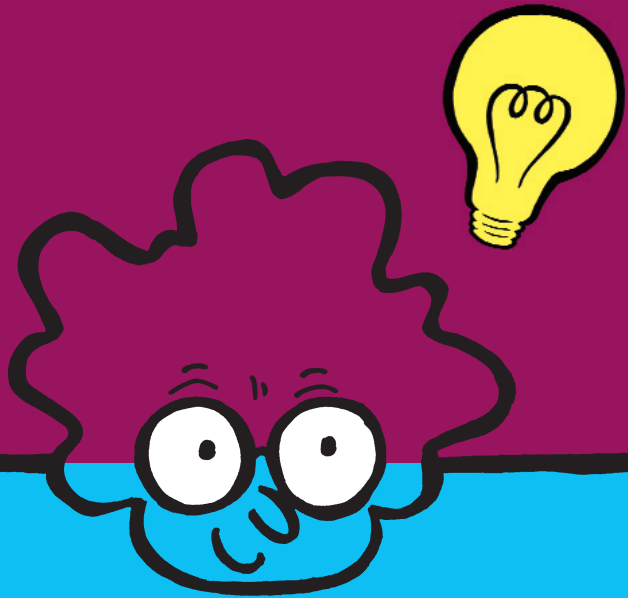


Efficient Complex Skill Training into Old Age

Exploring the Benefits of Cognitive Load Theory



Pascal van Gerven

Efficient Complex Skill Training into Old Age

Exploring the Benefits of Cognitive Load Theory

Efficient Complex Skill Training into Old Age

Exploring the Benefits of Cognitive Load Theory

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus,
prof. dr. A.C. Nieuwenhuijzen Kruseman,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op vrijdag 21 juni 2002 om 16.00 uur

door

Pascal van Gerven

Promotores

Prof. dr. H.G. Schmidt

Prof. dr. J.J.G. van Merriënboer (Open Universiteit Nederland)

Co-promotor

Dr. F. Paas (Open Universiteit Nederland)

Beoordelingscommissie

Prof. dr. P.A. Kirschner (voorzitter)

Prof. dr. J. Jolles

Dr. R.M.J.P. Rikers (Erasmus Universiteit Rotterdam)

Prof. dr. J. Sweller (University of New South Wales, Sydney, Australia)

Prof. dr. J. Vermunt

Colophon

Cover design: © Fenne Roefs, Breda

Production: Datawyse | Maastricht University Press

© Pascal van Gerven, Rotterdam 2002

ISBN 90-9015834-0

The studies presented in this dissertation were funded by the Netherlands Organization for Scientific Research (NWO, The Hague, project no. 575-22-006). They were conducted at Maastricht University and at the Open University of The Netherlands in Heerlen. The conference presentations emerging from this project were financed by grants provided by NWO and the Limburg University Fund (SWOL).

Because we don't know when we will die, we get to think of life as an inexhaustible well. Yet everything happens only a certain number of times, and a very small number, really. How many more times will you remember a certain afternoon of your childhood, some afternoon that's so deeply a part of your being that you can't even conceive of your life without it? Perhaps four or five times more, perhaps not even that. How many more times will you watch the full moon rise? Perhaps twenty. And yet it all seems limitless.

Paul Bowles — *The Sheltering Sky*

Contents

<i>Acknowledgements</i>	9
<i>Preface</i>	11
1 — Cognitive Load Theory and Aging: An Introduction	15
2 — Worked Examples versus Conventional Problems	31
3 — Multimedia-Aided Cognitive Skill Training	49
4 — Cognitive Load and the Pupillary Response: An Intermezzo	63
5 — Training Variability and the Modality Effect	77
6 — Concluding Remarks and Future Research	95
<i>References</i>	103
<i>Summary</i>	111
<i>Samenvatting</i>	119
<i>Curriculum Vitae</i>	127

Acknowledgements

Reliable scientific research has shown that young people, who strongly activate the stereotype of an elderly person, tend to behave accordingly.¹ If this were true for forgetfulness, one could think that, after four years of doing research with elderly participants, I would definitely have trouble remembering the names of all the people I am indebted to. However, since the stereotype of old age in the negative sense was hardly confirmed during my project, I am quite convinced that I will not forget a single soul.

First of all, I am deeply indebted to my supervisors: Fred Paas for his mostly relaxing, but sometimes perplexing soberness and clarity, Jeroen van Merriënboer for his thoroughness and perseverance, and Henk Schmidt for his thumbs-up. Furthermore, I owe a lot to my first colleagues in Maastricht, who instantly created a joyful climate for the new recruit. I especially thank Remy Rikers for his mentorship and incomparable humor, Marianne van den Hurk for her warmth, and Wilco te Winkel for his earth-shaking presence. Furthermore, I thank my former roommates, Jettie Hoonhout and Nicole van Kesteren, as well as my present roommate, Anique de Bruin, for their very pleasant company.

It needs little explanation that the realization of this dissertation was no one-man performance. I am grateful to my assistants, Maaïke Hendriks and Marijn Bruinink, for running two of the experiments. I also thank Marcel Hendrickx, Johan Gielissen, and Ron Hellenbrand for combining technical support with great kindness. I thank Marieke Werrij and Saskia Marcano-Bakker for their cooperation in recording the auditory material. Furthermore, I thank John Sweller and Paul Chandler for their enthusiastic comments on the introductory chapter. Fren Smulders and Gerard van Breukelen are acknowledged for their sometimes puzzling, but mostly helpful methodological advices. Pieter Unema and Jukka Hyönä are acknowledged for their advices on pupillometric measurement. I am grateful to Astrid Quist, Jelle Jolles, and Peter Houx of the Maastricht Aging Study (MAAS) for providing a great deal of the elderly participants, and Paul Kirschner and the Sintermeertencollege in Heerlen for providing part of the young participants. I thank Marja Hortulanus and Ellen Blaauw for helping me to take all kinds of practical and financial obstacles. Warm thanks go to Fenne Roefs for her sparkling ideas in designing the cover of this thesis.

¹ Delightful reports are provided by Bargh, Chen, and Burrows (1996) and Dijksterhuis, Aarts, Bargh, and Van Knippenberg (2000).

I would like to express special gratitude to my colleagues and friends in Rotterdam, Maastricht, and Heerlen, who made my move to Rotterdam one of mixed feelings. Special thanks also go to my parents, family, and friends for their unlimited support and patience. Moreover, I am indebted to my sister, Peggy van Gerven, and precious friend, Maarten Bernards, for standing by me.

Above all, however, I thank Anne. Simply for being there.

Rotterdam, April 2002

Preface

The strong increase of the aging population in the Western world has long ago passed the stage of novelty. As we know by now, a relatively small population growth combined with a still increasing life expectancy due to advanced medical care is expected to boost the top segment of what once was a population pyramid for the next fifty years. This demographic transformation is affecting both the working and the non-working population. A typical European illustration in this respect is the Netherlands. According to the Social and Cultural Planning Office¹ — a governmental organization conducting scientific research in diverse social areas — the Dutch working population incorporates a fast growing segment of people between 50 and 65 years of age (De Klerk, 2001). With professional environments developing at an accelerating pace, the need for retraining is increasing just as well. Indeed, the percentage of older employees that participates in the broad spectrum of professional training is growing steadily. However, the greatest increase in the number of older students takes place in the non-professional realm. Obviously, this is a matter of motivation. That is, the drive of older people to take courses is not so much the acquirement of a certificate, but rather a combination of leisure time and pure interest.

Whether elderly people participate in education from a professional or from a non-professional background is not the issue here, however. The starting-point of this thesis is the notion that the elderly show a growing interest in all sorts of education. Although experience and wisdom accumulate during the lifespan, it is commonly known that learning and memory show a substantial degree of decline with old age. This has profound implications for the manner in which education should be extended. More precisely, elderly learners have their own needs with respect to the subject matter and — more importantly — the pace at, and the format in which the information is presented to them. This is especially the case in the more complex subject domains, such as information and communication technology (ICT), mathematics, and science. In this thesis, we are particularly interested in the acquisition of skills within such complex domains. It is assumed that learning complex skills puts high demands on the cognitive system, specifically on working memory. Furthermore, it is believed that working-memory capacity is limited, especially in elderly people. Thus, acquiring complex skills is a delicate matter, the success of which

¹ Sociaal en Cultureel Planbureau (SCP), Den Haag.

largely depends on the level of cognitive load,² which is imposed on the learner (Paas & Van Merriënboer, 1994a).

The studies described in the forthcoming chapters seek to apply *cognitive load theory* (Sweller, Van Merriënboer, & Paas, 1998) to the domain of educational gerontology. Cognitive load theory provides a theoretical basis for the design of instructional formats that stimulate an optimal use of the available cognitive resources while engaged in learning complex subject matter. The central idea of this thesis is that cognitive load theory can be particularly helpful for elderly learners, because their limited cognitive resources should be handled with extra care. To explore this idea, cognitive load theory will be linked with modern theories on cognitive aging (e.g., Hasher & Zacks, 1988; Salthouse, 1996). A sketch of a synthesizing framework will be presented in Chapter 1. Subsequently, specific aspects of this framework will be tested in Chapters 2, 3, and 5 (Chapter 4 is a digression in this respect).

As we will see, the crucial part of an instructional format is the training phase. In the training phase the learner is confronted with a series of practice problems. Practice problems lead to the kind of knowledge that enables a person to solve test problems he or she has never encountered before. In this respect, the learner's proficiency is determined by the depth at which the training problems are processed. In its turn, depth of processing is determined by the quality, variability, and the format in which the practice problems are presented. As will become clear, these training characteristics can be manipulated to arrive at an optimal learning result.

In Chapter 2, the use of worked examples in training a complex problem-solving skill will be tested against practicing with conventional problems. Conventional problems, which require learners to solve a series of training problems by themselves, are assumed to impose an excessive cognitive load on the learner. A large part of this excessive load is caused by a high-demanding problem-solving strategy, namely means-ends analysis. Means-ends analysis entails a great deal of extraneous cognitive operations, which do not contribute to the learning process (Sweller, 1988). In principle, means-ends analysis is a very effective strategy for solving novel problems. This is only true, however, if sufficient cognitive resources are available. This is the case in experts, who have already acquired a rich knowledge base that helps them to process new information such that their working memory does not get overloaded. Novices, on the other hand, who do not yet possess such a rich knowledge base, are rather susceptible to working-memory overload, especially when they have entered old age. Worked examples, on the other hand, entail little extraneous processes and focus the learner's attention on problem states and operations that are

² In the literature the term *mental workload* seems to prevail (for an overview, see Nygren, 1991). For the sake of uniformity within this thesis, however, we will stick to the term *cognitive load*.

needed to proceed towards a solution. It is hypothesized that worked examples lead to better results than conventional problems, both in terms of experienced cognitive load in the training phase and performance on a test. Following the framework presented in Chapter 1, this effect is expected to be stronger for the elderly than for the young.

In Chapter 3, the quality of worked examples will be addressed. More precisely, we will focus on the surplus value of presenting worked examples in a multimedia-based format. In such a format, part of the information is presented in a visual and the other part in an auditory form. There is a lot to say in favor of multimedia-based worked examples. One of the greatest advantages is that they prevent unnecessary visual search, which is a cognitively high-demanding. In a purely visual format, visual search is inevitable, because sources of information that refer to each other, such as diagrams and associated texts, are physically separated. If, however, textual information is presented auditorily, learners can concentrate on the diagrams and visual search is reduced to a minimum. An additional benefit of multimedia-based worked examples is that they address both the visual and the auditory component of working memory, which can prevent one of these components to become overloaded (this idea is specifically explored in Chapter 5). It is hypothesized that multimedia-based worked examples lead to better results than purely visual worked examples. Again, this effect is expected to be stronger for elderly learners than for their young counterparts.

Chapter 4 can be considered as an intermezzo, because it does not focus on the application of cognitive load theory. It does relate to the other chapters, however, in that it tests an alternative measure of cognitive load. Where in Chapters 2, 3, and 5 cognitive load is measured by means of a 9-point scale on which the participants rate their perceived cognitive load, Chapter 4 explores the possibilities of measuring pupil size as a more objective correlate of cognitive load. As we will see, early research has shown that the pupil does not only respond to variations in light intensity, but also to the intensity of cognitive processes, albeit far more subtly. Although pupil dilation has revealed its quality as one of the most sensitive physiological measures of cognitive load (e.g., Ahern & Beatty, 1979), there is reason to suspect that it is less appropriate for measuring cognitive load in elderly people. This is the central issue to be investigated in Chapter 4. For this purpose, pupil size will be measured while participants are engaged in a memory-search task (Sternberg, 1966).

In Chapter 5, the advantage of audiovisual worked examples, which in the literature is referred to as the modality effect, will be further explored. In addition, the role of germane cognitive load will be investigated. Germane cognitive load is the portion of instructional load that has a positive effect on the learning process (Sweller et al., 1998). A type of germane load concerns the depth of processing, which was mentioned earlier in this pref-

preface. Depth of processing can be manipulated by altering the variability of the training problems (e.g., Paas & Van Merriënboer, 1994b). A highly variable sequence of training problems will prevent learners from developing routine in processing these problems. As a consequence, they have to reanalyze every single problem in order to comprehend it. This stimulates deep processing and thus enhances the learning process. It is hypothesized that audiovisual worked examples that are presented in a high-variable sequence lead to the best results. Yet again, this will be especially true for elderly learners.

Finally, in Chapter 6, general conclusions will be drawn from the first five chapters. In addition, an extensive set of possible future research directions will be given. For a significant part, these will be based on a critical evaluation of cognitive load theory.

As may have become clear, the order of the chapters is a reflection of the chronological order in which the reported studies were performed. In fact, all chapters (except for Chapter 6) can be read in isolation, since they are either published or submitted as separate papers. They all contain a general introduction. Thus, there may be some overlap with respect to their contents. The chapters refer to one another through common references as well as chapter numbers.

As the author of this thesis, I do not pretend to have given a final answer to the question whether cognitive load theory is an appropriate guideline for the design of instructions that are aimed at elderly learners. Yet, I do believe it is. As I will point out in the final chapter, however, care should be taken concerning the experimental method that is used to infer its underlying cognitive mechanism. I hope that this thesis, as well as the papers it has yielded, will initiate a line of research, in which the potential power of cognitive load theory for elderly learners will eventually show to full advantage.

1 Cognitive Load Theory and Aging: An Introduction¹

The purpose of this introductory chapter is to explore the advantages of instructional formats based on cognitive load theory for elderly learners engaged in the acquisition of complex cognitive skills. A great body of research has demonstrated that cognitive aging is accompanied by a reduction of working-memory capacity, a general slowing of mental processes, and a decline of the ability to repress irrelevant information. The core idea of cognitive load theory is that working-memory capacity is limited and should therefore be managed with great care and discretion. Cognitive load theory claims that this can be achieved by minimizing the level of extraneous cognitive load, which is the portion of load that does not contribute to schema acquisition, and maximizing the level of germane cognitive load, which directly contributes to the construction of cognitive schemata. Since instructions based on cognitive load theory deal with cognitive limitations in that they lead to an efficient use of the available resources, it was hypothesized that they are especially effective when elderly people are involved. This idea was analyzed by means of a framework merging cognitive load theory with the aforementioned research findings concerning cognitive aging. It was concluded that cognitive load theory, enabling elderly people to acquire new complex skills, can be regarded as an essential guide for educational gerontology.

It has become a platitude to proclaim that our society is developing at an incredible speed, that its complexity is becoming enormous, and that we are daily overloaded with information. In the same light, it has become a truism that old people are more affected by this immensity than young people. Where, in the worst case, young people seem to be suffering from growing information quantity and complexity, old people are likely to be totally overwhelmed by it. Not surprisingly, a frequently uttered complaint of elderly people being confronted with, for instance, computer-based work (Morris, 1994), is that they are overburdened with numerous options, inadequately presented information, and jargon, which not only leads to poor performance, but also leaves them with a negative attitude towards the subject matter. Fortunately, this observation is broadly recognized as a problem, not least because the elderly represent a relatively large and still growing part of advanced industrial populations.

¹ A version of this chapter was published as: Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., & Schmidt, H.G. (2000). Cognitive load theory and the acquisition of complex cognitive skills in the elderly: Towards an integrative framework. *Educational Gerontology*, 26, 503-521.

Cognitive overload frequently plays a role in education. Research into some aspects of cognitive processing and instructional design has provided solutions to this problem. Much of this research seizes upon the issue of what can be done to manage cognitive load in such a way that people achieve optimal learning and performance. In this chapter we present a framework that integrates research findings in the field of cognitive load and cognitive aging. This framework is based on the assumption that using knowledge of factors determining cognitive load can lead to more effective instructions, and thus to enhanced performance of the learner. One of the theories that focuses on cognitive load and provides important guidelines for improving instructional material is Sweller's (1988, 1989, 1994) cognitive load theory (CLT). The question raised in this chapter is whether CLT-based modifications of conventional instructional material, having been extensively tested on relatively young people, are also effective with elderly people in a problem-solving context. In fact, we claim that the assertions of CLT perfectly match the cognitive declines associated with aging. Therefore, we hypothesize that elderly people gain relatively more from CLT-based instructions than young people (Van Gerven, Paas, & Van Merriënboer, 1998). Such instructions should enable older people to acquire skills that are otherwise far too intricate for them to grasp. Regarding the growing interest in life-long learning, this can have a great impact on educational gerontology.

In working towards a framework of cognitive load and cognitive aging, we first shed some more light on the concept of cognitive load, its causal factors, and its relation to learning. Subsequently, we discuss the cognitive changes that are associated with aging. Next, we give a brief description of cognitive load theory and its implications for instructional design. Guided by the above-mentioned framework, we finally explain the implications of cognitive load theory for elderly learners.

Cognitive Load and Learning

A comprehensive overview of factors determining the level of cognitive load is offered by Paas and Van Merriënboer (1994a; see Figure 1.1). They distinguish causal and assessment factors. *Causal factors* include characteristics of the person (e.g., age and cognitive abilities), the task (e.g., task complexity and time pressure), the environment (e.g., noise and temperature), and their mutual relations.

With respect to *assessment factors*, Paas and Van Merriënboer distinguish mental load, mental effort, and performance as the three measurable dimensions of cognitive load. *Mental load* is regarded as the portion of cognitive load that is imposed exclusively by the task and environmental demands. The *mental effort* dimension refers to the amount of cognitive

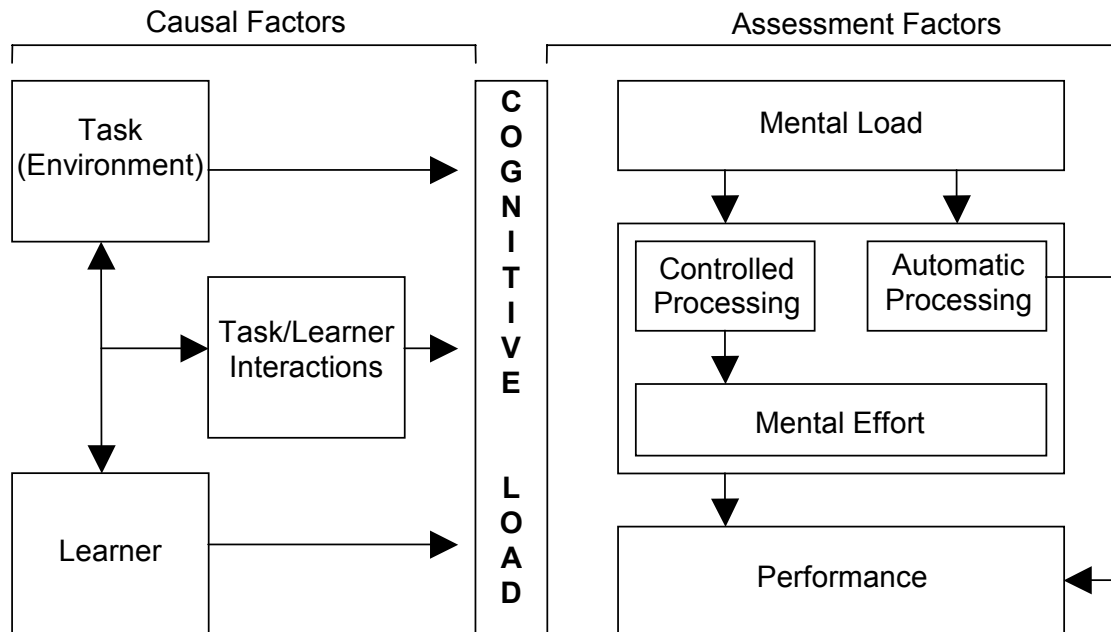


Figure 1.1 Cognitive load: causal and assessment factors (Paas & Van Merriënboer, 1994a).

capacity that is actually allocated to the fulfillment of the task. It pertains to the part of processing that is under mental control (as opposed to automatic processing). A person's *performance*, finally, is a reflection of mental load, mental effort, and the aforementioned causal factors.

Learning, which is reflected by performance change, requires working-memory capacity. That is, it imposes a *germane cognitive load* on the learner (Sweller, Van Merriënboer, & Paas, 1998). Germane cognitive load is required for an essential aspect of learning, namely the construction and storage of so-called *schemata* into long-term memory. Schemata evolve when information elements and production rules are integrated and chunked into wholes. This process saves storage capacity and optimizes cognitive functioning. Sweller (1994, p. 296) defines a schema as a “cognitive construct that organizes the elements of information according to the manner with which they will be dealt.” Following this definition, several types of schemata can be distinguished that relate to several types of situations. Schemata dealing with solving problems play a major role in this thesis. According to Sweller (1994), a ‘problem-solving schema’ is a flexible instrument that captures a problem into a specific category and provides the *rules* for solving it.

An important feature of schemata is that they enable *transfer*. Transfer is the application of existing schemata to new problems, which can be tackled similarly to related, earlier encountered problems. Transfer is essential for the performance of complex tasks, in which the encountered problems often deviate from the standard. Another essential feature of schemata is *automation*. Schemata that are automated do not require controlled proc-

essing anymore (see Figure 1.1). As a result, even more than the original schemata, automated schemata reduce mental effort and thus cognitive load (Van Merriënboer & Paas, 1990).

The construction of adequate and rich schemata is especially important in complex learning tasks. Complex learning tasks require more effort, because the elements contained by the to-be-learned material are highly interconnected (Sweller, 1994). In this respect, Sweller refers to *intrinsic cognitive load*, which is the portion of load that is imposed by the intrinsic characteristics of the task or subject matter. Tasks imposing a high intrinsic load often strain the cognitive system to its limits. In the next section, we will see that this is especially the case when the cognitive system has declined as a result of aging.

Characteristics of Cognitive Aging

Cognitive aging is accompanied by important cognitive changes. These concern both crystallized and fluid abilities (e.g., Horn & Cattell, 1967). Where crystallized abilities enhance and accumulate as a result of aging, fluid abilities tend to decline. This decline is threefold: (1) working-memory capacity decreases, (2) processing speed goes down, and (3) there is a reduced ability to distinguish relevant from irrelevant information. In the following subsections we will elaborate on these declines.

Reduced Working-Memory Capacity

Decrease of working-memory capacity seems to be the most obvious cognitive change associated with aging. The concept of *working memory* categorically differs from the somewhat obsolete expression “short-term memory.” That is, the term is not simply defined as the ability to temporarily store a certain number of information elements, but rather as the capacity “to preserve information while simultaneously processing the same or other information” (Salthouse & Babcock, 1991, p. 763). Whereas simple short-term storage capacity barely declines with aging, working memory does. Wingfield, Stine, Lahar, and Aberdeen (1988), for instance, found no difference between young and old people in a digit-span test and only a small difference in a word-span test. In a loaded-word span test, however, in which people had to make a false/true judgment about and recall the last words of progressively longer sets of statements, the elderly were substantially affected. Salthouse and Babcock (1991) obtained similar results. Where they found a relatively weak relation between age and digit/word-span, they found a gradual age-related decline in listening and computation span.

Reduced working-memory capacity becomes critical when the task demands are high and a relatively large working-memory capacity is needed.

A considerable body of research illustrates the importance of working-memory capacity for problem-solving skill. A study by Gilinsky and Judd (1994), for instance, demonstrates the role of reduced working-memory capacity in solving syllogistic problems. The difficulty of solving syllogistic problems depends on the number of 'solution models' that has to be evaluated in working memory. Old participants were more affected than their young counterparts when this number was high. Similar relations between age and task complexity were found in dual-task performance (Lorsbach & Simpson, 1988), computer-based work (Czaja & Sharit, 1993), and prospective tasks (Einstein, Smith, McDaniel, & Shaw, 1997; Kidder, Park, Hertzog, & Morrell, 1997).

Reduced Processing Speed

A second cognitive change associated with aging is a reduction of processing speed. This can be considered as an additional decline, but it can also serve as an alternative explanation for the age-related reduction of working-memory capacity. Salthouse and Babcock (1991), for instance, explain the phenomenon as a decrease in the rate at which information is activated. The number of information elements that can be processed equals the number of elements that can be activated before the activation of the first element has decayed below a threshold. Hence, the lower the activation speed, the less elements can be processed. This view was the precursor of Salthouse's (1996) *processing-speed theory*. In this theory, Salthouse postulates two mechanisms underlying an age-related decline of cognitive performance. First, there is the limited-time mechanism, which is based on the assumption that the time occupied by early processes puts restrictions on the time available for later processes. Second, Salthouse posits the simultaneity mechanism, holding that products of early processes can be lost by the time later processes are executed. That is, products of different processing stages, which have to be combined at another stage of processing, are not always available simultaneously. Most importantly, Salthouse hypothesizes that general cognitive slowing, rather than a specific deficit, such as a reduced working-memory capacity, underlies age-related declines of cognitive functioning. The general nature of cognitive slowing implies that there are high correlations between different speed measures. Thus, a measure of perceptual speed, for instance, should be a reliable predictor of motor speed and vice versa. Furthermore, speed measures should highly correlate with measures of working memory. Finally, the theory predicts that statistical control of cognitive speed will attenuate an age-related difference of working-memory capacity. Such an attenuation was found repeatedly (e.g., Fisk & Warr, 1996; Salthouse, 1993; Salthouse & Babcock, 1991).

Myerson, Hale, Wagstaff, Poon, and Smith (1990) proposed a model that is consistent with Salthouse's simultaneity mechanism. In their *infor-*

information-loss model they state that information processing occurs in discrete steps and that a person's response latency equals the sum of the step latencies. The duration of a step depends on the amount of information that is available at the onset of that step (cf. Salthouse, 1996). Unlike processing-speed theory, the information-loss model explicitly incorporates task complexity. That is, the more complex the task, the more processing steps are supposed to be involved (though Myerson et al. admit that this is an oversimplification). More processing steps will lead to proportionally more information loss and thus to lower performance. The number of processing steps required for a particular task does not depend on age. What increases with age is the duration per step as well as the proportion of information lost per step. As a consequence, elderly learners display a larger performance slide (i.e., longer latencies and a lower decision accuracy) than young people, when confronted with complex tasks.

Reduced Ability to Distinguish Relevant from Irrelevant Information

A third cognitive change emerging from old age is a decline of the ability to discriminate between relevant and irrelevant information. This phenomenon is generally ascribed to increased neural noise (e.g., Welford, 1985). This view is supported by studies showing that elderly people are poor performers on visual search tasks in which targets are surrounded by non-target objects (Allen, 1990; Allen, Madden, Groth, & Crozier, 1992; Madden, Connelly, & Pierce, 1994).

A view that may be in line with the neural-noise approach is the idea of reduced inhibition (Hartman & Hasher, 1991; Hasher & Zacks, 1988). In this view, elderly people are assumed to have difficulty inhibiting irrelevant information in favor of relevant information. Again, this approach can be seen as an alternative for the reduced-capacity view. That is, the problem may not be that the potential of working memory has decreased, but rather that working memory tends to get overloaded with irrelevant information.

Cognitive Load Theory and Instructional Design

The above studies are all concerned with task performance. We are interested in a special task, namely the processing of instructions, which should lead to the acquisition of new and complex problem-solving skills that require transfer. The extent to which new skills are learned is determined by the effectiveness of the instructions. In the light of the aforementioned aging studies, this is particularly true when elderly learners are involved. Instructions aimed at elderly learners should be designed in such a way that their cognitive limitations are taken into account as much as possible. For this purpose, *cognitive load theory* (Sweller, 1988, 1989, 1994) provides

some important guidelines. In cognitive load theory human memory is assumed to consist of a limited working memory connected to an unlimited long-term memory. It is further assumed that working memory is not a single entity, but that it embodies at least two mode-specific components: the *visuospatial sketchpad* and the *phonological loop*. Furthermore, it incorporates a coordinative component: the *central executive* (e.g., Baddeley, 1992). As we will see, this architecture can play a crucial role in managing cognitive load.

