Sweller (1990) investigated the effect of using means ends analysis during geometry problem solving. It was found that most errors occurred during the calculation of the sub-goal preceding the goal in either two or three step geometry problems. The authors stated that means ends analysis is often used not only when calculating the goal of the problem, but also in the sub goal prior to the goal. The use of means ends analysis might be beneficial for some problem learners when dealing with unfamiliar problems, because it can increase the chance of completing the goal of the problem, however, it does not necessarily facilitate learning. In addition, the difficulty level of the problem (or the intrinsic cognitive load) many contribute to the

use of means ends analysis. Ayres and Sweller, in their experiment, confirmed this fact and demonstrated that after the unfamiliar problems were altered to reduce the use of means ends analysis by constructing configurations that had a clear solution path thus encouraging a forward strategy, the calculation error location is random. This means that the use of means ends analysis can be minimised by tailoring the configuration of the problems. The authors concluded that the use of means ends analysis imposes a heavy cognitive load and can be minimised.

The means ends analysis strategy creates sub-goals or analogous problems by breaking down a problem into smaller sub problems, and testing the effectiveness of each step (Ayres & Sweller, 1990). In other words, to search for a solution, problem solvers move backwards from the goal to the problem state, creating sub-goals to be found in the process. This strategy may result in a problem solution, but can create heavy demands on working memory and direct cognitive resources away from schema construction (Ayres & Sweller, 1990; Sweller et al., 2011; Sweller et al., 1998).

Steps of Means Ends Analysis
1. Looking at the initial problem state
2. Looking at the current problem state
3. Looking at the goal state
4. Defining differences between these states
5. Finding moves to reduce those differences
6. Considering sub-goals that may lead to a solution

Figure 8. Steps of Means Ends Analysis

The following example of problem solution illustrates the use of means ends analysis strategy using a geometry problem: Finding a measure of an angle, which consists of eight steps of means ends analysis. It should be noted that this strategy is not effective and should only require two steps only to solve by applying satisfying knowledge if possessed by the problem solver.

Find a value for angle X
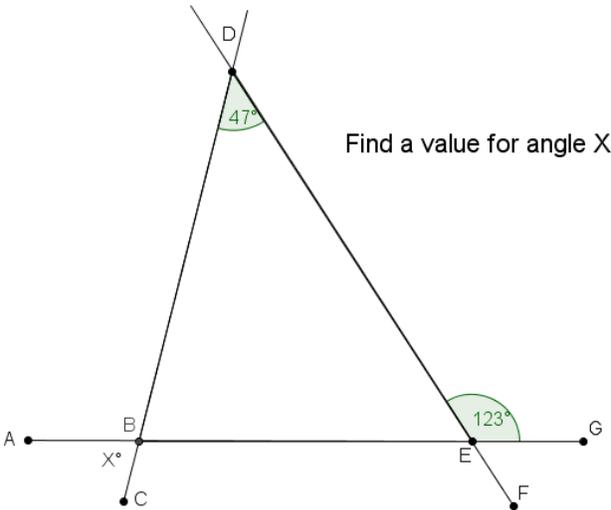
Step 1. Identify the goal : angle X^o
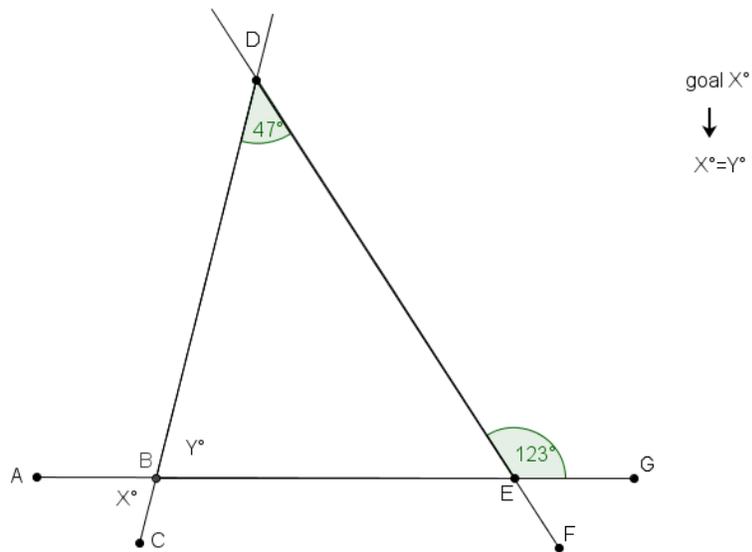
Step 2. Found that $X^o = Y^o$
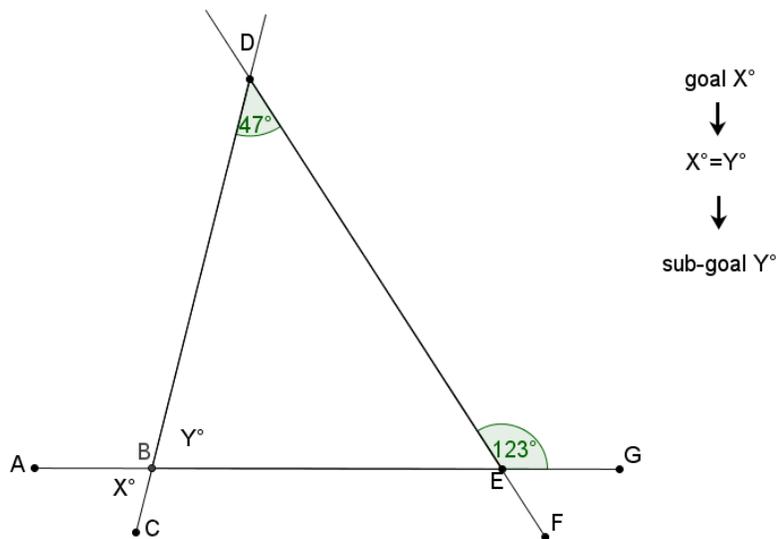
Step 3. Creating a sub-goal Y^o
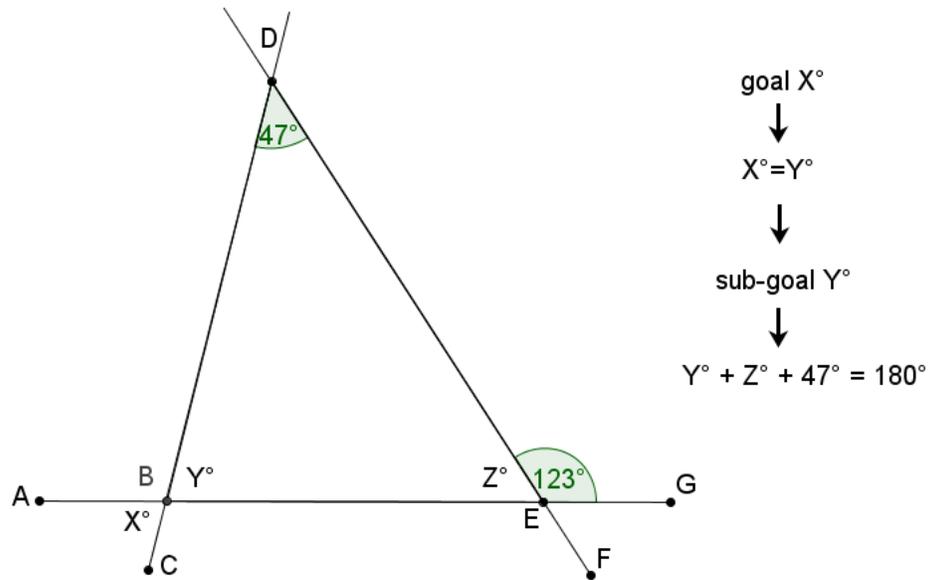
Step 4. Found that $Y^o + Z^o + 47^o = 180^o$
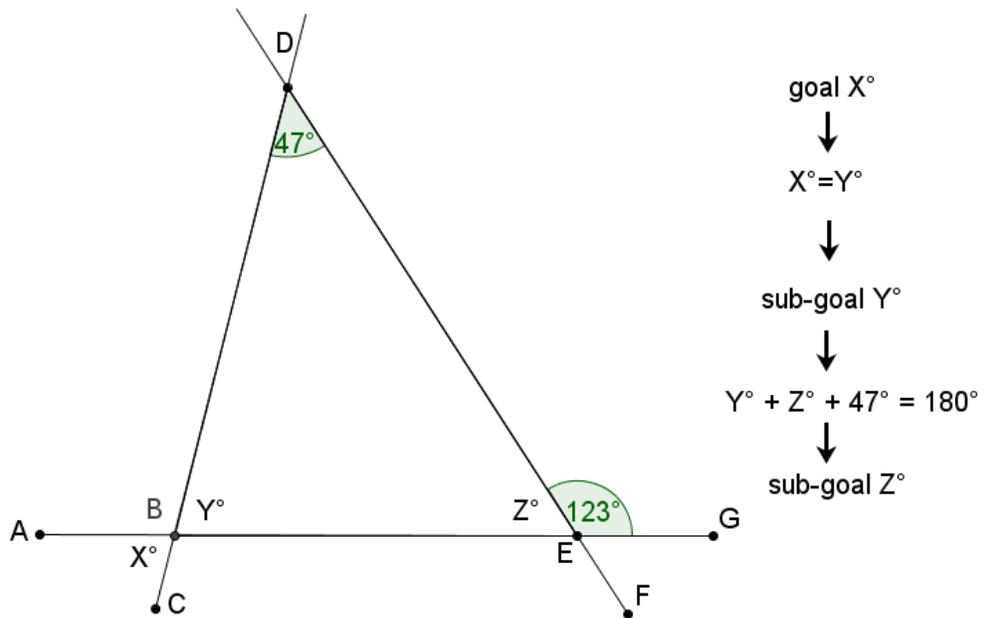
Step 5. Create sub-goal Z^o
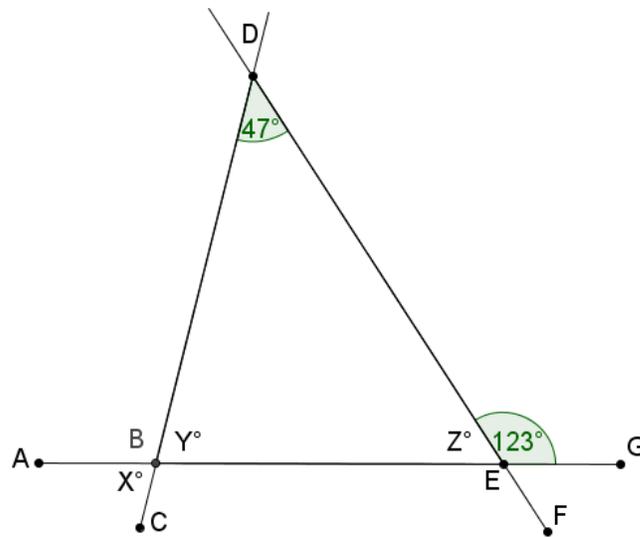
Step 6. Calculating angle Zᵒ
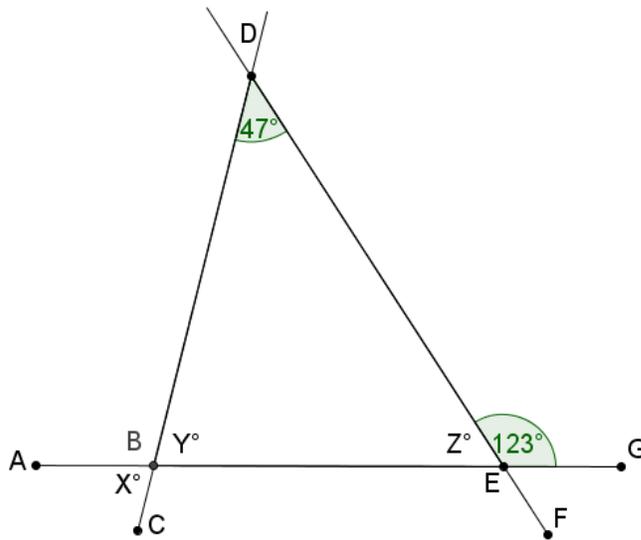
Step 7. Back to angle Yᵒ
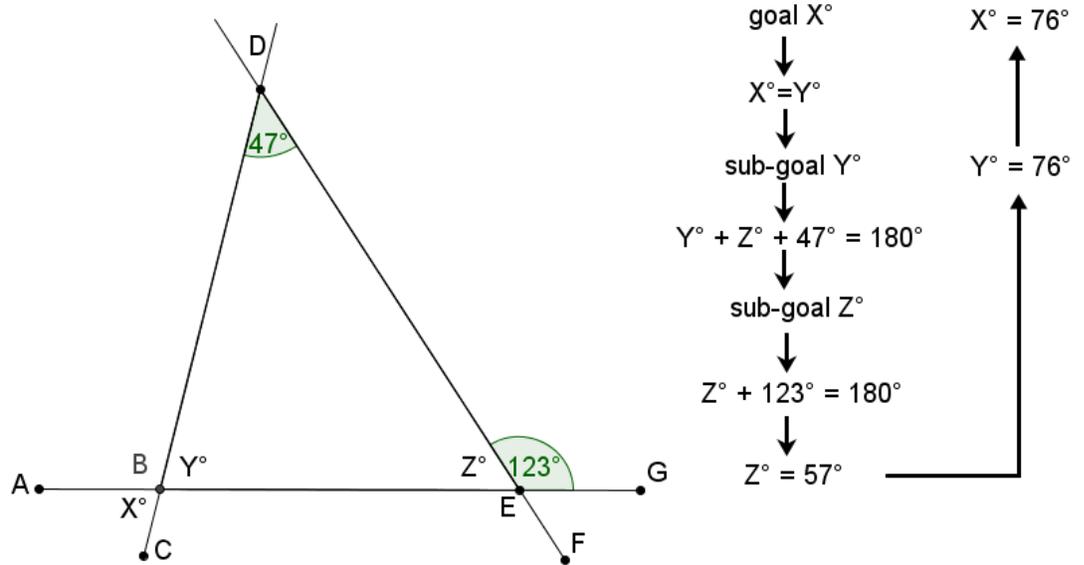
Step 8. Back to angle X°

Figure 9. Eight steps of means-ends analysis for solving a geometry problem

Summary

One of the aims of learning mathematics is to master problem solving
strategy; therefore it is crucial to understand the strategy of an expert problem solver
in a specific domain of mathematics. Cognitive psychologists have attempted to study
how learners can be an expert problem solver. Experts and novices in a specific
domain are distinguished by differences in their schematic knowledge. Experts can
categorise and formulate a solution to a problem based on its deep structure, and
solve problems in a forward-working strategy. In contrast, novices can only
categorise a problem based on its surface structure and most likely try to solve the

problem in a backwards-working strategy using means-ends analysis. As a

consequence, if a learner aims to construct knowledge of how to solve problems like

expert mathematicians, they should apply the strategies of problem solving used by

the mathematics experts. In order to acquire strategies of problem solving used by the

mathematics experts, the learner must possess schematic knowledge. The following

chapters describe cognitive load theory and the generated effects which may explain

how learners develop schematic knowledge effectively.

Chapter IV   Cognitive Load Theory

Human cognitive architecture, discussed in Chapter II, informs us of the central components of human cognition, their nature when processing information and how expertise in a specific domain can be built. It also explains how knowledge is acquired and how problem solving skills are developed. Cognitive load theory argues that an understanding of human cognition provides a useful framework for designing effective learning environments (Sweller et al., 2011).

Generally cognitive load theory is concerned with the limitations of working memory when learning novel information, and that the central role of learning is to facilitate the knowledge acquisition and automation of knowledge held in long term memory. Cognitive load theory emphasises that learning is reduced when the presentation of the to-be-learned material causes a cognitive overload.