Basically, cognitive load theory is concerned with the limitations of working-memory capacity. Sweller and others (e.g., Paas, 1992; Paas & Van Merriënboer, 1994a, 1994b) state that, especially in conventional instructional design, the strict boundaries of working memory are rarely taken into account. Conventional instructions tend to impose an *extraneous cognitive load* on working memory and ignore the simple fact that to actually learn something from the instructions an additional germane cognitive load, and thus additional working-memory capacity is indispensable. Thus, the limitation of working memory should impel instructional designers to develop training materials that impose a minimal extraneous load on working memory in order to reserve capacity for processes that have first priority, namely schema acquisition and automation. Therefore, a second important goal of instructions should be to stimulate the learner to actually dedicate the capacity gained to the schema-acquisition process.

The aim of conventional instructions is usually twofold. First, the learner is confronted with general information about the subject matter presented as text and, where applicable, figures or diagrams. Second, a great deal of conventional instructions, especially those aimed at problem solving (e.g., in mathematics and physics), incorporate a practice session, the purpose of which is to improve performance and to provide feedback. According to Sweller and collaborators, it is exactly this practice session, which, at least in the initial phase of skill acquisition, is ineffective. Paradoxically, this ineffectiveness is especially evident when clear goals are specified. Goal-specific problems are usually solved by applying a so-called *means-ends strategy*. Besides forward reasoning, a means-ends strategy typically entails backward reasoning. That is, a goal state is set, which can be attained by solving all underlying subgoals in a backward order. This way, a student is constantly trying to find the means needed to attain a subgoal, while keeping in mind the higher goals in the hierarchy, including the main goal. The problem is that a means-ends strategy puts a relatively heavy burden on working memory, while not contributing to the construction of schemata (Sweller & Levine, 1982). The more complex the practice problems, the more working-memory capacity is demanded by a means-ends strategy, and the less capacity is available for schema-acquisition. Apparently, practice does not make perfect in this context.

But what alternatives do we have? The answer to this question lies in the notion that instructions and training should be primarily aimed at the acquisition of cognitive schemata (Sweller, 1989). Once schemata are acquired and automation has taken place, problems are categorized and solved more rapidly and with less effort. Several CLT-based alternatives for conventional instructions have been proposed and evidence has been provided that these alternatives are superior. What CLT-based instructions have in common is that they restrict the amount of extraneous cognitive load to a minimum and, at the same time, stimulate the learner to devote the available capacity to the construction of schemata. The most important CLT-based instructional formats are discussed in the following section (for an exhaustive overview, see Sweller, Van Merriënboer, & Paas, 1998).

Important CLT-Based Instructional Formats

Goal-Free Problems

Contrary to conventional goal-specific problems, goal-free problems do not allow learners to employ means-ends strategies (Sweller, 1989). The instruction “calculate the value of variable X ,” for instance, can be reformulated as “calculate the value of as many variables as you can.” The specific goal is thus altered into a nonspecific goal. Instead of reasoning backward, the learner is now likely to focus on forward reasoning, so that there is no need to keep the main goal and its subgoals in mind. This saves valuable working-memory capacity, which can be used for the acquisition of schemata. Moreover, goal-free reasoning leads to a more exhaustive exploration of the subject domain and, therefore, to the construction of extensive schemata which integrate more interconnected elements. Such schemata enable the solution of a broader range of problems and increase the probability of achieving transfer. Instructional formats containing goal-free problems have been proved to result in better performance than instructions containing goal-specific problems (Owen & Sweller, 1985; Sweller & Levine, 1982).

Worked Examples

Intuitively, practicing seems the ultimate means of getting insight into complex material. After all, by solving practice problems, the student is actively processing the subject matter. Moreover, the satisfaction of working towards and finding a solution can be an important motivational factor. Nevertheless, a number of studies have shown that a more effective alternative for problem solving is studying worked examples (Paas, 1992; Paas & Van Merriënboer, 1994b; Sweller & Cooper, 1985; Ward & Sweller, 1990). Unlike conventional practice problems, worked examples, being

problems accompanied by their worked-out solutions, do not compel students to apply a capacity-demanding means-ends strategy. Instead, they can use their whole cognitive capacity to achieve a full comprehension of the subject matter.

Avoiding Split Attention

In general, instructions should be presented in such a way that the learner is not obliged to first convert them into a comprehensible form. One of the weaknesses of conventional instructions is that they often comprise mutually referring sources of information, which force learners to split their attention. An example is a diagram accompanied by a caption. Because, generally, neither of these two information sources can be understood on their own, they have to be mentally integrated. According to Sweller (1989), this integration process imposes a high extraneous load. Sweller argues that one of the means to reduce this unnecessary mental load is to physically merge the two information sources into a whole. The physical integration of interdependent information sources has been extensively tested against the traditional, non-integrated format and found to be superior (Bobis, Sweller, & Cooper, 1993; Chandler & Sweller, 1992; Sweller & Chandler, 1994; Sweller, Chandler, Tierney, & Cooper, 1990).

Distributing Information over Different Modalities

As we have seen, cognitive load theory assumes that working memory comprises at least two independent, modal channels: a visuospatial sketchpad and a phonological loop (e.g., Baddeley, 1992). This view yields another tool for avoiding split-attention. That is, if we assume that the visuospatial sketchpad and the phonological loop operate independently and in parallel, two sources of mutually referring information can be presented simultaneously via the visual and the auditory channel. In this dual-coding approach (e.g., Clark & Paivio, 1992), mental working-memory capacity is spared, so that more resources are available for schema acquisition. Dual coding was found to lead to better comprehension in different domains (Mayer & Moreno, 1998; Mousavi, Low, & Sweller, 1995; Tindall-Ford, Chandler, & Sweller, 1997; see Penney, 1989, for an extensive overview of modality effects).

Leaving Out Redundant Information

It is not always necessary to mentally integrate mutually referring information sources. In the case where two different sources of information can be understood on their own, one of them is often redundant. The usual assumption with respect to different instances of similar information in, for instance, computer manuals, is that in the worst case redundant information is simply ignored by the learner. It has been demonstrated, however, that discarding redundant information from instructional material has

beneficial effects on comprehension and performance (Bobis et al., 1993; Chandler & Sweller, 1992).

Towards an Integrative Framework

How can we now reformulate the above findings concerning cognitive aging and cognitive load theory into an explanatory and predictive framework? We propose a framework, which on the one hand incorporates and explains cognitive load theory, and on the other hand predicts the effects of CLT-based instructions under conditions of high task complexity and reduced cognitive abilities. A schematic representation of the framework is depicted in Figure 1.2. The left side of the figure contains three independent bipolar variables: instructions, age, and task complexity. These variables establish the basic task and learner characteristics. Instructions are either “conventional” or “CLT-based,” age is either “young” or “old,” and task complexity is either “low” or “high.” The boxes in the figure contain dependent variables. The arrows represent causal relations. Minus signs indicate negative connections; plus signs indicate positive connections.

The variable instructions, being either “conventional” or “CLT-based,” in fact stands for the efficiency of the instructional format. How is this instructional efficiency reflected by other variables in the framework? An approach is to view efficiency as the ratio between extraneous and germane cognitive load (CL) imposed on the learner. That is, the more efficient the instructions, the lower the level of extraneous CL and the higher the level of germane CL. Paas and Van Merriënboer (1993) determine instructional efficiency by representing an instructional format as a point in

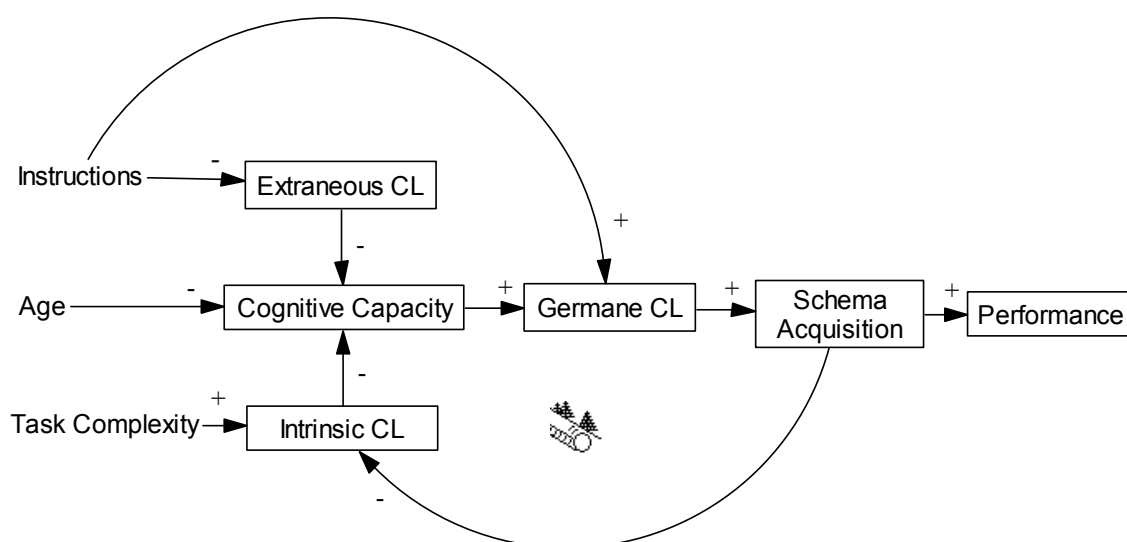


Figure 1.2 An integrative framework of cognitive load theory and cognitive aging.

a diagram (see, e.g., Figures 2.4 and 3.2 in the forthcoming chapters), the position of which is determined by the z scores of the dimensions mental effort (M , on the x-axis) and performance (P , on the y-axis). Efficiency (E) is now defined as the perpendicular distance from this point to the line $P = M$, which can be expressed as:

$$E = \frac{P - M}{\sqrt{2}}.$$

If performance P equals mental effort M , then efficiency E equals zero, which is to be interpreted as a neutral score. If $P > M$, then the instructions are efficient in that they apparently lead to a level of schema acquisition that minimizes the amount of mental effort needed for the task and at the same time maximizes the performance on that task. If $P < M$, however, then the performance is lower than can be expected on the basis of the invested amount of mental effort. In this case, the instructions are not efficient, because obviously they do not lead to a level of schema acquisition that is at least sufficient for an average performance. Mental effort is not incorporated in the present framework, but is reflected by the level of germane CL that is imposed on the available cognitive capacity.

Besides motivation (which is not included in the framework), the level of germane CL depends on the quality of the instructions. The variable instructions has a positive connection with germane CL and a negative connection with extraneous CL. If, in the case of conventional instructions, the efficiency is low, then extraneous CL is high and germane CL has a moderate level. If, however, in the case of CLT-based instructions, the efficiency is high, then the level of extraneous CL is minimized, whereas the level of germane CL is maximized. These are optimal conditions for the construction of cognitive schemata.

The variable age determines cognitive capacity. Cognitive capacity includes working-memory capacity, processing speed, and the ability to inhibit irrelevant information. These are all aspects of working memory. Thus, *cognitive capacity* can be defined as the maximum number of task-relevant items that can be stored and processed in working memory within a certain unit of time. In the light of this definition, it is important that, especially in the case of reduced cognitive capacity, working memory is only occupied with effective operations on relevant information elements. After all, the amount of cognitive capacity that can be dedicated to schema acquisition is reduced by extraneous CL and intrinsic CL, so that it is essential to make maximal use of it. Thus, the cognitive capacity in the framework is the maximal capacity that can actually be claimed by germane CL.

In the framework, age neither affects extraneous CL, nor intrinsic CL. According to Paas and Van Merriënboer (1994a), however, age is a deter-

mining factor of cognitive load (see Figure 1.1). In their view, the concept can be interpreted as a sort of overall cognitive load, which can be assessed by means of a subjective scale (e.g., Paas, Van Merriënboer, & Adam, 1993). This subjective or experienced cognitive load is high when the task and instructional demands (i.e., intrinsic CL and extraneous CL) are high and only little capacity is left for schema construction.

The level of intrinsic CL is determined by the complexity of the task. A task is complex when it contains relatively many interconnected elements. According to Sweller (1994), intrinsic CL has a fixed magnitude. That is, a task has a stable structure that requires a fixed set of cognitive operations. However, when schemata are constructed and become more refined and automated, less mental effort is needed for the same task (see Figure 1.1). In other words, intrinsic CL should gradually decrease. The system should thus contain a loop that affects the level of intrinsic CL. To track this loop, let us return to germane CL.

germane CL is not only determined by the instructions. Its maximum value is confined by the available cognitive capacity. The instructions determine the extent to which this cognitive capacity is used. The more cognitive capacity, the more germane CL can be imposed on the learner. And the more germane CL is imposed, the more efficient is the Schema-Acquisition process. The loop is closed when the acquired schemata enhance the learner's performance as a result of decreased intrinsic CL. Thus, we have an iterative process, which accounts for the accumulation (symbolized by the rolling snowball in Figure 1.2), refinement, and automation of schemata, and, as a result of that, for a reduction of mental effort necessary for solving problems in a particular domain.

Implications of Cognitive Load Theory for Elderly Learners

Since CLT-based instructions deal with cognitive limitations, they are likely to accommodate the mental abilities of elderly learners. In fact, CLT-based instructions optimally compensate for age-related cognitive declines. First, since limited working-memory capacity is the basic constraint CLT-based instructions deal with, more capacity becomes available for learning. Second, CLT-based instructions compensate the loss of information due to decreased processing speed. That is, if extraneous cognitive load is minimized, more relevant information elements can be processed within the same unit of time, so that less information will get lost. Moreover, the simultaneous activation of relevant information elements increases the chance of integration, and thus schema acquisition. Third, the problem of reduced inhibition can be by-passed, because redundant information will be left out as much as possible. Finally, CLT-based instruc-

tions present information such that the learner is directed to its essentials (e.g., by applying worked examples).

What can we now predict following the framework in Figure 1.2? To answer this question, let us perform a thought experiment with a group of young and a group of old adults. One half of both groups is presented with conventional instructions, the other half is presented with CLT-based instructions. Let us assume that the old adults have lost 20% of their original cognitive capacity. Let us further assume that both age groups initially experience equal levels of intrinsic CL in both conditions and that the level of extraneous CL, being equal for the age groups,² is reduced to 50% in the CLT condition. It thus follows that after deducting intrinsic CL and extraneous CL from the Total cognitive capacity, the elderly have less cognitive capacity left than the young, especially in the conventional condition. As a consequence, the elderly can devote less cognitive capacity to germane CL and thus to schema acquisition. Imagine that in the CLT condition the gain of cognitive capacity equals 50% of the initial extraneous-CL level for both age groups. Since the elderly have only 80% of the cognitive capacity of the young, their gain in cognitive capacity is proportionally more. The real benefit, however, lies in the proportion of germane CL. Formally, this can be expressed as:

$$\frac{\Delta GCL}{TCC_{old}} > \frac{\Delta GCL}{TCC_{young}},$$

where ΔGCL symbolizes the absolute gain of germane CL and TCC the total cognitive capacity, that is the cognitive capacity ‘before’ deducting intrinsic CL and extraneous CL. Given the assumption that ΔGCL proportionally takes effect in the performance of the learner, we can predict that the old group’s performance will improve relatively more than the young group’s performance.

In the critical situation where the level of cognitive load imposed by very complex tasks in combination with conventional instructions exceeds the total cognitive capacity of elderly learners, CLT-based instructions can reduce extraneous CL such that enough capacity becomes available for a

² It is not inconceivable that the levels of extraneous CL and intrinsic CL are in fact unequal for the age groups. If we consider, for instance, the inequality in the ability to reduce irrelevant information and we are dealing with a task in which such a competence is required, combined with an instructional format which contains a lot of redundant information, it becomes imaginable that the elderly experience higher levels of both intrinsic CL and extraneous CL. As a matter of fact, this would strengthen the effect of CLT-based instructions relative to conventional instructions, because now there is even more extraneous CL to reduce in the elderly, providing them an extra advantage relative to the young. For the sake of simplicity and due to a lack of empirical literature, however, we have refrained from including a connection between Age and extraneous or intrinsic CL in the framework.

full comprehension of the subject matter. Moreover, enough resources may come available for schema acquisition. In young learners it is more likely that the level of cognitive load remains within the limits of the total cognitive capacity. They also benefit from CLT-based instructions, but in their case it is not a question of whether or not the material is grasped, but rather to what extent it is grasped.

The construction of schemata will eventually reduce the effort needed for a task, because a broader variety of problems can be tackled and the required mental operations are chunked into wholes (Paas & Van Merriënboer, 1994b). Although Sweller's (1994) definition of intrinsic CL may suggest invariability, it will eventually decrease, because its magnitude depends on the extent to which schemata are acquired. For elderly learners, a decrease of intrinsic CL will yield a substantial increase of the available cognitive capacity, paving the way for higher levels of germane CL and schema acquisition. Of course, both age groups will eventually exhibit enhanced performance, but since this enhancement is proportionally larger for elderly learners, the performance level of this group will grow towards the young group's performance level.

Finally, it should be noted that we have to differentiate between the acquisition of new knowledge and skills on the one hand, and the life-long accumulation of knowledge and practicing of skills (i.e., crystallized abilities) on the other hand (e.g., Rabbitt, 1993). Despite a decline in the ability to acquire new schemata, the elderly have had many more years to acquire and automate their schemata. These 'old' schemata can of course be used to more than overcome their working-memory declines. After all, this is why the world values experience.

Conclusion

With cognitive load theory we have an important set of tools at our disposal, which can be of great help in developing instructional material suiting the cognitive abilities of elderly learners. First, CLT is aimed at optimizing schema acquisition by stimulating an efficient use of working-memory. Second, CLT makes it possible for elderly people to learn complex transfer-demanding skills, which are normally beyond their capabilities. Third, since CLT especially benefits the cognitive abilities of the elderly, it can lift their performance level towards the performance level of young adults, and thus reduce the gap between the age groups. Finally, because cognitive limitations especially play a role in the initial phase of acquisition, it is wise to present information with great care at least at this stage. That is, especially when the cognitive demands are high, extraneous cognitive load should be minimized and learners should be motivated to invest mental effort for the acquisition of schemata. Once the instructions have

persuaded the learner to take the plunge and rudimentary schemata have been acquired, subsequent steps will require less effort, motivation will grow, and knowledge will accumulate. For this reason, cognitive load theory can be considered as an essential guide.

2 Worked Examples versus Conventional Problems¹

Cognitive load theory (CLT) is aimed at developing training material that efficiently makes use of the available cognitive processing capacity and stimulates the learner's ability to use acquired knowledge and skills in new situations. It is claimed that CLT-based training formats meet the cognitive abilities of elderly learners particularly well. That is, cognitive aging brings about several declines of working memory, which impede the acquisition of complex cognitive skills. By making an optimal use of the 'remaining' cognitive resources, learning can be enhanced. For that purpose, CLT provides a promising range of training formats that have proven their effectiveness relative to conventional formats in young adults. This chapter presents an experimental study ($N = 54$) aimed at the efficiency of worked examples as a substitute for conventional practice problems in training both elderly and young adults. According to CLT, studying worked examples is a more efficient means of training complex skills than solving conventional problems. As predicted, the results show that — with respect to the elderly — the efficiency of studying worked examples is higher than the efficiency of solving conventional problems in that less training time and cognitive load leads to a comparable level of performance.

Cognitive aging entails both growth and decline. Growth concerns so-called crystallized abilities (Horn & Cattell, 1967; Rabbitt, 1993), which emerge from a lifelong accumulation of knowledge and experience. Decline, on the other hand, refers to fluid abilities, which relate to diverse aspects of working memory. The decline of working memory impedes the acquisition of complex skills in at least three ways. First, due to a reduction of working-memory capacity (e.g., Salthouse & Babcock, 1991; Wingfield, Stine, Lahar, & Aberdeen, 1988), less information elements can be processed, which undermines comprehension and learning of complex material. Second, the overall processing speed decreases (Cerella, 1990; Fisk & Warr, 1996; Myerson, Hale, Wagstaff, Poon, & Smith, 1990; Salthouse, 1993, 1996), which hinders the simultaneous activation of mutually referring information elements. A third decline concerns the ability to inhibit irrelevant or distracting information (e.g., Allen, Madden, Groth, & Crozier, 1992; Hasher & Zacks, 1988), which is likely to withhold older

¹ A version of this chapter was published as: Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., & Schmidt, H.G. (2002a). Cognitive load theory and aging: Effects of worked examples on training efficiency. *Learning and Instruction*, 12, 87-105.

people from the proper processing of information that is relevant for the task at hand.

If we consider the major role of working memory in learning new material, it is obvious that the above three declines dramatically reduce the ability of elderly people to acquire complex skills. It is true that, compared to young people, the elderly have a much richer and more integrated body of general and, dependent on their professional background, specific knowledge at their disposal. It can be expected, however, that this knowledge is likely to fail when it comes to learning entirely new material. In that case, people are forced to fully rely on their working memory abilities, rather than on prior knowledge and experience.

Regarding the aforementioned age-related working memory declines, this brings us to the assumption that in order to optimize the acquisition of complex material in elderly people information should be presented in a highly efficient manner. We claim that Sweller's (1988, 1994) cognitive load theory offers important tools for this purpose. Cognitive load theory (CLT) is directly concerned with the limits of working memory. In a nutshell, CLT claims that an optimal use of working memory requires a maximum number of mental operations that directly contribute to the learning process and a minimum number of operations that do not contribute to the learning process. An essential role in this respect is allocated to the training material. Training should be designed in such a way that the learner is encouraged to spend as much as possible working-memory capacity to relevant operations and is not forced to waste resources on operations that are not relevant or even detrimental to the learning process.

Cognitive Load Theory and the Design of Training

Working memory plays an important role in the storage of information into long-term memory and the acquisition of new skills. Since comprehension and learning are determined by the potential of working memory (e.g., Just & Carpenter, 1992), it can be argued that working memory should be merely occupied by task-relevant operations, especially when dealing with complex material. Cognitive load theory acknowledges this point in that it pleads for a proper use of working memory by means of efficient training. According to cognitive load theory, efficient training has two important characteristics. First, it should impose an as low as possible *extraneous cognitive load*. This is the portion of load that does not contribute to the learning process. Second, it should optimize the level of *germane cognitive load* (Sweller, Van Merriënboer, & Paas, 1998). This is the portion of load that directly contributes to the learning process. Learning is here primarily conceived as the construction of cognitive *schemata*. Schemata enable a person to categorize and solve problems in a specific domain (Sweller,

1994). Schemata reduce the amount of mental effort or working-memory capacity needed for the performance of a particular task, especially when they become automated (Paas & Van Merriënboer, 1994a; Van Merriënboer & Paas, 1990). Stimulating the acquisition of schemata through well-designed training can lead to *transfer* (Paas, 1992), which is the application of a schema to a problem that more or less deviates from problems that were encountered during training. Furthermore, an efficient training is characterized by a favorable effort-performance ratio (Paas & Van Merriënboer, 1993). That is, training is efficient when a relatively low mental effort results in a relatively high performance.

Implications of CLT for Elderly Learners

In diverse experimental settings and problem domains a so-called *complexity effect* has been demonstrated, indicating that relative to the young, the performance of the elderly is especially impaired when the complexity of the task is raised (e.g., Czaja & Sharit, 1993; Gilinsky & Judd, 1994; Lorschach & Simpson, 1988). In other words, the cognitive declines associated with aging are particularly manifest when the demands of the task are high and a relatively heavy burden is imposed on the cognitive system, such as in transfer tasks. In this light, it is expected that the elderly perform relatively poorly if a transfer problem deviates considerably from previously encountered problems (i.e., far transfer) than if a transfer problem closely resembles earlier problems (i.e., near transfer).

As we have seen, training based on CLT takes the limits of working memory into account. Since these limits are especially evident in elderly persons, it is assumed that CLT-based training accommodates the cognitive abilities of this group particularly well. Basically, CLT-based training keeps extraneous cognitive load as low as possible. This offers the opportunity to increase germane cognitive load, allowing elderly people to acquire schemata at a relatively early stage and paving the way for a more extensive schema construction. If this process recurs, the result is an accumulation of schemata and a lowering of the perceived task complexity (Van Gerven, Paas, Van Merriënboer, & Schmidt, 2000, Chapter 1).

How do CLT-based training formats counterbalance the aforementioned three aspects of cognitive aging? The answer to this question demonstrates the perfect match between CLT and cognitive aging. First, since limited working-memory capacity is the basic constraint CLT deals with, more capacity becomes available for learning. Second, CLT-based training compensates the loss of information due to decreased processing speed. That is, if extraneous cognitive load is minimized, more relevant information elements can be processed within the same unit of time and less information will get lost. Moreover, the simultaneous activation of relevant

information elements increases the chance of integration and chunking, and thus the construction of schemata. Third, the problem of reduced inhibition can be by-passed, because redundant (i.e., extraneous) information will be left out as much as possible. In sum, CLT-based training formats optimize learning by making an efficient use of the available cognitive capacity, thereby directing the learner to the essentials of the subject matter and stimulating the construction of cognitive schemata.

Worked Examples versus Conventional Problems

We attempted to test the above claims with one of the most important CLT-based training formats, namely the use of *worked examples* as a substitute for *conventional problems*. Unlike worked examples, conventional problems force learners to apply a capacity-demanding means-ends analysis, in which the goal state of the problem can only be attained by subdividing it into a hierarchy of goals and subgoals, which have to be achieved in a backward order by finding the appropriate operators. For novices, who have not yet acquired extensive cognitive schemata, problem solving turns out to be a rather ineffective means of learning, because it imposes an extremely high extraneous cognitive load (Sweller, 1988), whereas the level of germane cognitive load remains relatively low. In other words, solving conventional problems brings about high costs, while yielding a low pay-off in terms of learning results. Worked examples, on the other hand, are more likely to lead to an effective construction of cognitive schemata, because they focus the learner's attention on problem states and operators, rather than on goals and subgoals. Numerous studies have shown that, in case of complex tasks, worked examples lead to superior performance and transfer relative to conventional problems. These results were obtained with young adults (e.g., Paas, 1992; Paas & Van Merriënboer, 1994b; Ward & Sweller, 1990) as well as with children (e.g., Pillay, 1994).

The aim of the present study is to test whether the application of worked examples leads to improved performance in elderly learners. According to Van Gerven et al.'s (2000) framework this should certainly be the case. The framework assumes that the cognitive capacity of elderly people is smaller than the cognitive capacity of young people, so that any gain in cognitive capacity due to a lowering of extraneous load is proportionally larger for the elderly than for the young. The same holds for any increase of germane cognitive load. Therefore, the elderly should benefit relatively more from worked examples in terms of training time, invested mental effort, and performance than the young.

According to Paas and Van Merriënboer (1993), the ratio between invested mental effort and test performance can be interpreted as the efficiency of a training format. That is, a training format is considered effi-

cient if a relatively low mental effort results in a relatively high performance, whereas a training format is considered inefficient if a relatively high mental effort results in a relatively low performance. To quantify training efficiency, Paas and Van Merriënboer developed a mathematical procedure involving the standardized scores of mental effort and performance (see the Results section for details). This formal definition of training efficiency, combined with Van Gerven et al.'s (2000) framework, brings us to state two hypotheses.

The first hypothesis (Hypothesis 1) predicts an *interaction* between age group and training format. We expect that training with worked examples is more efficient than training with conventional problems, especially for the elderly. That is, for both age groups there is a more optimal relationship between mental effort during training and transfer performance (i.e., lower mental effort combined with higher performance) for worked examples than for conventional problems. However, this relation is more optimal for the elderly than for the young. With respect to training time, the interaction effect should be reversed. That is, worked examples should require less training time than conventional problems, because they are more efficient. Again, this difference should be larger for the elderly than for the young.