In more recent developments, Sweller has compared cognitive architecture to the theory of evolution by natural selection (Sweller, 2003, 2004). He asserts that the information processing system is not unique to human cognitive architecture. Biological evolution provides an example of a natural information processing system that has an identical basic framework to the information processing system in human cognitive architecture. Influential in this conceptualisation has been the work of Geary.

Geary separates knowledge into two categories (Geary, 1995, 2002, 2007, 2008). The first is biologically primary knowledge, which consists of skills that humans have specifically evolved to acquire, and the second is biologically secondary knowledge, which consists of specific cultural skills that humans need to acquire. This categorisation suggests that instructional procedures should also be constructed depending on which knowledge is to be acquired. According to these recent theoretical developments, Sweller (2003, 2004) reconceptualised cognitive load theory into five basic principles which reflect the core characteristics of both human cognition and biological evolution.

This chapter is divided into two sections. The first section discusses the framework of cognitive load theory from an evolutionary education perspective. It encompasses the distinction between biologically primary and secondary knowledge and Sweller's five principles of cognitive load theory. The second section examines the sources of cognitive load, and describes a number of strategies that reduce cognitive load.

Human Cognitive Architecture in Evolutionary Perspective

Biologically Primary and Secondary Knowledge

Geary argues that different cognitive processes occur when dealing with biologically primary knowledge and biologically secondary knowledge as a

consequence of the evolution of human cognitive architecture (Geary, 1995, 2002, 2007, 2008). He assumes that cognition has evolved as a function of adaptation, reproduction and survival, and is influenced by biological and cultural demands.

Geary has proposed a detailed distinction between biologically primary knowledge and biologically secondary knowledge, which can be clearly contrasted by how they are differently acquired. It is argued that the human brain evolved over many generations to acquire biologically primary knowledge; that is, information required to survive in life, such as finding a path from one place to another, speaking to others, understanding facial expression, negotiating, making decisions, and listening to voices. This knowledge has grown as a tool to survive in everyday life. However, human cognition has also evolved to assimilate novel knowledge considered to be biologically secondary knowledge. This knowledge is culturally built and consists of the information needed for success in modern society, such as driving a car, writing a book, speaking a second language, baking a cake, playing a game, or solving mathematical problems. This knowledge, which is very recent compared to primary knowledge, has only developed to fulfil the cultural needs formed by society.

Biologically primary knowledge may be considered general knowledge because it is applicable across domains. However, biologically secondary knowledge is domain specific. For example, the knowledge of basic algebraic equations has been acquired to solve various problems in the domain of algebra, but cannot be

guaranteed to solve other problems in other domains. However, a heuristic problem solving strategy such as means-ends analysis (discussed briefly in Chapter III) is considered as biologically primary knowledge. It is a primary skill that can be commonly applied to many problem solving situations, regardless of the domain, when the required secondary skill is not available (see Youssef, Ayres, & Sweller, 2012), even though it may not successfully solve the problem.

In terms of cognitive effort, biologically primary knowledge is acquired easily, rapidly, automatically or unconsciously by immersion into a functioning society (e.g., family, community, social group). In contrast, biologically secondary knowledge requires conscious cognitive effort, and is usually acquired through formal educational and professional organisations.

Learning to listen to and read our native language provides an example of primary and secondary skills respectively. Our skill to listening is most likely acquired unconsciously despite being comprised of various sound recognition skills. We may learn the sound of soft, charming, strong, intimidating or fearful voices through our daily life; and we have accumulated them rapidly, yet effortlessly, since we are able to hear. Interestingly, this skill grows without explicit instruction because we automatically learn to listen as our life is surrounded by sounds and voices. There are biological and cultural demands influencing our ability to understand voices. On the other hand, we need to deliberately acquire skills to read the alphabet simply because it is a fairly recent addition to society, although very important, as Geary

(1995) notes. Without direct instruction for acquiring knowledge of reading, failure will most certainly occur. In addition, the learning process is conscious and demands cognitive effort.

Sweller et al. (2011) argue that both knowledge categories are learnable because human cognition evolves to construct knowledge. However, biologically primary knowledge cannot be explicitly taught because human cognition has evolved to acquire this knowledge automatically. On the contrary, biologically secondary knowledge is teachable, and should be taught using direct, explicit instruction because this knowledge is not acquired naturally (Geary, 1995). Lastly, it is important to note that biologically primary and secondary knowledge requires different contexts for their acquisition. Primary skills can be learned in the natural environment, but secondary skills need to be learned in well-organised environments, such as schools.

Because of the critical differences between primary and secondary knowledge, cognitive load theory is mostly concerned with instructional development for the acquisition of biologically secondary knowledge rather than biologically primary knowledge (Sweller et al., 2011).

The following section discusses five principles linking human cognitive architecture and biological evolution. Using the analogy of human cognition as biological evolution has provided cognitive load theory a framework for analysing the efficiency of instructional designs as part of natural occurrence. (Sweller, 2003, 2004; Sweller et al., 2011; Sweller & Sweller, 2006). These five principles are: (1) the

information store principle, (2) the borrowing and recognising principle, (3) the randomness as genesis principle, (4) the narrow limits of change principle, and (5) the environmental organising and linking principle.

Basic Principles of the Information Processing System

An information processing system requires at least four defining characteristics to be successful (Sweller et al., 2011). Firstly, it should be creative, in that it is able to generate novel information to overcome the complexity of information in the environment. Secondly, the effectiveness of created novel information needs to be tested and effective information retained and subsequently used. Thirdly, stored information can be used to direct the activity of the system. Lastly, effective information can be transferred across space and time.

Both human cognition and biological evolution are examples of sophisticated natural information processing systems and can be characterised as successful information processors (Sweller et al., 2011). Both consist of a set of natural entities that function to organise and process information. The following table (Table 1) describes aspects of human cognition that are comparable to aspects of biological evolution and their function.

Table 1. *Natural information processing system principles (Adapted from Sweller&Sweller, 2006, p. 436)*

	Principles	**Cognitive case**	**Evolutionary case**	**Function**
1	***Information store principle***	Long term memory	Genome	Store information for indefinite periods
2	***Borrowing and reorganizing principle***	Transfer information to long term memory	Transfer information to a genome	Permit the rapid building of an information store
3	***Randomness as genesis principle***	Create novel ideas	Create novel genetic codes	Create novel information
4	***Narrow limit of change principle***	Working memory	Epigenetic system handling environmental information	Input environmental information to the information store
5	***Environmental organizing and linking principle***	Long term working memory	Epigenetic system handling genetic information	Use information stored in the information store

First principle: The information store principle

This principle states that all natural information processing systems must have a central store of information to accommodate the huge and complex variations of information available in the natural environment in which the system functions (Sweller & Sweller, 2006). Long term memory and the human genome provide an information store in human cognition and biological evolution systems respectively. Both have an unlimited capacity.

In the biological system, Sweller and Sweller (2006) maintained that a genome stores a large amount of genetic-based information and governs biological activity. Additionally, the amount of information in a genome must be large in order to survive since it functions in an environment with a very wide range of information.

In human cognitive architecture, long term memory stores a huge amount of information and similar to a genome, the organised information in long term memory controls the activity of human cognition. As discussed in Chapter 2, the pioneering studies of de Groot (1978), and Chase and Simon (1973), demonstrated extremely the large capacity of long-term memory.

Second principle: The borrowing and reorganising principle

This principle affirms the manner in which information is obtained and amassed into an information store (Sweller & Sweller, 2006). According to Sweller, biological information is acquired by a genome via asexual or sexual reproduction. During asexual reproduction, information in a genome is copied and repeatedly passed to its offspring. During this reproduction, genetic information in the parent cell is exactly copied into the new cells. During sexual reproduction, an equal amount of genetic information is borrowed from two sexually different genomes, and then reorganised in such a way that results in a new unique genome.

Information acquisition in human cognition is considered identical to the reproduction mechanisms in the genome (Sweller & Sweller, 2006). Similar to

asexual reproduction, almost all of the secondary knowledge stored in long term memory is a result of borrowing the secondary knowledge from the long term memory of others. Repeatedly this acquired knowledge can be transmitted to others. Sweller argues that this mechanism is supported by primary knowledge, such as our skills of communicating with other people by listening to explanations, reading printed information or imitation.

Moreover, Sweller suggests that human cognition rarely imitates others exactly, because the transmitted information will most likely be reconstructed to fit the information store. The schema theory, discussed in Chapter 2, provides evidence that borrowed information is reorganised by the Piagetian processes of assimilation, accommodation and equilibration in order to construct a better representation of knowledge in long term memory.

Third principle: The randomness as genesis principle

The borrowing and reorganising principle above demonstrates a method of acquiring and reconstructing information from others. The information is new to the borrower (learner) but old to the lender (teacher, peer and so forth), but does not explain how totally novel information is initially created and retained in the information store. The randomness as genesis principle describes how new information is initially created through a random generation and test of effectiveness procedure.

Sweller and Sweller (2006) point out that mutation is a mechanism used by natural systems to generate new variations of genetic material in genomes to increase their chance of survival. Further, Sweller et al. (2011) argue that although random mutation is the source of all variation and novelty in biology, without tests of effectiveness, mutations are worthless. Only those that are effective for survival and reproduction are retained in the genome. Drawing parallels with biological evolution, it can be proposed that human cognition initially creates new information using random generation and effectiveness test procedures (Sweller & Sweller, 2006).

Evidence for the random generation and test strategy is found in the use of general problem solving strategies. As indicated in Chapter 2, when humans are presented novel tasks to solve, in the absence of relevant prior knowledge they rely on a general problem solving strategy, such as means-ends analysis. New moves can be created using such strategies, although most cause dead-ends because they are applied without a sufficient knowledge base. However, if a successful move is acquired, often after many generations and testing cycles, it will be stored in long term memory as new knowledge, whereas failed moves are likely to be rejected and forgotten. Sweller et al. (2011) argue that the randomness as genesis principle underlies all human creativity.

Fourth principle: The narrow limits of change principle

In a novel environment, the randomness as genesis principle through a random generation and test procedure allows the information system to generate all

possible combinations of information, and randomly and repeatedly, selects a combination to test its effectiveness until a successful new solution is found and new information acquired. The borrowing and reorganising principle also creates new information but as a result of the alteration of previously stored information (Sweller et al., 2011). Sweller and Sweller (2006) describe how the epigenetic system can facilitate or inhibit the occurrence of mutations in some parts of the genome determined by the condition of the environment. Consequently, large numbers of mutations do not occur simultaneously.

In human cognition, Sweller and Sweller (2006) argue that working memory can be considered analogous to the epigenetic system in biological evolution. As highlighted in Chapter 2, working memory's central function is to process information, but it has a severely limited capacity and duration for processing simultaneous information (Cowan, 2000; Miller, 1956). Due to the random generation and test procedure, potentially many combinations of information elements might need to be considered. To keep such combinations manageable, the natural information processing system relies on a limited capacity working memory. The narrow limits of change principle provides an explanation of why evolution has led to a restricted working memory capacity. Without such limitations, it would be impossible to handle all the possible combinations generated by the randomness as genesis principle.

Fifth principle: The environment organising and linking principle

According to Sweller et al. (2011) the previous four principles explain how a natural information processing system acquires information. The last principle explains how stored information is translated into activities and used in a specific environment.

In biological evolution, the epigenetic system transfers genetic material from the genome in order to respond to changes in environment, as well as to guide the functioning of the organism in response to environmental input (Sweller & Sweller, 2006).

Analogous to the epigenetic case, the environment organising and linking principle permits working memory to obtain unlimited amounts of organised information from long term memory. It can be recalled from Chapter 2 that experts have a superior cognitive system because of their ability to retrieve lots of information from long term memory. In contrast to the limited processing of novel information, due to the generation and test strategy, working memory can process vast quantities of information if the organised knowledge is stored in long-term memory in the form of schemas. The demands of the environment provide the cue for the relevant information to be transferred to working memory. Prior to this activation, schematic knowledge lies dormant in long term memory.

It is this interaction between long term memory and working memory which underpins cognitive load theory. Working memory creates the vital link between the environment and long-term memory.

Cognitive Load

It has been discussed that working memory can only manage a limited amount of novel information and as a consequence of this limit, knowledge acquisition is very much affected by the demands placed on working memory. Therefore, cognitive load theory is particularly focused on the level of cognitive load or mental activity imposed on working memory when dealing with novel information (Pass, Renkl, & Sweller, 2004; Sweller et al., 2011; Sweller et al., 1998).

Cognitive load can be defined as the amount of information that working memory processes at any one time (Sweller, 1988). In order to explain instructional effectiveness, cognitive load theory has defined two categories of cognitive load according to their function: intrinsic and extraneous cognitive load (Sweller, 1994, 2010). Both categories are determined by the level of element interactivity associated with the learning materials. Whereas intrinsic cognitive load is imposed by the element interactivity generated by the intrinsic structure of the learning material, the extraneous cognitive load is caused by the element interactivity generated by the presentation of the learning material. These two categories of cognitive load are

considered additive and the total determines the working memory resources required to process the information (Sweller et al., 2011; Sweller et al., 1998).

Element interactivity

Learning material consists of elements or chunks of elements of information that need to be processed in order for learning to occur. Elements may be considered single items of information, or simple information structures. Logical connections between elements determine the level of element interactivity of the learning material (Sweller & Chandler, 1994). In other words, elements are connected in such a way that they make a meaningful construction. Element interactivity refers to the extent to which elements interact with each other.

If elements in learning material do not interact then these elements can be learned separately. Such material is considered as having zero or very low element interactivity. On the other hand, some materials are considered high in element interactivity because many elements interact and can only make logical meaning if processed together. It can be recalled that prior knowledge, stored in long term memory, can be used to chunk information together. As a consequence, prior knowledge will also determine how many interacting elements can be chunked together in working memory. Accordingly, some materials that are high in element

interactivity for beginner learners can be low in element interactivity for more advance learners.

An example of low element interactivity can be illustrated when learning about various geometrical shapes. A student can learn about a square without necessarily learning about a triangle at the same time, because in this context, considerations of interactivity between the shapes are unnecessarily to understand the basic concepts connected to a square. Therefore, the properties of each shape (square and triangle) can be learned in isolation from each other. However, learning to calculate the volume of a prism, which may require knowledge about triangles and squares, can be considered high in element interactivity because there are many elements that need to be processed simultaneously, such as the property of the prism, the base area, the height, the volume formula as well as the unit used in the calculation. Without simultaneously attending to these elements, the task cannot be understood or completed correctly. However, attending to all these elements simultaneously will increase the demands on working memory. But as knowledge about prisms builds, interacting elements will become chunked into larger single elements and decrease the burden on working memory. Element interactivity can be used to describe understanding and task difficulty discussed next.

Understanding. Sweller (1994) proposed that the term *understanding* is more suitably applied to learning about materials with high element interactivity. As high element interactivity materials involve interconnected elements, total understanding

will only be reached if the elements and the connections are considered at the same time. An understanding of high element interactivity material is achieved by making meaning of the connections between elements. As a consequence failure to connect some elements together will most likely cause a misunderstanding of the learning materials. It is argued that low element interactivity materials can be learned serially, because they do not necessarily relate to each other (Sweller et al., 1998). Any lack of knowledge about an element, or a failure to remember an element will not directly cause a misunderstanding of the other elements.

Task difficulty. Additionally, Sweller and Chandler (1994) argued that element interactivity can be used to determine why some material can be difficult to learn. Tasks can be considered complex if they are high in element interactivity (see Sweller et al., 2011; Sweller & Chandler, 1994), and hence difficult to understand because of the number of interacting elements that need to be considered. Nevertheless, learning material consisting of low element interactivity can also be difficult to learn. For example, learning new vocabulary may be difficult to learn simply because of the total number of elements (new words) involved, even though the elements are independent of each other and may not interact.

Intrinsic cognitive load

Intrinsic cognitive load refers to the intrinsic nature of the learning materials themselves (Sweller, 1994; Sweller & Chandler, 1994; Sweller et al., 1998). Material that is low in element interactivity imposes a low intrinsic cognitive load. Conversely, material that is high in element interactivity imposes a high intrinsic cognitive load.

Intrinsic cognitive load is usually considered fixed, although it can be reduced by altering the complexity of the learning task, such as by reducing the number of interacting elements (Ayres, 2006; Pollock, Chandler, & Sweller, 2002). Without changing the complexity of the task, the intrinsic nature of the material or the element interactivity level, remains unchanged (Sweller & Chandler, 1994). However, as learners obtain prior knowledge about specific materials, schemas enable interacting elements to be chunked together as more advanced single elements, thus reducing element interactivity (Mayer & Moreno, 2003). Hence, for more advanced learners, specific materials may be considered quite simple; however, for less advanced learners the same materials may be considered complex and difficult to understand.

Extraneous cognitive load

The manner in which learned material is presented is the primary factor determining extraneous cognitive load (Sweller, 1994; Sweller & Chandler, 1994; Sweller et al., 1998). In contrast to intrinsic cognitive load, which is created by the

actual materials to-be-learned, extraneous cognitive load is created by the teacher or instructional designer. For example, if students are asked to learn about a science topic but are given poorly formatted diagrams and instructions, more processing may be required to understand the diagrams and explanations rather than focusing on the actual learning content. In this case, precious working memory resources are taken up by processing information that is irrelevant, or extraneous to learning. In such cases, the extraneous processing can directly interfere with learning.

Sweller (2010) has argued that extraneous cognitive load can also be described in terms of element interactivity. Whereas element interactivity can indicate the intrinsic characteristics of the learning materials, element interactivity can also indicate the connectivity of elements presented by the instructional materials. If the instructional materials are low in elementary interactivity then learning is more likely to occur because fewer working memory resources are needed. On the other hand, if the instructional materials are high in element interactivity then learning will be more likely interfered with, as more working memory resources will be needed. Just as intrinsic load can have different levels of complexity, so can extraneous load.

The majority of research on cognitive load theory has been to investigate strategies to decrease extraneous cognitive load (Mayer & Moreno, 2003; Sweller et al., 2011; Sweller et al., 1998). Effective instructional designers will try to lower extraneous cognitive load by modifying the presentation of learning materials accordingly. Cognitive load theory is particularly concerned with decreasing

extraneous cognitive load when the learning materials impose high intrinsic cognitive load, in order to promote learning.

Germane cognitive load

In addition to intrinsic and extraneous cognitive load, the working memory load actually invested in schema acquisition (learning) has been defined as the germane cognitive load. Originally conceptualised by Sweller et al. (1998) as an independent load, more recently germane load has been linked directly to intrinsic cognitive load. The working memory resources used to deal directly with the intrinsic cognitive load are now considered germane load as the cognitive process are directed towards a learning goal (Sweller, 2010; Sweller et al., 2011). Using working memory resources to deal with the extraneous load is not germane because schema acquisition is not directly facilitated.

Germane cognitive load involves activities that are relevant to learning, such as eliciting self explanation (Paas & Van Gog, 2006). Paas & Van Gog (2006) suggested that requiring learners to generate explanations underlying the solution steps can stimulate them to invest working memory load for activities relevant to learning.

It is worth noting that during the earlier development of cognitive load theory, the total cognitive load was calculated by adding the three loads together (Sweller et

al., 1998). This belief has been recently reformed, and now the total load is defined as the amount of cognitive load generated by the intrinsic and extraneous loads added together (Kalyuga, 2011; Sweller, 2010; Sweller et al., 2011). If the total cognitive load required is below the working memory limit, then the freed resources can be allocated to the germane cognitive load to help schema construction (Paas & Van Gog, 2006; Paas & van Merriënboer, 1994; Pass et al., 2004).

Summary

This chapter discussed the theoretical framework of cognitive load theory from the perspective of evolutionary educational psychology. It showed how human cognitive architecture underpins cognitive load theory and uses an analogy with natural information processing systems. Geary's knowledge categorisation into biologically primary and secondary knowledge was also discussed. This distinction is used to show which knowledge humans have evolved to acquire (primary), and which knowledge requires well-structured learning environments (secondary).

Subsequently, five principles underlying cognitive load theory were described. Generally, these principles demonstrate: (1) why the cognitive processes require an unlimited long term memory; (2) how human cognition constructs schematic knowledge by borrowing from others; (3) how human cognition interacts with unfamiliar information and as a consequence constructs new knowledge; (4)

why working memory has a limited capacity; and (5) how human cognition interacts with the environment, and the importance of the interactions between working memory and long term memory.

Two categories of cognitive load, intrinsic and extraneous, were also detailed using the concept of element interactivity. Germane cognitive load was also defined as the load directly invested in schema acquisition, and specifically linked to intrinsic cognitive load. Intrinsic and extraneous cognitive load are additive and form the total cognitive load. For the most effective learning to occur, the total cognitive load must not exceed the working memory capacity of the learner. While the intrinsic cognitive load is unchangeable due to its innate nature, unless the task is altered in some way, the extraneous cognitive load, which is generated by the instructional designer, must be kept at a low level in order to create the most effective learning environment.

Chapter V   Cognitive Load Effects

Cognitive load theory was developed in the early 1980s and the initial

research was focused on the search for alternatives to conventional problem solving

strategies. Considerable evidence was collected which indicated that asking students

to acquire knowledge through problem solving was ineffective due to a heavy

reliance on means-ends analysis. As previously described, means-ends analysis is a

general problem solving strategy that is used when there is lack of prior knowledge.

However, means-ends analysis creates extraneous cognitive load and thus inhibits

learning. This chapter reports on the cognitive load theory research that has

investigated the different types of extraneous load and the strategies used to reduce its

impact.

How Means-Ends Analysis Increases Extraneous Cognitive Load

Means-ends analysis has been discussed in previous chapter and in this

section is discussed in detail how it causes extraneous cognitive load during learning.

Initial evidence for the use of means-ends analysis was accumulated using

maze puzzle problems (Mawer & Sweller, 1982; Sweller, Mawer, & Howe, 1982).

Sweller and Levine (1982) found that means-ends analysis was used by problem

solvers who were given the final goal location of a maze. Although they could often

find the solution to the maze, they learnt little about the problem structure.

Additionally, they made more errors when solving the problem and were less able to

transfer their knowledge about the maze compared to problem solvers who could not

observe the maze's goal location. Sweller and Levine argued that providing

information about the finish location of the maze caused problem solvers to pay more

attention to the goal rather than learn about the problem structure. Therefore, they

concluded that means-ends analysis did not necessarily facilitate learning. Much

stronger learning was achieved by removing the goal from the sight of the problem

solvers.

Means-ends analysis leads problem solvers to try to reduce the distance

between the problem state and problem goal by generating a system of sub-goals that

have to be considered to find the final solution (Ayres & Sweller, 1990; Sweller,

1988). Sweller (1988) argued that during means-ends analysis, problem solvers have

to pay more attention to a sub-goal by working backwards from the goal, rather than

applying previously learnt knowledge about solution paths. This is contradictory to

schema acquisition, where more attention needs to be directed towards problem states

and associated moves. Further, during means-ends analysis, problem solvers must

simultaneously consider the problem state, the goal state, the relation between these,

and the relation between problem solving operators, while also considering a

sequence of sub-goals. Simultaneously handling a large number of elements requires a heavy use of working memory capacity, and thus learning is hindered.

In sum, presenting novel problems, such as the maze problem previously noted, causes the use of means-ends analysis, which requires a heavy use of cognitive processing. By asking learners to solve problems without direct instruction, instructional designers or teachers are creating extraneous cognitive load. To prevent the use of means-ends analysis and the generation of extraneous cognitive load, Sweller and colleagues devised and tested two alternative learning strategies: (1) goal-free problems, and (2) worked examples. The following section discusses these two strategies.

The Goal-Free Effect

Goal-free problems are also known as no-goal problems. Considerable evidence has shown that a goal free strategy is a superior learning strategy to conventional problem solving, and its effect is called the goal-free or goal-specificity effect (Sweller et al., 2011). To use the goal-free strategy, acquisition problems are presented without a specific goal. By removing the goal, means ends analysis becomes impossible because problem solvers cannot work backwards from a goal, as there is no goal. Instead, learners are directed to solve the problem using a forward strategy based on the problem statement.

Sweller and Levine (1982) identified the goal-free effect in their maze-tracking research. As discussed previously, when problem solvers were presented with the goal, they tended to use means-ends analysis and did not learn about the problem structure. On the other hand, when problem solvers were not presented with the goal, they attempted to use a solution rule they had learned and hence developed knowledge about the problem structure.

In a series of four experiments, Sweller, Mawer, and Ward (1983) tested if goal free problems would eliminate the use of means ends analysis. They predicted that substituting the instruction to find a specific variable with goal free instructions, as well as removing the goal, would reduce the use of a means ends strategy. Using geometry problems, where students had to find angles using specific theorems, during acquisition a goal free group was given the instruction "Calculate the value of as many angles as possible" for problems where the goal had been removed. In contrast, a goal group had a clearly defined goal (angle X), and were given the specific instruction "Calculate the value of X" (Sweller, et al., 1983, p. 653). Overall the results found a goal free effect, as students who were provided with goal free problems during the acquisition phase, performed better on later goal-specific tests than students who were given conventional goals during acquisition (Sweller et al., 1983).
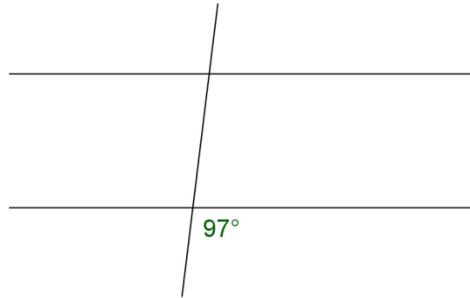
Find the measure of as many angles as possible

Figure 10. Example of goal-free problem
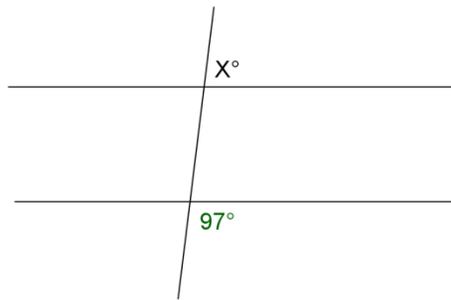
Find the measure of angle X°

Figure 11. Example of goal-given problem

Similarly, Owen and Sweller (1985) investigated goal free problems in trigonometry, where triangle sides needed to be calculated. Again during the acquisition phase the goal was removed for a goal-free group who were asked to "find the length of all unknown sides", while the goal group used conventional problem solving to find a fixed goal (a given side of the triangle). Owen and Sweller

found that the goal-free group made fewer errors than the goal group, and had better transfer performance.

In an attempt to capture error patterns when solving novel problems, Ayres and Sweller (1990) conducted a series of experiments using multi-step geometric problems for high school students (calculating angles). It was confirmed that most errors occurred during sub-goal calculations as a result of using means ends analysis. By removing the goal, and modifying the goal statement into "find all unknown angles", instead of "find x" (Ayres, 1993, p. 378), it was found that goal-free problems prevented the use of means ends analysis and led to superior learning.

Further evidence for the goal free effect has also been demonstrated by several studies. Bobis, Sweller, and Cooper (1994) used primary school children learning about geometrical paper folding; Vollmeyer, Burns, and Holyoak (1996) used university students studying biology; Paas, Camp, and Rikers (2001) used elderly people who were learning computerised mazes; and Wirth, Künsting, and Leutner (2009) used high school students examining computerised physics problems.

There are some possible limitations to the goal-free effect. The applicability of goal-free problems in classroom learning with time restrictions, where finding many irrelevant 'unknowns' (e.g., sides and angles can be found), may result in many unnecessary calculations that detract from focusing on the most important structures (Sweller et al., 2011). Additionally, when the aim is to learn about an application of a specific procedure, a defined goal may be more suitable (Wirth et al., 2009).

Therefore, although the use of goal-free problems does reduce the use of backward-working strategies, this can come at a cost.

The worked example strategy was developed to overcome these limitations. The following section discusses the worked example effect and the required format for worked examples to be successful.

The Worked Example Effect

In a conventional learning environment, a worked example is commonly used to demonstrate how to solve a type of problem, which is then followed by practice on a number of similar and/or transfer problems. Mathematics or other computational learning domains have used worked examples for this purpose. Moreno (2006) describes worked examples as instructional devices for learning a specific problem solving skill. Worked examples usually include the problem statement and step-by-step moves leading to the final solution (Ayres & Sweller, 2013). Further, as R. K. Atkinson et al. (2000) commented worked examples should show an expert's problem solving model for learners to study and imitate.

How worked examples reduce extraneous cognitive load

Cognitive load theorists have advocated worked examples as an effective strategy to reduce extraneous cognitive load and facilitate effective schema acquisition. Essentially worked examples eliminate the use of means-ends analysis (Sweller et al., 1998) by presenting a solution to study rather than asking students to find one. According to cognitive load theory, when learners who are given worked examples during acquisition perform better on subsequent tests than learners who are given the same problems to solve during acquisition, the *worked example effect* occurs (Sweller et al., 1998). Sweller et al. argued that when learners are provided with worked examples to study, they are directed to pay attention to the problem states and the various associated moves, rather than focusing on the goal. Knowledge acquired from worked examples also decreases the chance of using means-ends analysis when given similar problems; hence when studying worked examples, working memory load is devoted to schema acquisition and automation.

Furthermore, Sweller (2006) illustrated that learning from worked examples is an example of the borrowing principle, and learning by solving novel problems is an example of the randomness as genesis principle. A worked example can be deemed the representation of knowledge from an expert's long term memory. Accordingly, worked examples can be used to acquire new knowledge via the borrowing principle. When relevant information is either inaccessible or does not exist, learning through problem solving occurs via the randomness as genesis principle. As was discussed in

Chapter 2, acquiring knowledge via the borrowing principle is more likely to be effective because it imposes a lower cognitive load compared to the randomness as genesis principle.

R. K. Atkinson et al. (2000) reported that research into the use of worked examples has been carried out for more than six decades; however, in the 1980s, researchers paid more attention to the strategy as an alternative to the ineffective problem solving methods (e.g., means-ends analysis). Most of the original research used controlled experiments to test the prediction that learning novel (complex) problems from worked examples is more advantageous compared to learning by problem solving only (Cooper & Sweller, 1987; Sweller & Cooper, 1985; Zhu & Simon, 1987). The following section discusses the empirical evidence in support of the worked example effect.

Initial evidence for the worked example effect

The initial evidence that worked examples can facilitate knowledge acquisition was provided by Sweller and Cooper (1985). They predicted that worked examples can be used to direct learners' attention away from the goal of a problem to the relations between problem states and associated moves. In a series of four experiments using a high school algebra topic, learning with worked examples was compared to learning through problem solving. In the initial acquisition phase, the

worked example group used pairs of worked examples and similar conventional problems. For each pair students were required to study a problem and then solve a similar problem. The problem solving group was given the identical problem pairs to those in the worked example condition, but all problems were presented as conventional problems which students had to solve. In the first two experiments, the acquisition phase was followed immediately by a test consisting of problems similar to those in the acquisition phase. The results indicated that the worked example group spent substantially less time solving the problems and made fewer errors.