The second hypothesis (Hypothesis 2) concerns *transfer distance*. As we have seen, a distinction can be made between near- and far-transfer problems. Near-transfer problems can be considered relatively simple in that they closely resemble the problems encountered during the training phase. Far-transfer problems, on the other hand, are relatively complex in that they considerably deviate from the training problems. It is expected that the chances for attaining far transfer are greater in the worked-examples than in the conventional-problems condition, because of a more extensive schema construction. Since the achievement of transfer is an essential characteristic of adequate schema construction (Paas, 1992), it is expected that the interaction effect for training efficiency, as predicted by Hypothesis 1, is stronger in the far- than in the near-transfer domain.

Method

Participants

A first group of participants consisted of 30 psychology students (18 women, 12 men) of Maastricht University, ranging in age from 18 to 30 years ($M = 20.47$, $SD = 2.71$). A second group consisted of 24 elderly participants (12 women, 12 men) ranging in age from 61 to 76 years ($M = 66.67$, $SD = 4.72$). The elderly participants were selected from a participants pool provided by the Maastricht Aging Study (MAAS, Jolles, Houx, Van Boxtel, & Ponds, 1995). Participants from both groups were randomly

assigned to the experimental conditions ($n = 15$ for the young, $n = 12$ for the elderly). All participants were in good health and had normal or corrected-to-normal vision. They received 10 Dutch guilders (about €4.50) payment per hour and a refund of their travel expenses.

Materials

Intelligence test. To check the intellectual comparability of the experimental groups, a shortened version of the Groningen Intelligence Test (GIT, Luteijn & Van der Ploeg, 1983; Snijders & Verhage, 1962) was administered. The test included four scales (out of nine): vocabulary (factor verbal comprehension), spatial abilities (factor visualization), mental arithmetic (factor number), and analogies (factor induction/deduction). The sum of the standardized scale scores was multiplied by $9/4$, yielding an estimate of the complete test score. This estimate was converted into an IQ score. The 95%-confidence interval of the shortened GIT is ± 10 points, against ± 6 points for the total test (Luteijn & Van der Ploeg, 1983).

Computation-span test. Working-memory abilities were assessed with Salthouse and Babcock's (1991) computation-span test, which is a variant of the well-known reading-span test (Daneman & Carpenter, 1980). The participants had to solve a range of simple arithmetic problems that were verbally presented to them, for example: "7 plus 8," "9 minus 4," et cetera. At the same time, they had to memorize the last digit of each problem, which is 8 and 4 in this example. The number of problems per trial ranged from 1 to 7, with three trials for each level. The participants had to select the correct answer out of three alternatives, which were presented in a booklet. Once all problems of a trial were presented and the corresponding answers were selected, the participant had to turn to the next page of the booklet and write down the recalled digits. The computation-span score was determined by the highest number of digits recalled in at least two trials.

Subjective cognitive-load scale. A 9-point symmetrical category scale was used as a subjective cognitive-load (SCL) measure (after Paas, Van Merriënboer, & Adam, 1994). The scale was presented on a touch screen (see Apparatus and Interface for details) as a slide control and was accompanied by the phrase: *I have experienced the foregoing as: ...* The scale had three labels, ranging from *not difficult at all* (1), via *moderately difficult* (5), to *very difficult* (9). Intermediate points had no labels. Paas (1992) found an internal consistency coefficient (Cronbach's α) of .90 using a comparable scale. Paas and Van Merriënboer (1994b) evaluated the scale as a highly reliable and sensitive instrument for the assessment of cognitive load ($\alpha = .82$).

Task domain. Some considerations preceded the selection of a suitable task. For the old participants, whose educational background varied more than that of the young participants, the task had to appeal to their com-

mon sense rather than to their abstract abilities. To prevent a floor effect in their performance, the task had to be straightforward enough to master in a relatively short time. For the young participants, on the other hand, the task had to be complex enough to prevent a ceiling effect. Luchins' (1942) classical water-jug problem appeared to meet these requirements. The water-jug problem is a Tower-of-Hanoi-like task, in which a certain amount of water has to be acquired by pouring jugs of different sizes and containing different amounts of water into each other. Water-jug problems can be constructed such that they put a substantial burden on working memory (Atwood, Masson, & Polson, 1980; Atwood & Polson, 1976). The task has mainly been used for studying the effect of *Einstellung* (e.g., Lippman, 1994, 1996; Luchins & Luchins, 1991), which is the tendency to persevere in applying a previously efficient solution strategy, even though this strategy is not efficient in a given situation. Ransopher and Thompson (1991) found that the strength of this solution rigidity depends on the number of previous problems that induces its occurrence. This effect was not influenced by age.

In the version of the water-jug task employed in this study, the goal amount of water had to end up in a target jug. For that purpose, an infinite amount of water was available. A limited set of operations could be performed with the so-called working jugs. First, a working jug could be filled, but only to the brim. Second, a working jug could be emptied (i.e., water could be wasted). A third operation was the most essential: pouring water from one jug into another. Jugs receiving water could never overflow during this operation. That is, pouring stopped as soon as the donating jug was empty or the receiving jug was full. When a receiving jug was full, a possible residual remained in the donating jug. This residual could be the target amount of water or could be used for further operations. The target jug could only be filled by pouring water in it from one of the working jugs. Furthermore, the target jug could be emptied in case its current content exceeded the target amount of water. No water could be poured from the target jug into one of the working jugs.

A division was made between near-transfer problems and far-transfer problems. The processing model depicted in Figure 2.1 clarifies this distinction. The model can be considered as an 'ideal' problem solver in the sense that it subsequently checks whether a progressively more complex solution strategy is applicable in a given situation. First, it checks whether one of the jugs has a capacity that equals the target amount of water. If this is the case, it directly fills that particular jug and pours its content into the target jug. Subsequently, it checks whether the sum of two or more jugs yields the target amount. If that is not the case, it checks whether the difference between the capacities of two or more jugs yields the desired amount. That is, if the contents of a jug are poured into a smaller jug, a residual remains in the bigger jug as soon as the smaller one

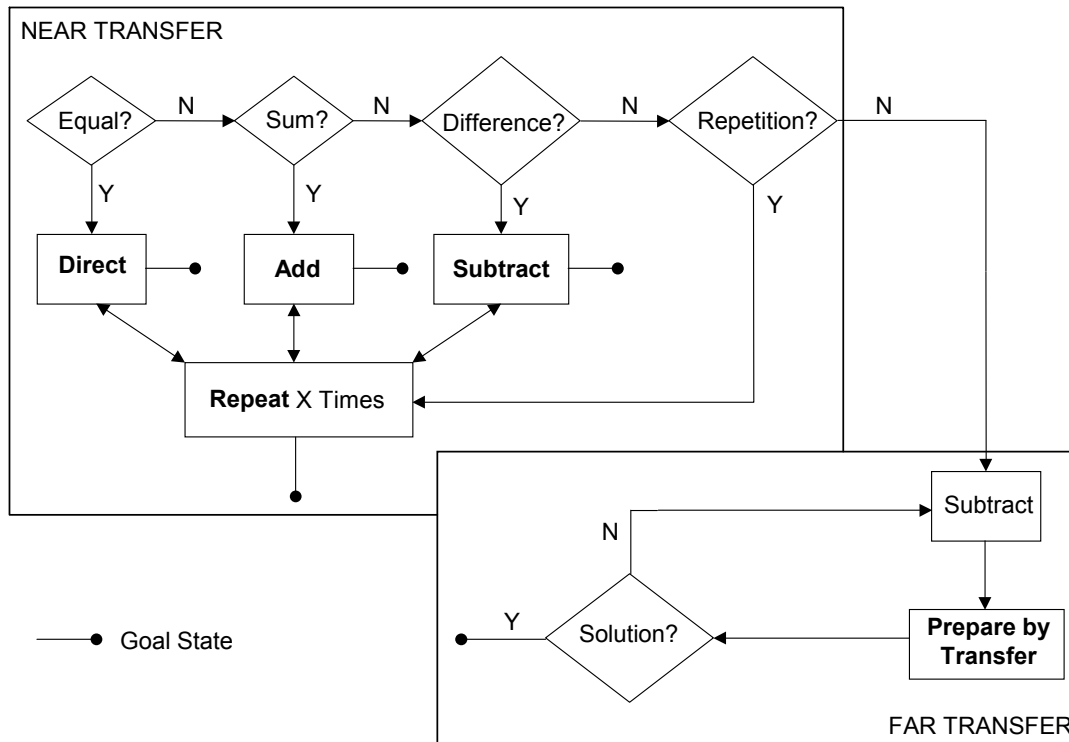


Figure 2.1 A general processing model for the water-jug problem. Note that this is not a cognitive model, but merely a clarification of the distinction between near- and far-transfer problems.

is full. This residual can be regarded as the outcome of subtracting the capacity of the small jug from the capacity of the larger one. If subtraction does not lead to a solution, the model examines whether a repetition or combination of one or more of the previous strategies does.

So far the solution strategies are rather straightforward, because they are only appropriate to tackle *near-transfer* problems. Near-transfer problems closely resemble the problems that have been encountered during the training phase. *Far-transfer* problems, on the other hand, require a more sophisticated solution strategy, which we call ‘preparation by transfer.’ That is, the content of a jug or the residual of a subtraction is transferred to a different jug to prepare for a further subtraction. Imagine, for example, that we have one jug with a capacity of 5 units and one jug with a capacity of 7 units (like in Figure 2.2). Suppose we fill the jug of 5 and transfer its content to the jug of 7. The jug of 7 can contain 2 more units. So, actually, we have created a jug with a capacity of 2 units. If we now refill the jug of 5 and pour its content into the jug of 7 again, 3 units of water remain in the jug of 5. This residual can be transferred again and the whole procedure can be repeated until a solution is achieved (see the loop in Figure 2.1).

Apparatus and interface. The experiment was run on an IBM-compatible computer. The software controlling the experiment was

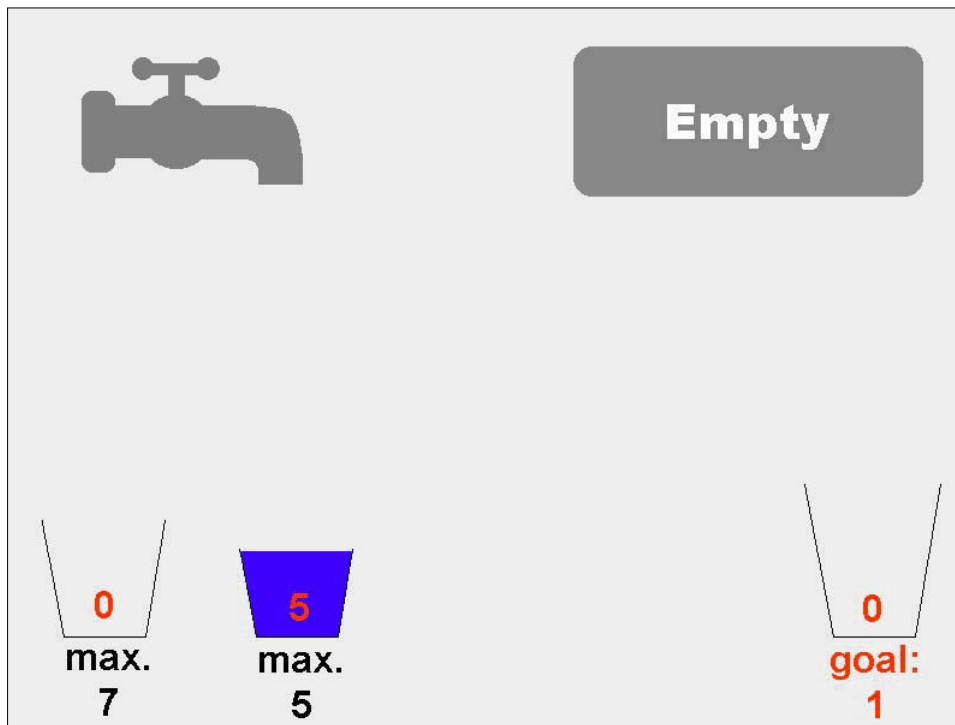


Figure 2.2 The interface of the computerized version of Luchins' (1942) water-jug problem. Depicted is the initial phase of the solution of a far-transfer problem. The second working jug has just been filled. (Originally, labels were in the Dutch language).

programmed in Authorware 3.5 (1996). The interaction with the computer took place by means of a Philips 17-inch monitor with an integrated Elo touch screen. The interface is depicted in Figure 2.2. There is a tap in the upper left corner and a button labeled *Empty* in the upper right corner. The three jugs in the bottom left corner are the working jugs. The jug in the bottom right corner is the target jug. The current contents are displayed inside the jugs. The capacities are displayed beneath the jugs (indicated as *max.*). The goal content is displayed beneath the target jug (indicated as *goal*). There were three basic operations. First, a working jug could be filled by first touching the tap and then the working jug to be filled. Second, a jug could be emptied by first touching the *Empty* button and then the jug to be emptied. And third, jugs could be poured into each other by first touching the donating jug and then the receiving jug. The software registered every legitimate step performed by the participant and whether or not a problem was solved.

Training formats. There were two training conditions: a conventional problems and a worked examples condition. In the conventional problems condition, participants had to solve four conventional problems, requiring basic strategies. In the worked examples condition, participants had to study these same basic problems presented in a worked-out fashion (see Figure 2.3). That is, for each solution step to appear, they had to push a

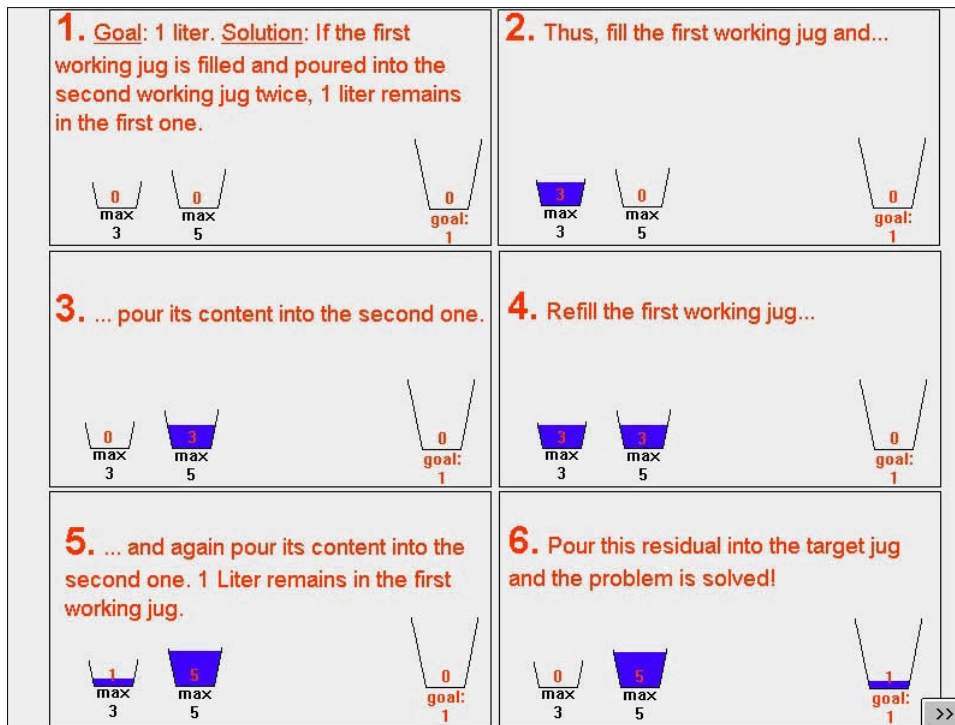


Figure 2.3 A worked example. Individual images represent the subsequent solving steps and appear on the screen one at a time by pushing the lower right button. (Originally, texts and labels were in the Dutch language.)

button on the touch screen. In this way, maximal six steps were displayed as a series of ‘mini-screens,’ enabling the participants to keep track of the strategy. The solution steps were accompanied by explanatory text.

Transfer test. A final test consisted of 23 water-jug problems, which gradually increased in complexity. The first 11 problems of the test were classified as near-transfer problems. These problems only slightly differed from the problems presented in the training phase and required basic solution strategies. Problems 12 to 23 were classified as far-transfer problems. The same basic rules applied to these problems. However, in order to solve them, the participants were required to make an essential strategic switch (problems 12 to 20; see Figure 2.1), or they had to obtain a certain distribution of water over the jugs (problems 20 to 23; after Atwood & Polson, 1976).

Design and Procedure

There were two independent between-groups variables: age group, with levels young and old, and training format, with levels conventional problems and worked examples. Furthermore, there was one independent within-groups variable: transfer distance, with levels near and far. 15 Young participants were randomly assigned to the conventional problems condition and 15 other young participants were assigned to the worked ex-

amples condition. Likewise, 12 elderly participants were randomly allocated to the conventional problems condition and 12 other elderly participants were allocated to the worked examples condition. All participants were confronted with both the near- and the far transfer test. This yields a 2×2 (between groups) $\times 2$ (within groups) design. During the training phase, the subjective cognitive load (SCL) and the time (in seconds) invested per problem (training time) were registered by the software. In the test phase, the proportion of solved problems was registered for the near- and the far-transfer conditions.

The experiment was performed in individual sessions. After a short verbal introduction by the experimenter, the intelligence and computation-span test were administered. Subsequently, the participant was presented with a general instruction explaining the basic rules of the water-jug problem as well as the interface. The use of the interface was practiced under supervision of the experimenter. The participant also got acquainted with the cognitive-load measure. After that, s/he was presented with the training problems. Finally, the participant was put to the test. Throughout the experiment, the experimenter answered simple questions by the participant that did not concern the solution strategy. These questions were not incorporated in the analysis. There was a time limit of ten minutes per problem. If requested, a problem could be skipped (the participant was informed about this possibility beforehand). The participant had two hours at most for the instructions, the training, and the test. About half an hour before the maximum time had passed, s/he was informed about the remaining time. After the maximum time had passed, the participant was asked to stop with the test.

Results

A 2×2 ANOVA, with age group (levels young and old) and training format (levels conventional problems and worked examples) as independent between-groups variables, was performed on the control variables IQ and Computation Span and the dependent variables training time, subjective cognitive load (SCL), performance, and training efficiency. Furthermore, a 2 (age group) $\times 2$ (training format) $\times 2$ (transfer distance) ANOVA was performed on the training efficiency scores with transfer distance (levels near and far) as the independent within-groups variable. Data means and standard deviations can be found in Table 2.1. Outcomes of the analyses are summarized in Table 2.2.

IQ and Computation Span

Within the age groups and between the training conditions, there were no significant intelligence differences. It is remarkable, however, that the

Table 2.1 Means and standard deviations of the dependent variables.

		Conventional problems		Worked examples	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Young (<i>n</i> = 30)	Training efficiency	0.86	0.46	0.55	0.61
	Near-transfer efficiency	0.74	0.35	0.32	0.99
	Far-transfer efficiency	0.83	0.48	0.54	0.57
	Training time	44.31	15.96	25.59	8.95
	Subjective cognitive load	2.22	0.85	2.77	1.13
	Near-transfer performance	1.00	.00	.99	.05
	Far-transfer performance	.89	.11	.86	.14
Old (<i>n</i> = 24)	Training efficiency	-1.34	1.41	-0.42	1.21
	Near-transfer efficiency	-1.18	1.40	-0.15	1.13
	Far-transfer efficiency	-1.38	1.41	-0.24	1.00
	Training time	206.58	97.14	81.24	32.95
	Subjective cognitive load	4.79	2.16	3.19	1.71
	Near-transfer performance	.95	.05	.97	.04
	Far-transfer performance	.40	.28	.55	.33

mean IQ score of the elderly participants ($M = 128.67$, $SD = 9.47$) was almost 10 points higher than the mean IQ score of the young participants ($M = 119.47$, $SD = 8.70$), $F(1, 50) = 13.66$, $MSE = 82.60$, $p = .001$.

With respect to Computation Span, there were no differences within the age groups and between the training conditions. However, computation spans differed significantly between the age groups, $F(1, 50) = 28.40$, $MSE = 1.45$, $p < .001$. Where the young participants had a mean score of 4.97 ($SD = 1.33$), the elderly scored no more than 3.21 ($SD = 0.98$). These results are in accordance with Salthouse and Babcock's (1991) study, in which a gradual age-related decrease of computation-span scores was found.

Table 2.2 *F* values, significance levels, degrees of freedom, and mean square errors (AG = age group, TF = training format, TD = transfer distance).

	AG	TF	AG × TF	AG × TF × TD	<i>df</i>	<i>MSE</i>
Training efficiency	36.27***	1.35	5.45*	0.10 ^{a, b}	1, 50	0.92
Near-transfer efficiency	18.77***	1.20	6.91*	-	1, 50	1.02
Far-transfer efficiency	36.05***	2.89	8.20**	-	1, 49 ^b	0.81
Training time	65.73***	28.72***	15.73***	-	1, 50	2408.00
Subjective cognitive load	13.43**	1.66	6.94*	-	1, 50	2.23
Near-transfer performance	10.98**	0.24	2.63	-	1, 50	.00
Far-transfer performance	44.09***	1.00	2.13	-	1, 49 ^b	.05

Note: * $p < .05$, ** $p < .01$, *** $p < .001$. ^a $df = 1, 49$, $MSE = 0.19$. ^b Degrees of freedom differ because one of the elderly participants was not able to reach the far-transfer part of the test.

Training Efficiency

Training efficiency was determined for each age group, training format, and transfer distance, using the subjective cognitive load scores of the training phase and the performance scores of the transfer test. With regard to the elderly, subjective cognitive load is substantially lower in the worked examples condition than in the conventional problems condition. More than that, the interaction between age group and training format, $F(1, 50) = 6.94$, $MSE = 2.23$, $p < .05$, suggests a disproportional ‘decrease’ in SCL for the elderly participants. Furthermore, there is an effect of age group, $F(1, 50) = 13.43$, $p = .001$, but no effect of training format, $F(1, 50) = 1.66$, $p > .05$. Not surprisingly, there is a huge performance difference between the age groups. This holds for both near-transfer performance, $F(1, 50) = 10.98$, $MSE = .00$, $p < .01$, and far-transfer performance, $F(1, 49) = 44.09$, $MSE = .05$, $p < .001$. However, there is neither an effect of training format nor an interaction (see Table 2.2).

Efficiency calculations were performed following Paas and Van Merriënboer (1993), who defined training efficiency as the ratio between mental effort (i.e., SCL) and performance (i.e., the proportion of solved problems). Mental-effort (M) and performance scores (P) were transformed into z scores. The means of these standardized scores are plotted in a coordinate system as dots representing the training formats (see Figure 2.4). Training efficiency (E) is now determined by the perpendicular distance from each of these dots to the diagonal $P = M$, which is calculated as follows:

$$E = \frac{P - M}{\sqrt{2}}.$$

According to this expression, E is positive if $P > M$, negative if $P < M$, and zero if $P = M$. A negative E implies that a relatively high mental effort is followed by a relatively low performance. A positive E implies that a relative low mental effort is followed by a relatively high performance. An E of zero means that mental effort and performance are balanced.

As can be seen in Figure 2.4, the efficiencies of both training formats are positive for the young participants and negative for the elderly participants. However, contrary to the young participants, for whom the worked examples condition appears to be relatively inefficient, the elderly display a substantial efficiency ‘gain’ in the worked examples condition relative to the conventional problems condition, the latter of which remains highly inefficient. The analysis of variance reveals an effect of age group, $F(1, 50) = 36.27$, $MSE = 0.92$, $p < .001$, as well as an interaction between age group and training format, $F(1, 50) = 5.45$, $p < .05$, which is in accordance with Hypothesis 1. However, there is no interaction between age group, training format, and transfer distance, $F(1, 49) = 0.10$, $MSE = 0.19$, $p > .05$,

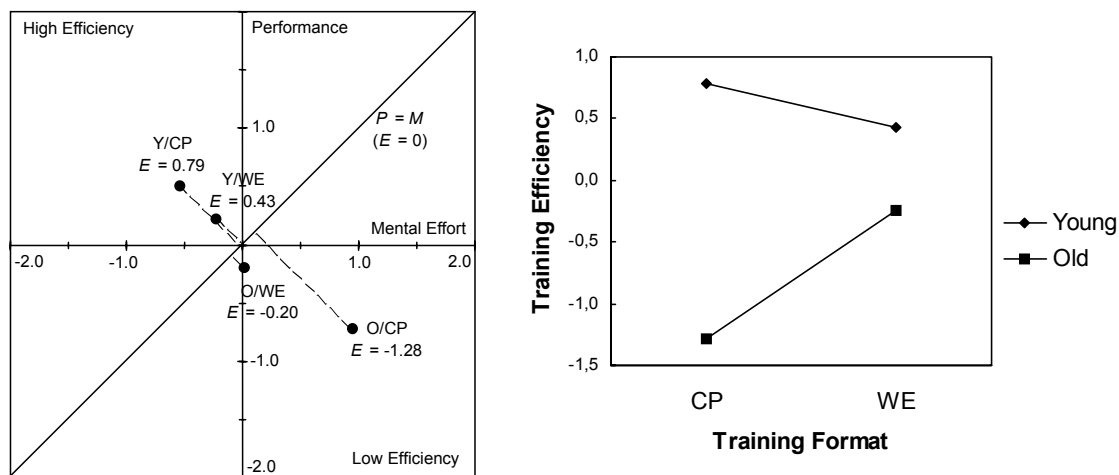


Figure 2.4 Mean overall training efficiency per experimental group, based on subjective cognitive load in the training phase and performance in the test phase. The left diagram is a representation after Paas and Van Merriënboer (1993), where Y = young, O = old, CP = conventional problems, and WE = worked examples. The right diagram clearly illustrates the obtained interaction.

which is not in line with Hypothesis 2. This means that the age group \times training format interaction is independent of transfer distance, which is confirmed by the age group \times training format interactions for near-and far-transfer efficiencies separately, although their significance levels differ: $F(1, 50) = 6.91$, $MSE = 1.02$, $p < .05$, against $F(1,49) = 8.20$, $MSE = 0.81$, $p < .01$, respectively.

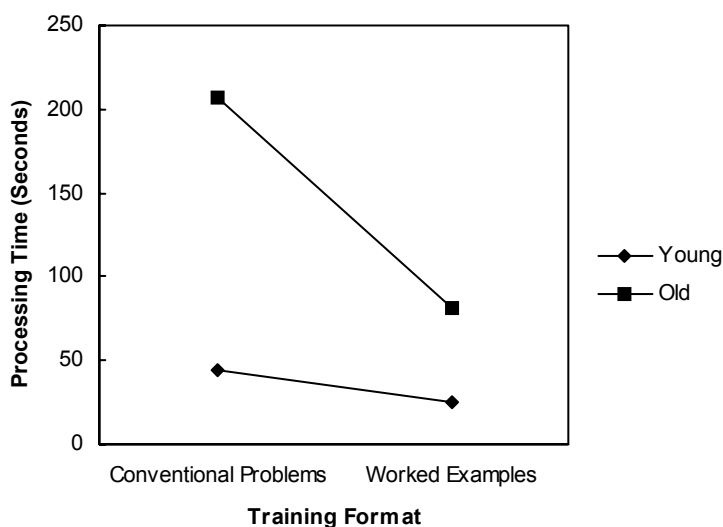


Figure 2.5 Mean training time per conventional problem and worked example.

Training Time

Figure 2.5 shows the mean training time in seconds per conventional problem and worked example. The analysis of variance reveals main effects for the factors age group, $F(1, 50) = 65.73$, $MSE = 2408.00$, $p < .001$, and training format $F(1, 50) = 28.72$, $p < .001$. There is also a significant interaction between age group and training format, $F(1, 50) = 15.73$, $p < .001$. The interaction pattern, depicted in Figure 2.5, suggests that the elderly participants are disproportionately favored when studying worked examples, which supports Hypothesis 1.

Discussion

The study reported in this chapter was aimed at testing the efficiency of worked examples relative to conventional problems in elderly and young learners. On the basis of Sweller's (1988, 1994) cognitive load theory, the literature on cognitive aging, and Van Gerven et al.'s (2000) framework, it was hypothesized that elderly learners benefit relatively more from worked examples in terms of training efficiency and training time than their young counterparts. That is, more than the young, the elderly were expected to experience less cognitive load and spend less time studying worked examples, while attaining at least an equal performance level (Hypothesis 1). Furthermore, because transfer is considered as an essential feature of learning, this interaction effect was expected to be stronger in the far- than in the near-transfer domain (Hypothesis 2).