To test whether worked examples facilitated transfer to dissimilar problems, the last two experiments included transfer problems in the test phase. No significant difference was found for these transfer problems. It was suspected that the relatively limited number of problem types used in the acquisition phase might have hindered transfer. Sweller and Cooper (1985) concluded that while the use of worked examples facilitated schema acquisition, the strategy was only beneficial on a restricted range of problems.

In a follow up study, Cooper and Sweller (1987) argued that their original work (1985) did not show transfer effects because the acquisition phase was too short and the learning material might have extended the working memory load too much to promote schema acquisition and automation. Cooper and Sweller (1987) predicted that if learners have sufficient time during acquisition, schemas would become more automated, leading to greater transfer. To test this prediction through worked

examples, they conducted a series of experiments using less complex material rather than those that they used previously (1985) and provided more learning time.

Cooper and Sweller (1987) found that the worked example group required less time during the acquisition phase compared with the conventional problem solving group. While there was no significant difference between groups in similar test results, a significant difference was found for the transfer problems, with less completion time needed and less errors made by the worked example group. Since the problems were less complex and sufficient time was provided during the acquisition phase, both conditions facilitated schema acquisition for similar test problems. However, the schema automation required for transfer problems was only facilitated under the worked example condition.

Both studies by Sweller and Cooper described above have become influential because they provided the initial evidence that learning by worked examples can be more effective than conventional problem solving. It was argued that worked examples eliminate the use of means-ends analysis that imposes a heavy cognitive load. As worked examples considerably reduce this extraneous cognitive load more working memory resources are available for schema acquisition and automation.

More evidence for the worked example effect in mathematics was provided by Zhu and Simon (1987) in their longitudinal study using a 3-year curriculum in algebra and geometry in a Chinese middle school. Zhu and Simon found that students

studying the worked examples could complete the three-year course in only two years. Research by Tarmizi and Sweller (1988) using geometry also found the worked example effect, once split-attention was avoided (more detail on split-attention is given later). Further evidence of the worked example effect were provided by Chi, Bassok, Lewis, Reimann, and Glasser (1989) using college physics, and Ward and Sweller (1990) using geometric optics and kinematics.

The growth of research on the worked example effect

Since this early research, a massive amount of research examining the effectiveness of worked examples have been completed, not only in mathematics and its applications which have more-structured procedures (for review, see R. K. Atkinson et al., 2000; Ayres & Sweller, 2013; Sweller et al., 2011), but also in less-structured tasks, such as learning musical notation (Owens & Sweller, 2007), visual art recognition (Rourke & Sweller, 2009), text interpretation (Oksa, Kalyuga, & Chandler, 2010), legal case reasoning (Nievelstein, van Gog, van Dijck, & Boshuizen, 2013), and essay writing (Kyun, Kalyuga, & Sweller, 2013). Worked example instructions have also been implemented in multimedia (e.g., Moreno & Mayer, 1999; Mousavi, Low, & Sweller, 1995), hypermedia (e.g., Gerjets, Scheiter, & Schuh, 2008), and web-based learning environments (e.g., Crippen & Earl, 2007).

All these studies provided strong evidence of the worked example effect, indicating that the effect is not limited to a particular domain or learning setting.

Cognitive load researchers have investigated variations of the format of worked examples in order to accommodate different learning materials and the characteristics of the learner. In turn, several factors that moderate the effectiveness of worked examples have been identified. These factors are usually explained in terms of whether the modified instruction format reduces or increases either extraneous or intrinsic cognitive loads. Cognitive load theory has derived instructional principles based on these findings, as the following discussion indicates.

Variations to Worked Examples

Most research on worked examples has demonstrated the effectiveness of learning from worked examples rather than learning by solving conventional problems. The worked example instruction usually combines worked examples and a similar problem solving task. However, researchers have implemented the instruction in different ways. Therefore, further cognitive load effects have been generated based on the variations of the instruction.

The alternation strategy: Study one - Solve one

An important consideration for teachers and instructional designers is how to structure worked examples. When testing the worked example effect for the first time, Sweller and Cooper (1985) presented pairs of worked examples and structurally identical problems to be solved. Sweller and Cooper "assumed that motivation, while reading a worked example, would be increased by the knowledge that a similar problem would need to be solve immediately afterwards" (1985, p. 69). Many researchers who have investigated the worked example effect adopted this alternation format in their study (Sweller et al., 2011). An example of alternating worked examples with similar problem can be seen in Figure 7.

Study this example.

Find the length of the hypotenuse, x.

7 cm   24 cm x
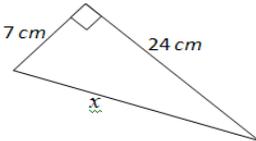
Solution:

$$x = \sqrt{7^2 + 24^2}$$

$$x = \sqrt{49 + 576}$$

$$x = \sqrt{625}$$

$$x = 25 \ cm$$

Solve this problem.

Find the length of the hypotenuse, x.

8 cm   15 cm x
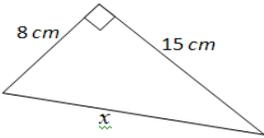
Solution:

Figure 7. An example of alternating worked examples with a similar problem to-be-solved.

Trafton and Reiser (1993) provided direct evidence of the effectiveness of the alternation format compared to a blocked format. As can be seen in Figure 8, *two alternating format* conditions were investigated; students were either given pairs of an example to study and a similar problem to solve, or pairs of two similar problems to solve. And there were *two blocked formats*; students were given a set of examples

to study, and then a set of similar problems to solve; or a set of problems to solve, followed by a set of similar problems to solve.

Alternating Example	Alternating Problem Solving	Blocked Example	Blocked Problem Solving
• Study example 1a • Solve similar problem 1b	• Solve problem 1a • Solve similar problem 1b	• Study example 1a • Study example 2a • and so on	• Solve problem 1a • Solve problem 2a • and so on
• Study example 2a • Solve similar problem 2b	• Solve problem 2a • Solve similar problem 2b	• Solve similar problem 1b • Solve similar problem 2b	• Solve similar problem 1b • Solve similar problem 2b
• and so on	• and so on	• and so on	• and so on

Figure 8. Alternating or blocked formats, adapted from Trafton and Reiser (1993)

Trafton and Reiser (1993) found that the most efficient strategy was the alternating example format, where a similar problem to solve was given immediately after each worked example was presented to study. The alternating problem solving or the blocked problem solving formats were not efficient strategies, and the blocked example format was found to be the least efficient overall.

More recently, van Gog, Kester, and Paas (2011) compared the alternation strategy, example–problem pairs, to the other three strategies: problem–example pairs, examples only and problems only. As found by Trafton and Reiser (1993), example–problem pairs were an effective strategy compared to the problem–example pairs and problems only. Additionally, they found that the effectiveness of the example–problem pairs was not significantly different from giving the examples

only. And also, the effectiveness of the problem–example pairs was not significantly different than the problems only. Furthermore, Reisslein, Atkinson, Seeling, and Reisslein (2006) observed that low prior knowledge students benefited from example–problem pairs; but, on the other hand, high prior knowledge students benefited from problem–example pairs (see the expertise reversal effect in following section). Consequently, it can be concluded that the original alternation format of Sweller and Cooper (1985) was an effective method for structuring worked examples, especially for learning novel materials.

The problem completion effect

When studying worked examples, teachers/ instructors have to insure that the learner is attending to the task. Although most worked examples provide full solution steps, Chi et al. (1989) found that most low prior knowledge students did not try to fully read and study all the solution steps provided in the examples. To overcome this potential problem, van Merriënboer (1990) suggested the use of completion problems that require learners to complete some key solution steps in the worked examples by themselves. Using an introductory computer programming task, van Merriënboer showed that this strategy can be just as effective as studying worked examples with complete solution steps, particularly when worked examples have many solution steps.

Using the learning domain of statistics, Paas (1992) investigated the effect of completion problems by comparing three groups: partially completed worked examples, worked examples and conventional problems. Results indicated that the completion and worked example conditions resulted in significantly higher performance than conventional problem solving condition. In addition, Paas found that the completion and worked example conditions required a significantly shorter study time than conventional condition. Paas also used a subjective scale to measure perceived mental effort during the tests, demonstrating that that the conventional group invested more mental effort than the other two groups. In other words, students in the problem solving group were learning very inefficiently as they had a lower level of test performance, but invested more mental effort (experienced higher cognitive load).

It has been argued that by partially completing the example, learners are guided to pay more attention to the problem state and the provided key solution steps while filling the incomplete solution steps (Sweller et al., 2011). Sweller et al. contend that completion problems might be considered a combination of worked examples and problem solving, and can be used as an alternative format to standard worked examples. In further research, completion problems have been used to generate the fading guidance effect (see following Section), as a consequence of the expertise reversal effect, which is discussed next.

The expertise reversal effect

The expertise reversal effect occurs when an instructional strategy that is effective for low prior knowledge is inefficient for high prior knowledge learners (see Kalyuga, Ayres, Chandler, & Sweller, 2003). It is argued that the prior knowledge of more experience learners will interact with the presented instructional materials, replicating the same information and leading to redundancy. Redundancy creates extraneous cognitive load for the learners and ultimately interferes with learning (the redundancy effect is discussed further in following Section).

Early evidence for the expertise reversal effect was found by Yeung, Jin, and Sweller (1997). Using English reading passages and the explanatory notes, Yeung et al. (1997) examined the affect of split attention and integrated formats on students with low and high expertise in the domain (see the split attention effect Section). In the comprehension test, they found that high prior knowledge students benefited from the split attention format but not from the integrated format, as it was redundant. In contrast, low prior knowledge students benefited from the integrated format, but not from the split attention format, which created extraneous load. Similar findings were found by Kalyuga, Chandler, and Sweller (1998) using electrical circuit problems where instructions based on split attention were compared to an integrated format. Again the integrated format, which was helpful in overcoming split-attention for low knowledge learners, was found to be ineffective for high knowledge learners.

More evidence of the expertise reversal effect was provided by Tuovinen and Sweller (1999). In this study the effectiveness of worked examples was tested against an exploration practice (conventional problem solving). Using database-programming tasks for college students, Tuovinen and Sweller found that only students with no prior experience with database programs benefited from worked examples. For students with some experience in the domain, the effectiveness of worked examples was negligible.

Kalyuga, Chandler, and Sweller (2001) investigated the interactions between worked examples, expertise, and problem complexity. When learning about more complex tasks, novice learners initially benefited from worked examples compared to a conventional problem strategy (exploratory learning). But after two training periods with worked examples, learner expertise increased and the worked example effect disappeared. In fact, the exploratory learning strategy became more effective than the worked example strategy.

As mentioned previously, Reisslein et al. (2006) found that example–problem pairs were more effective for low prior knowledge students since the examples assisted them with initial knowledge acquisition. Additionally, the high prior knowledge students may have received an advantage from the problem–example pairs because they already had sufficient prior knowledge to solve the initial problem, and the subsequent example may have provided useful feedback.

The expertise reversal effect was also demonstrated by Pollock et al. (2002) in a study designed to lower intrinsic cognitive load using worked examples. Pollock et al. (2002) reduced the element interactivity of complex tasks by isolating their element before introducing the tasks with fully *interacting elements.* This 2-stage strategy benefited less knowledgeable learners, but not more knowledgeable learners. It was argued that high knowledge learners were already able to maintain and process tasks that consisted of fully interacting elements, but not low knowledge learners. Further, Ayres (2006) also showed that low prior knowledge students benefited from an isolated-element (or partial) approach, whereas high prior knowledge students benefited only from fully interacting-element tasks (see also Ayres, 2012).

Much evidence in support of the expertise reversal effect has been accumulated by cognitive load theorists (see Ayres & Paas, 2007; Sweller et al., 2011). Overall the findings highlight the importance of considering levels of prior knowledge when designing instructions. In the case of worked example instructions, if the information presented in a worked example has already been acquired by learners, the worked example will be redundant and will lead to the expertise reversal effect. In such situations, learners would be able to learn through problem solving. Considerations of the expertise reversal effect were used to develop the guidance fading strategy.

The guidance fading strategy

The guidance fading strategy uses a combination of worked examples, completion problems, and problem solving which are presented sequentially, and designed to facilitate a smooth transition from novice to more experienced learners (R. K. Atkinson, Renkl, & Merrill, 2003; Renkl & Atkinson, 2003; Renkl, Atkinson, Maier, & Staley, 2002). Underlying this strategy is the expertise reversal effect, because, as expertise develops, less direct guidance from worked examples is required.

Renkl et al. (2002) suggested two fading techniques, backwards and forwards. In a series of *backward fading* techniques, the first worked example is fully completed, the second worked example has the solution to the final step removed, the third has the two last steps removed, and so forth, until the final example presents the whole problem to-be-solved only. Learners are expected to fill in the steps, whose number increases as expertise develops. For the *forward fading* technique, the series occurs in the opposite direction. The first step of the worked out solution is incomplete, then the second step is removed, and so forth, in a forward direction until the full incomplete problem is presented.

According to Renkl et al. (2002) the backward fading technique is more favourable for low prior knowledge learners since it provides a full worked example at the beginning of the learning phase, which is critical in assisting initial knowledge

acquisition. Renkl et al. (2002) found the guidance fading strategy to be an effective technique to facilitate the transition from novice to expert, compared to a series of fully worked examples.

In addition to Renkl et al.'s findings, Reisslein, Sullivan, and Reisslein (2007) reported that slow fading strategy was more advantageous for low prior knowledge students transitioning from worked example stage to independent problem solving. Slow fading strategy uses a backward fading technique that provides students with a longer phase of knowledge acquisition. In contrast,a fast fading strategy,was found to be more advantageous for high prior knowledge students.

Extraneous Cognitive Load Caused by Worked Example Designs

Cognitive load theory has been particularly concerned that the instructional design of worked example is aligned with the learner's cognitive capacity (Sweller et al., 2011). Two general cognitive effects have been identified as a source of extraneous cognitive load which impact on worked example designs: the split attention and redundancy effects.

The split attention effect

Split attention occurs when multiple sources of information that cannot be understood in isolation are presented separately in terms of space or time, and as a result, learners are required to split their attention while mentally integrating the different sources of information (Ayres & Sweller, 2005; Sweller et al., 2011; Sweller et al., 1998). Mental integration in this case involves searching and matching information from the different sources as well as linking the relationships between the information. This process increases extraneous cognitive load and reduces learning. Extraneous cognitive load can be reduced by integrating the sources of information, as the amount of searching and matching can be lowered. For example in the case of explanatory text and diagrams the related sources of information should be placed near each other on the page. The split attention effect occurs when split-attention based instruction produces significantly lower learning outcomes compared to integrated based instruction (Sweller et al., 2011). Specifically, split attention caused by information separated in space is called the spatial contiguity effect, and when separated by time (sequential presentation) is called the temporal contiguity effect (Mayer, 2001).

This effect was initially investigated by Tarmizi and Sweller (1988) using circle geometry in a series of five experiments. After failing to find the worked example effect in the first three experiments, Tarmizi and Sweller modified their worked examples to an integrated format by placing the associated explanatory text

within the geometric diagram, rather than below the diagram as traditionally

presented. Having done this, the integrated format was found to be more effective

compared to the traditional text-diagram format which caused split attention.

The midpoint N of XY, where X is (4,3) and Y is (8,7), is given by

$$N = \left(\frac{4 + 8}{2}, \frac{3 + 7}{2} \right)$$

$$= (6, 5).$$

The midpoint M, of interval AB, where A is (x_1, y_1) and B is (x_2, y_2), is given by

$$M = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right).$$
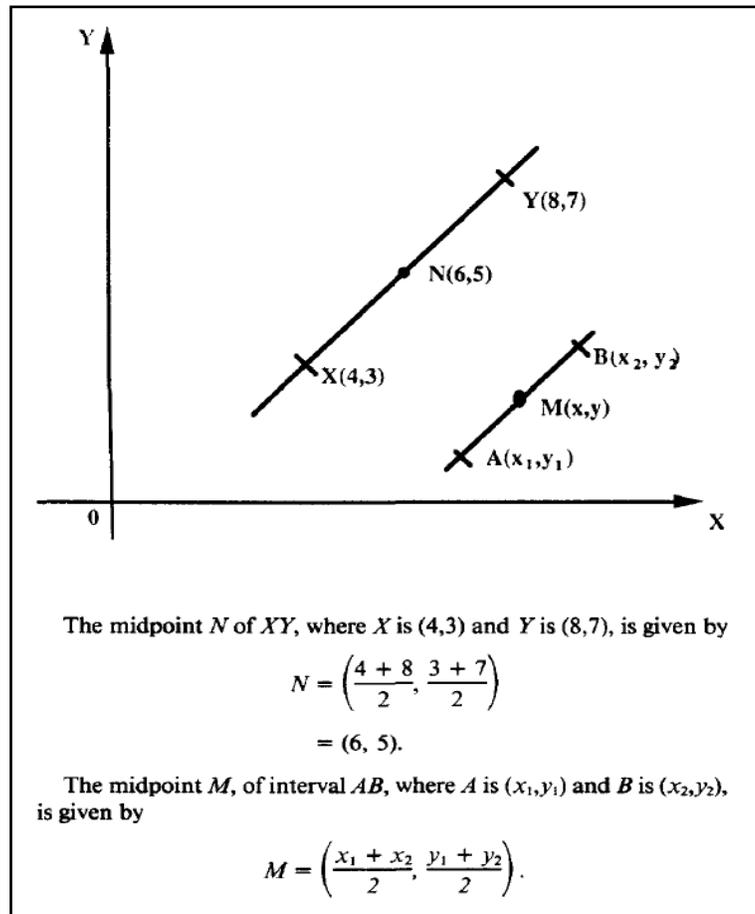
(a)
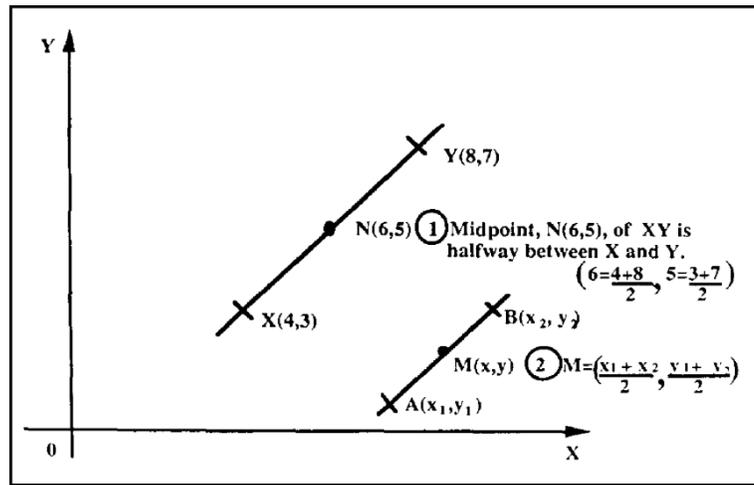
(b)

Figure 9. (a) An example of split attention caused by the diagram and related explanation and (b) the integrated format, adapted from Sweller, Chandler, Tierney, and Cooper (1990)

Split attention often appears in traditional geometry and coordinate geometry textbooks, where a diagram and the associated explanation are presented separately. For example, Sweller et al. (1990) illustrated that the coordinate diagram and explanation to find the coordinates of a point are usually separated (see Figure 9.a). Sweller et al. found the integrated format, as shown in Figure 9.b, was more useful for learning than the split-attention format.

Significant evidence has been found in support of the split attention effect (for review, see Ayres & Sweller, 2005; Ginns, 2006). A multimedia alternative to avoid split-attention materials is to use a combination of both auditory and visual sources of information (Jeung, Chandler, & Sweller, 1997; Mayer, 2001; Mayer & Moreno,

2003; Mousavi et al., 1995). This strategy is known as the *modality effect*, and is successful because it allows learners to examine the picture or diagram while simultaneously listening to an explanation. The search processes caused by split-attention are consequently reduced.

In summary, split-attention materials will impact negatively on learning, including the use of worked examples, unless learners have a high degree of prior knowledge. The effect can be avoided by integrating the different sources of information.

The redundancy effect

Redundancy occurs when multiple sources of information that can be understood in isolation are presented simultaneously. In other words, different sources of information repeat the same information. A common example of redundancy can be found when a speaker reads to the audience, word-for-word, the information already presented on a power-point slide. Consequently, studying worked examples with redundant information will also be disadvantageous (Sweller et al., 2011; Sweller et al., 1998). The negative impact occurs when learners attend to the different sources of information and attempt to establish relations between them (see Sweller & Chandler, 1994). A direct consequence is that the redundant information must be omitted to avoid excessive extraneous cognitive load. The redundancy effect

occurs when learning materials containing no redundant information are found to be more effective than the materials containing redundant information. Nevertheless, like most cognitive load effects, this effect occurs mostly when the material has a high element interactivity (Sweller & Chandler, 1994).

Initial evidence for the redundancy effect was demonstrated by Chandler and Sweller (1991) using electrical engineering and biology materials. To avoid split-attention, Chandler and Sweller designed an integrated instructional format where the integration of a diagram and text was not actually required because the diagram was self-explanatory, and compared this strategy with a single source of instruction containing only the diagram. The results indicated that the design which included only a diagram was superior to the dual-mode design. It was argued that the explanatory material provided in the integrated instructional group was disadvantageous because it imposed extraneous cognitive load caused by unnecessary processing.

Since this initial study was conducted, considerable evidence in support of the effect has been collected (for example, see Diao & Sweller, 2007; Kalyuga, Chandler, & Sweller, 1999).

Additionally, van Gog, Paas, and van Merriënboer (2006) found that adding lengthy textual explanations to worked examples may cause the redundancy effect. Distinguishing between a *process-oriented* worked example, where additional information about why and how the solution steps are chosen are added to the

example, and a *product-oriented* worked example which only shows solution steps, van Gog et al. (2006) found the product-oriented worked examples to be superior. For electrical circuit problems, the process-oriented worked example created redundancy because the learners had to process the essential information (solution steps) and the additional information (process information), which was not necessarily related to the understanding of the solution steps. As a consequence, the process-oriented worked examples imposed a heavier cognitive load and inhibited learning.

At a later date, van Gog, Paas, and van Merriënboer (2008) revisited process and product oriented worked examples by investigating four different sequences: product–product, product–process, process–product and process–process. It was found that the sequence of process–product oriented worked examples was most advantageous for low prior knowledge learners, but not for high prior knowledge learners. Arguably, low knowledge learners benefited by process oriented worked examples at the initial stage of knowledge acquisition but once expertise increased, process oriented worked examples became redundant and caused extraneous cognitive load. This finding added more evidence in support of the expertise reversal effect discussed above.

Promoting Germane Cognitive Load

Briefly, as previously noted, germane cognitive load can be described as the working memory resources devoted to dealing with the intrinsic cognitive load presented by the learning material. This cognitive load is considered a 'good' or effective load since it directs working memory to activities that support schema acquisition and automation (see Chapter 2). Two cognitive load theory effects have been generated from the implementation of worked example instructions aimed at specifically promoting germane cognitive load: namely, the variability and the self-explanation effects.

The variability effect

It is not unusual for instructors to provide multiple worked examples, consisting of very similar examples, or a wide range of different examples. High variability worked examples present a wide range of different problems that utilise the same concept or procedure. According to the variability effect, worked examples with highly variable features improve learning compared to worked examples with more similar features (Sweller et al., 2011; Sweller et al., 1998).

Using a geometry task in computer numerically controlled machinery programming, Paas and van Merriënboer (1994) examined the effectiveness of studying worked examples under a number of conditions, including a set of problem-

example pairs with low and high variability. Worked examples with low variability had only differences in numerical values, but in those with high variability, there were differences in both values and problem formats, including problem goals and problem settings. The results suggested that high variability worked examples were superior to those with low variability. Paas and van Merriënboer argued that problem situation variability could motivate students to improve schema acquisition and automation while they are learning to recognise the key feature between modified problems, hence imposing a high germane cognitive load.

Quilici and Mayer (1996) presented worked examples in statistics word problems, with high variability in terms of structural and surface features. They argued that worked examples can be used to assist learners solving similar problems through analogical reasoning, where learners can extract and map the *surface* or *structural* feature of the worked example to the problem they are attempting to solve. Surface variability was determined by the attributes of the objects illustrated by the cover story of the word problems, while structural variability was defined by variations of relations among objects within the solution procedure. Their findings indicated that the high variability structure, emphasising examples facilitated better learning than the surface emphasising examples. Catrambone (1994) further found that transfer skills would be facilitated when variations in sub-goals were emphasised. Similarly, Renkl, Stark, Gruber, and Mandl (1998) demonstrated that a block of multiple examples were superior to a block of uniform examples.

The self-explanation effect

Earlier research on worked examples conducted by Chi et al. (1989) discovered that self-explanation methods helped high achievers process worked examples in a more meaningful manner. For this reason, researchers argue that asking learners to self-explain, like more knowledgeable learners (e.g., prompting them to elaborate solution steps), improves learning from worked examples since it directs germane working memory resources to deal with essential elements that constitute the new knowledge demonstrated in the worked out solution (Chi, De Leeuw, Chiu, & Lavancher, 1994; Renkl, 1997b, 2002).

The self-explanation effect occurs when adding self-explanations improves learning from worked examples compared to not adding self-explanations. The studies of Catrambone (1994) and Renkl et al. (1998) have shown that eliciting self-explanations, especially with high variability, tended to improve the effectiveness of worked examples. Moreover, R. K. Atkinson et al. (2003) found that low ability students benefited from self-explanation prompts that were provided with each solution step of the worked examples, compared to students who were not given self-explanation prompts.

Renkl et al. (1998) explored the benefit of training students to self-explain by comparing two groups: explicit training and general training. In explicit training, the

participants were given a short training session covering essential components in self-explanation, an example of self-explanation as well as coached practice. In general training, the participants were given thinking-aloud training. It was found that the explicit self-explanation training was superior and, in particular, benefited the low knowledgeable learners.

Despite some positive results in favour of self-explanations, there have also been some negative effects. Using *molar* and *modular* worked examples, Gerjets, Scheiter, and Catrambone (2006) found that self-explanation instruction was not advantageous. Gerjets et al. (2006) illustrated that the main distinction between molar and modular is that molar examples emphasise problem categories while modular examples emphasise procedural solutions. It was found that modular examples were more effective and self-explanation was not advantageous for both. Gerjets et al. argued that the examples had written explanations and therefore may be already intelligible; hence the self-explanation prompts forced learners to process redundant information contained within the examples and their elaborations.

Additionally, Große and Renkl (2006) found that self-explanation instruction did not enhance learning using either the *multiple solutions* or *uniform solutions* of worked examples. Moreover, when learners were given incorrectly worked out solutions and asked to self-explain, the quality of self-explanations decreased, especially for more able learners (Große & Renkl, 2007).

Self-explanation on one hand may allocate a learner's germane resource to deal with the intrinsic cognitive load presented by worked examples, and may enhance learning. However, on the other hand, self-explanation instruction may create unnecessary extraneous cognitive load and reduce learning. In reviewing the self-explanation effect generated by research published before 2000, R. K. Atkinson et al. (2000) indicated that more evidence for the self-explanation effect during studying worked examples was needed. More recently, Wittwer and Renkl (2008) found that instructional explanations often did not adapt sufficiently to the learner's characteristics. Furthermore, their effectiveness may also depend on the knowledge domain, whether it is more conceptual, procedural or reasoning, and the educational setting, whether learning takes place in the classroom, with peer tutors or in small group discussion. Consequently, adding self-explanations to worked examples still needs considerable research to identify the conditions under which it might be most effective.

Summary

This chapter discussed a number of cognitive load effects generated by cognitive load theory, especially the goal-free and worked example effects. Both strategies (goal-free problems and worked examples) were established by the theory to prevent the use of means-ends analysis when learning through conventional

problem solving. Means-ends analysis creates high extraneous cognitive load and hampers learning.

The effectiveness of goal-free problems has been shown using a number of various learning tasks. The goal-free effect is obtained by removing the goal state of a problem during an acquisition period. Learning is facilitated through goal-free problems as cognitive capacity is directed to the problem state and associated moves, rather than the problem goal and the search to reduce the distance between the problem state and the problem goal induced by means-ends analysis.

The effectiveness of worked examples for learning about novel information has been shown across many domains. Learning from a set of worked examples and problem solving pairs, or completion problems (partial worked examples), rather than trying to solve problems without guidance, has been shown to be highly effective in the initial stages of knowledge acquisition. Nevertheless, the expertise reversal effect has demonstrated that worked examples can be ineffective for high prior knowledge learners. To overcome the expertise reversal effect, the guidance fading effect provides a gradual transition from full worked examples to full problem solving instructions as expertise increases. Lastly, it is noted that formatting worked examples requires careful consideration, as other sources of extraneous cognitive load can be created through split-attention and redundant materials. Finally, two strategies, variability and self-explanations, were discussed as they have been used with worked examples to promote germane cognitive load and improve learning.

Chapter VI   Collaborative Learning

Collaborative learning is an example of a social context that commonly allocates three or more students into small groups (Levine & Moreland, 2012) where they mutually work together and learn from each other while attempting to accomplish a problem solving task (Van den Bossche, Gijselaers, Segers, & Kirschner, 2006).

The idea of dividing a classroom into small work groups has been applied by many teachers for decades. Recently, school curricula in some countries have recommended teachers use group learning. For example, the mathematics curriculum used nationally in the USA that was developed by the NCTM (National Council of Teachers of Mathematics) in 2000, NCTM (2000, p. 10) stated in its teaching principles that teachers should encourage "students' discussion and collaboration" as well as encouraging students to "construct mathematical arguments and respond to others' arguments". The learning principle in the curriculum further stated that:

"Learning with understanding can be further enhanced by classroom interactions … social interaction can be used to promote the recognition of connections among ideas and the reorganization of knowledge … in such

settings, procedural fluency and conceptual understanding can be developed through problem solving, reasoning and argumentation" (NCTM, 2000, p. 13).

In a study comparing traditional curricula and NCTM curricula, Latterell (2005) found that the currently used NCTM curricula emphasise the use of group work and group discussion methods. Latterell (2005, p. 96) asserted that "[t]he curricula are often set up so that the teacher introduces a topic then students are responsible for working with each other …". Latterell also observed that the NCTM curricula are widely used in many countries, so it can be assumed that many other countries have applied this method at schools. In Indonesia, collaboration has been a respected value in daily life and it is commonly called "gotong royong" which means working together. Recently, the National Curriculum of 2013 has included collaborative learning or cooperative learning as highly recommended learning method in mathematics classrooms. Accordingly, this chapter discusses cognitive psychological aspects need to be considered when instructing students to learn in small groups.

In the research literature, a group of students working in a collaborative learning environment is often called different names, such as: group work, group study, group learning, small groups, or group discussions. In a strictly designated setting, it is also known as cooperative learning (Johnson & Johnson, 1994) where

studying in groups is based on a problem solving approach, applying specific

grouping rules and rewards, and usually requiring longer learning periods.

Historically, the more usual setting is face-to-face where group members

physically meet, and communicate directly with each other. Grouping settings can

also occur in a virtual environment, where group members are physically separated,

and collaboration is facilitated by communication technology devices. More recently,

Kim and Baylor (2006) proposed learning by an interaction framework using

pedagogical agents, such as a digital character whom is created using a computer

program, and interacts with learners.