Between the training conditions, groups were highly comparable in terms of intelligence and computation span. Between the age groups, however, there were substantial differences. Where the elderly achieved higher IQ scores than the young, they scored significantly lower on the computation-span task, which is in accordance with the results of Salthouse and Babcock (1991). Together with the finding that the elderly spent more time on training and experienced higher levels of cognitive load relative to the young, this strongly supports the view that working memory rather than intellectual ability plays an essential role in learning new skills.

The relative drawback of the elderly due to cognitive aging was manifest in all dependent variables. The old participants were less efficient, performed more poorly, and showed higher levels of cognitive load and training time than the young (see Tables 2.1 and 2.2). Performance was especially poorer in the more complex far-transfer domain, which is a confirmation of earlier research findings (Czaja & Sharit, 1993; Gilinsky & Judd, 1994; Lorbach & Simpson, 1988).

The interactions between age group and training format with regard to training time and cognitive load strongly suggest that the elderly benefit comparatively more from worked examples than their young counterparts.

This support is strengthened by the efficiency measures, which indicate that, contrary to the young, the elderly achieve an equal performance level by investing less mental effort when studying worked examples. Again, this suggests that the elderly gain relatively more from worked examples than the young, which strongly supports Hypothesis 1. Moreover, these findings imply that worked examples offer the opportunity to add germane cognitive load, for example, by increasing the example variability (Paas & Van Merriënboer, 1994b). The similarity between the interaction patterns of cognitive load and training time furthermore suggests that training time might as well be used as a component in calculating training efficiency (see Pollock, Chandler, & Sweller, 2002).

A rather peculiar finding is that the young did not profit from the worked examples used in this study. Their mean scores even suggest a slight negative effect of worked examples on training efficiency. However, both worked examples and conventional problems required relatively little cognitive load and led to almost the same performance level, which is reflected by the positive efficiency scores. In retrospect, a differential efficiency effect was not likely to occur, since the young seem to have reached their upper performance limit in both conditions. Apparently, the water-jug task was too easy for them.

In fact, the employed worked examples were not effective enough to evoke a significant performance enhancement in either age group, which is contrary to previous research (Paas, 1992; Paas & Van Merriënboer, 1994b; Pillay, 1990; Ward & Sweller, 1990). With respect to the elderly, a possible explanation for this finding is the frequently observed phenomenon that old adults display far more statistical variance than young adults on various cognitive tasks (e.g., Anstey, Stankov, & Lord, 1993), so that subtle differences between conditions tend to get overshadowed. Indeed, the standard deviations of the far-transfer scores were substantially greater for the elderly than for the young (see Table 2.1). In this light, the contrast between the conditions in this study may have been too weak to evoke an effect. To enhance transfer performance, germane cognitive load could be increased. This germane load could either be directed at schema construction (e.g., by increasing the variability of the examples) or at schema automation (e.g., by increasing the number of examples with the same variability). However, this would only work for the elderly, because the young seem to have reached their ceiling.

The absence of a significant interaction between age group, training format, and transfer distance with respect to training efficiency is not in line with Hypothesis 2. This means that the older participants' gain from worked examples was independent of transfer distance. One possible explanation for this result relates to the complexity of the near- and far-transfer problems. It seems that the predicted interaction was not obtained, because the near-transfer problems, which were supposed to be

rather simple, were in fact experienced as rather complex by the elderly. Although the near- and far-transfer problems reflected different levels of complexity, the complexity of the near-transfer problems could have been already sufficiently high to eliminate differential effects.

In sum, worked examples were found to be a more efficient means of training the elderly than conventional problems. Moreover, the elderly took more advantage of worked examples than their young counterparts. Although, contrary to our predictions, this effect was not stronger in the far- than in the near-transfer domain, it can be concluded that, with regard to the elderly, worked examples make a more efficient use of the available working-memory capacity than conventional problems. In future research, worked examples aimed at elderly learners could be designed such that they address different modalities and their corresponding working-memory components (see Baddeley, 1992). For this purpose, an effective use of multimedia techniques, in which images or animations are combined with narration, offers great opportunities (e.g., Mayer & Moreno, 1998, 2002; Tindall-Ford, Chandler, & Sweller, 1997). Such an approach would further reduce the amount of extraneous cognitive load caused by split attention and thus raise the possibility to increase germane cognitive load. Eventually, the overall cognitive-load measure, which does not discriminate between extraneous and germane load, may reach a very high level. The ratio between these two portions of load, however, is then likely to be far more beneficial.

3 Multimedia-Aided Cognitive Skill Training¹

On the basis of a multimodal model of working memory, cognitive load theory predicts that a multimedia-based instructional format leads to a better acquisition of complex subject matter than a purely visual instructional format. This study investigated the extent to which age and instructional format had an impact on training efficiency among both young and old adults. It was hypothesized that studying worked examples that are presented as a narrated animation (multimedia condition) is a more efficient means of complex skill training than studying visually presented worked examples (unimodal condition) and solving conventional problems. Furthermore, it was hypothesized that multimedia-based worked examples are especially helpful for elderly learners, who have to deal with a general decline of working-memory resources, because they address both mode-specific working-memory stores. The sample consisted of 60 young (mean age = 15.98 years) and 60 old adults (mean age = 64.48 years). Participants of both age groups were trained in either a conventional, a unimodal, or a multimedia condition. Subsequently, they had to solve a series of test problems. Dependent variables were perceived cognitive load during the training, performance on the test, and efficiency in terms of the ratio between these two variables. Results showed that for both age groups multimedia-based worked examples were more efficient than the other training formats in that less cognitive load led to at least an equal performance level. Although no difference in the beneficial effect of multimedia learning was found between the age groups, multimedia-based instructions seem promising for the elderly.

The strong increase of the aging population in Western societies has provoked a true avalanche of research aimed at numerous aspects of both physical and psychological aging. For that matter, educational research has surely not remained an unexplored realm. Although aging is often associated with infirmity, it can be expected that the proportion of independent, active, and eager-to-learn elderly people is growing just as well. It is therefore not surprising that educational research has put forward *lifelong learning* as one of its central objectives. Lifelong learning can be achieved by an instructional intervention that efficiently addresses the available cognitive capacity. For that purpose, principles of instructional design

¹ A version of this chapter is submitted for publication as: Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., Hendriks, M. & Schmidt, H.G. (2002). The efficiency of multimedia learning into old age.

have to be combined with outcomes of cognitive aging research. It is peculiar, however, that little of the work done in this field is actually based on such an integrative theory (exceptions are Czaja, 1996; Morrel & Echt, 1996, 1997; Morrell & Park 1993). Van Gerven, Paas, Van Merriënboer, and Schmidt (2000, Chapter 1) propose a framework in which cognitive load theory (Sweller, 1988; Sweller, Van Merriënboer, & Paas, 1998) is merged with general views on cognitive aging (e.g., Hasher & Zacks, 1988; Salthouse 1996; Salthouse & Babcock, 1991).

Cognitive load theory (CLT) is an instructional theory that starts from the idea that human working memory is limited with respect to the amount of information it can hold and the number of operations it can perform on that information. CLT furthermore assumes that working memory plays an essential role in skill learning, because it serves as a device for transforming instructional information into useful *cognitive schemata*. Cognitive schemata are information structures in long-term memory that enable someone to solve a certain category of problems and at the same time save working-memory capacity by chunking information elements and production rules into a whole. Moreover, cognitive schemata facilitate transfer performance (e.g., Paas & Van Merriënboer, 1994a), which is the use of acquired knowledge in problem situations that more or less deviate from problem situations that were encountered during training. The central idea of CLT is that a learner should be stimulated to use his or her limited working memory efficiently, especially when acquiring a complex skill. That is, in order to achieve optimal learning, working memory should be occupied by a minimum number of irrelevant operations and a maximum number of relevant operations. This can be achieved by a proper instructional design for which CLT promotes a series of guidelines (see Sweller, Van Merriënboer, & Paas, 1998, for an overview).

A theory concerned with cognitive limitations may especially be applicable to the acquisition of complex cognitive skills on behalf of learners whose cognitive capacity is particularly limited, such as the elderly. Healthy elderly people have to cope with at least three cognitive declines. First, there is a decrease of working-memory capacity (e.g., Salthouse & Babcock, 1991; Wingfield, Stine, Lahar, & Aberdeen, 1988), which is the ability to store and manipulate a relatively small amount of information for a relatively short period. Second, there is a decline of the rate at which information is activated and processed (Myerson, Hale, Wagstaff, Poon, & Smith, 1990; Salthouse, 1996). This reduced cognitive speed can lead to poor comprehension and learning as a result of what Salthouse (1996) calls the “simultaneity mechanism:” If mutually dependent information elements are not activated quickly enough, activation of one element can have dissipated before activation of the other has reached a certain threshold. A third decline concerns the failure to suppress information that is not relevant for a task at the cost of capacity available for relevant

processing (e.g., Allen, Madden, Groth, & Crozier, 1992; Hasher & Zacks, 1988; McDowd, 1997).

According to Van Gerven et al. (2000, Chapter 1), CLT offers the principles to optimally deal with these three declines. The decreased-capacity problem can be largely compensated by preventing the learner from getting engaged in task-irrelevant processes. The same holds for the decline of cognitive speed: The probability of attaining simultaneous activation of interdependent information elements can be increased by merely allowing task relevant information to become active in working memory. Problems with suppressing irrelevant information, finally, can be avoided by presenting a minimal number of redundant stimuli. These principles may seem rather obvious. In conventional instructional settings, however, they are hardly taken into account.

This is especially true for the practice phase of an instruction. Complex skills are usually trained by solving a relatively large number of practice problems. Problem solving as a means of training has a significant drawback, however. It imposes a high cognitive load on the learner, which is, as we will see, especially problematic in an early stage of learning. This high level of cognitive load is caused by a commonly used strategy for solving relatively unfamiliar problems, namely *means-ends analysis*. Means-ends analyses take place in two phases. First, a range of goals and subgoals is set, working backward from the main goal. When a level of attainable subgoals is reached, the problem solver works forward from these subgoals towards the main goal. It may be clear that keeping in mind the whole hierarchy of goals and subgoals as well as finding the right operators to attain them puts a considerable burden on working memory. When cognitive resources are sufficient, means-ends analysis may be regarded as an effective problem-solving strategy. When resources are limited, however, problem solving through means-ends analysis is hardly effective.

In both cases, the learning result is far from optimal, because too many resources are used for irrelevant mental operations (Sweller, 1988). The irrelevant or, in terms of CLT, *extraneous* part of means-ends analysis concerns the backward-processing or goal-setting phase, because it does not contribute to the acquisition of cognitive schemata. For the construction of cognitive schemata, problem states have to be associated with useful operators. According to CLT, this process should not be obstructed by any irrelevant activity in working memory.

Instead, cognitive resources should be saved by preventing the learner from using 'weak' solving strategies, such as means-ends analyses, which require backward processing. A CLT-based alternative in this respect is training with *goal-free problems*. In goal-free problems no clear-cut goal state is requested. In the domain of mathematics, for instance, the learner could be asked to solve an equation for as many variables as possible instead of one particular variable. In this way, forward rather than back-

ward processing is encouraged, which is beneficial for the learning process. Training with goal-free problems was found to result in better performance than conventional problem solving in young adults (Bobis, Sweller, & Cooper, 1994; Owen & Sweller, 1985; Sweller & Levine, 1982) and recently also in older adults (Paas, Camp, & Rikers, 2001).

Another important CLT-based alternative for conventional problem solving is the use of *worked examples*. Worked examples are problems accompanied by the subsequent solution steps that lead to a particular goal state. According to CLT, worked examples impose relatively little cognitive load on the learner by directly focusing attention on problem states and their corresponding solution operators. This is especially beneficial in an early stage of learning, in which specific cognitive schemata do not yet exist. Without cognitive schemata, the learner will rely on conscious processing. Conscious processing of complex material is rather demanding in terms of cognitive resources, so that overload is likely to occur. Provided that the learner is optimally motivated, studying worked examples quickly leads to the construction of rudimentary, but indispensable schemata. This saves processing capacity, which can be used for transforming these undeveloped schemata into more elaborate ones. Superior performance using worked examples as opposed to conventional problems was repeatedly demonstrated (Paas, 1992; Paas & Van Merriënboer, 1994b; Sweller & Cooper, 1985; Ward & Sweller, 1990). Van Gerven, Paas, Van Merriënboer, and Schmidt (2002a, Chapter 2) found worked examples to be a more efficient training method for elderly learners in terms of invested mental effort.

There are numerous ways to design worked examples. The most straightforward one is a combination of images and explanatory text. These two sources of information are often physically separated (like the figures and texts in this chapter). Since images and text are often interdependent in that they cannot be understood on their own, they have to be mentally integrated, which not only requires visual search, but also extraneous activity in working memory. This *split attention* phenomenon can largely be prevented by physically integrating the text with the visual material, for instance by imposing text fragments on those parts of the image that they refer to. Research has shown that this integrated format leads to better performance than conventional non-integrated formats (Bobis, Sweller, & Cooper, 1993; Chandler & Sweller, 1992; Sweller & Chandler, 1994; Sweller, Chandler, Tierney, & Cooper, 1990).

The Modality Effect

An alternative approach to deal with split attention is to address different sensory modalities during training. More specifically, visual material can

be presented through the visual channel and explanatory text via the auditory channel, which can easily be achieved with modern multimedia techniques. There is a long tradition of research exploring the effects of information presented via different modal channels (see Penney, 1989, for a review). Much of the theory in this area is based on the idea that the visual and auditory channels correspond to partly independent working-memory stores (e.g., Baddeley, 1992). An additional idea is the concept of “dual coding,” which emphasizes the benefits of encoding information in both visual and auditory form (e.g., Clark & Paivio, 1991). CLT has adopted these views and claims that addressing different sensory modalities reduces extraneous cognitive load in that it decreases visual search and the need for mental integration. Cognitive load is furthermore reduced, because incoming information is distributed over different limited-capacity stores. The overall reduction of instruction-related cognitive load increases the likelihood of grasping complex subject matter.

Although some researchers are skeptical concerning the value of “multi-representational” learning environments (e.g., Ainsworth, Wood, & Bibby, 1997), evidence in favor of multi-modal instructional environments is overwhelming. In the context of CLT, the modality effect was first demonstrated by Mousavi, Low, and Sweller (1995) with high school students in the domain of geometry. Subsequently, the effect was obtained by Mayer, Bove, Bryman, Mars, and Tapangco (1996) in the domain of meteorology, by Tindall-Ford, Chandler, and Sweller (1997) in the domain of electrical engineering, and most recently by Kalyuga, Chandler, and Sweller (1999, 2000) in other technical domains. A related set of studies by Mayer and associates (Mayer & Anderson, 1991; Mayer & Moreno, 1998; Moreno & Mayer, 1999) stresses the importance of *temporal contiguity* in the design of multimedia training. Temporal contiguity refers to the synchronization of pictorial and auditory information. Overall, it was found that adequate timing of narrated text plays an essential role in comprehension and recall of multimedia-based training material (see Mayer & Moreno, 2002, for an overview).

The aim of the present study was to investigate the modality effect with animated worked examples in both young and older adults by combining the modality effect, which springs from cognitive load theory, and the contiguity effect, which arises from Mayer’s cognitive theory of multimedia learning. This makes it a sequel to an earlier study (Van Gerven et al., Chapter 2), which investigated the relative efficiency of purely visual worked examples in the domain of Luchins’ (1942) water-jug problem. It was found that worked examples, which were presented as a set of pictures representing the subsequent solving steps of a problem, were more efficient relative to conventional problems in that less training time and less invested cognitive load led to a comparable performance level. These *unimodal* worked examples could still have caused a substantial extrane-

ous load, however. First, they required the learners to combine textual and pictorial information within a single picture. Second, they required integrating the information of up to six pictures within an example. Unnecessary search and integration processes could thus have been the result. A final source of extraneous load could have been the repetition of visual information, which is intrinsic to a step-by-step representation and is likely to cause unnecessary distraction.

A *multimedia* version of these worked examples, in which pictorial information is presented in an animated, and the text in an auditory form, will not have all the aforementioned disadvantages. For elderly learners multimedia-based worked examples are expected to have additional benefits. Optimal contiguity of visual and auditory information will minimize the probability of simultaneity problems (Salthouse, 1996). In addition, animations will avoid the repetition of invariable pictorial elements, which are likely to distract attention and cause search problems (McDowd, 1997). These notions bring us to formulate the following two hypotheses.

On the basis of Sweller et al.'s cognitive load theory and Mayer's cognitive theory of multimedia learning, Hypothesis 1 states that worked examples presented in a multimedia format are more efficient than both conventional problems and unimodally presented worked examples in that at least an equal level of performance can be obtained with less effort. Following Van Gerven et al.'s (2000, Chapter 1) framework, Hypothesis 2 states that elderly learners will profit more from multimedia-based worked examples relative to unimodal worked examples and conventional problems than their young counterparts. Moreover, both hypotheses imply that unimodal worked examples are superior to conventional problems, which was already demonstrated by Van Gerven et al. (2002a, Chapter 2).

Method

Participants

Sixty secondary school students (24 men and 36 women, mean age = 15.98 years, $SD = 0.77$) were recruited from a high school in Heerlen, The Netherlands. Sixty elderly participants (34 men and 26 women, mean age = 64.48 years, $SD = 4.92$) were recruited in Maastricht and the surrounding area via advertisements in local newspapers. All participants were in good health and had normal or corrected-to-normal vision. They received 40 Dutch guilders (about €18) payment for their participation.

On a shortened version of the Groningen Intelligence Test (GIT, Luteijn & Van der Ploeg, 1983), the elderly participants ($M = 126.83$, $SD = 8.38$) scored about 11 IQ points higher than their young counterparts ($M = 115.68$, $SD = 8.73$), $t(118) = -7.14$, $p < .001$. The GIT is a commonly used Dutch estimate of formal IQ. The shortened version includes four subtasks

(out of nine), which are supposed to give the best approximation of the full-scale IQ: vocabulary [woordenlijst], mental rotation [legkaart], mental arithmetic [cijferen], and analogies [woordmatrijzen]. The 95%-confidence interval of the shortened GIT is ± 10 points.

Working memory abilities were assessed with Salthouse and Babcock's (1991) computation-span test, which is a variant of the widely used reading-span test. During the computation-span test, the participant has to work out simple adding and subtraction sums, which were presented verbally (e.g., "7 plus 8," "9 minus 4," etc.). At the same time, the last digit of each sum (e.g., 8, 4, etc.) has to be memorized and reproduced after a series. A series comprises up to 7 sums and there are three trials for each. The computation-span score is determined by the maximum number of digits recalled in at least two trials. As expected, the computation span of the older participants ($M = 3.60$, $SD = 1.24$) was significantly lower than the computation span of the young participants ($M = 4.62$, $SD = 1.32$), $t(118) = 4.36$, $p < .001$.

Apparatus and Materials

Apparatus. The experiment was run on an IBM-compatible computer. Interaction with the computer took place via a Philips 17-inch monitor with an integrated touch screen (Elo TouchSystems, Fremont, CA). The software controlling the experiment was programmed in Authorware 3.5 (1996).

Subjective cognitive-load scale. A 9-point symmetrical category scale was used as a subjective cognitive-load (SCL) measure (after Paas, Van Merriënboer, & Adam, 1994). The scale was presented on the touch screen as a slide control and was accompanied by the phrase: *I have experienced the foregoing as: ...* The scale had three labels, ranging from *not difficult at all* (1), via *moderately difficult* (5), to *very difficult* (9). Intermediate values had no labels. Paas (1992) found an internal consistency coefficient (Cronbach's α) of .90 using a comparable scale. Paas and Van Merriënboer (1994b) evaluated the scale as a highly reliable and sensitive instrument for the assessment of cognitive load ($\alpha = .82$).

Water-jug problem. A computerized version of Luchins' (1942) water-jug problem was selected as the task domain. The goal of the water-jug problem is to acquire a certain amount of water that had to end up in a target jug. For that purpose, the participants had an infinite amount of water at their disposal and a set of so-called "working jugs" of different sizes. The working jugs could be filled and poured into each other, but they could never overflow. That is, when a bigger jug was poured into a smaller one, pouring stopped once the smaller one was full. Thus, a residual remained in the bigger jug. The goal could be reached by adequately pouring working jugs into each other and strategically making use of the residuals (for more details, see Van Gerven et al., 2002a, Chapter 2).

Interface. The interface of the water-jug problem is depicted in Figure 2.2 (Chapter 2). The display contains several objects: a tap in the upper left corner, a button labeled *Empty* in the upper right corner, a set of working jugs in the lower left corner, and a target jug in the lower right corner. Contents are indicated inside the jugs, capacities beneath the jugs. Touching meaningful combinations of two objects resulted in three possible operations. Touching the tap in combination with one of the working jugs resulted in filling that particular working jug. Touching the *Empty* button and one of the working jugs or the target jug resulted in emptying that jug. Touching a working jug and then touching another working jug or the target jug resulted in pouring water from the former to the latter.

Training problems and conditions. Four basic water-jug problems were selected as training problems. These were presented in three different forms, corresponding to the three experimental training conditions. In the conventional problems (CP) condition, the participant was presented with the initial state of a problem, after which the goal state had to be attained independently. In the unimodal worked examples (UWE) condition, the training problems were presented in a cartoon-like fashion: The initial state, the intermediate states, and the goal state were represented as a sequence of pictures. Explanatory text was imposed on the images. The CP and the UWE conditions were identical to the conventional-problems and worked-examples condition in the Van Gerven et al. (2002a, Chapter 2) study. In the multimedia-based worked examples condition (MWE), finally, the training problems were presented in an animated form, which was accompanied by a narrated explanation. The contents of this verbal explanation resembled the text in the UWE condition as much as possible. The animations and corresponding narration were recorded with Lotus ScreenCam (1997). In accordance with Moreno and Mayer's (1999) contiguity principle the events in the animations were temporally synchronized with the narrative's content.

Test problems. The test consisted of 12 water-jug problems. The problems were sequenced such that the test gradually increased in difficulty. This was done to prevent discouragement in the elderly participants.

Design and Data Collection

Participants of both age groups were randomly assigned to one of the three experimental training conditions ($n = 20$). This yielded a basic 2 (age group) \times 3 (training format) between-groups design, with age group comprising the levels young and old, and training format comprising the levels conventional problems (CP), unimodal worked examples (UWE), and multimedia-based worked examples (MWE).

The core dependent variable was training efficiency, which reflects the ratio between subjective cognitive load (SCL) and performance. A training format can be considered inefficient if SCL during training is relatively

high and performance during the test is relatively low. In contrast, a training format can be considered efficient if SCL is relatively low and performance is relatively high. Paas and Van Merriënboer's (1993) procedure was followed to convert SCL and performance scores into efficiency scores. In this procedure, a training format is represented as a dot in a co-ordinate system with the mean z score of SCL on the x-axis and the mean z score of performance on the y-axis (see Figure 3.2a). The z scores are calculated with the grand means and standard deviations of an age group. The training efficiency (E) score is determined by the perpendicular distance between a dot and the diagonal $E = 0$, where SCL and performance are in balance. Thus, E is calculated as follows:

$$E = \frac{z_{\text{performance}} - z_{\text{SCL}}}{\sqrt{2}}.$$

Other dependent variables were subjective cognitive load (SCL) during training, training time (in seconds), and performance (proportion of solved problems).

Procedure

The experiment was administered in individual sessions. A session consisted of a test block, comprising the intelligence and computation-span test, and the experiment itself. The order of these two parts was counter-balanced to control for fatigue: One half of the participants first went through the test block and then through the experiment, the other half first went through the experiment and then through the test block. The experiment consisted of a general introduction, presented as text and illustrations on the computer screen, a practice session to get acquainted with the interface, a training session (depending on the training format condition), and a final test. During the experiment, the dependent variables training time, SCL, and performance were recorded.

Results

The data were analyzed with a 2×3 (age group \times training format) between-groups MANOVA. In addition, all contrasts (i.e., CP-MWE, UWE-MWE, and CP-UWE) were analyzed with separate 2×2 between-groups ANOVAs (applying a Bonferroni correction by dividing the .05 significance level by three: $\alpha = .017$). The means and standard deviations of the dependent variables are given in Table 3.1. As described in the Method section, training efficiency is a composite of SCL and performance. Before reporting the ANOVA results of training efficiency, an account is given of

Table 3.1 Means and standard deviations per age group and training format.

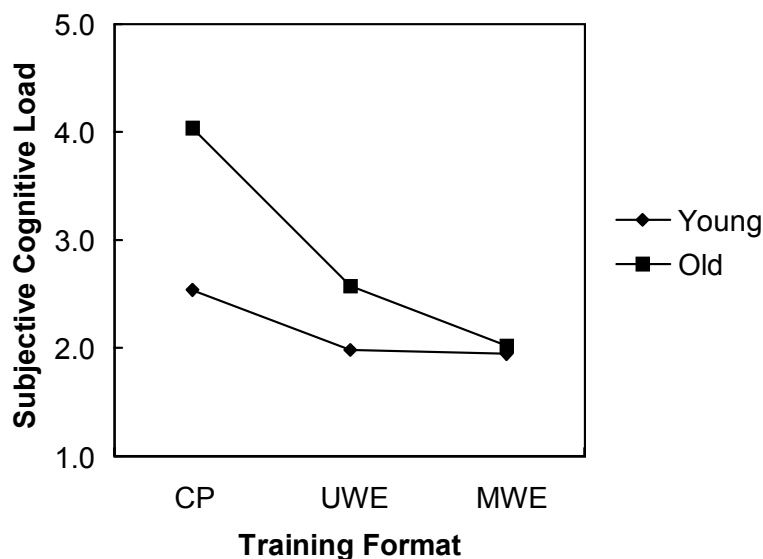
		Training Format					
		CP		UWE		MWE	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Young (<i>n</i> = 60)	SCL	2.54	1.16	1.98	1.00	1.94	1.46
	Performance	.90	.10	.84	.18	.91	.14
	Training efficiency	-0.10	0.82	-0.11	1.36	0.29	1.19
	Training time	49.70	25.01	30.33	10.63	70.88	3.91
Old (<i>n</i> = 60)	SCL	4.04	1.81	2.58	1.51	2.01	1.26
	Performance	.46	.26	.53	.20	.54	.22
	Training efficiency	-0.62	1.50	0.17	0.92	0.45	0.98
	Training time	123.06	68.48	110.60	74.23	77.88	10.58

Note. Performance is indicated in the proportion of solved problems. Training time in seconds.

these constituent variables. Finally, the results of training time are presented.

Subjective Cognitive Load

With respect to SCL, a main effect of age group, $F(1, 114) = 8.15$, $MSE = 1.94$, $p < .01$, indicates that the elderly participants experienced higher levels of cognitive load during training than the young (see Figure 3.1). A main effect of training format, $F(2, 114) = 9.78$, $p < .001$, signifies that the level of SCL during training with worked examples is lower than the level of SCL in the conventional-problems condition. There is no significant age group \times training format interaction, $F(2, 114) = 2.68$, $p = .073$, suggesting

**Figure 3.1** Mean subjective cognitive load.

that the effect of training format is equal for the age groups (although the low p value indicates a trend towards an interaction). For the contrast CP-MWE, there are both effects of age group, $F(1, 76) = 5.94$, $MSE = 2.09$, $p = .017$, and training format, $F(1, 76) = 16.51$, $p < .001$. Furthermore, there is a trend towards an interaction, $F(1, 76) = 4.87$, $p = .03$, indicating that the elderly show a greater decrease in SCL between the conventional and the multimedia condition than the young. The contrast UWE-MWE, on the other hand, reveals neither a main effect, nor an interaction. This means that the multimedia condition does not reduce the level of SCL relative to the unimodal condition. It is interesting, though, that neither the unimodal, nor the multimedia condition reveals any significant age difference in SCL. The CP-UWE contrast, finally, shows a main effect of age group, $F(1, 76) = 11.15$, $MSE = 1.98$, $p < .01$, as well as training format, $F(1, 76) = 10.37$, $p < .01$. This is a replication of Van Gerven et al.'s (2002a, Chapter 2) finding that worked examples impose less cognitive load on the learner than conventional problems. However, unlike Van Gerven et al., no interaction effect was found.

Performance

With regard to performance, there is only a main effect of age group, $F(1, 114) = 114.93$, $MSE = .04$, $p < .001$, indicating that the young performed better than the elderly. However, there is neither an effect of training format, nor an interaction. The contrast UWE-MWE again reveals an effect of age group, $F(1, 76) = 89.10$, $MSE = .03$, $p < .001$, but no effect of training format or an interaction. Similar results are found for the contrasts UWE-MWE and CP-UWE. In both cases, there is only an effect of age group: $F(1, 76) = 66.70$, $MSE = .04$, $p < .001$, and $F(1, 76) = 74.70$, $MSE = .04$, $p < .001$, respectively. The other effects are non-significant.