To give a broad sense of collaborative learning, the theoretical frameworks

proposed by prominent theorists are initially summarised in this chapter. A number of

positive and negative factors that contribute to cognitive performance (i.e.,

knowledge acquisition and transfer) are also identified.

Why Collaborative Learning: Some Theoretical Frameworks

Collaborative learning has been widely used (Gillies, 2003), well-researched

(Levine & Moreland, 2012) and advocated by many leading educators and

organisations (e.g., NCTM, 2000; Rosenshine, 2010). In particular, social

constructivist theory is frequently used to emphasise that learning should be facilitated through social and collaborative activities where students construct knowledge by interactions with others (Johnson & Johnson, 1994; Schreiber & Valle, 2013).

Underlying social constructivism are the cognitive developmental perspectives of Piaget and Vygotsky (Blatchford, Kutnick, Baines, & Galton, 2003; E. G. Cohen, 1994; Schreiber & Valle, 2013). Based on Piaget's theory (see Chapter 2), cognitive disequilibrium stimulates learners to interact within the social context to assimilate, modify and accommodate knowledge into more developed constructions. According to Piagetian theory, learners are the main actors in knowledge construction (Daniels, 2001). In other words, learners have to construct knowledge by themselves, and hence teachers should only provide the social context and materials that support discussions aiming at cognitive conflict resolution (Geary, 1995).

On the other hand, Vygotsky's theory assumes that learning is enhanced within social and cultural contexts because these contexts influence how learners interpret and understand concepts (Daniels, 2001). It is argued that interactions within social contexts facilitate knowledge construction and, as a consequence, teachers should create a collaborative environment where learners can actively communicate and contribute towards constructing meaning (Schreiber & Valle, 2013). Vygotsky proposed the concept of the *zone of proximal development* in which learners require scaffolding (assistance) from instructors or more able peers to understand meaning

that learners cannot comprehend by themselves (Daniels, 2001; Schreiber & Valle, 2013).

It is further argued that the zone of proximal development can also take place in collaborative contexts, consisting of relatively similar levels of expertise (e.g., peers), as long as active collaboration is maintained (Schreiber & Valle, 2013). Moreover, Mayer (1999) noted that, according to Vygotsky, collaborative learning should be situated in the real world of the learners, thus creating more authentic and meaningful learning. Nevertheless, both Gillen (2000) and Mayer (1999) argued that the implementation of Vygotsky's theory in classroom practices may create a number of challenges because not all lessons can occur in natural settings.

From a social cognitive learning perspective, Bandura (1986) argued that learning is determined by triadic interactions between cognitive ability (e.g., attention, retention, reproduction), social behaviour (e.g., motivation, self-efficacy) and environment (e.g., learning situations, social systems). In particular, Bandura suggested that people have evolved to learn from the observation of other people's behaviour. Consequently, collaborative learning creates a collective behaviour that, to some extent, contributes to individual motivation, which determines whether learners acquire observed skills or not.

According to Schmidt et al. (2007), group discussion in a problem based learning environment (PBL –a learning strategy advocated by many social

constructivists) has two functions. First, it activates prior knowledge among group members to deal with the learning task. Second, it facilitates sharing expertise. Prior knowledge activation and sharing expertise are essential to begin collaboration and for learning new problem solving skills. Both these points are consistent with a cognitive load theory approach to learning. Access to prior knowledge reduces working memory load, and by working together in a group, the intrinsic cognitive load may also be reduced because of cognitive sharing among group members (Hoogveld, Paas, & Jochems, 2003; Schmidt et al., 2007).

It is clear that many educationalists believe that there are strong theoretical foundations to support collaborative learning, as it has been widely advocated and implemented. It is argued that the role of social interactions in learning are vital to foster multiple perspectives and representations of knowledge (Schreiber & Valle, 2013). Moreover, Blatchford et al. (2003) suggested that group settings are pedagogically beneficial for students because their dynamic and dialogic features can be expected to affect student engagement in the learning processes. Nevertheless, Gillies and Boyle (2010) found that the implementation of collaborative learning in classrooms was not always successful. In addition, the National Mathematics Advisory Panel, of the US Department of Education (2008) reported that the implementation of collaborative learning in mathematics classrooms and curricula needed further scientific testing. So even though there is strong support for

collaborative learning, there are several issues associated with its successful implementation, which are discussed next.

Improving Collaborative Learning

Many studies have been conducted to identify the factors that improve collaborative learning (for reviews, see E. G. Cohen, 1994; Kreijns, Kirschner, & Jochems, 2003; Schreiber & Valle, 2013; Van den Bossche et al., 2006; Webb, 2009; Weinberger, Stegmann, & Fischer, 2007). It is generally agreed that collaborative learning requires active social interactions, and simply putting students together in a group does not guarantee that effective collaborative learning will occur.

Despite the broad variety of research conducted, many studies indicate inconsistent findings when learning outcomes are measured (e.g., Barron, 2000; Kester & Paas, 2005). Consequently, efficient procedures that can be simply followed by collaborative learning instructors have not been easily specified. As stated by Webb (2009, p. 21), "... to what extent the teacher's role in promoting collaborative dialogue depends on specific features of the classrooms and the students in them is largely unknown". Furthermore, research and theory relevant to collaborative learning in authentic classroom conditions has been rather limited (Blatchford et al., 2003).

One factor thought to be crucial in facilitating effective collaborative environments is task quality. The E. G. Cohen (1994) review found that it was important to choose a suitable task to maintain task-related interactions. According to Cohen, the most suitable group task is a task that cannot be carried out by individuals. Further, the use of open-ended problems, discovery tasks, or complex problems was thought to be necessary in order to stimulate active interactions since they require multiple resources and can be solved using different strategies and methods. Moreover, Johnson and Johnson (1994), Laughlin, Zander, Knievel, and Tan (2003), argued that complex problem solving improves interactions because it promotes the exchange of ideas and the discovery of underlying principles.

In addition to providing complex problems to solve, it is asserted that group members should be informed that they should not complete the task alone, but have a responsibility to help other members of the group also complete the task (Johnson & Johnson, 1994). The perception of a group member that they have to successfully work together with the other group members is called *positive interdependence*. It is argued that by having positive interdependence, group members are forced to provide mutual support while working together to maximise the learning process (Johnson & Johnson, 1994, 2002). It is further suggested that positive interdependence can be improved by assessing not only group performance, but also individual performance as well as providing group rewards based on both individual and group achievements.

An important consideration for effective collaboration is how the groups are formed. Some research has shown that allocating close friends to a group produces better learning outcomes (Andersson & Rönnberg, 1995; Hanham & McCormick, 2009; Weldon & Bellinger, 1997). Further, heterogeneous groups consisting of mixtures of low ability and medium ability students, and high ability and medium ability students, have also demonstrated significant achievement (Webb, 1991). In contrast, homogenous groupings of high ability or low ability students have not been found to be significantly related to better achievement (Saleh, Lazonder, & de Jong, 2007). Overall, heterogeneous ability groupings, consisting of a balanced number of high, medium and low ability students, is favoured by many researchers (Johnson & Johnson, 1994; Webb, 1991).

Assigning students to groups with little direction or support does not guarantee success. Receiving or giving elaborated explanations is considered relevant to improve collaborative learning (Webb, 1991; Webb & Mastergeorge, 2003). Therefore, Webb suggested that students should be able to request help in such a manner that they will receive detailed explanations instead of final correct answers. However, to be effective help seekers, training might be needed (Webb, 1991, 2009; Webb & Mastergeorge, 2003), since research on giving or listening to explanations demonstrated that it did not necessarily improve individual performance (Renkl, 1997a; Webb & Mastergeorge, 2003). Such training might be advantageous;

however, it is also argued that when students do not have sufficient prior-knowledge of the to-be-learned material, giving explanations remains a difficult task.

Additionally, Johnson and Johnson (1994) argued that students should be trained specifically on basic cooperation skills, while Schmidt et al. (2007) suggested training was needed in typical collaborative skills. Furthermore Laughlin et al. (2003) found that providing initial information both about the requirement of the group task and the expected group processes was necessary to facilitate effective information processing strategies during complex group tasks.

One well-known negative effect of social interaction is called *social loafing*, which is a tendency to exert less effort when working with others. For example, Ingham, Levinger, Graves, and Peckham (1974) investigated group performances in a physical task (rope pulling), finding that individual productivity decreased when co-workers were provided. Latane, Williams and Harkins (1979) reported that individual performance in clapping and shouting tasks decreased when they worked with others, either face-to-face or perceptually (i.e., they were blindfolded and told that they performed together with others, but actually they performed alone). Social loafing is also considered a motivational or coordination loss, and the larger the size of the group, the higher the tendency for social loafing (Petty, Harkins, & Williams, 1980). Petty et al. found that when in groups, students performed fewer positive evaluations than students who performed the evaluations alone.

Although exerting lower effort when working with others could be seen as individual efficiency, however it is important to note that the learning process is an individual construction of knowledge which requires individual responsibility to learn. If individual accountability (see Johnson & Johnson, 1994) during working in groups is decreased through social loafing, it may also lower the individual's learning as insufficient mental effort is made towards schema acquisition. The aim of learning in collaboration is not only to complete the group task, but also to assist each group member in mastering the group task.

Furthermore, Arterberry, Cain, and Chopko (2007) discovered that social loafing increased when there was no assessment of the learning process. Furthermore, Harkins and Petty (1982) suggested that social loafing can be eliminated by giving more difficult tasks to the group or assigning each student to perform different tasks. Similarly, Andersson and Rönnberg (1995) also suggested the use of complex tasks to reduce negative effects during collaborative work.

Nevertheless, while possessing collaborative skills is important, it should be noted that improving collaborative learning is meant to improve the quality of individual performance. However, it has been reported that much of the research into the effectiveness of collaborative learning has not directly tested the performance of individuals, but has focused more on group processing aspects, such as motivation or self-process attributes as well as the whole group performance (F. Kirschner et al., 2009a; Paas & Sweller, 2012). Hence, it is suggested that the research should give

more emphasise to the measurement of performance of each individual after learning in collaborative contexts (F. Kirschner et al., 2009a; Paas & Sweller, 2012).

Cognitive Perspective on Collaborative Learning

This chapter is concerned with instructional designs for collaborative learning in accord with human cognitive architecture. Although cognitive load theory has been largely used to test and establish instructional procedures for individual learning, with limited data collected on collaborative learning instruction, the theory provides a strong theoretical base. Specifically, the recent evolutionary educational psychology view of human cognitive architecture can be used to explain some of the fundamental underpinnings of collaborative learning and thus design effective learning group environments (Paas & Sweller, 2012; Sweller et al., 2011).

Recently, cognitive load theory researchers have seen collaborative learning as an alternative strategy for learning about more complex materials that are difficult to learn individually due to working memory restrictions. As such a new cognitive load theory effect has been proposed (Sweller et al., 2011). The *collective working memory effect* occurs when individuals obtain higher learning outcomes after learning in collaborative contexts compared to individuals who learned alone (F. Kirschner et al., 2009a). It is assumed that the intrinsic cognitive load is distributed across group members during collaborative learning, freeing up more working memory capacity at

the individual group member level. This does not occur when students are engaged in individual learning and have to deal with all the working memory load themselves (for reviews, see Paas & Sweller, 2012; Sweller et al., 2011).

The effectiveness of collaborative learning can be seen from the evolutionary perspective of cognitive load theory. Paas and Sweller (2012) suggested that collaborative learning demonstrates an example of the borrowing and reorganising principle. This principle indicates that the most effective way to obtain new information is by directly borrowing it from an other's long term memory (see Chapter 2). As discussed previously, humans have evolved to communicate in everyday life, to share and obtain information from each other. Consequently, collaborative learning will facilitate learning, as students can share information and learn from each other, just like in everyday life (Sweller et al., 2011).

However, it should be noted that previous studies have shown that effective collaboration does not always occur (for example, see Kreijns et al., 2003). Specifically, Paas and Sweller (2012, p. 31) note that collaborative interaction requires not only "general communication and coordination" like in a natural social context, but also requires "task-specific communication and coordination" which is more related to assigned learning. Similarly, Geary (1995, 2008) argues that collaborative learning has similar features to social contexts where students learn biologically primary knowledge. Therefore, in collaborative learning, students tend to automatically develop general communication and coordination skills (biologically

primary knowledge), which might not be related to learning, rather than allocating more attention to the assigned biologically secondary knowledge (Geary, 1995, 2008). Arguably, task-specific communication and coordination is more useful; however, it is argued that students need to learn about this directly through training (Paas & Sweller, 2012). As a consequence, the interaction process in collaborative learning demands some cognitive load from each group member, and is known as the *transaction cost* (F. Kirschner et al., 2009a). This can be extraneous load (i.e., when it is directed to off-task activities) or germane load (i.e., when it is directed to the learning task) (Janssen, Kirschner, Erkens, Kirschner, & Paas, 2010). Because working memory is limited when learning complex materials, any transaction cost extraneous to learning must be kept to a minimum and any transaction cost germane to learning must be invested to achieve the expected learning outcomes.

It is argued that providing students with a complex learning task, divided among group members, will promote task-specific collaboration and facilitate the collective working memory effect (Paas & Sweller, 2012). When the material is shared among group members, task-specific interactions will be required to integrate and fully understand the material, and hence direct students' attention to the task. Consequently, unnecessary extraneous transaction costs (e.g., off-task conversation, social loafing) will be less likely to occur.

When the extraneous transaction cost is kept to a minimum, collaboration will facilitate the learning of more-complex materials that are hard to learn individually

121

(Paas & Sweller, 2012). Complex materials impose a high intrinsic cognitive load, and novice learners have limited working memory capacity to learn such material. However, when the learning material is shared among several group members, an individual is required to process less task-relevant information, which is much lower in intrinsic cognitive load because of reduced element interactivity. The remaining working memory resources can then be allocated to learning about the important aspects of the materials by processing relevant information communicated from other group members. Although all group members may share all the thoughts discussed, the actual information processing will be sub-divided (see P. A. Kirschner, Kirschner, & Janssen, 2014). Hence, through collaboration, individuals are more able to learn about complex materials.

Evidence for the collective working memory effect

According to Paas and Sweller (2012), research into the collective working memory effect should have two characteristics. Firstly, it should be conducted in controlled and randomised experimental conditions by isolating the cognitive effects of task complexity and minimising the effects of social and motivational factors on collaborative learning. Secondly, it should be conducted in a traditional face-to-face collaborative learning context. Interaction in this context is assumed to require

biologically primary knowledge and hence impose lower cognitive demand. The research reported below followed these requirements.

Initial evidence in support of the collective working memory effect was found by F. Kirschner, Paas, and Kirschner (2009b), using a high-school biology topic, where an individual learning condition was compared to a collaborative learning condition (consisting of three group members). During the learning phase, students were given problem solving tasks to solve individually or collaboratively. For the collaborative learning condition, every member of a group had information about one third of the whole task only, and hence sharing was required to complete the task; whereas in the individual condition, one student was given the whole task to solve. Following the learning phase, all students were tested individually with retention and transfer tasks. Cognitive load (mental effort) and efficiency measures were also collected (see Paas & van Merriënboer, 1994). No significant differences were found between learning conditions on the performance tests or cognitive load measures. However, there was a significant interaction effect between the learning condition and the test type on the efficiency measure. On the transfer test, the collaborative learning condition had higher efficiency than the individual learning condition.

F. Kirschner, Paas, and Kirschner (2011) continued their investigation by examining the impact of collaborative learning on low-complexity and high-complexity tasks. Again using high school biology, low and high-complexity problem solving tasks were used to compare learning individually with learning

collaboratively (using triad grouping and distributing one third of the task information to each member). Interaction effects were found between task-complexity and learning conditions on performance, mental effort and efficiency measures for the transfer test. For the low-complexity task, there was no significant difference between learning conditions; however, for the high-complexity task, the collaborative learning condition was superior to the individual learning condition.

The effect of high-complexity tasks on collaborative learning was also investigated by Zhang, Ayres, and Chan (2011) using a quasi-experimental design. Two collaborative conditions were formed by grouping students together to complete a take-home assignment (i.e., designing a personal homepage). One condition was task-based (a theme was assigned) and the other an open-ended project (the group decided their own theme, content and arrangement). Each group member was required to develop at least five web pages and then they worked collaboratively to link all the web pages together in a unified homepage. An individual context was also created, by assigning a cohort of students to complete the same assignment individually (the assigned theme). The open-ended collaborative learning context demonstrated higher performance and lower cognitive load compared to the individual learning context and the task-based collaborative learning context. Notably information about the task was not subdivided across group members as used by F. Kirschner, et al. (2009b, 2011) (consequently, there was no need of combining prior-knowledge among group members). Nevertheless, it can be argued that in the Zhang

et al. study active collaboration was achieved since students were not only required to complete the assignment in the given period, but there was sufficient time for students to develop social interactions outside the classrooms, which might be a more natural setting. To reinforce the need for collaboration, students working in groups were also required to give presentations to their peers at the end of the learning period.

More evidence of the importance of task complexity was collected by F. Kirschner, Paas, Kirschner, and Janssen (2011). In this study, biology tasks were studied in both individual and collaborative learning contexts either by using worked examples or through conventional problem solving instructional formats. The collaborative groups were structured as in the previous studies (Kirschner, et al., 2009b, 2011), where each group member was presented with one third of the material (one-third of the information of the problem solving task for the problem solving group, and one-third of the worked examples for the worked example group) and hence sharing of information was required. The individual learners were given either full worked examples or problem solving tasks. During the learning phase, students in the worked example conditions were asked to study three worked examples only (with no paired exercise problems), and equivalently, students in the problem solving conditions were asked to solve the three problems without instruction. The results indicated that overall, there was a main effect of social context: the collaborative conditions led to significantly higher performance and efficiency than the individual

125

conditions on the test. There was an interaction effect between the instructional format and the social context on both performance and efficiency. It was shown that for the collaborative groups, learning by problem solving was more efficient than worked examples. For individuals, learning by worked examples was more effective than problem solving. The authors argued that high task complexity imposed not only by the intrinsic nature of the task, but also by the instructional format should be taken into account to improve collaborative learning. These results confirmed previous findings by F. Kirschner, Paas, and Kirschner (2011), that the efficiency of collaborative learning was increased by presenting high-complexity problem solving.

A previous study by Retnowati et al. (2010) compared worked examples with problem solving strategies in collaborative settings using a high school geometry task. Retnowati et al. found that worked examples benefited both individuals and collaborative groups for both numeric and reasoning scores on similar and transfer tests. In addition, a marginal interaction effect between worked examples and collaborative learning was found for the reasoning score. The results showed that the effect of worked examples on reasoning was stronger for collaborative learning. However, it is important to note that there was no task-distribution in Retnowati et al.'s study, as group members were not required to share discrete sections of knowledge. In other words, collaborative learning could benefit by worked examples when each group member had the same worked examples to study together (i.e., to discuss the same materials with essentially the same knowledge base). Overall, this

126

study, consistent with the worked example effect, found no evidence for the effectiveness of problem solving strategy when learned individually or collaboratively.

It is notable that the evidence for the collective working memory effect was found through the implementation of very structured and scripted collaborative groupings. Firstly, each group member received only a portion of the learning material. This setting is not common in regular classrooms where students usually receive the same learning material. Secondly, the group members were not allowed to use pencil/pen and paper while learning to prevent offloading any working memory burden, thus controlling this factor. The collaboration process relied heavily on verbal communication. Consequently, F. Kirschner, Paas, Kirschner, et al. (2011, p. 597) concluded that "it is not clear to what extent the results obtained in this study can be generalised to real classroom settings".

Nevertheless, recent research has examined which collaborative structure is more effective. It was found that when collaboration among group members is a must because of each group member has to share their knowledge to learn the given material then the individual performance is better compared to when collaboration is simply encouraged as their support of learning (Retnowati, Ayres, & Sweller, 2015). The "must share knowledge" group structure used a jigsaw approach in structuring the collaborative group (i.e. group members are divided and have different knowledge base), and the "encouraged" group structure used the common structure (i.e. all

127

students have the same knowledge base). In this research, the learning material is

Year 7 geometry called the application of Pythagoras' theorem to find the area of a

triangle.

Task: Find the area of these figure (not to scale)
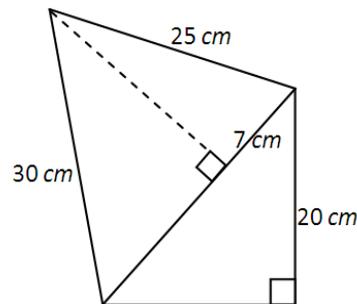
Figure 12. Example of problem solving requires Pythagoras' theorem and the concept of triangle's area

The learning phase was divided into two stages. In the jigsaw groupings, half

students learned the basic knowledge of triangle area and the other half the basic

knowledge of Pythagoras' theorem for the first learning stage. The learning material

was developed using a worked example approach, where students were individually

studying by examples and completing similar problem solving. In the common

groupings, all students learned both material but the quantity of the practice was a

half of these in the jigsaw group. In the second learning stage, all students learned a

more complex material which is the application of the Pythagoras' theorem to

calculate the area of triangle, in groups. The learning material was based on problem

solving. In the jigsaw group, students with Pythagoras' theorem knowledge base had to share with students with the triangle area knowledge base. Without this, they would not be able to handle the learning material in the collaborative groups.

Summary

This chapter described some of the theoretical foundations in support of collaborative learning. Many leading educators encourage its use in classrooms, although it is acknowledged that the research does not consistently demonstrate its effectiveness. Even though, as F. Kirschner et al. (2009a) point out, there is a lack of empirical studies supporting group effectiveness on individual learners, collaborative learning seems highly appealing.

Collaborative learning is an example of a social context that is formed in a classroom by allocating students to study a learning topic together. Research suggests that the effectiveness of working together in a group occurs under certain conditions, such as when group members show mutual support by elaborating explanations. Further, it is suggested that teachers or instructors play an important role in preparing suitable group compositions. In addition to this, providing complex problem solving is necessary for encouraging positive interdependence as well as reducing the negative effect of working in groups, such as the social loafing effect.

From a cognitive load theory perspective, social interaction is also seen as a possible context to specifically learn complex tasks, as found in many mathematics topics, which an individual would find difficult to learn alone because of a limited working memory capacity. By working in groups working memory load can be offset by sharing the task. The collective working memory effect occurs when a student performs better after learning in a collaborative context rather than learning individually. The main evidence for this effect has been found when members are required to share discrete sections of information and combine them to accomplish the group task. More importantly, the collective working memory effect has been shown to be effective when complex problem solving is used.

REFERENCES

Amirault, R. J., & Branson, R. K. (2006). Educators and expertise: A brief hystory of theories and models. In A. K. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The cambridge handbook of expertise and expert performance* (pp. 69-88). New York, NY: Cambridge University Press.

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 48*(1), 35-44.

Andersson, J., & Rönnberg, J. (1995). Recall suffers from collaboration: Joint recall effects of friendship and task complexity. *Applied Cognitive Psychology, 9*(3), 199-211. doi: 10.1002/acp.2350090303

Arterberry, M. E., Cain, K. M., & Chopko, S. A. (2007). Collaborative problem solving in five year old children: Evidence of social facilitation and social loafing. *Educational Psychology, 27*(5), 577-596. doi: 10.1080/01443410701308755

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In W. S. Kenneth & S. Janet Taylor (Eds.), *Psychology of Learning and Motivation* (Vol. 2, pp. 89-195). San Diego, CA: Academic Press.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short term memory. *Scientific American, 225*, 82-90.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from

 Examples: Instructional principles from the worked examples research.

 Review of Educational Research, 70(2), 181-214. doi:

 10.3102/00346543070002181

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying

 examples to solving problems: Effects of self-explanation prompts and fading

 worked-out steps. *Journal of Educational Psychology, 95*(4), 774-783.

Ayres, P. (1993). Why goal-free problems can facilitate learning. *Contemporary*

 Educational Psychology, 18(3), 376-381. doi: 10.1006/ceps.1993.1027

Ayres, P. (2006). Impact of reducing intrinsic cognitive load on learning in a

 mathematical domain. *Applied Cognitive Psychology, 20*(3), 287-298. doi:

 10.1002/acp.1245

Ayres, P. (2012). Worked example effect. In N. M. Seel (Ed.), *Encyclopedia of the*

 sciences of learning (1st ed., pp. 3467-3471). Berlin: Springer.

Ayres, P., & Paas, F. (2007). Can the cognitive load approach make instructional

 animations more effective? *Applied Cognitive Psychology, 21*(6), 811-820.

 doi: 10.1002/acp.1351

Ayres, P., & Sweller, J. (1990). Locus of difficulty in multistage mathematics

 problems. *The American Journal of Psychology, 103*(2), 167-193. doi:

 10.2307/1423141

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning.

In R. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (1st ed., pp.

135-146). New York, NY: Cambridge University Press.

Ayres, P., & Sweller, J. (2013). Worked examples. In J. A. C. Hattie & E. M.

Anderman (Eds.), *The International Handbook of Student Achievement* (pp.

408-410). London, UK: Routledge.

Baars, B. J., & Franklin, S. (2003). How conscious experience and working memory

interact. *Trends in Cognitive Sciences, 7*(4), 166-172. doi: 10.1016/S1364-

6613(03)00056-1

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556-559.

Baddeley, A. D. (1996). The fractionation of working memory. *Proceedings of the*

National Academy of Sciences of the United States of America, 93(24), 13468-

13472. doi: 10.2307/40924

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory.

Trends in Cognitive Science, 4(11), 417-423.

Baddeley, A. D. (2001). The concept of episodic memory. *Philosophical*

Transactions: Biological Sciences, 356(1413), 1345-1350. doi:

10.2307/3067094

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist,*

7(2), 85-97.

Baddeley, A. D. (2007). Working memory, thought, and action. New York, NY: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In H. B. Gordon (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47-89). San Diego, CA: Academic Press.

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

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *The Journal of the Learning Sciences, 9*(4), 403-436. doi: 10.2307/1466763

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. London: The Cambridge University of Press.

Blatchford, P., Kutnick, P., Baines, E., & Galton, M. (2003). Toward a social pedagogy of classroom group work. *International Journal of Educational Research, 39*, 153-172. doi: 10.1016/S0883-0355(03)00078-8

Bobis, J., Sweller, J., & Cooper, M. (1994). Demands imposed on primary school students by geometric models. *Contemporary Educational Psychology*(19), 108-117.

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

Catrambone, R. (1994). Improving examples to improve transfer to novel problems. *Memory & Cognition, 22*(5), 606-615. doi: 10.3758/bf03198399

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55-81. doi: 10.1016/0010-0285(73)90004-2

Chi, M. T. H. (2006). Two approaches to the study of expert characteristics. In A. K. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The cambridge handbook of expertise and expert performance* (pp. 21-30). Cambridge: Cambridge University Press.

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

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121-152.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J.

    Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7-75).

    Hillsdale, NJ: Erlbaum.

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect:

    Is the reduction of extraneous cognitive load accompanied by an increase in

    germane cognitive load? *Computers in Human Behavior, 25*(2), 315-324. doi:

    10.1016/j.chb.2008.12.020

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small

    groups. *Review of Educational Research, 64*(1), 1-35. doi: 10.2307/1170744

Cohen, J. (1988). *Statistical power analysis for the behavorial sciences* (2nd ed.).

    Hillsdale, NJ: Erlbaum.

Cooper, G., & Sweller, J. (1987). The 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. (2000). The magical number 4 in short-term memory: A reconsideration

    of mental storage capacity. *Behavioral and Brain Sciences, 24*(01), 87-185.

    doi: 10.1017/S0140525X01003922

Cowan, N. (2005). *Working memory capacity*. New York, NY: Psychology Press.

Crippen, K. J., & Earl, B. L. (2007). The impact of web-based worked examples and

    self-explanation on performance, problem solving, and self-efficacy.

Computers & Education, 49(3), 809-821. doi:

10.1016/j.compedu.2005.11.018

Daniels, H. (2001). *Vygotsky and pedagogy*. New York, NY: RoutledgeFalmer.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the

sperling partial report procedure: Evidence for brief auditory storage.

Cognitive Psychology, 3(2), 255-267. doi: 10.1016/0010-0285(72)90007-2

De Groot, A. D. (1978). *Thought and choice in chess* (2nd ed.). The Hague, the

Netherlands: Mouton. (Original work published in 1946).

Departemen Pendidikan Nasional. (2003). *Pelayanan profesional kurikulum 2004*

[translation: Professional service for curriculum 2004]. Jakarta, Indonesia:

Pusat Kurikulum, Balitbang Depdiknas.

Diao, Y. L., & Sweller, J. (2007). Redundancy in foreign language reading

comprehension instruction: Concurrent written and spoken presentations.

Learning and Instruction, 17(1), 78-88. doi:

10.1016/j.learninstruc.2006.11.007

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings.

Memory & Cognition, 7, 149-158.

Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (Eds.). (2006). *The*

cambridge handbook of expertise and expert performance. New York, NY:

Cambrisge University Press.

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

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.

Feltovich, P. J., Pritetula, M. J., & Ericsson, A. K. (2006). Studies of expertise from psychological perspectives. In A. K. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The cambridge handbook of expertise and expert performance* (pp. 41-67). New York, NY: Cambridge University Press.

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: implications for mathematical development and instruction. *American Psychologist, 50*(1), 24-37. doi: 10.1037/0003-066X.50.1.24

Geary, D. C. (2002). Principles of evolutionary educational psychology. *Learning and Individual Differences, 12*(4), 317-345. doi: 10.1016/S1041-6080(02)00046-8

Geary, D. C. (2007). Educating the evolved mind: Conceptual foundations for an evolutionary educational psychology *Educating the evolved mind: Conceptual foundations for an evolutionary educational psychology* (pp. 1-99). Charlotte, NC: Information Age Publishing.

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

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction, 16*(2), 104-121. doi: 10.1016/j.learninstruc.2006.02.007

Gerjets, P., Scheiter, K., & Schuh, J. (2008). Information comparisons in example-based hypermedia environments: supporting learners with processing prompts and an interactive comparison tool. *Educational Technology Research and Development, 56*(1), 73-92. doi: 10.1007/s11423-007-9068-z

Gillen, J. (2000). Versions of vygotsky. *British Journal of Educational Studies, 48*(2), 183-198.

Gillies, R. M. (2003). Structuring cooperative group work in classrooms. *International Journal of Educational Research, 39*(1–2), 35-49. doi: 10.1016/S0883-0355(03)00072-7

Gillies, R. M., & Boyle, M. (2010). Teachers' reflections on cooperative learning: Issues of implementation. *Teaching and Teacher Education, 26*(4), 933-940. doi: 10.1016/j.tate.2009.10.034

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*, 511-525.

Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. xv-xxviii). Hillsdale, NJ: Erlbaum.