Training Efficiency

Figure 3.2a represents training efficiency as the mean z scores of SCL and performance in a co-ordinate system (following Paas & Van Merriënboer, 1993). Figure 3.2b depicts the mean values of training efficiency. The analysis of variance reveals no main effect of age group. There is an effect of training format, however, $F(2, 114) = 3.97$, $MSE = 1.33$, $p < .05$. This effect can be fully attributed to the contrast between CP and MWE, because this is the only contrast for which the 2×2 ANOVA reveals a significant main effect, $F(1, 76) = 7.96$, $MSE = 1.33$, $p < .01$. That is, the multimedia training format is significantly more efficient than the CP format. Neither the overall ANOVA, nor the contrast ANOVAs reveal a significant age group \times training format interaction, indicating that the effect of training format did not differ between the age groups.

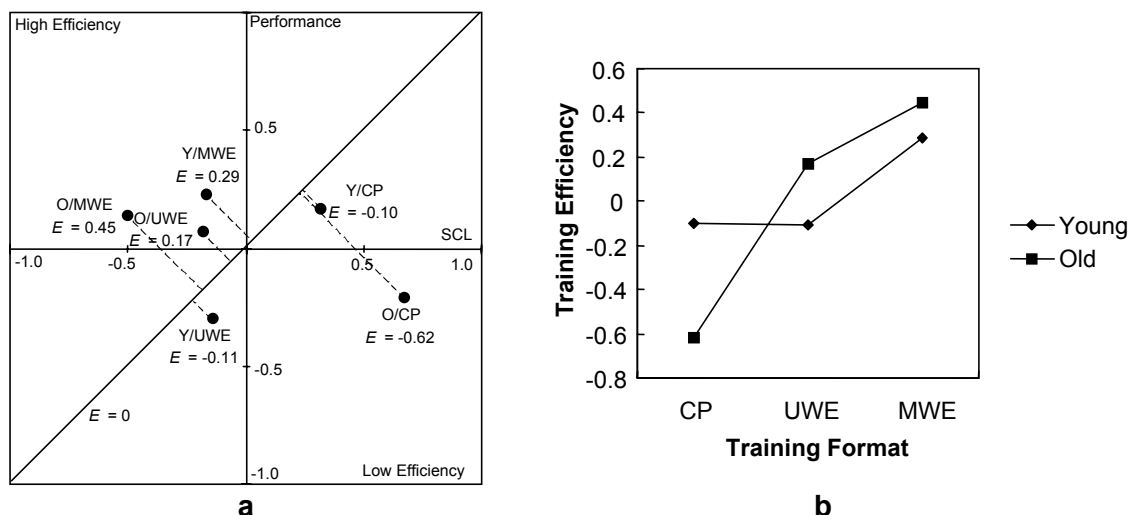


Figure 3.2 Mean training efficiency.

Training Time

Mean training time is depicted in Figure 3.3. There is a main effect of age group, $F(1, 114) = 47.33$, $MSE = 1087.00$, $p < .001$: the elderly needed more training time than their young counterparts. Furthermore, there is a main effect of training format, $F(2, 114) = 11.01$, $p < .001$, and an age group \times training format interaction, $F(2, 114) = 10.17$, $p < .001$. The young participants show a strong increase in training time between the UWE and the MWE condition, which is due to the fixed length of the animations in the MWE condition. That is, the young were compelled to spend a minimum

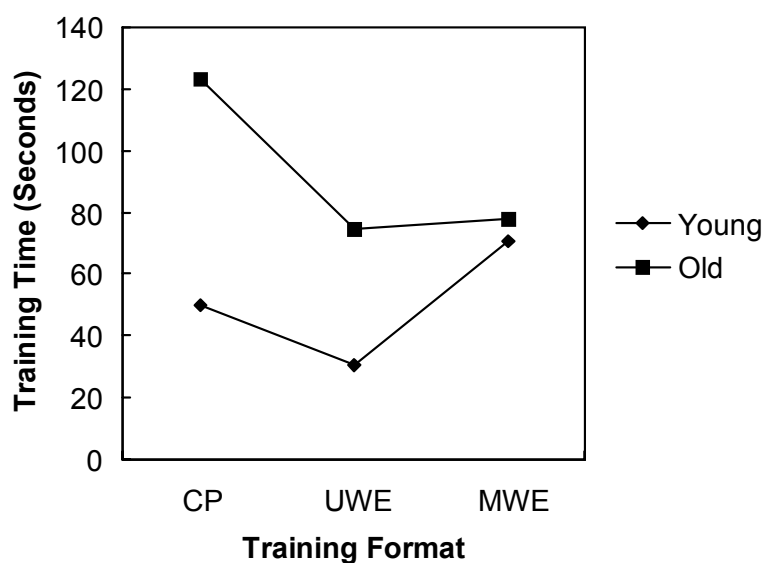


Figure 3.3 Mean time spent per training problem in seconds.

training time in this condition. The main effect of training format is also found in the contrast UWE-MWE, $F(1, 76) = 32.29$, $MSE = 302.32$, $p < .001$. The interaction is significant for the contrast CP-MWE, $F(1, 76) = 16.18$, $MSE = 1360.00$, $p < .001$, indicating that the elderly save more training time in the multimedia condition than the young. There is also an interaction effect for the contrast UWE-MWE, $F(1, 76) = 22.52$, $p < .001$, which suggests that the young were indeed required to spend more time in the multimedia condition than in the unimodal condition. The elderly, on the other hand, seem to retain the same amount of training time.

Discussion

The purpose of this study was to investigate the efficiency of training with multimedia-based worked examples relative to training with unimodal worked examples and conventional problems in both young and older adults. Two hypotheses were stated. On the basis of cognitive load theory it was hypothesized that multimedia-based worked examples are more efficient than unimodal worked examples and that unimodal worked examples are more efficient than conventional problems (Hypothesis 1). These effects were expected to be stronger for the elderly than for the young, since for the former the enhancement would be proportionally larger than for the latter (Hypothesis 2; Van Gerven et al., 2000, Chapter 1).

Participants studying worked examples of both formats invested significantly less mental effort than participants solving conventional problems, which is in line with Hypothesis 1 and Van Gerven et al. (2002a, Chapter 2). Furthermore, the interaction trend for the contrast between conventional problems and multimedia-based worked examples suggests a disproportional advantage for the elderly participants, which supports Hypothesis 2. The analyses of the efficiency scores, which reflected the ratio between invested cognitive load and performance, did not yield this interaction, however, which does not support Hypothesis 2. There was a main effect of training format, though. Contrast analyses revealed that this effect was fully attributable to the difference between the conventional and multimedia condition. A remarkable additional finding was that in the contrast analysis of the unimodal and multimedia condition there was no effect of age group, suggesting that the age groups experienced comparable levels of cognitive load. For both age groups, training time was lower in the unimodal worked-examples condition than in the conventional-problems condition. Between the unimodal and multimedia-based worked examples there was no difference in training time for the elderly participants. The young participants, however, were compelled to spend more time in the multimedia condition.

The key question to be answered is whether multimedia-based worked examples add to the efficiency of skill training for the elderly. If the definition of efficiency is restricted to the ratio between cognitive load and performance (Paas & Van Merriënboer, 1993), the answer is negative. In a broader perspective, however, there are at least three reasons to take the affirmative view. First, multimedia-based worked examples led to a significantly lower perceived cognitive load for the elderly than for the young. Second, the difference in training efficiency turned out to be significant only for the contrast between conventional problems and multimedia-based worked examples, although this applies to both the old and the young participants. Third, with respect to training time, there is an interaction between age group and training format for the contrast between the conventional and multimedia condition. That is, the elderly need significantly less time to process multimedia-based worked examples, whereas the young take significantly more training time. In part, this effect is due to the minimum time the young participants were compelled to spend in the multimedia condition. For the greater part the effect seems to be attributable to the fact that the elderly did not choose to take more time in this condition, in spite of the opportunity to repeat the animations. Pollock, Chandler, and Sweller (2002) even suggest that training time can be used as an additional ingredient of training efficiency.

The ultimate purpose of a training design is not only to make an efficient use of the available processing capacity, but also to realize superior performance. Although the present study does not provide direct support for that, it strongly suggests that enhanced performance induced by unimodal or multimedia-based worked examples is eventually attainable. As we have seen, the multimedia condition imposes significantly less cognitive load than the conventional condition. This offers the opportunity to add so-called *germane* cognitive load (Sweller, Van Merriënboer, & Paas, 1998), which contributes to the construction of cognitive schemata. In future research, this can be tested by increasing the complexity and variability of the training examples (see Paas & Van Merriënboer, 1994b), which could eventually increase efficiency by enforcing a performance difference. In its turn, increased efficiency is likely to boost motivation, making life-long learning not only a necessary, but also an enjoyable experience.

4 Cognitive Load and the Pupillary Response: An Intermezzo¹

The differential effects of memory load on task-evoked pupillary responses (TEPRs) among 16 young adults and 16 old adults were investigated. Mean pupil dilation, peak dilation, peak latency, and reaction time were determined in a memory-search task, which involved six levels of memory load. The reaction times revealed a classic interaction pattern in that they disproportionately increased as a function of memory load in the elderly. All pupillary responses varied considerably as a function of age. In contrast to the young participants, whose mean dilation increased as a function of memory load, the older participants' mean dilation was not sensitive to memory load. Peak dilation and peak latency did not vary with memory load in both age groups. The results suggest that the usefulness of pupillary responses as correlates of memory load diminishes with old age.

Numerous attempts have been made to find reliable and sensitive physiological correlates of cognitive load in various 'real life' situations, such as studying instructional material and flying an airplane (e.g., Paas, Van Merriënboer, & Adam, 1994; Wierwille, Rahimi, & Casali, 1985). In addition to the widely used subjective scaling techniques (e.g., Nygren, 1991; Paas & Van Merriënboer, 1994b; Van Gerven, Paas, Van Merriënboer, & Schmidt, 2002a, Chapter 2), there is still a need for these more 'objective' measures. One of the most promising physiological measures, both in terms of reliability and sensitivity, is the pupillary response. The aim of the present study is to test whether the pupillary response, regarding its liability to physical change, retains its quality as a measure of cognitive load into old age.

Pupillometry had its bloom in the 1960s and early 70s. Although its history actually begins a little earlier, pupillometric research aimed at processes that nowadays would be attributed to working memory was initiated by Hess and Polt (1964). They found pupil diameter to become larger in people who were engaged in solving multiplication problems of increasing difficulty. Kahneman and Beatty (1966) were the first to relate pupil size to memory load. Participants were verbally presented with an increasing number of digits, which had to be reproduced after a short pause. It was found that pupil size increases during presentation and de-

¹ A version of this chapter is submitted for publication as: Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., & Schmidt, H.G. (2002b). Memory load and task-evoked pupillary responses in aging.

creases during recall. During the entire procedure, however, pupil diameter was larger when the number of digits was higher. Peavler (1974) investigated what happens to the size of the pupil when memory is overloaded by an excessive number of digits. Similar to earlier research, he found that the pupil widened as the number of digits increased. However, when the number of digits exceeded nine, the diameter of the pupil stabilized, which in his view indicates a “momentary suspension of processing effort” due to the perceived risk of overload.

It is peculiar that after this successful initial period (see Janisse, 1977, for a review), which established the power of pupillometry as a tool for measuring mental activity, it subsequently seemed to pass into oblivion. Strong attempts to inject new life into it were done by Ahern and Beatty (1979) and Beatty (1982). Only throughout the last decade, however, pupillometry as an instrument for measuring cognitive load seems to have regained a great deal of its initial appeal. This period marks a number of new directions in such areas as physiology and pharmacology (Matthews, Middleton, Gilmartin, & Bullimore, 1991), ergonomics (Backs & Walrath, 1992), computational modeling (Hoeks & Levelt, 1993), language processing (Hyönä, Tommola, & Alaja, 1995; Just & Carpenter, 1993), visual perception (Verney, Granholm, & Dionisio, 2001), and abnormal psychology (Granholm, Morris, Sarkin, Asarnow, & Jeste, 1997). There is also a renewed interest in pupil size as an indicator of resource limits and memory overload. Granholm, Asarnow, Sarkin, and Dykes (1996), for instance, found a progressive increase of pupil size with memory load, which stabilized at resource limits (i.e., nine digits). Contrary to the results of Peavler (1974), however, pupil size declined after the processing demands exceeded the available resources.

Thus, it seems that the new boost of research has consolidated the value of pupillometry as a psychometric instrument. That is, within the limits of cognitive functioning it has proven to be a reliable and accurate measure of mental activity. However, little is known about its value for measuring mental load in elderly people. It is widely accepted, though, that pupil size substantially decreases with old age, a phenomenon commonly referred to as *senile miosis*. Much of the predominantly medical literature on senile miosis (e.g., Bitsios, Prettyman, & Szabadi, 1996; Korczyn, Laor, & Nemet, 1976; Loewenfeld, 1979; Sekuler & Owsley, 1983; Winn, Whitaker, Elliott, & Phillips, 1994) focuses on its biological basis and its consequences for peripheral functions, such as the pupillary light reflex. The core message is that senile miosis results from a degenerating dilation muscle (the dilator pupillae) in the iris, which causes a linear decline in pupil size with age. It is important to note that the dilator pupillae is controlled by sympathetic pathways from the central nervous system (in contrast to the constriction muscle — the sphincter pupillae — which is controlled by the parasympathetic system). Since the sympathetic nervous

system is associated with activation, there is reason to believe that task-evoked dilation of the older pupil — if detectable — is an attenuated correlate of mental activity. However, none of the consulted literature mentions an effect of senile miosis on the cognitive pupillary response.

The same holds for the more psychologically oriented literature. There are hardly any age-related studies aimed at pupillary responses that can be attributed to cognitive activity. One of the few comes from Morris, Granholm, Sarkin, and Jeste (1997), who related pupil dilation to memory performance in both healthy and schizophrenic participants ranging in age from 26 to 75 years. They presented their participants with series of four, six, eight, and ten digits and measured the pupillary response at last-digit presentation and at digit retrieval. The participants were required to recall as many digits as possible. Because Morris and colleagues were primarily interested in performance differences between the schizophrenics and the controls at high memory-load levels, they only report data of the eight- and ten-digits condition. Differences between these conditions were not statistically tested, however. Moreover, the reported means are based on the whole age range. Regression analyses revealed a decline of pupillary response with increasing age. This holds for both pupil dilation at last-digit presentation and pupil dilation at digit retrieval.

One of the most recent pupillometric studies comparing young and older adults comes from Kim, Beversdorf, and Heilman (2000). These researchers were interested in pupillary responses induced by novel stimuli. They required both young (mean age = 30 years) and older participants (mean age = 60 years) to attend to a digit presented in the center of a screen. At certain moments a new digit appeared somewhere in the periphery of the screen. The subsequent arousal response was recorded after the participant's gaze had shifted to the new stimulus. The data revealed a typical response pattern, comprising three phases: a brief constriction of the pupil, followed by a dilation — the “arousal response” — and a second constriction or “habituation response.” It was found that the interval between the arousal response and the habituation response is longer in the elderly than in the young participants. Remarkably, however, no significant age-related difference was found in the magnitude of the dilation response.

Apparently, there is no research that has shown an effect of different levels of ‘normal’ memory load on pupil size in elderly persons. What we do know is that the pupil in older people dilates as a result of excessive memory load and stimulus novelty. We also know that, generally, the older pupil shows less dilation than the young pupil as a result of a physiological decline of the dilator pupillae. The question remains whether in spite of this decline, subtle variations in memory load — that is, within the classical limits of working memory (Miller, 1956) — result in a detectable pupillary response.

An appropriate task in this respect is Sternberg's (1966) memory-search task. In this task, sequences of one to six digits have to be memorized. After memorizing a sequence, a series of single-digit probes is presented. For every probe, the participant has to decide whether it belongs to the memory set or not. A positive probe belongs to the memory set, a negative probe does not. Response latencies typically increase as the size of the memory set is extended. Sternberg did not find effects of the serial position of a positive probe in the memory set. This suggests that the search process is exhaustive, rather than self-terminating (i.e., the search process stops as soon as the probe is found). If, however, people were asked to indicate this position, reaction times increased as the probe was located further in the sequence (Sternberg, 1967). The Sternberg task entails two different working-memory activities: encoding and searching. Encoding requires more effort as the memory set becomes larger. Thus, higher levels of pupil dilation can be expected for larger memory sets (e.g., Granholm et al., 1996). In the same light, searching larger memory sets is more effortful and can thus be expected to lead to more pupil dilation.

Beatty and Lucero-Wagoner (2000) describe three useful task-evoked pupillary responses (TEPRs), which can be treated similarly to event-related potentials. First, *mean dilation* during a critical post-stimulus time interval. Second, *peak dilation*, which is defined as the maximum dilation during this interval. And third, *peak latency*, which is the time elapsed between probe onset and the appearance of the peak dilation. The magnitude of these TEPRs typically rises as the burden on working memory becomes more intense. Mean dilation is the most commonly used TEPR. Peak dilation and peak latency bear the restriction that they are only useful in case of a single stimulus. Therefore, these TEPRs will only be used in the search phase of the Sternberg task, which involves single-digit probes.

On the basis of the literature discussed, it can be expected that the elderly will show less dilation than the young. Regarding Morris et al.'s (1997) study, in which working memory was loaded to the limit and the young still exhibited more pupil dilation than the elderly, it can furthermore be expected that the elderly's pupil dilation will not exceed that of the young, even though the elderly will probably experience higher levels of cognitive load. Furthermore, because peak dilation is assumed to be an indicator of the decision process, it is hypothesized that there is a positive correlation between peak latency and reaction time in both age groups.

Modern theories on age-related cognitive slowing (e.g., Salthouse, 2000) predict that the reaction times of the elderly will be much longer than those of the young. Furthermore, there is the generally found *complexity effect*, which implies an interaction between age and task complexity in that performance differences between young and old people become larger when the complexity of the task increases (e.g., Craik & Byrd, 1982; Perfect & Maylor, 2000). The complexity effect is likely to occur in the re-

response latencies resulting from the Sternberg task. The question rises whether the complexity effect will also be reflected by the TEPRs. That is, will the elderly show relatively more dilation and longer peak latencies in the higher memory sets? Regarding the senile-miosis literature, the effect will at least show a different pattern. That is, the young will display larger TEPRs than the elderly, but in the larger memory sets the TEPRs of the elderly will grow towards the TEPRs of the young. That is, for both age groups the TEPRs will increase, but this increase will be stronger for the elderly. Hence, the gap between the age groups will become smaller. If there is a positive correlation with reaction time, peak latency is the most probable TEPR to reveal such an interaction effect.

Method

Participants

Participants were 16 health-science, medicine, and psychology students of Maastricht University (1 man and 15 women, mean age = 20.38 years, $SD = 2.16$) and 16 elderly adults (11 men and 5 women, mean age = 68.63 years, $SD = 4.69$), who were recruited via advertisements in local newspapers. All participants were in good health and had normal or corrected-to-normal vision. None of them took medication or had ever suffered an eye disorder that could influence their pupillary response. They received 10 Dutch guilders (about €4.50) for their participation.

Memory-Search Task

In the memory-search task (Sternberg, 1966, 1967), participants are required to memorize a certain number of randomly selected digits (ranging from 0 to 9), in which none of the digits occurs more than once. The number of digits in this so-called *memory set* varies from one to six. In a series of trials, the participant is presented with a random probe, which is a single digit. The participant has to decide as quickly as possible whether this probe is in the memory set (positive probe) or not (negative probe).

Apparatus

Pupil size was measured with the SMI (SensoMotoric Instruments, Berlin, Germany) Remote Eyetracking Device (RED) with an angular resolution of less than 0.5° . This device consists of an infrared source, which is aimed at the participant's eye at a distance of about 90 cm, and an infrared camera. Rays from the infrared source are reflected by the retina and picked up by the camera, which generates a high-resolution image of the pupil on a computer screen. The participant was seated in an adjustable chair and his or her head was stabilized by a chin and forehead rest. At a frequency of 50 Hz, that is every 20 ms, both the horizontal and vertical pupil diame-

ters were measured in screen pixels, which subsequently were transformed into millimeters. Stimuli were presented by an IBM-compatible computer on a 21-inch monitor, which was placed at a distance of about 110 cm from the participant. The size of the stimulus digits on the screen was approximately 2.5×1.5 cm. The software controlling the stimulus presentation was programmed in ERTS (Experimental Run Time System, Beringer, 1996). The participant responded by pressing one of two response buttons. Reaction times were measured in milliseconds.

Design and Procedure

Age group (levels young and old) was the independent between-groups variable and memory set (levels 1 to 6) was the independent within-groups variable. The function of the response buttons was counterbalanced. That is, half of the participants had to press the left button if a negative probe was presented and the right button if a positive probe was presented; the other half of the participants had to press the opposite keys. The order of the memory-set presentation was counterbalanced by presenting one half of the participants with an increasing number of digits (1 to 6), whereas the other half was presented with a decreasing number of digits (6 to 1). The number of positive and negative probes was balanced as well.

The experiment was performed in individual sessions. After calibrating the RED, the participant was presented with a concise onscreen instruction concerning the memory-search task. There was a practice session with a memory set of two digits and eight response trials. Subsequently, the first experimental memory set was presented. The digits of a memory set were presented simultaneously. After memorizing a set, the participant had to push a button to continue. He or she was then presented with a sequence of 24 trials, each comprising a random probe. A 1000 Hz warning tone was presented 1000 ms before each probe onset. The inter-trial interval was 5000 ms, which was considered sufficient for the pupil to return to baseline level. The participant did not receive feedback about the correctness of a response.

Data Processing

Eye blinks were filtered out on the basis of sudden drops in vertical pupil diameter (since this is the direction in which the eyelids move). The criterion for detecting a drop was the ratio between the vertical (d_v) and horizontal diameter (d_h). For a perfectly circular pupil, the ratio d_v/d_h may be expected to lie around 1. During a blink or semi-blink, however, this ratio quickly drops towards zero. All data with a ratio differing more than one standard deviation from the mean were replaced by linear interpolation. Subsequently, the pupil area (A) was calculated as the area of an ellipse: $A = \frac{1}{4}\pi d_h d_v$. The baseline pupil area was determined for each trial by the pupil's average size during a period of 500 ms preceding the warning tone.

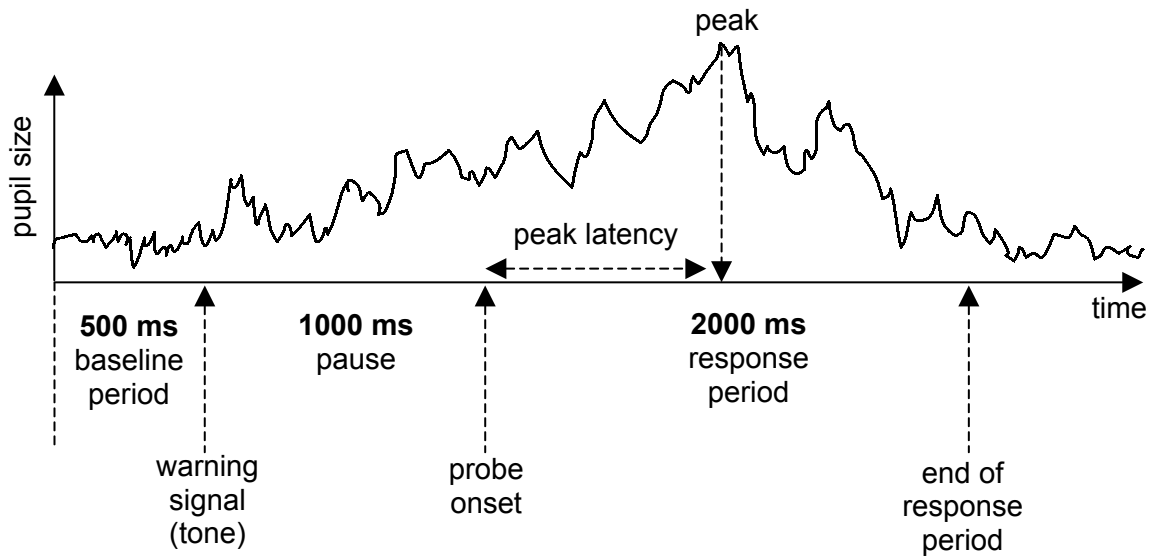


Figure 4.1 The time course of an experimental trial.

Pupil dilation or constriction was calculated by subtracting this baseline area from the pupil area during the response period, which was a period of 2000 ms following probe onset (see Figure 4.1).

Results

With respect to the pupillary responses, three dependent variables were determined: mean dilation, peak dilation, and peak latency (after Beatty & Lucero-Wagoner, 2000). For the encoding phase only mean dilation was calculated. For the search phase dilation-by-time plots revealed that the dilation curves pertaining to the different memory sets started to diverge around 650 ms after probe onset, which roughly corresponds to the average pupillary response latency found in earlier research (see, e.g., Beatty, 1982). Therefore, mean dilation was determined as the average dilation during a time window of 650 to 2000 ms after probe onset. Mean dilation was also determined during the encoding phase. Peak dilation was the maximum dilation during the response period. Peak latency was determined as the time elapsing between probe onset and the peak (see Figure 4.1). If more than one peak of the same magnitude occurred during the response period, the first peak was selected to determine peak latency.

Mean dilation, peak dilation, and peak latency were calculated by averaging the data across trials. Values corresponding to an incorrect response were replaced by the mean score of the participant within a memory set. In this way, 3.8% of the young and 2.5% of the old participants'

Table 4.1 Means and standard deviations (indicated in parentheses) of RT, mean dilation (MD), peak dilation, and peak latency.

		Memory set					
		1	2	3	4	5	6
Young (<i>n</i> = 16)	RT	507.03 (75.60)	584.02 (75.95)	612.68 (72.36)	682.26 (102.31)	736.86 (86.01)	763.95 (128.95)
	MD	-0.73 (1.27)	-0.62 (1.50)	-0.13 (2.10)	-0.09 (1.81)	0.58 (1.91)	-0.27 (2.24)
	MD	0.96 (1.41)	0.84 (1.25)	1.03 (1.26)	1.24 (1.40)	1.24 (1.85)	1.45 (1.84)
	searching						
	Peak	2.40 (1.84)	2.20 (1.46)	2.43 (1.40)	2.45 (1.63)	2.53 (1.98)	2.71 (1.74)
	dilation						
	Peak	915.00 (364.75)	951.25 (270.40)	957.50 (321.24)	1011.25 (290.38)	990.00 (463.38)	1153.75 (526.61)
	latency						
Old (<i>n</i> = 16)	RT	711.86 (144.55)	768.48 (124.31)	854.85 (139.99)	972.81 (245.36)	1089.13 (275.97)	1210.00 (275.78)
	MD	-0.38 (1.07)	-0.23 (1.22)	0.24 (1.47)	0.61 (1.31)	0.53 (1.49)	-0.16 (1.65)
	MD	0.36 (0.29)	0.48 (0.32)	0.36 (0.48)	0.33 (0.70)	0.34 (0.74)	0.25 (0.86)
	searching						
	Peak	0.98 (0.36)	1.14 (0.48)	1.07 (0.48)	1.08 (0.54)	1.03 (0.54)	0.99 (0.65)
	dilation						
	Peak	1156.25 (338.41)	1253.75 (407.90)	1138.75 (432.08)	1262.50 (456.24)	1222.50 (605.61)	1021.25 (648.87)
	latency						

Note. Mean dilation and peak dilation are indicated in square millimeters. Reaction time and peak latency are given in milliseconds.

data was substituted. The corrected data sets were submitted to an analysis of variance. In addition, correlations were calculated between reaction time and peak latency. Means and standard deviations of the TEPRs and reaction times can be found in Table 4.1. Results of the ANOVAs are summarized in Table 4.2.

Encoding

Analysis of mean dilation in the encoding phase reveals a main effect of memory set, $F(5, 150) = 2.63$, $MSE = 2.33$, $p < .05$. It is striking, however, that there is neither an effect of age group nor an interaction (see Table 4.2). Although the elderly pupil seems to dilate more than the young pupil

Table 4.2 *F* values, significance levels, and mean square errors resulting from the analyses of variance.

	AG	MS	AG × MS	MSE_{between}	MSE_{within}
RT	35.49***	70.24***	8.25***	111,197.97	9478.39
MD encoding	0.92	2.63*	0.15	4.17	2.33
MD searching	4.36*	0.72	2.34*	6.58	0.29
Peak dilation	11.90**	0.47	1.35	7.98	0.27
Peak latency	3.47 ^a	0.32	1.36	444,582.08	146,009.75

Note. * $p < .05$, ** $p < .01$, *** $p < .001$, ^a $p = .07$.

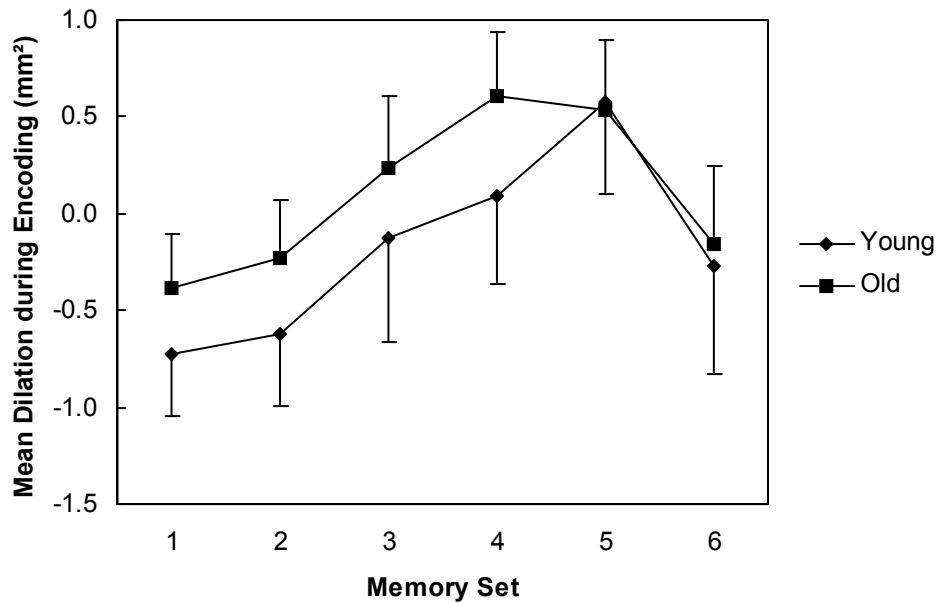


Figure 4.2 Mean pupil dilation during encoding as a function of memory set. Values are differential scores relative to baseline pupil size. Bars indicate standard errors of the mean.

(see Figure 4.2), both seem to be similarly affected by memory set. Remarkably, both curves show a downward bend at memory set 6. After removing memory set 6, the test of contrasts reveals a significant linear main effect of memory set, $F(1, 30) = 14.12$, $MSE = 2.03$, $p = .001$.

Memory Search

Reaction time. Mean reaction times yield a complexity effect (see Figure 4.3). That is, both age groups show an increase of reaction time as a function of memory set, but the elderly show higher reaction times and a higher slope than the young. This classic pattern is confirmed by the analyses. There is a highly significant main effect of age group, $F(1, 30) = 35.49$, $MSE = 111,197.97$, $p < .001$. Furthermore, there is a main effect of memory set, $F(5, 150) = 70.24$, $MSE = 9478.39$, $p < .001$. A significant age group \times memory set interaction effect, $F(5, 150) = 8.25$, $p < .001$, complements the pattern.

Mean dilation. Since the mean reaction times are typical in that they suggest a complexity effect, where the elderly are more affected by increased memory load than the young, the question remains whether these effects are reflected by the mean dilation data. The analyses show that this is partly true. As Figure 4.4 already indicates, there is a significant main effect of age group, $F(1, 30) = 4.36$, $MSE = 6.58$, $p < .05$, where the young participants show a higher pupil dilation than their old counterparts. This is in line with the hypothesis. However, no main effect of mem-

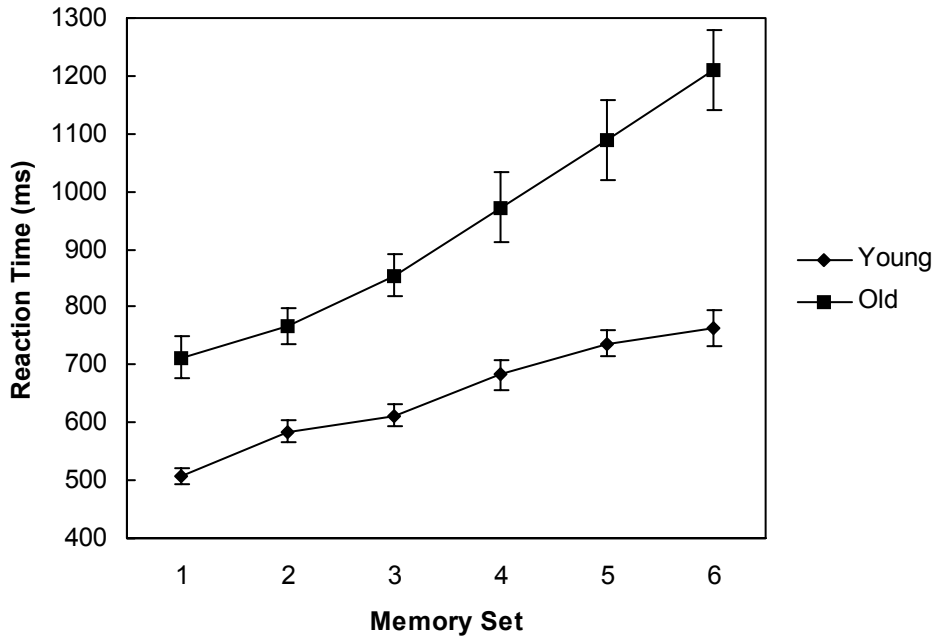


Figure 4.3 Reaction time as a function of memory set. Bars indicate standard errors of the mean.

memory set could be detected, $F(5, 150) = 0.72$, $MSE = 0.29$, $p > .05$. There is a significant interaction, however, $F(5, 150) = 2.34$, $p < .05$. Regarding the pattern in Figure 4.4, this elicits the suspicion that unlike the old participants, who more or less show a flat line, the young participants seem to be affected by an increasing memory set. This suspicion is intensified when the data of the young are analyzed separately. In this reduced data set a trend towards a main effect of memory set is observable, $F(5, 75) = 2.07$, $MSE = 0.39$, $p = .08$.

Peak dilation and peak latency. Similar to average dilation, the analysis of peak dilation yields a main effect of age group, $F(1, 30) = 11.90$, $MSE = 7.98$, $p < .01$. However, no main effect is observed with respect to memory set, $F(5, 150) = 0.47$, $MSE = 0.27$, $p > .05$. Moreover, there is no significant interaction, $F(5, 75) = 1.35$, $p > .05$. Analysis of the peak latency data does not reveal any significant effects. The same holds for the separate analyses of the negative- and positive-probe data (see Table 4.2). A trend can be observed towards an effect of age group, however, $F(5, 150) = 3.47$, $MSE = 146,009.75$, $p = .07$. This suggests that on the average the elderly pupil responds slower than the young pupil.

Finally, in line with the hypothesis, there is a positive correlation between reaction time and peak latency. The overall Pearson correlation coefficient is $.28$, $p < .01$. It is remarkable, however, that the significance of this correlation evaporates after controlling for age group. Apparently, there is a substantial difference between the age groups. Indeed, separate

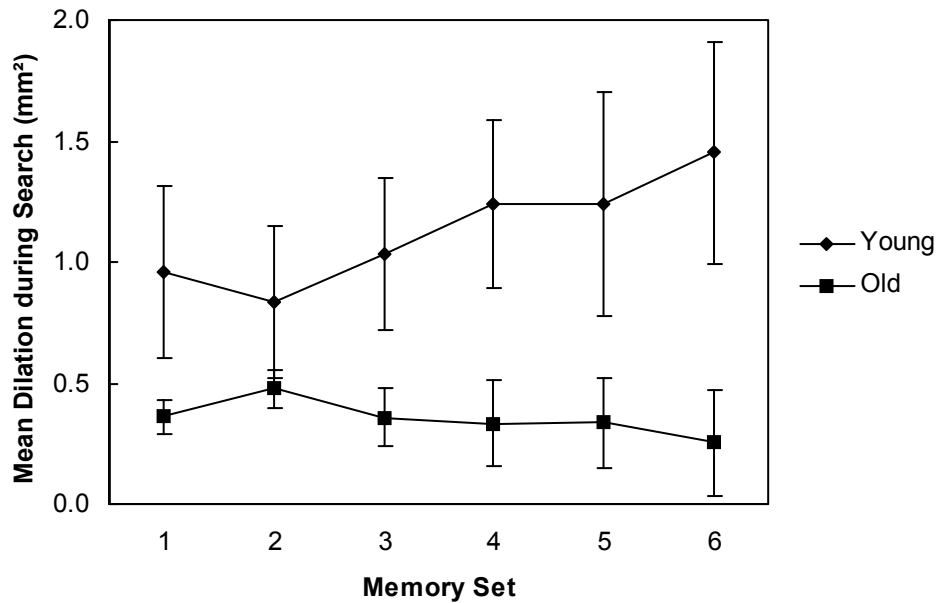


Figure 4.4 Mean pupil dilation during memory search as a function of memory set. Values are differential scores relative to baseline pupil size. Bars indicate standard errors of the mean.

analyses reveal a correlation of .37, $p < .01$, for the young and a zero correlation for the elderly participants.

Discussion

This study attempted to detect three task-evoked pupillary responses (TEPRs) in older adults: mean dilation, peak dilation, and peak latency. A memory-search task (Sternberg, 1966, 1967), involving six levels of memory load, was used to evoke these responses. The task comprised two phases: an encoding phase and a search phase. In the encoding phase there were no age differences in pupil dilation. In the search phase the young participants showed significantly more pupil dilation than their old counterparts. It is conceivable that the elderly's pupil already reached its maximum size in the encoding phase. An alternative explanation is that the elderly reached the limits of their cognitive resources in this phase. This would imply, however, that the elderly performed poorly on the memory-search task, especially under high memory load. This was not the case: The elderly's performance in terms of the percentage correct responses was comparable to the performance of the young. Moreover, resource limits in elderly people are usually observed at much higher levels of memory load (Morris et al., 1997). It is true that the response times revealed a complexity effect in that the elderly were more affected by an increase of memory

load than the young, but this is probably due to a different speed-accuracy tradeoff, not to capacity constraints. In the search phase, the elderly showed lower peak dilations than the young. Moreover, the latencies to the peak were somewhat longer, which can be interpreted as a delayed activation of cognitive resources (cf. Salthouse & Babcock, 1991).

The central question was whether TEPRs can differentiate between moderate levels of memory load. With respect to mean dilation this seemed to be the case. In the encoding phase mean dilation progressively increased with memory set in both age groups. In the search phase, however, no additional dilation could be detected in the elderly. That is, the elderly did show dilation responses, but their magnitude was more or less equal over the whole range of memory load. Possibly, these responses reflected basic arousal levels induced by the mere appearance of the stimuli (cf. Kim et al., 2000). Although it is disputable whether such responses have a cognitive origin, it is still peculiar that different levels of memory load did not elicit different levels of arousal. It is remarkable that both age groups show a decline in mean pupil dilation at memory set 6. It is unlikely that this decline is due to memory overload (cf. Granholm et al., 1996). Regarding the nature of the Sternberg task, it is more likely that it marks a strategic switch: Instead of memorizing a positive set of six digits, participants could have focused on the negative set of only four digits. This could have reduced memory load and thus mean pupil dilation.

With respect to peak dilation and peak latency, there were no effects of memory set in both age groups. Furthermore, there was no correlation between reaction time and peak latency in the elderly participants. It is possible that for a positive probe the latency of the peak depends on the position of the probe in the memory set. However, this possibility is refuted by the absence of an effect for the negative-probe data, where — within a memory set — search times were theoretically equal. This is in line with Sternberg's (1966) finding that the search process is exhaustive rather than self-terminating.

The ultimate test of pupillometry is its usefulness as an 'online' method for tracing cognitive load in a wide variety of tasks (see Beatty, 1982). The question is whether pupillometry bears enough advantages in comparison to 'simpler' methods, such as secondary-task performance. First and foremost, pupil dilation offers a more or less continuous monitor of cognitive load. Discrete measuring instants and time windows only need to be defined for capturing peak dilation and peak latency. Moreover, pupil dilation seems to be one of the most sensitive psychophysiological correlates of mental load. In the elderly, however, pupillometry seems to encounter its limitations. In this group, task-evoked pupillary responses cannot be detected under all circumstances. Whether this is due to cognitive limitations or physiological decay — such as senile miosis — remains an open question.

In conclusion, pupillometry seems only moderately suitable for measuring 'normal' levels of cognitive load in older people. In spite of the dilation effects found for both age groups in the encoding phase, no dilation effect could be detected for the elderly participants in the search phase of the Sternberg task. Regarding the laboriousness and intrusiveness of the pupillometric procedure, simple reaction-time methods seem to be preferable.

5 Training Variability and the Modality Effect¹

This study addressed the role of sensory modality and practice variability in training both young and elderly learners to solve complex problems. Modality was manipulated by presenting the training problems either purely visually (unimodal condition) or both visually and auditorily (bimodal condition). According to cognitive load theory a bimodal presentation of training material should lead to better performance than a unimodal presentation, because it addresses both mode-specific working-memory components and reduces extraneous cognitive load. Practice variability was manipulated by presenting the training problems either in blocks of the same category (blocked condition) or at random (random condition). According to cognitive load theory a random sequence of training problems should lead to better results, because it requires deeper processing and thus higher germane cognitive load. It was expected that a combination of bimodal and random training would lead to optimal results. Such a combination was expected to be especially helpful to elderly learners. Testing these hypotheses by training a group of young and a group old participants ($N = 80$) revealed that bimodal training leads to lower levels of perceived cognitive load than unimodal training. Furthermore, a trend was found towards a positive effect of practice variability on performance. However, the absence of an interaction suggests that there was no combined effect of modality and variability. Furthermore, no differential effects were found for the elderly participants.

Acquiring a complex cognitive skill makes high demands on working memory. Part of these demands relates to the complexity of the skill. Another part pertains to the format of training. Two aspects of the training format can be distinguished: irrelevant aspects that can be detrimental to the learning process and relevant aspects that directly contribute to the learning process. Logically, it follows that a training format should incorporate as few as possible irrelevant aspects and as many as possible relevant aspects. This idea is expressed by cognitive load theory (Sweller, Van Merriënboer, & Paas, 1998). Cognitive load theory (CLT) was developed from the notion that working memory has a limited capacity. Even though working memory is capable of grasping complex material, it is vulnerable to overload, which is devastating to the learning process. Therefore, CLT

¹ A version of this chapter is submitted for publication as: Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., & Schmidt, H.G. (2002c). Modality and variability as factors in training the elderly.

promotes training formats that minimize the chances of overload by avoiding irrelevant processes in working memory. At the same time, these training formats stimulate relevant processes within the limits of working-memory capacity. This combination should lead to an optimal construction of rich cognitive schemata (see Van Merriënboer, Schuurman, De Croock, & Paas, 2002). Ultimately, these schemata enable a person to achieve *transfer*, which is the application of acquired knowledge in new problem situations (e.g., Paas & Van Merriënboer, 1994a).

The philosophy of cognitive load theory especially seems applicable in cases where working-memory capacity is particularly limited, such as in the elderly. Research has demonstrated that elderly people have to cope with three declines of working memory. First, tests aimed at multiple aspects of working memory suggest that the capacity of working memory becomes smaller with old age (e.g., Salthouse & Babcock, 1991; Wingfield, Stine, Lahar, & Aberdeen, 1988). Second, a general slowing of cognitive functions, including working memory, has been observed (e.g., Fisk & Warr, 1996; Salthouse, 1996). A final decline concerns the ability to discriminate between relevant and irrelevant information. It was found that elderly people often fail to inhibit information that is not relevant for the task at hand (e.g., Hasher & Zacks, 1988; Stoltzfus, Hasher, & Zacks, 1996)). This failure tends to result in an overload of irrelevant information. Taken together, these declines have a great impact on the elderly's ability to process new information. As Van Gerven, Paas, Van Merriënboer, and Schmidt (2000, Chapter 1) pointed out, however, training formats that are based on cognitive load theory can compensate for these declines to a large extent. That is, reducing the number of irrelevant aspects of a training format and maximizing the number of relevant aspects leads to an optimal use of working memory, which in its turn leads to an optimal learning result.

Ingredients of Optimal Training

Within cognitive load theory, reducing the number of irrelevant aspects of a training format is referred to as minimizing the portion of *extraneous* cognitive load it imposes on the learner. There are numerous ways to reduce extraneous cognitive load (see Sweller, Van Merriënboer, & Paas, 1998, for an overview). This study will focus on the so-called *modality effect* (Kalyuga, Chandler, & Sweller, 2000; Moreno & Mayer, 1999; Tabbers, Martens, & Van Merriënboer, 2001; Tindall-Ford, Chandler, & Sweller, 1997). Theoretically, this effect emerges from the assumption that working memory consists of two mode-specific components, each of which has a limited capacity (Baddeley, 1992). Basically, the modality effect comes down to the phenomenon that addressing both the visual and auditory modality

leads to better learning than merely addressing the visual modality. In conventional instructional formats, visual material, such as a diagram, is accompanied by explanatory text that is placed at some distance from the visual material. Learners are therefore compelled to split their attention between these two sources of information and perform unnecessary visual search in order to integrate this information. According to CLT, this leads to extraneous cognitive load. If, however, the textual material is presented through the auditory modality, then there is no split attention and unnecessary visual search is precluded. Moreover, working-memory capacity is used optimally in such a format, because information is not only processed in the visual component, but also in the auditory component. Thus, the chance that the visual component gets overloaded is reduced.

When extraneous cognitive load is reduced, more capacity becomes available for raising the number of relevant aspects of a training format. In terms of cognitive load theory, these aspects are referred to as the *germane* portion of cognitive load that is imposed on the learner. One way to optimize the amount of germane cognitive load is to increase the *variability* of the training problems (Paas & Van Merriënboer, 1994b; Renkl, Stark, Gruber, & Mandl, 1998; Van Merriënboer et al., 2002). Instead of using training problems of the same category, the use of training problems covering a whole range of categories is likely to lead to the acquisition of rich cognitive schemata. Rich cognitive schemata increase the chances of achieving transfer. Increased training variability leads to *contextual interference* (e.g., Battig, 1972, 1979), which in its turn results in higher levels of cognitive load. The nature of this increased load is considered positive, however. That is, contextual interference prevents learners to develop routine in processing the training problems. Instead, they are compelled to re-examine every new training problem, which leads to a deeper processing of the problem at hand. Thus, higher levels of cognitive load, causing longer training times and, in the case of a time limit, fewer attended training problems, eventually lead to better transfer performance (e.g., De Croock, Van Merriënboer, & Paas, 1998). This ostensible inconsistency is known as the *transfer paradox* (Van Merriënboer, De Croock, & Jelsma, 1997). Increased training variability establishes one of the basic ingredients of modern theories on complex skill training (e.g., Van Merriënboer, 1997).

A CLT-based training format in which a decrease of extraneous cognitive load is combined with an increase of germane cognitive load is the use of worked examples instead of conventional problems. Where conventional problems require learners to solve a range of practice problems, worked examples require them to study the same problems with their worked-out solutions. Although the latter activity may seem relatively passive, it has repeatedly proven to lead to better results than a conventional format for novice learners (e.g., Pirolli & Anderson, 1985; Van Merriënboer & Paas, 1990; Ward & Sweller 1990; Zhu & Simon, 1987). CLT explains these find-

ings by assuming that conventional problems elicit both forward processing and backward processing, whereas worked examples only evoke forward processing. Backward processing, which is induced by ‘weak’ solving strategies, such as means-ends analysis, is associated with extraneous cognitive load, because it does not promote insight in the problem structure. Instead, it focuses on reducing the difference between the current state and the goal state of a problem. Forward processing is associated with germane cognitive load, because it focuses on problem states as well as the operators that lead to the goal state. Van Gerven, Paas, Van Merriënboer, and Schmidt (2002a, Chapter 2) found worked examples to be a more efficient training format for both young and elderly learners in that a lower level of invested cognitive load led to a comparable level of performance. This effect was replicated in a study where worked examples were presented in a multimedia format (Van Gerven, Paas, Van Merriënboer, Hendriks, & Schmidt, 2002, Chapter 3). In these studies the level of germane cognitive load was not manipulated as a separate factor. The current study seeks to unravel the role of germane cognitive load in training efficiency. Manipulating germane load will have an impact on the total amount of cognitive load. To detect this impact, an accurate measurement of cognitive load is indispensable.

Stating the Problem

The aim of the current study was to vary both extraneous cognitive load and germane cognitive load in training strategies for solving water-jug problems (Luchins, 1942). Water-jug problems represent a well-defined problem space in which jugs of different sizes and containing different amounts of water are strategically poured into each other to acquire a certain goal amount of water (see the Method section for details). There is a limited number of basic strategies that can be combined or repeated to achieve a goal state. Hence, there is a limited number of problem categories that can be trained in order to acquire general schemata for dealing with water-jug problems. This offers the opportunity to manipulate training variability. As was described before, increasing training variability can be achieved by increasing the number of problem categories. Variability can also be achieved by applying different sequencing schemas. Van Merriënboer (1997), for instance, suggests a blocked versus a randomized sequence of training problems. In a *blocked* sequence problems of different categories are presented in discrete packages that do not merge in time. A typical blocked sequence would be: C1P1, C1P2, C1P3, C2P1, C2P2, C2P3, in which C indicates the category and P the problem number. In a *random* sequence, on the other hand, problems of different categories are intermixed. Thus, a possible random sequence would be: C2P3, C1P2, C1P1,

C2P2, C1P3, C2P1. Unlike the blocked sequence, the random sequence will elicit a higher contextual interference and increase depth of processing, because it prevents learners from running through the training problems on a routine basis.

In this study training problems were presented in a worked-out fashion, where images of problem states were accompanied by explanatory text. The modality effect was pursued by presenting the explanatory texts either visually or auditorily. If the texts were presented visually, only the visual component of working memory was addressed. This was the *unimodal* condition. If, on the other hand, the texts were presented auditorily, both the visual and the auditory component were addressed. This was the *bimodal* condition. Unlike Van Gerven et al. (2002, Chapter 3), who also manipulated the dynamics of the visual presentations (i.e., by means of animation), the visual appearance of the problem states did not differ between the modality conditions. A possible effect could thus exclusively be attributed to the use of different modalities. It is expected that the bimodal condition leads to a lower extraneous cognitive load than the unimodal condition.

Considering the two central factors in this study, modality and variability, four hypotheses can be stated with respect to the total amount of cognitive load. First, in the unimodal/random condition the total amount of cognitive load will reach the highest level. That is, both extraneous and germane cognitive load will be relatively high and will possibly exceed the available cognitive resources. Second, in the bimodal/blocked condition, the total amount of cognitive load will reach the lowest level, since both extraneous and germane cognitive load remain relatively low. Third, in the unimodal/blocked condition the total amount of load will reach an intermediate level. Extraneous load will have the upper hand, however. The fourth and final hypothesis concerns the bimodal/random condition. This condition will also result in an intermediate level of load. However, here the germane portion of load will prevail.

These hypotheses have direct consequences for the expected performance outcomes. Performance will be highest if there is no overload and germane cognitive load is maximal. This is theoretically the case in the bimodal/random condition. The efficiency of this condition will be high in that a moderate level of load will lead to a relatively high performance (see Paas & Van Merriënboer, 1993). The lowest level of performance is expected in the unimodal/random condition, but only in the case of overload. In this case efficiency is low, because a high level of cognitive load leads to a low performance. If there is no overload, performance may equal the performance in the bimodal/random condition. Efficiency is then somewhat higher. In the blocked conditions, performance will reach a moderate level, just like efficiency. In a unimodal format, this will only be the case if the total amount of load does not exceed the cognitive resource limits.

What consequence do the conditions have for the elderly? The cognitive load outcomes will probably resemble the pattern of the young, although on the average cognitive load will be higher. However, starting from the simple assumption that the capacity of working memory is smaller for the elderly than for the young, the elderly are more susceptible to cognitive overload than their young counterparts. As a consequence, relative to the young, their performance is likely to be impaired in the unimodal/random condition. In the bimodal/random condition, however, they can theoretically achieve a performance level that is comparable to the performance level of the young (see also Paas, Camp, & Rikers, 2001). In the unimodal/blocked and in the bimodal/blocked condition their performance will reach a low to moderate level.

Measuring Cognitive Load

Testing the above hypotheses requires an adequate cognitive load measure. There are roughly two types of cognitive-load measures: subjective and objective measures (for an overview, see Meshkati, Hancock, Rahimi, & Dawes, 1995). Subjective measures lean on the assumption that people are capable of evaluating their experienced level of cognitive load. Generally, this seems to be the case (e.g., Paas & Van Merriënboer, 1994b). Objective measures can be subdivided into physiological measures and non-physiological measures. An important drawback of physiological measures is that most of these techniques are quite laborious and intrusive to the participant. Moreover, physiological measures often lack sensitivity (e.g., heart rate variability: see Paas, Van Merriënboer, & Adam, 1994). A positive exception in this respect is pupillometry (e.g., Beatty, 1982). In elderly people, however, this technique seems not to be a sensitive correlate of cognitive load under all circumstances (Van Gerven, Paas, Van Merriënboer, & Schmidt, 2002b, Chapter 4).

Non-physiological measures mainly concern dual- and secondary-task techniques (for an overview, see Heuer, 1996). In both techniques persons are involved in a primary and a secondary task. In the dual-task paradigm the decrement of performance on the primary task due to a claim of cognitive capacity by the secondary task is seen as evidence for the role of the central executive (Baddeley, 1992). The rationale behind the secondary-task technique is that performance on the secondary task provides information about the experienced mental load on the primary task. In other words, if a high-priority task imposes an increasing load on the cognitive system, this will be reflected by a decreasing performance on a low-priority task. Performance on a secondary task is often expressed in reaction time. Reaction times in a secondary task typically decrease if the load imposed by a primary task is raised.

In the context of cognitive load theory, Chandler and Sweller (1996) are among the few to apply a secondary-task paradigm (the others are Sweller, 1988, and Marcus, Cooper, & Sweller, 1996). As a primary task, they required participants to learn about a technical computer program. In a secondary task the participants were regularly presented with a tone, immediately followed by a letter that was depicted on a separate screen. On these occasions they had to recall the letter that was presented previously, while encoding the new letter. Performance on the secondary task was expressed as the percentage of correct responses. As expected, this percentage turned out to be relatively low when the participants were involved in a more complex primary task.

In the current study a relatively simple secondary-task paradigm was used. Simplicity of the secondary task was considered important, because in the elderly performance on concurrent tasks is generally found to be disproportionately impaired relative to the young, especially when the constituent tasks become more complex (e.g., Lorbach & Simpson, 1988; McDowd & Craik, 1988; Salthouse, Fristoe, Lineweaver, & Coon, 1995). Signal detection seemed most appropriate in this light. Simple visual or auditory signal detection is a commonly used secondary task (e.g., Berent & Perfetti, 1993). Since an auditory signal was not possible because it would interfere with the auditory materials used in the primary task, a simple visual stimulus was chosen.

Method

Participants

Forty students (10 men and 30 women, mean age = 23.33 years, $SD = 2.99$) of Maastricht University represented the young group. Forty older adults (24 men and 16 women, mean age = 65.05 years, $SD = 4.51$), who were recruited through advertisements in local newspapers, represented the old group. All participants were in good health and had normal or corrected-to-normal vision. They received 35 Dutch guilders (about €16) for their participation. Participants were randomly assigned to the experimental conditions.

A shortened version of the Groningen Intelligence Test (GIT, Luteijn & Van der Ploeg, 1983) was used to acquire an indication of the participants' intellectual abilities. The GIT is a commonly used Dutch estimate of formal IQ. The shortened version includes four subscales (out of nine), which give the best approximation of the full-scale IQ: vocabulary [woordenlijst], mental rotation [legkaart], mental arithmetic [cijferen], and inductive/deductive reasoning [woordmatrijzen]. The 95%-confidence interval of the shortened GIT is ± 10 IQ points. The young participants had a mean

IQ of 120.85 ($SD = 7.71$). The elderly's mean score was significantly higher: 126.58 ($SD = 11.35$), $t(78) = -5.73$, $p = .01$.