Große, C. S., & Renkl, A. (2006). Effects of multiple solution methods in

    mathematics learning. *Learning and Instruction, 16*(2), 122-138. doi:

    10.1016/j.learninstruc.2006.02.001

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can

    this foster learning outcomes? *Learning and Instruction, 17*(6), 612-634. doi:

    10.1016/j.learninstruc.2007.09.008

Hanham, J., & McCormick, J. (2009). Group work in schools with close friends and

    acquaintances: Linking self-processes with group processes. *Learning and*

    Instruction, 19(3), 214-227. doi: 10.1016/j.learninstruc.2008.04.002

Harkins, S. G., & Petty, R. E. (1982). Effects of task difficulty and task uniqueness

    on social loafing. *Journal of Personality and Social Psychology, 43*(6), 1214-

    1229.

Hoogveld, A. W. M., Paas, F., & Jochems, W. M. G. (2003). Application of an

    instructional systems design approach by teachers in higher education:

    Individual versus team design. *Teaching and Teacher Education, 19*, 581-590.

Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. (1974). The ringelmann

    effect: Studies of group size and group performance. *Journal of Experimental*

    Social Psychology, 10(4), 371-384.

Janssen, J., Kirschner, F., Erkens, G., Kirschner, P. A., & Paas, F. (2010). Making the

    black box of collaborative learning transparent: Combining process-oriented

and cognitive load approaches. *Educational Psychology Review, 22*(2), 139-154. doi: 10.1007/s10648-010-9131-x

Jeung, H.-J., Chandler, P., & Sweller, J. (1997). The roles of visual indicators in dual sensory mode instructions. *Educational Psychology, 17*(3), 329-343. doi: 10.1080/0144341970170307

Johnson, D. W., & Johnson, R. T. (1994). *Learning together and alone: Cooperative, competitive and individualistic learning*. USA: Allyn and Bacon.

Johnson, D. W., & Johnson, R. T. (2002). Learning together and alone: Overview and meta-analysis. *Asia Pacific Journal of Education, 22*(1), 95-105. doi: 10.1080/0218879020220110

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

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23-31.

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), 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*(4), 351-371. doi: 10.1002/(sici)1099-0720(199908)13:4<351::aid-acp589>3.0.co;2-6

Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of

    instructional guidance. *Educational Psychology, 21*(1), 5-23. doi:

    10.1080/01443410124681

Kantowski, M. G. (1977). Process involved in mathematical problem solving.

    Journal for Research in Mathematics Education, 8(3), 163-180.

Kester, L., & Paas, F. (2005). Instructional interventions to enhance collaboration in

    powerful learning environments. *Computers in Human Behavior, 21*(4), 689-

    696. doi: 10.1016/j.chb.2004.11.008

Kim, Y., & Baylor, A. (2006). A Social-Cognitive Framework for Pedagogical

    Agents as Learning Companions. *Educational Technology Research and

    Development, 54*(6), 569-596. doi: 10.1007/s11423-006-0637-3

Kirschner, F., Paas, F., & Kirschner, P. A. (2009a). A cognitive load approach to

    collaborative learning: united brains for complex tasks. *Educational

    Psychology Review, 21*(1), 31-42. doi: 10.1007/s10648-008-9095-2

Kirschner, F., Paas, F., & Kirschner, P. A. (2009b). Individual and group-based

    learning from complex cognitive tasks: Effects on retention and transfer

    efficiency. *Computers in Human Behavior, 25*(2), 306-314. doi:

    10.1016/j.chb.2008.12.008

Kirschner, F., Paas, F., & Kirschner, P. A. (2011). Task complexity as a driver for

    collaborative learning efficiency: The collective working-memory effect.

    Applied Cognitive Psychology, 25(4), 615-624. doi: 10.1002/acp.1730

Kirschner, F., Paas, F., Kirschner, P. A., & Janssen, J. (2011). Differential effects of problem-solving demands on individual and collaborative learning outcomes. *Learning and Instruction, 21*(4), 587-599. doi: 10.1016/j.learninstruc.2011.01.001

Kirschner, P. A., Kirschner, F., & Janssen, J. (2014). Collaboration principle in multimedia learning. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning (2nd Revised Edition)*. New York, NY: Cambridge University Press.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86. doi: 10.1207/s15326985ep4102_1

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Computers in Human Behavior, 19*(3), 335-353. doi: 10.1016/S0747-5632(02)00057-2

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*(3), 385-408. doi: 10.1080/00220973.2012.727884

Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*(6), 822-832.

Latterell, C. M. (2005). *Math wars*. Westport, CT: Praeger.

Laughlin, P. R., Zander, M. L., Knievel, E. M., & Tan, T. K. (2003). Groups perform

    better than the best individuals on letters-to-numbers problems: Informative

    equations and effective strategies. *Journal of Personality & Social*

    Psychology, 85(4), 684-694. doi: 10.1037/0022-3514.85.4.684

Leppink, J., Paas, F., Vleuten, C. M., Gog, T., & Merriënboer, J. G. (2013).

    Development of an instrument for measuring different types of cognitive load.

    Behavior Research Methods, 45(4), 1058-1072. doi: 10.3758/s13428-013-

    0334-1

Levine, J. M., & Moreland, R. L. (2012). A history of small group research. In A. W.

    Kruglanski & W. Stroebe (Eds.), *Handbook of the hystory of social*

    psychology (pp. 383-405). New York, NY: Psychology Press.

Mawer, R. F., & Sweller, J. (1982). Effects of subgoal density and location on

    learning during problem solving. *Journal of Experimental Psychology:*

    Learning, Memory, & Cognition, 8(3), 252-259.

Mayer, R. (1977). *Thinking and problem solving: An introduction to human cognition*

    and learning. Glenview, Illinois: Scott, Foresman and Company.

Mayer, R. (1999). *The promise of educational psychology: Learning in the content*

    areas (Vol. II). Upper Saddle River, NJ: Prentice Hall.

Mayer, R. (2001). *Multimedia learning*. Cambridge, UK: Cambridge University

    Press.

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

Miller, R. (1956). The magic number of seven plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Moreno, R. (2006). When worked examples don't work: Is cognitive load theory at an Impasse? *Learning and Instruction, 16*(2), 170-181. doi: 10.1016/j.learninstruc.2006.02.006

Moreno, R., & Mayer, R. (1999). Cognitive principles of multimedia learning: the role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358-368.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presention modes. *Journal of Educational Psychology, 87*(2), 319-334.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2011). Chapter 8: Clasroom instruction. *TIMSS 2011 International Results in Mathematics* (pp. 325-411). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction, 16*(2), 173-199. doi: 10.2307/3233720

National Mathematics Advisory Panel. (2008). *Foundations for success: Final report of the National Mathematics Advisory Panel*. Washington, DC: United States Department of Education. Retrieved from http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf.

NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

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

Oberauer, K., & Hein, L. (2012). Attention to information in working memory. *Current Directions in Psychological Science, 21*(3), 164-169. doi: 10.1177/0963721412444727

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

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

Owens, P., & Sweller, J. (2007). Cognitive load theory and music instruction. *Educational Psychology, 28*(1), 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.

Paas, F., Camp, G., & Rikers, R. (2001). Instructional compensation for age-related
  cognitive declines: Effects of goal specificity in maze learning. *Journal of
  Educational Psychology, 93*(1), 181-186. doi: 10.1037/0022-0663.93.1.181

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory:
  using the human motor system and collaboration to support the learning of
  complex cognitive tasks. *Educational Psychology Review, 24*(1), 27-45. doi:
  10.1007/s10648-011-9179-2

Paas, F., & Van Gog, T. (2006). Optimising worked example instruction: Different
  ways to increase germane cognitive load. *Learning and Instruction, 16*, 87-91.

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

Pass, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional
  implications of the interaction between information structures and cognitive
  architecture. *Instructional Science, 32*(1-2), 1-8.

Pawley, D., Ayres, P., Cooper, M., & Sweller, J. (2005). Translating words into
  equations: A cognitive load theory approach. *Educational Psychology, 25*(1),
  75-97. doi: doi: 10.1080/0144341042000294903

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal

    items. *Journal of Experimental Psychology, 58*(3), 193-198. doi:

    10.1037/h0049234

Petty, R. E., Harkins, S. G., & Williams, K. D. (1980). The effects of group diffusion

    of cognitive effort on attitudes: An information-processing view. *Journal of*

    Personality & Social Psychology, 38(1), 81-92.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information.

    Learning and Instruction, 12(1), 61-86. doi: 10.1016/S0959-4752(01)00016-0

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to

    categorize statistics word problems. *Journal of Educational Psychology,*

    88(1), 144-161.

Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the

    expertise reversal effect with a computer-based environment on electrical

    circuit analysis. *Learning and Instruction, 16*(2), 92-103. doi:

    10.1016/j.learninstruc.2006.02.008

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*(1), 45-56. doi:

    10.1002/j.2168-9830.2007.tb00914.x

Renkl, A. (1997a). *Learning by explaining or better by listening*. Paper presented at

    the Annual Meeting of the American Educational Research Association,

    Chicago, IL.

Renkl, A. (1997b). Learning from worked-out examples: A study on individual

    differences. *Cognitive Science, 21*(1), 1-29. doi:

    10.1207/s15516709cog2101_1

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning

    by self-explanations. *Learning and Instruction, 12*(5), 529-556. doi:

    10.1016/S0959-4752(01)00030-5

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to

    problem solving in cognitive skill acquisition: A cognitive load perspective.

    Educational Psychologist, 38(1), 15-22. doi: 10.1207/S15326985EP3801_3

Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to

    problem solving: Smooth transitions help learning. *The Journal of

    Experimental Education, 70*(4), 293-315. doi: 10.1080/00220970209599510

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out

    examples: The effects of example variability and elicited self-explanations.

    Contemporary Educational Psychology, 23(1), 90-108. doi:

    10.1006/ceps.1997.0959

Retnowati, E., Ayres, P., & Sweller, J. (2010). Worked example effects in individual and group work settings. *Educational Psychology, 30*(3), 349-367. doi: 10.1080/01443411003659960

Rosenshine, B. (2010). *Principles of instruction*. Belgium: The International Academy of Education (IAE).

Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognise designers' styles. *Learning and Instruction, 19*(2), 185-199. doi: 10.1016/j.learninstruc.2008.03.006

Saleh, M., Lazonder, A. W., & de Jong, T. (2007). Structuring collaboration in mixed-ability groups to promote verbal interaction, learning, and motivation of average-ability students. *Contemporary Educational Psychology, 32*(3), 314.

Schmidt, H. G., Loyens, S. M. M., Van Gog, T., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 91-97. doi: 10.1080/00461520701263350

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1-66.

Schreiber, L. M., & Valle, B. E. (2013). Social constructivist teaching strategies in
the small group classroom. *Small Group Research, 44*(4), 395-411. doi:
10.1177/1046496413488422

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval process in long term
memory. *Psychological Review, 76*(2), 179-193.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human
information processing: II. Perceptual learning, automatic attending and a
general theory. *Psychological Review, 84*(2), 127-190.

Simon, H. A., & Gilmartin, K. (1973). A simulation of memory for chess positions.
Cognitive Psychology, 5(1), 29-46. doi: 10.1016/0010-0285(73)90024-8

Sperling, G. (1960). The information available in brief visual presentations.
Psychological Monographs, 74(11, Whole No. 498), 1-29. doi:
10.1037/h0093759

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

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

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

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

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science, 32*(1-2), 9-31.

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction, 16*(2), 165-169. doi: 10.1016/j.learninstruc.2006.02.005

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

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer.

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

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load and selective attention as factors on the structuring of technical material. *Journal of Experiment Psychology: General, 119*, 176-192.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59-89.

Sweller, J., & Levine, M. (1982). Effects of goal specifity on means-ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*(8), 463-483.

Sweller, J., Mawer, R., & Ward, M. (1983). Development of expertise in
mathematical problem solving. *Journal of Experiment Psychology*(112), 634-
656.

Sweller, J., Mawer, R. F., & Howe, W. (1982). Consequences of history-cued and
means-end strategies in problem solving. *The American Journal of
Psychology, 95*(3), 455-483. doi: 10.2307/1422136

Sweller, J., & Sweller, S. (2006). Natural information processing systems.
Evolutionary Psychology, 4, 434-458.

Sweller, J., van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and
instructional design. *Educational Psychology Review, 10*(3), 251-296. doi:
10.1023/a:1022193728205

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving.
Journal of Educational Psychology, 80(4), 424-436.

Tindale, R. S., Smith, C. M., Dykema-Engblade, A., & Kluwe, K. (2012). Good and
bad group performance: Same process--different outcomes. *Group Processes
Intergroup Relations, 15*, 603-618. doi: 10.1177/1368430212454928

Trafton, J. G., & Reiser, B. J. (1993). *The contributions of studying examples and
solving problems to skill acquisition.* Paper presented at the Proceedings of
the fifteenth annual conference of the cognitive science society, Hillsdale, NJ.

Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*(2), 334-341. doi: 10.1037/0022-0663.91.2.334

Van den Bossche, P., Gijselaers, W. H., Segers, M., & Kirschner, P. A. (2006). Social and cognitive factors driving teamwork in collaborative learning environments: Team learning beliefs and behaviors. *Small Group Research, 37*(5), 490-521. doi: 10.1177/1046496406292938

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

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

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

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

van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2008). Effects of studying

    sequences of process-oriented and product-oriented worked examples on

    troubleshooting transfer efficiency. *Learning and Instruction, 18*(3), 211-222.

    doi: 10.1016/j.learninstruc.2007.03.003

van Merriënboer, J. J. G. (1990). Strategies for programming instruction in high

    school: Program completion vs. program generation. *Journal of Educational*

    Computing Research, 6(3), 265-285. doi: 10.2190/4NK5-17L7-TWQV-1EHL

Vollmeyer, R., Burns, B., & Holyoak, K. (1996). The impact of goal specificity on

    strategy use and the acquisition of problem structure. *Cognitive Science*(20),

    75-100. doi: 10.1016/S0364-0213(99)80003-2

von Wright, J. M. (1972). On the problem of selection in iconic memory.

    Scandinavian Journal of Psychology, 13(1), 159-171. doi: 10.1111/j.1467-

    9450.1972.tb00064.x

Wadsworth, B. J. (1978). *Piaget for the classroom teacher*. New York and London:

    Longman.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition*

    and Instruction, 7(1), 1-39. doi: 10.1207/s1532690xci0701_1

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72*,

    89-104.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22*(5), 366-389.

Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology, 79*(1), 1-28. doi: 10.1348/000709908x380772

Webb, N. M., & Mastergeorge, A. (2003). Promoting effective helping behavior in peer-directed group. *International Journal of Educational Research, 39*, 73-97. doi: 10.2307/3233805

Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concept and assessment. *Learning and Instruction, 17*, 416-426. doi: 10.1016/j.learninstruc.2007.03.007

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1160-1175.

Wirth, J., Künsting, J., & Leutner, D. (2009). The impact of goal specificity and goal type on learning outcome and cognitive load. *Computers in Human Behavior, 25*(2), 299-305. doi: 10.1016/j.chb.2008.12.004

Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist, 43*(1), 49-64. doi: 10.1080/00461520701756420

Yeung, A. S., Jin, P., & Sweller, J. (1997). Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology, 23*(1), 1-21. doi: 10.1006/ceps.1997.0951

Youssef, A., Ayres, P., & Sweller, J. (2012). Using general problem-solving strategies to generate ideas in order to solve geography problems. *Applied Cognitive Psychology, 26*(6), 872-877. doi: 10.1002/acp.2888

Zhang, L., Ayres, P., & Chan, K. (2011). Examining different types of collaborative learning in a complex computer-based environment: A cognitive load approach. *Computers in Human Behavior, 27*(1), 94-98. doi: 10.1016/j.chb.2010.03.038

Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction, 4*(3), 137-166. doi: 10.1207/s1532690xci0403_1