Working memory abilities were assessed with Salthouse and Babcock's (1991) computation-span test, which is a variant of the widely used reading-span test. During the computation-span test, the participant has to work out simple adding and subtraction sums, which are presented verbally (e.g., "7 plus 8," "9 minus 4," etc.). At the same time, the last digit of each sum (e.g., 8, 4, etc.) has to be memorised and reproduced after a series. A series comprises up to 7 sums and there are three trials for each. Computation span is determined by the maximum number of digits recalled in at least two trials. The young participants' mean computation span was 5.33 ($SD = 1.29$). The elderly participants' mean span was only 3.48 ($SD = 1.32$). The difference was highly significant, $t(78) = 6.34$, $p < .001$.

Apparatus and Software

The experiment was run on an IBM-compatible computer. The interaction with the computer took place through a Philips Brilliance 107 17-inch monitor with an integrated touch screen (Elo TouchSystems, Fremont, CA). The software controlling the experiment was programmed in Authorware 3.5 (1996). Reaction times were registered with a response box containing one large (approximately 2×2 inch) jukebox button, which could be illuminated. The response box was controlled by a second IBM-compatible computer, which was linked to the stimulus computer through a parallel port connection. The response box was placed in front of the touch screen. The software controlling the reaction time registration was programmed in ERTS (Experimental Run Time System, Beringer, 1996).

Materials

Subjective cognitive-load scale. A 9-point symmetrical category scale was used as a subjective cognitive-load (SCL) measure (after Paas, Van Merriënboer, & Adam, 1994). The scale was presented on the touch screen as a slide control and was accompanied by the phrase: *I have experienced the foregoing as: ...* The scale had three labels, ranging from *not difficult at all* (1), via *moderately difficult* (5), to *very difficult* (9). Intermediate values were not labeled. Paas (1992) found an internal consistency coefficient (Cronbach's α) of .90 using a comparable scale. Paas and Van Merriënboer (1994b) evaluated the scale as a highly reliable and sensitive instrument for the assessment of cognitive load ($\alpha = .82$).

Water-jug problem. The water-jug problem (Luchins, 1942) is a Tower-of-Hanoi-like task that requires the participant to obtain a certain amount of water in a target jug by pouring jugs of different sizes and filled with varying amounts of water into each other. Generally, the key for a solution to this problem lies in the rule that jugs cannot overflow. If water is

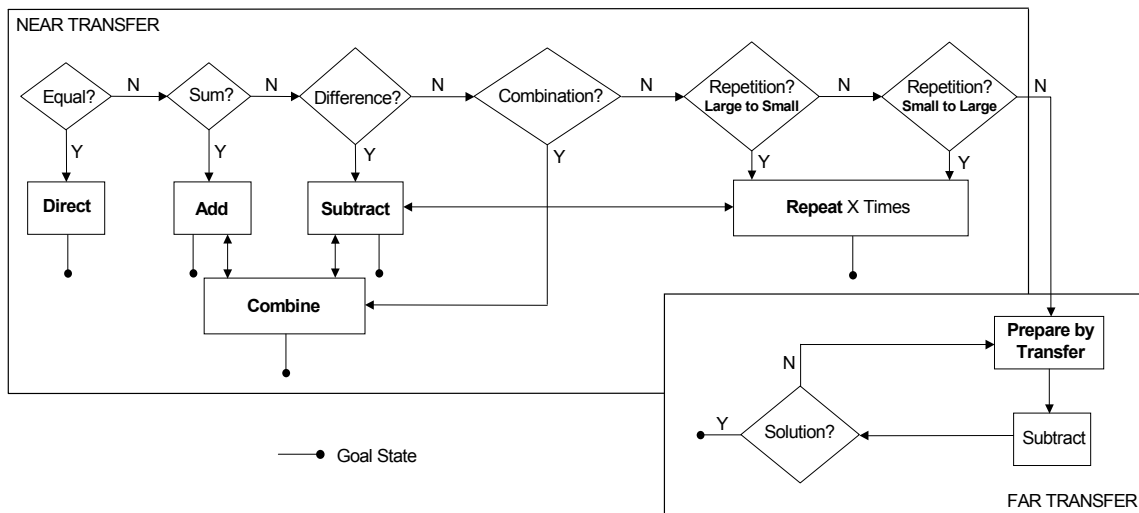


Figure 5.1 A revised version of the general processing model for the water-jug problem proposed by Van Gerven et al. (2002a, Chapter 2, Figure 2.1).

poured from one jug into another and the contents of the donating jug exceed the free space of the receiving jug, pouring stops as soon as the receiving jug is full. In that case a residual remains in the donating jug. In some cases, this residual may equal the goal amount of water. In other cases it can be used for further operations. In this experiment a computerized version of the task was used (see Van Gerven, Paas, Van Merriënboer, & Schmidt, 2002a, Chapter 2, for details).

Worked examples. Ten water-jug problems were selected as training problems. These were presented as worked examples. The worked examples represented the five categories of basic strategies for solving water-jug problems. Thus, there were two worked examples per category. The five categories are incorporated in the processing model depicted in Figure 5.1. It is a revised version of the general processing model used by Van Gerven et al. (2002a, Chapter 2, Figure 2.1). As Van Gerven et al. already pointed out, it is not a cognitive model, but merely an overview of the subsequent steps that an ‘ideal’ problem solver might run through when confronted with a water-jug problem.

The model subsequently checks whether a progressively more complex solution strategy is applicable in a given situation. First, it checks whether one of the jugs has a capacity that equals the target amount of water. If this is the case, it directly fills that particular jug and pours its content into the target jug. Subsequently, it checks whether the sum of two or more jugs yields the target amount or whether the target amount is a multiple of a jug’s capacity. If this is not the case, it checks whether the difference between the capacities of two or more jugs yields the desired amount. If subtraction does not lead to a solution, the model examines whether a repetition or combination of one or more of the previous strategies does. Subsequently, it checks whether subtraction by repetition leads to a solu-

tion. Subtraction by repetition can be achieved in two ways: A jug can repeatedly be poured into a smaller one or a jug can repeatedly be poured into a larger one. If a large jug is repeatedly poured into a small jug, the small jug is emptied after each pouring operation. During each pouring operation, the contents of the large jug are decreased by the capacity of the small jug. Thus, the contents of a full jug with a capacity of 8 units, which are repeatedly poured into a jug with a capacity of 2 units, are decreased in steps of 2 units. The residuals either or not equal the goal amount of water. If, on the other hand, a small jug is repeatedly poured into a large one, the small jug is filled with water before each pouring operation. The contents of the large jug increase by the capacity of the small one, until it is nearly filled to the brim. At that point, a new pouring operation leads to a residual in the smaller jug. Thus, repeatedly pouring a jug of 4 units into a jug of 7 units eventually leads to a residual of 1 unit. Finally, if all the previous strategies do not lead to a solution, a more sophisticated strategy is required, which Van Gerven et al. call “preparation by transfer.” In this strategy, the result of a subtraction operation is transferred to another jug, which leads to a problem state that forms the basis for yet another subtraction operation. This sequence can be repeated until a solution is found.

The worked example categories included the adding, the subtracting, the combination, and the two subtracting-by-repetition strategies. The ‘direct’ strategy was considered too straightforward for training. The prepare-by-transfer strategy was retained for the transfer test.

Transfer test. Twelve water-jug problems were selected as test problems. These were sorted by increasing complexity on the basis of the item scores of Van Gerven et al.’s (2002a, Chapter 2) study. The test included both near-transfer problems and far-transfer problems (see Figure 5.1). The near-transfer problems were variations of the training problems. The far-transfer problems required the preparation-by-transfer strategy.

Design and Procedure

There were three independent factors: age group, with levels young and old, modality with levels unimodal and bimodal, and variability, with levels blocked and random. This resulted in a $2 \times 2 \times 2$ between-subjects factorial design. Eighty participants, 40 young and 40 old, were randomly assigned to the experimental conditions ($n = 10$). Dependent variables were subjective cognitive load during training, reaction time during training, the proportion of missed signals during training, and performance (the proportion of solved test problems).

The experiment took place in individual sessions. After a short verbal introduction by the experimenter, the intelligence and computation-span tests were administered. Subsequently, the participant was presented with a general instruction explaining the basic rules of the water-jug problem as well as the interface. The use of the interface was practiced under

supervision of the experimenter. The participant also got acquainted with the subjective cognitive-load measure and the response box.

During training, worked examples were presented as a sequence of full-screen images representing the solution steps. Consistent with research by Tabbers, Martens, and Van Merriënboer (2001), it was decided to make the presentation system-paced. During each worked example light signals were given at random instants. Three signals (i.e., illuminations of the response button) were presented per worked example. The signal extinguished as soon as the participant pushed the response button. The reaction time was then registered. Signals were presented for maximally 2000 ms. Missed signals were registered as such.

In the unimodal condition explanatory text was imposed on the images. In the bimodal condition the images were accompanied by an auditory version of the same text. The intervals between the pictures were based on the duration of the narrations in the bimodal condition. Thus, there was no difference in duration between the conditions. In the blocked condition worked examples of the same category were presented in blocks of two. Within blocks the order of the two problems was counterbalanced. The order of the blocks (i.e., the order of the categories) was counterbalanced by applying a 5×5 Latin square. In the random condition the order of the problems was randomized such that problems of the same category did not follow each other.

The training session was succeeded by the transfer test. Throughout the experiment, the experimenter answered simple questions of the participant providing that they did not concern the solution strategy. These questions were not incorporated in the analysis. There was a time limit of six minutes per problem. Before the start of the training the participants were informed about the possibility to skip problems on request. Generally, participants scarcely made use of this option, however.

Results

A $2 \times 2 \times 2$ between-subjects MANOVA was performed on all dependent measures. The means and standard deviations of the dependent measures are summarized in Table 5.1. The results of the analyses are given in Table 5.2.

Subjective Cognitive Load

With regard to subjective cognitive load (SCL), the analysis of variance yielded an effect of age group, $F(1, 72) = 4.02$, $MSE = 1.72$, $p < .05$. The effect is small, however. The young participants report an average SCL of 1.94 ($SD = 1.00$), whereas the elderly report an SCL of 2.53 ($SD = 1.60$). Both means are low given the 9-point scale on which they were measured.

Table 5.1 Means and standard deviations of the dependent measures.

			Unimodal		Bimodal	
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Young	Blocked	SCL	2.47	2.47	2.09	0.96
		RT	549.74	110.28	633.98	217.24
		Missed	.01	.02	.02	.05
		Performance	.84	.13	.89	.11
		Efficiency (SCL)	-0.41	0.94	0.08	0.90
		Efficiency (RT)	0.08	0.85	0.00	1.18
	Random	SCL	1.55	0.60	1.66	0.52
		RT	581.93	200.73	564.41	251.33
		Missed	.00	.01	.02	.04
		Performance	.82	.19	.87	.20
		Efficiency (SCL)	0.13	0.79	0.27	1.01
		Efficiency (RT)	-0.15	0.88	0.14	1.48
Old	Blocked	SCL	2.76	1.67	3.18	2.05
		RT	821.79	156.10	809.99	275.21
		Missed	.10	.08	.07	.12
		Performance	.46	.27	.61	.22
		Efficiency (SCL)	-0.35	1.17	-0.08	1.07
		Efficiency (RT)	-0.43	1.20	0.07	1.28
	Random	SCL	1.76	1.23	2.42	1.20
		RT	703.95	218.72	734.45	172.98
		Missed	.07	.15	.07	.09
		Performance	.52	.17	.57	.26
		Efficiency (SCL)	0.27	0.69	0.16	1.04
		Efficiency (RT)	0.14	1.01	0.22	0.93

There is also an effect of modality, $F(1, 72) = 7.05$, $p = .01$. That is, in the unimodal condition ($M = 2.63$, $SD = 1.58$) higher levels of SCL were reported than in the bimodal condition ($M = 1.85$, $SD = 0.97$). Again, the effect is rather small (see Figure 5.2a). The factor variability did not yield a significant effect. None of the possible interactions between age group, modality, and variability was significant (see Table 5.2). This means that

Table 5.2 F values, significance levels, and mean square errors resulting from the multivariate analysis of variance (A = age group, M = modality, V = variability).

	AG	MOD	VAR	A × M	A × V	M × V	A × M × V	<i>MSE</i>
SCL	4.03*	7.05*	0.48	0.12	1.33	0.39	0.05	1.71
RT	6.08***	1.56	0.21	0.71	0.07	0.10	0.61	42,585.77
Missed	11.70**	0.50	0.01	0.06	0.54	0.13	0.08	1.72
Performance	48.85***	0.02	2.93 ^b	0.17	0.36	0.26	0.26	0.04
Efficiency (SCL)	0.01	3.42 ^a	0.86	0.02	0.30	0.73	0.00	0.92
Efficiency (RT)	0.01	0.42	0.60	0.67	0.14	0.00	0.62	1.26

Note. * $p < .01$, ** $p < .01$, *** $p < .001$, ^a $p = .07$, ^b $p = .09$.

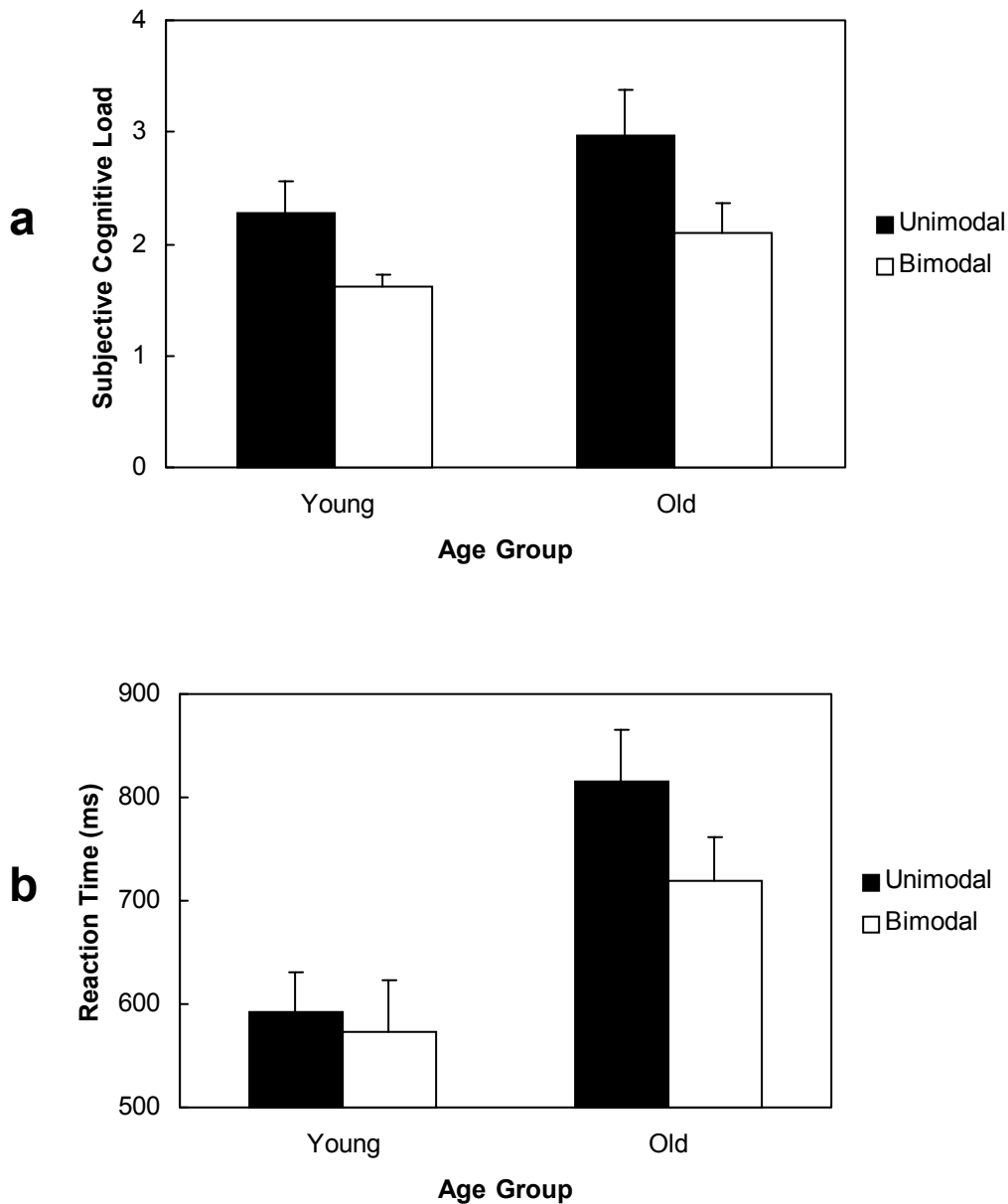


Figure 5.2 Cognitive load measures. Panel a depicts subjective cognitive load as a function of age group and modality. Panel b shows reaction time (in ms) in the secondary task as a function of age group and modality. Error bars indicate one standard error of the mean.

there was no combined effect of modality and variability. At the same time, none of the CLT-based conditions had a disproportional effect on the perceived cognitive load of the elderly participants.

Reaction Time in the Secondary Task

Not surprisingly, there is a significant effect of age group regarding reaction time on the secondary task, $F(1, 72) = 16.08$, $MSE = 42,585.77$, $p < .001$. Where the mean reaction time of the young was 582.51 ms ($SD =$

196.50), the elderly participants' average response delay was 767.55 ms ($SD = 208.76$). However, no effects were found for modality and variability. With respect to modality, the means clearly differ in the hypothesized direction, however. Where in the unimodal condition the average response latency is 703.87 ms ($SD = 224.94$), the mean latency in the bimodal condition is 646.19 ms ($SD = 217.82$). It is important to note that this effect was mainly present in the elderly participants (see Figure 5.2b). Their mean latency was 815.89 ms ($SD = 217.85$) in the unimodal condition against 719.20 ms ($SD = 192.56$) in the bimodal condition. All interactions were non-significant (see Table 5.2). Again, this indicates that there was neither a combined effect of modality and variability, nor a significant disproportional effect for the elderly participants. Nevertheless, in the young group the highest mean reaction time was found in the unimodal/random condition (see Table 5.1).

Proportion of Missed Signals

On the average, only 4.6% of all signals was missed. This indicates that the proportion of missed signals in this experiment is probably not a proper indicator for discriminating between the experimental conditions. Indeed, only an effect of age group was found, $F(1, 72) = 11.70$, $MSE = .0071$, $p = .001$. Where the young participants on average missed 1% ($M = .01$, $SD = .04$) of the signals, the elderly missed 8% ($M = .08$, $SD = .11$).

Performance

Performance was registered as the proportion of solved problems during the test. There was a significant effect of age group, $F(1, 72) = 48.85$, MSE

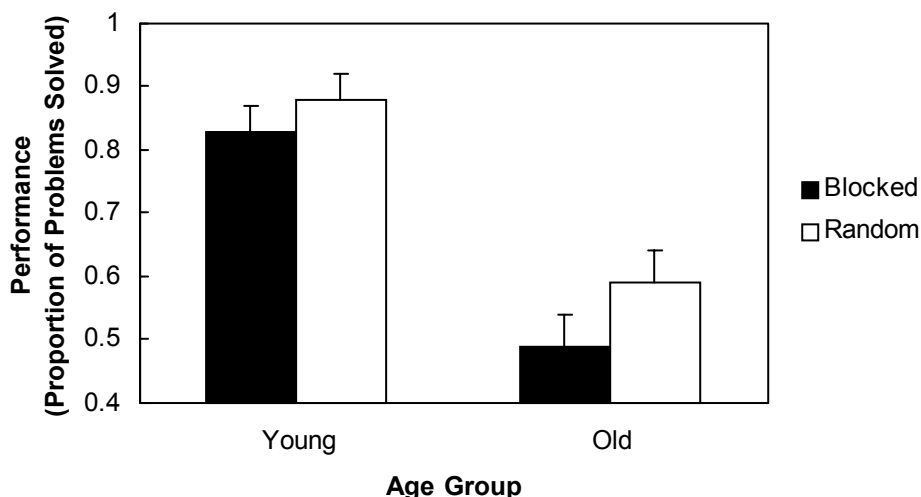


Figure 5.3 Performance (proportion of problems solved) as a function of age group and variability. Error bars indicate one standard error of the mean.

= .041, $p < .001$. On the average, the young participants solved 85% ($M = .85$, $SD = .16$) of the problems, whereas the elderly participants solved only 54% ($M = .54$, $SD = .23$). No significant effect was found with regard to modality and variability. Overall performance scores favored the random condition, however. Here the mean score was .74 ($SD = .25$) against .66 ($SD = .26$) in the blocked condition. The analysis of variance reveals a slight trend for this effect, $F(1, 72) = 2.93$, $p = .091$. Moreover, the effect is stronger for the elderly than for the young (see Figure 5.3). Where the young participants score .83 ($SD = .16$) in the blocked condition and .88 ($SD = .16$) in the random condition, the elderly participants score .49 ($SD = .22$) in the blocked condition against .59 ($SD = .24$) in the random condition. None of the analyses revealed an interaction (see Table 5.2).

Efficiency

Following Paas and Van Merriënboer (1993), a combined score of performance and cognitive load could be calculated. The magnitude of this score is determined by the ratio between cognitive load and performance. If the invested cognitive load during training is relatively high and the performance on the test is relatively low then the efficiency score is negative. If, on the other hand, cognitive load is relatively low and performance is relatively high then the efficiency score is positive. If there is a perfect balance between cognitive load and performance then the efficiency score equals zero. Efficiency was calculated as follows (see Paas & Van Merriënboer, 1993, for details):

$$Efficiency = \frac{Performance - Cognitive\ Load\ Measure}{\sqrt{2}}$$

In this formula not the raw scores, but the z scores of performance and the cognitive-load measure were used. These z scores were based on the grand means and standard deviations of the age groups. Since in this experiment two different measures of cognitive load were collected, that is subjective cognitive load and reaction time, two different efficiency scores were computed.

Analysis of the efficiency scores based on reaction time did not reveal any significant effects (see Table 5.2). The same holds for the analysis of the SCL-based efficiency scores. However, there was a trend towards an effect of modality, $F(1, 72) = 3.42$, $MSE = 0.92$, $p = .07$. While the unimodal condition yields a mean efficiency of -0.19 ($SD = 1.00$), the bimodal condition generates a mean efficiency of 0.21 ($SD = 0.86$). The absence of a trend towards an interaction suggests that the effect does not differ between the age groups.

Discussion

On the basis of cognitive load theory (Sweller, Van Merriënboer, & Paas, 1998), it was hypothesized that bimodally presented worked examples lead to lower levels of cognitive load and higher performance than unimodally presented worked examples. Worked examples presented in a random order were expected to lead to a higher level of cognitive load and a higher level of performance than worked examples presented in blocks. A combined effect of bimodal worked examples presented in a random order was expected. That is, a moderate level of cognitive load was expected to lead to relatively high levels of performance. Finally, on the basis of Van Gerven et al.'s (2000, Chapter 1) framework it was hypothesized that these effects would be stronger for the elderly participants than for their young counterparts.

The computation-span scores showed that the elderly participants' working-memory capacity is indeed smaller than the young participants' working-memory capacity. At the same time, the IQ scores of the elderly were significantly higher than the IQ scores of their young counterparts. The same pattern of results was found by Van Gerven et al. (2002a, Chapter 2) and Van Gerven et al. (2002, Chapter 3). The measures of subjective cognitive load revealed that the elderly experienced higher levels of cognitive load than the young. A similar effect was found in the reaction times on the secondary task. Moreover, the elderly missed significantly more signals in the secondary task than the young, which suggests that the primary task put higher demands on their cognitive capacity. These results, combined with the finding that the elderly performed poorer on the transfer test, suggest that working memory rather than intelligence seems to be the critical factor in processing complex new material.

Both the young and the old participants reported rather low levels of subjective cognitive load in all conditions, which may be due to a negative response bias. Subjective cognitive load levels were higher in the unimodal than in the bimodal condition. A trend in the same direction was found with respect to reaction time. The absence of an interaction suggests that this effect did not differ between the age groups. That is, contrary to the hypothesis, the elderly did not experience a larger 'drop' of cognitive load in the bimodal condition relative to the unimodal condition than the young.

Interestingly, there was a trend towards a performance effect of variability. That is, in the random condition performance was higher than in the blocked condition. This 'effect' seems to be mainly attributable to the elderly group. Apparently, the random arrangement of training problems induced a more elaborate processing than the blocked format. This implies a higher level of germane cognitive load, which in its turn raises the total amount of cognitive load. However, the absence of a modality \times variability

interaction with respect to subjective cognitive load does not support this line of reasoning. The finding that the young participants had the highest mean reaction time in the unimodal/random condition points in the hypothesized direction, however. Finally, a trend towards a general effect of modality on efficiency (based on SCL) was observed. That is, the efficiency of the bimodal condition was higher than the efficiency of the unimodal condition, which is in line with the predictions of cognitive load theory. The random condition seems to have a mediating effect in this respect. That is, the gap between the unimodal and the bimodal condition seems to be filled in by the random condition. Apparently, the positive effect of the random condition on performance raises the efficiency of the unimodal condition and smoothens the jump to the higher efficiency levels of the bimodal condition.

What do we learn from the above results? It is obvious that the data did not indisputably establish the surplus value of the CLT-based instructional formats (bimodal/random) relative to the conventional formats (unimodal/blocked) for the elderly. This may be due to a low statistical power. Most effects are in the hypothesized direction, however. These findings provoke preliminary conclusions in favor of CLT and encourage further explorations. First, there is the tendency of the elderly participants to perform better in the random than in the blocked condition. A 'performance effect' induced by a CLT-based training format was not found in earlier research with elderly participants (Van Gerven et al., 2002, Chapter 3; Van Gerven et al., 2002a, Chapter 2). Second, especially in the elderly reaction times tend to be shorter in the bimodal condition than in the unimodal condition. Apparently, extraneous load is lower in the bimodal than in the unimodal condition. Third, there is an overall effect of modality with respect to subjective cognitive load. However, both age groups rate their perceived cognitive load rather low. This was also the case in previous research (Van Gerven et al., 2002, Chapter 3; Van Gerven et al., 2002a, Chapter 2). If this low rating is not due to a response bias, there is a clear-cut conclusion: The level of germane cognitive load should be raised, that is, not only by manipulating the variability but also the complexity of the worked examples. Raising the burden could lead to overload, however. Therefore, it would be wise to adjust the level of germane cognitive load to the capacity of the learner, for instance, by selecting training problems on the basis of previous performance and experienced cognitive load (see Camp, Paas, Rikers, & Van Merriënboer, 2001).

Regarding the elderly's overall performance, it seems that the gap between the training problems and the problems in the transfer test was too large. Reducing this gap by making the training problems more complex is delicate, however, because this could lead to a ceiling effect, which is already observable in the young participants. Given that the water-jug task is self-explaining to a certain extent, the chances of a ceiling effect are

substantial if the gap becomes too small. Besides, reducing the gap can invalidate the transfer test. After all, a substantial training-test discrepancy is what makes a test a transfer test. Since achieving transfer is one of the main objectives of CLT-based training formats this quality remains essential.

An alternative approach would be the choice of a less intuitive task domain. As was already pointed out, the water-jug task is possibly self-explaining to an important extent. That is, the task invites learners to explore the problem space by trial and error. For some learners this could lead to an additional learning effect. Others are trapped in cycles of operations that lead to the same undesired result over and over again. Yet another group of learners will attempt all possible operation sequences, which unintentionally could lead to a solution. A less intuitive domain would discourage this behavior. Such a domain could, for instance, lean more on knowledge than on common sense. However, common sense was considered a desirable factor by Van Gerven et al. (2002a, Chapter 2), because a domain that incorporates a substantial amount of new knowledge would disproportionately put the elderly at a disadvantage. This problem could be overcome partly by matching the educational backgrounds of the elderly participants with those of the young.

To conclude, the results of this study have revealed the advantages of cognitive load theory in general and its possibilities as an instructional theory for elderly learners in particular. At the same time they have established promising directions for future research. Although it is disputable whether the water-jug problem remains an appropriate problem domain in this line of research, it seems to be a reliable starting point for a target group that represents people of such diverse backgrounds as the elderly. However, regarding the educational needs of elderly people in a professional context, it would surely be interesting to test the principles of cognitive load theory in a less contrived domain.

6 Concluding Remarks and Future Research

In this thesis, a first attempt was made to explore the benefits of training formats based on cognitive load theory (CLT, Sweller, Van Merriënboer, & Paas, 1998) for elderly learners. For this purpose, three experiments were performed. The first experiment tested the efficiency of training with worked examples against the efficiency of training with conventional problems (Chapter 2). The second experiment tested the benefits of multimedia-based worked examples relative to purely visual worked examples (Chapter 3). A third experiment specified the research question of the second experiment by exploring the advantages of addressing two modalities relative to addressing only one modality (Chapter 5). Furthermore, this experiment explored the benefits of adding germane cognitive load by manipulating the variability of the training problems. As a ‘digression,’ a fourth experiment was performed to test the value of pupil dilation as a correlate of cognitive load (Chapter 4). In this final chapter the outcomes of the aforementioned experiments will be discussed in terms of the benefits they entail for elderly learners. Finally, suggestions will be given for future research.

What do we learn from this thesis? First of all, this thesis has demonstrated that CLT-based training formats generally impose less cognitive load on the learner than conventional formats. This holds for training with worked examples in general and, more specifically, for training with multimedia-based and bimodal worked examples. Second, CLT-based training formats bring forth a comparable performance level as conventional formats. At first sight, this may seem a disappointing result. However, the finding that active practice (i.e., solving conventional problems) does not automatically lead to better performance than ‘passive’ practice (i.e., studying worked examples) can be regarded an even more disappointing result for the conventional format. In this respect, the aphorism “practice makes perfect” has certainly become a relative assertion. A third major outcome concerns the ratio between cognitive load and performance. Since CLT-based training formats impose a relatively low cognitive load on the learner and lead to a comparable level of performance as conventional formats, they make a more efficient use of the available cognitive resources. High-variability training, finally, seems to have a positive effect on performance.

Do Elderly Learners Benefit from Cognitive Load Theory?

The main objective of this thesis concerns the question whether *elderly* learners benefit from CLT-based training formats. The answer to this question is a plain “yes.” With respect to the level of experienced cognitive load, the elderly profited from CLT-based training formats to at least a comparable extent as the young. The same holds for training efficiency. Apart from the experiment described in Chapter 5, however, no performance gains were observed for either age group. A solution to this problem still lies in a proper way to increase germane cognitive load such that it obeys cognitive resource limits.

The ultimate question arises from the major hypothesis formulated in Chapter 1: Do the elderly gain more from CLT-based training formats than the young? The answer to this question is mixed. The experiment in Chapter 2 shows that the elderly gain *more* from worked examples in terms of efficiency than their young counterparts. This result was replicated by the experiment reported in Chapter 3, but only for the contrast between conventional problems and multimedia-based worked examples, and only with respect to subjective cognitive load. The experiment described in Chapter 5 failed to replicate the effect. This study revealed a slight effect of random training on performance, however.

In conclusion, cognitive load theory has proven to be a promising guide for designing training formats aimed at elderly learners. However, in order to establish its true value, additional explorations are certainly required. For this purpose, the next and final section provides some suggestions.

Suggestions for Future Research

Let us return to the integrative framework described in Chapter 1 (see Figure 1.2). Evidently, the present thesis does not provide a complete account for this framework. It does underline the relation between age and cognitive capacity, however. Furthermore, it highlights the possibility of manipulating extraneous cognitive load through instructions. One of the most important issues raised in this thesis is the manipulation of germane cognitive load. The experiment described in Chapter 5 provides an indication of how germane cognitive load can lead to better performance by manipulating training variability. This seizes upon the issue of how to manage the total level of cognitive load during training.

Controlling Germane Cognitive Load

The studies discussed in Chapters 2, 3, and 5 revealed rather low levels of perceived cognitive load. Since these were general measures, they concerned both extraneous and germane cognitive load (intrinsic load is con-

sidered a uniform baseline here). Differences in the total level of cognitive load were ascribed to a manipulation of extraneous load. Controlling extraneous load is only one side of the picture, however. Low levels of general load imply that enough cognitive resources are available to raise the total amount of load. In line with cognitive load theory, the best way to use free mental capacity is to stimulate the cognitive system to undertake operations that contribute to the construction of cognitive schemata. As may have become clear, the notion of free mental space tends to be a post-hoc affair: The level of perceived cognitive load — and thus the indication of free cognitive capacity — seems unpredictable. It pops up after testing the experimental training conditions. Furthermore, cognitive load depends on individual differences, not only in age, but also in cognitive abilities. In this light, instant fine-tuning of cognitive load would be the ultimate challenge. This can be achieved by adjusting the level of germane cognitive load to the individual and momentary resources of the learner. A nice example in this respect is a study by Camp, Paas, Rikers, and Van Merriënboer (2001). They sequenced training problems on the basis of performance on previous problems as well as the level of cognitive load imposed by these problems. The result was an enhanced test performance, which resulted in a higher training efficiency relative to a condition in which the training problems were presented in a fixed sequence. Dynamic training-problem selection seems an appropriate method for enhancing performance in the elderly, especially regarding the notion that this group generally shows much more individual differences on such variables as intelligence and computation span than their young counterparts (to get a picture, one only needs to compare the standard deviations of these variables in the previous chapters).

Selecting Measurement Techniques

A second major issue raised in this thesis is the measurement of cognitive load. Selecting an appropriate technique for measuring cognitive load seems to be a matter of expediency, since there is no measure that satisfies all essential criteria. But what are these criteria?

First, there is *sensitivity*. A cognitive-load measure should be sensitive in that it can detect small differences in ‘normal’ cognitive load. Subjective measures seem to meet this criterion, but are prone to response biases. Physiological measures seem rather insensitive to different levels of load (e.g., Paas, Van Merriënboer, & Adam, 1994). An exception is the pupillary response (see Beatty & Lucero-Wagoner, 2000). Chapter 4 has shown, however, that with respect to the elderly the pupillary response is not sensitive under all conditions. That is, a linear relation between memory load and pupil dilation was found when older participants were engaged in memorizing sequences of one to six digits. However, when they were re-

quired to search these sequences for a particular probe, the pupillary response (as a function of sequence length) showed a flat line.

A second criterion is *objectivity*. Providing that participants are not, or only to a limited extent able to give an unbiased rating of their experienced cognitive load, the criterion of objectivity is only met by non-subjective measures. Although subjective ratings of cognitive load have proven to be both reliable and sensitive (e.g., Gopher & Braune, 1984; Paas, 1992; Paas & Van Merriënboer, 1994b), one could still consider it doubtful whether they provide a valid picture of both the absolute and relative magnitudes of cognitive load.

A final criterion concerns the *intrusiveness* of the measurement procedure. It is almost impossible to probe cognitive load without disturbing the participant in an ongoing task. Subjective measures and secondary task measures, for instance, require the participant to either interrupt the task or to divide attention between the task and the measurement procedure. Physiological measures are often physically intrusive. The pupillary measurement technique described in Chapter 4 is an exception in this respect. This procedure involved a remote eye-tracking device in which the only physical contact concerned supporting the participant's head by means of a chin and forehead rest. Especially for elderly participants, both cognitive and physical intrusiveness are essential factors to take into account when selecting a procedure for tracking cognitive load.

Since none of the techniques discussed above meet the criteria of sensitivity, objectivity, and intrusiveness, it seems wise to combine different techniques such that advantages are added up and weaknesses are cancelled out. Combining the sensitivity of a subjective technique with the objectivity of a secondary-task method, for instance (see Chapter 5), is likely to give a more reliable picture of cognitive load than a subjective technique on its own. In the same light, it would be interesting to combine a subjective technique with pupillometry. Although the latter technique seems not to be fully appropriate for elderly persons, its true value will only come to light by testing it in a broader variety of tasks. Combining a subjective with an objective method would furthermore acknowledge Paas & Van Merriënboer's (1994a) definition of cognitive load as consisting of a mental-load (i.e., objective) and a mental-effort (i.e., subjective) component (see Chapter 1, Figure 1.1; see also Paas, Tuovinen, Tabbers, & Van Gerven, 2002).

Different Kinds of Load

In this thesis, instructional cognitive load was conceptualized as consisting of three components: intrinsic cognitive load, extraneous cognitive load, and germane cognitive load. Intrinsic cognitive load was associated with the characteristics of the subject matter or problem domain (e.g., Luchins' water-jug problem). Extraneous cognitive load was ascribed to unneces-

sary aspects of the training format (e.g., split attention). Germane cognitive load was defined as the load generated by mental operations that directly contribute to the construction of cognitive schemata (e.g., associating problems states with the appropriate operators). Now, let us again return to the framework described in Chapter 1. An essential part of this framework concerns the connections between the variables instructions, extraneous cognitive load, and germane cognitive load. It is assumed that both extraneous and germane cognitive load can be manipulated by the instructions (i.e., the training format). That is, the more the instructions are based on the principles of cognitive load theory, the lower the level of extraneous cognitive load and the higher (i.e., more optimal) the level of germane cognitive load. In addition, it is assumed that as the schema-acquisition process develops, the level of intrinsic cognitive load decreases. These assumptions bring us to a delicate issue. Defining different kinds of cognitive load as well as postulating specific statements around them inevitably leads to the obligation to operationalize these types of load and demonstrate their associated effects. Only in this way, cognitive load theory can transcend from a clever set of instructional guidelines, which has strongly proven its effectiveness, to a cognitive model whose underlying mechanism is fully accounted for.

It has to be noted that in this thesis — as well as in the whole range of literature on cognitive load theory — there is a discrepancy between theory and research in that the multi-dimensional concept of cognitive load is reduced to a general operationalization, which is reflected by a single measure. How can we unravel this general concept of cognitive load into its constituents? The notion that three kinds of cognitive load are simultaneously imposed on the learner may even provoke the more disturbing question of whether it is possible at all to empirically disentangle these constituent types of load. To answer these questions, a drastic reconsideration of the empirical method within CLT research is needed. Some suggestions will be given below.

Intrinsic cognitive load may be operationalized on the basis of a detailed task or skill analysis. Skill analysis typically entails decomposition of the skill into its elementary procedures, which are described in terms of substeps and logical rules (e.g., Van Merriënboer, 1997). A drawback of task analysis is that it does not necessarily reflect the cognitive operations that are required to perform a particular task. This would call for a careful mapping of task-analysis results on, for instance, a detailed verbal-protocol analysis (e.g., Ericsson & Simon, 1980). Furthermore, task analysis does not give information about how *individuals* experience intrinsic cognitive load. Hence, the results of a task analysis have to be associated with measures of individual capacities as well as prior knowledge. Since cognitive load theory offers clear definitions of which cognitive operations are relevant and which are not, task and protocol analysis may as well

provide specific operationalizations of extraneous and germane cognitive load.

There is still one fundamental issue to be addressed, however: Does cognitive load simply equal the sum of a skill's constituents or does one constituent play a more dominant role than the other? In other words, should a constituent skill be weighed with respect to the amount of load it imposes on the cognitive system? According to Sweller (1994), the intrinsic cognitive load imposed by a task is not so much determined by the number of elements it entails, but rather by the extent to which these elements are interdependent. Interdependent elements have to be simultaneously active in working memory. The larger the cluster of mutually dependent elements, the more elements have to be activated in working memory at the same time. A standardized procedure for determining element interdependency — or as Sweller puts it, “element interactivity” — would give a useful indication of the amount of cognitive load a task imposes on the cognitive system.

A straightforward alternative to skill analysis would be the construction of a questionnaire, which comprises CLT-specific scaling items. That is, simply ask participants to indicate their perceived cognitive load with respect to task (i.e., intrinsic cognitive load) and instructional characteristics (i.e., extraneous and germane cognitive load) on separate scales. This method has the advantage of being far less laborious. It raises the question, however, whether people are able to differentiate between the different types of task and training characteristics and their associated types of cognitive load. This method would at least require considerable training of the respondents.

It may be clear that it is desirable to operationalize the constituent types of cognitive load by the same method, since this will provide an indication of their relative magnitude. Moreover, it offers the opportunity to predict the total amount of load as measured by the aforementioned subjective and objective measurement techniques. If, for example, a CLT-based task analysis successfully predicts that the total amount of cognitive load increases by adding germane cognitive load, while keeping the other types of load constant, this is an important result. The same holds for manipulating the other types of cognitive load. A systematic analysis of a series of such experiments would uncover the specific cognitive mechanism that underlies the success of CLT-based instructional formats. At the same time, it would establish the true importance of cognitive load theory for applied psychology in general and educational psychology in particular.

What about Schema Construction?

Yet another operationalization issue concerns the central role of schemata within cognitive load theory. The framework described in Chapter 1 shows that germane cognitive load promotes schema construction and that

schema construction in its turn promotes performance. Because schema construction is assigned this explicit role, it is necessary to provide a detailed, *task-specific* definition of a cognitive schema. Although the exact representation of a schema will always be shrouded in mystery, nothing stands in the way of at least formulating a working definition. It is peculiar, therefore, that only little of the research done in the field of cognitive load theory comprises a dependent variable that directly refers to the construction of cognitive schemata. Most research — including the present thesis — focuses on performance in terms of the proportion of correctly solved test problems, the mean time needed for solving the problems, or the number of errors. These variables are *hypothetically* influenced by the construction of cognitive schemata, because cognitive schemata are supposed to facilitate transfer performance and boost problem-solving speed through rule automation. This hypothesized causal relation is scarcely accounted for. That is, the actual construction of cognitive schemata is rarely demonstrated experimentally. An exception is a study by Cooper and Sweller (1987), who probed schema construction by collecting verbal protocols during problem solving. They found that participants who had been deeply engaged in studying worked examples not only performed better than participants who had been solving conventional problems, but also reported more schema-related knowledge. Although the verbal protocols played a relatively minor role in Cooper and Sweller's study, they represent an important finding, namely that cognitive schemata — not some other factors — are responsible for enhanced performance. Thus, in the light of the framework presented in Chapter 1, incorporating one or more dependent variables that relate to task-specific cognitive schemata would further test the theoretical fundamentals of cognitive load theory.

There is More to Cognitive Load Theory

It needs little explanation that there is more to cognitive load theory than the instructional formats evaluated in this thesis. Worked examples, bimodal presentation formats, and training variability represent only a limited selection out of the broad arsenal of training strategies that cognitive load theory advances (see Sweller, Van Merriënboer, & Paas, 1998, for a complete overview). Strong conclusions concerning the value of cognitive load theory for elderly learners would thus give only little credit to its potential power. One only needs to think of a simple phenomenon like the redundancy effect (e.g., Bobis et al., 1993) and combine it with Hasher and Zacks' (1988) reduced-inhibition view to pave the way for a new and promising line of research.

As pointed out in Chapter 1, every instructional format that is designed according to the 'rules' of cognitive load theory is likely to benefit the elderly, because it stimulates an efficient use of the available cognitive resources. For that matter, the studies presented in this thesis have pro-

vided only a glimpse of CLT's possibilities. However, before plunging into a sea of instructional-design options, future research should strengthen the basic principles of cognitive load theory. It should first and foremost address a range of fundamental questions. How can we differentiate between the different kinds of cognitive load? How does the ratio between extraneous and germane cognitive load account for test performance? Do factors that we relate to germane cognitive load, such as training variability, truly contribute to schema construction? Does schema construction account for enhanced performance? Does schema construction reduce intrinsic cognitive load? Only if we can formulate a full answer to these questions, the theoretical pillars of cognitive load theory can be consolidated.

The Pragmatic Approach

The foregoing section may have evoked the uncomfortable feeling of scientific fundamentalism. One could easily make the opposing statement that if an instructional design is successful in terms of experienced cognitive load and performance, it is worth reporting as an achievement for educational science. Moreover, such a finding would serve society more than demonstrating some deeper cognitive mechanism. So why make a fuss if it simply works? Surely, there is a lot to say in favor of such a pragmatic approach. But what about educational *psychology*? Does not the label "psychology" require a more thorough explanation than simply demonstrating that it works? It is true that cognitive load theory provides a psychological explanation for the phenomena it predicts. But is it demonstrated empirically? As may have become clear, it is not or at least not conclusively.

So what about the elderly? Are they not particularly served by the pragmatic approach? A confirmatory answer to this question would be all too obvious. As a concluding thought, consider what Perfect and Maylor (2000) refer to as the "dull hypothesis." The dull hypothesis simply states that young people outperform old people in a broad range of cognitive tasks. The authors' message is that if cognitive-aging research wants to have anything to offer, it should attempt to reject this dull hypothesis. In other words, experiments that demonstrate that older persons perform less than young persons are not exactly astonishing. Truly noteworthy are experiments, which demonstrate that older people are able to achieve a performance level that is comparable to the performance level of young people. In statistical terms, this means that interactions are far more interesting than main effects. In a somewhat modified form, rejection of the dull hypothesis should definitely be pursued by research on cognitive load theory and aging. Thus, if we adopt the pragmatic approach, CLT-based training formats should not only lead to better results for elderly learners than conventional formats. Above all, it should lift the elderly's performance level towards that of the young. This can be considered the ultimate challenge of educational gerontology.

References

- Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science*, *205*, 1289-1292.
- Ainsworth, S.E., Wood, D.J., & Bibby, P.A. (1997, August). *Evaluating principles for multi-representational learning environments*. Paper presented at the 7th conference of the European Association for Research on Learning and Instruction (EARLI), Athens, Greece.
- Allen, P.A. (1990). Influence of processing variability on adult age differences in memory distribution of order information. *Cognitive Development*, *5*, 177-192.
- Allen, P.A., Madden, D.J., Groth, K.E., & Crozier, L.C. (1992). Impact of age, redundancy, and perceptual noise on visual search. *Journal of Gerontology: Psychological Sciences*, *47*, 69-74.
- Anstey, K., Stankov, L., & Lord, S. (1993). Primary aging, secondary aging, and intelligence. *Psychology and Aging*, *8*, 562-570.
- Atwood, M.E., Masson, M.E.J., & Polson, P.G. (1980). Further explorations with a process model for water jug problems. *Memory & Cognition*, *8*, 182-192.
- Atwood, M.E., & Polson, P.G. (1976). A process model of water jug problems. *Cognitive Psychology*, *8*, 191-216.
- Authorware 3.5 [Computer software]. (1996). San Francisco: Macromedia.
- Backs, R.W., & Walrath, L.C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*, *23*, 243-254.
- Baddeley, A. (1992). Working memory. *Science*, *255*, 556-559.
- Bargh, J.A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230-244.
- Battig W.F. (1972). Intratask interference as a source of facilitation in transfer and retention. In R.F. Thompson & J.F. Voss (Eds.), *Topics in learning and performance* (pp. 131-159). New York: Academic Press.
- Battig, W.F. (1979). The flexibility of human memory. In L.S. Cermak & F.I.M. Craik (Eds.), *Levels of processing in human memory* (pp. 23-44). Hillsdale, NJ: Erlbaum.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*, 276-292.
- Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J.T. Cacioppo, L.G. Tassinary, & G.G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142-162). Cambridge, MA: Cambridge University Press.
- Berent, I., & Perfetti, C.A. (1993). An on-line method in studying music parsing. *Cognition*, *46*, 203-222.
- Beringer, J. (1996). Experimental Run Time System (ERTS). [Computer software]. Frankfurt, Germany: BeriSoft.
- Bitsios, P., Prettyman, R., & Szabadi, E. (1996). Changes in autonomic function with age: A study of pupillary kinetics in healthy young and old people. *Age and Ageing*, *25*, 432-438.
- Bobis, J., Sweller, J., & Cooper, M. (1993). Cognitive load effects in a primary-school geometry task. *Learning and Instruction*, *3*, 1-21.
- Bobis, J., Sweller, J., & Cooper, M. (1994). Demands imposed on primary-school stu-

References

- dents by geometric models. *Contemporary Educational Psychology*, *19*, 108-117.
- Camp, G., Paas, F., Rikers, R., & Van Merriënboer, J.J.G. (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort, and mental efficiency. *Computers in Human Behavior*, *17*, 575-595.
- Cerella, J. (1990). Aging and information-processing rate. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 201-221). San Diego: Academic Press.
- Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, *62*, 233-246.
- Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*, *10*, 151-170.
- Clark, J.M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*, 149-210.
- Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, *79*, 347-362.
- Craik, F.I.M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F.I.M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 191-211). New York: Plenum Press.
- Czaja, S.J. (1996). Aging and the acquisition of computer skills. In W.A. Rogers, A.D. Fisk, & N. Walker (Eds.), *Aging and skilled performance: Advances in theory and applications* (pp. 201-220). Mahwah, NJ: Erlbaum.
- Czaja, S.J., & Sharit, J. (1993). Age differences in the performance of computer-based work. *Psychology and Aging*, *8*, 59-67.
- Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450-466.
- De Croock, M.B.M., Van Merriënboer, J.J.G., & Paas, F. (1998). High versus low contextual interference in simulation-based training of troubleshooting skills: Effects on transfer performance and invested mental effort. *Computers in Human Behavior*, *14*, 249-267.
- De Klerk, M.M.Y. (Ed.). (2001). *Rapportage ouderen 2001: Veranderingen in de leef-situatie* [Report on the elderly 2001: Changes in social situation]. The Hague, The Netherlands: Social and Cultural Planning Office.
- Dijksterhuis, A., Aarts, H., Bargh, J.A., & Van Knippenberg, A. (2000). On the relation between associative strength and automatic behavior. *Journal of Experimental Social Psychology*, *36*, 531-544.
- Einstein, G.O., Smith, R.E., McDaniel, M.A., & Shaw, P. (1997). Aging and prospective memory: The influence of increased task demands at encoding and retrieval. *Psychology and Aging*, *12*, 479-488.
- Ericsson, K.A., & Simon, H.A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215-251.
- Fisk, J.E., & Warr, P. (1996). Age and working memory: The role of perceptual speed, the central executive, and the phonological loop. *Psychology and Aging*, *11*, 316-323.
- Gilinsky, A.S., & Judd, B.B. (1994). Working memory and bias in reasoning across the life span. *Psychology and Aging*, *9*, 356-371.
- Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors*, *26*, 519-532.
- Granholm, E., Asarnow, R.F., Sarkin, A.J., & Dykes, K.L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*, 457-461.
- Granholm, E., Morris, S.K., Sarkin, A.J., Asarnow, R.F., & Jeste, D.V. (1997). Pupillary responses index overload of working memory resources in schizophrenia. *Journal of Abnormal Psychology*, *106*, 458-467.

- Hartman, M., & Hasher, L. (1991). Aging and suppression: Memory for previously relevant information. *Psychology and Aging, 6*, 587-594.
- Hasher, L., & Zacks, R.T. (1988). Working memory, comprehension, and aging: A review and a new view. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 193-225). San Diego: Academic Press.
- Hess, E.H., & Polt, J.M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143*, 1190-1192.
- Heuer, H. (1996). Dual-task performance. In O. Neumann & A.F. Sanders (Eds.), *Handbook of perception and action* (Vol. 3, pp. 113-153). San Diego: Academic Press.
- Hoeks, B., & Levelt, W.J.M. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments, & Computers, 25*, 16-26.
- Horn, J.L., & Cattell, R.B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26*, 107-129.
- Hyönä, J., Tommola, J., & Alaja, A.M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Quarterly Journal of Experimental Psychology, 48A*, 598-612.
- Janisse, M.P. (1977). *Pupillometry: The psychology of the pupillary response*. Washington, DC: Hemisphere.
- Jolles, J., Houx, P.J., Van Boxtel, M.P.J., & Ponds, R.W.H.M. (Eds.). (1995). *The Maastricht Aging Study: Determinants of cognitive aging*. Maastricht, The Netherlands: Neuropsych Publishers.
- Just, M.A., & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.
- Just, M.A., & Carpenter, P.A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology, 47*, 310-339.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583-1585.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.
- Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*, 126-136.
- Kidder, D.P., Park, D.C., Hertzog, C., & Morrell, R.W. (1997). Prospective memory and aging: The effects of working memory and prospective memory task load. *Aging, Neuropsychology, and Cognition, 4*, 93-112.
- Kim, M., Beversdorf, D.Q., & Heilman, K.M. (2000). Arousal response with aging: Pupillo-graphic study. *Journal of the International Neuropsychological Society, 6*, 348-350.
- Korzyn, A.D., Laor, N., & Nemet, P. (1976). Sympathetic pupillary tone in old age. *Archives of Ophthalmology, 94*, 1905-1906.
- Lippman, L.G. (1994). Rule-governed performance and sensitivity to contingencies: What's new? *Journal of General Psychology, 121*, 353-360.
- Lippman, L.G. (1996). A short report concerning the Einstellung effect. *Journal of General Psychology, 123*, 233-235.
- Loewenfeld, I.E. (1979). Pupillary changes related to age. In H.S. Thompson, R. Daroff, L. Frisen, J.S. Glaser, & M.D. Sanders (Eds.), *Topics in neuro-ophthalmology* (pp. 124-150). Baltimore: Williams & Wilkins.
- Lorsbach, T.C., & Simpson, G.B. (1988). Dual-task performance as a function of adult age and task complexity. *Psychology and Aging, 3*, 210-212.
- Luchins, A.S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs, 54*, Whole No. 248.

References

- Luchins, A.S., & Luchins, E.H. (1991). Task complexity and order effects in computer presentation of water jar problems. *Journal of General Psychology, 118*, 45-72.
- Luteijn, F., & Van der Ploeg, F.A.E. (1983). *Groninger intelligentie test: Handleiding* [Groningen intelligence test: Manual]. Lisse, The Netherlands: Swets & Zeitlinger.
- Madden, D.J., Connelly, S.L., & Pierce, T.W. (1994). Adult age differences in shifting focused attention. *Psychology and Aging, 9*, 528-538.
- Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 88*, 49-63.
- Matthews, G., Middleton, W., Gilmartin, B., & Bullimore, M.A. (1991). Pupillary diameter and cognitive load. *Journal of Psychophysiology, 5*, 265-271.
- Mayer, R.E., & Anderson, R.B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83*, 484-490.
- Mayer, R.E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology, 88*, 64-73.
- Mayer, R.E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312-320.
- Mayer, R.E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*, 107-119.
- McDowd, J.M. (1997). Inhibition in attention and aging. *Journal of Gerontology: Psychological Sciences, 52*, 265-273.
- McDowd, J.M., & Craik, F.I.M. (1988). Effects of aging and task difficulty on divided attention performance. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 267-280.
- Meshkati, N., Hancock, P.A., Rahimi, M., & Dawes, S.M. (1995). Techniques in mental workload assessment. In J. R. Wilson & E. N. Corlett (Eds.), *Evaluation of human work: A practical ergonomics methodology* (pp. 749-782). Philadelphia: Taylor & Francis.
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.
- Moreno, R., & Mayer, R.E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358-368.
- Morrell, R.W., & Echt, K.V. (1996). Instructional design for older computer users: The influence of cognitive factors. In W.A. Rogers, A.D. Fisk, & N. Walker (Eds.), *Aging and skilled performance: Advances in theory and applications* (pp. 241-265). Mahwah, NJ: Erlbaum.
- Morrell, R.W., & Echt, K.V. (1997). Designing written instructions for older adults: Learning to use computers. In A.D. Fisk & W.A. Rogers (Eds.), *Handbook of human factors and the older adult* (pp. 335-361). San Diego: Academic Press.
- Morrell, R.W., & Park, D.C. (1993). The effects of age, illustrations, and task variables on the performance of procedural assembly tasks. *Psychology and Aging, 8*, 389-399.
- Morris, J.M. (1994). Computer training needs of older adults. *Educational Gerontology, 20*, 541-555.
- Morris, S.K., Granholm, E., Sarkin, A.J., & Jeste, D.V. (1997). Effects of schizophrenia and aging on pupillographic measures of working memory. *Schizophrenia Research, 27*, 119-128.
- Mousavi, S.Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*, 319-334.
- Myerson, J., Hale, S., Wagstaff, D., Poon, L.W., & Smith, G.A. (1990). The information-loss model: A mathematical theory of age-related cognitive slowing. *Psychological Review, 97*, 475-487.

- Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, *33*, 17-33.
- Owen, E., & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Educational Psychology*, *77*, 272-284.
- 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*, 181-186.
- Paas, F., Tuovinen, J.E., Tabbers, H., & Van Gerven, P.W.M. (2002). Cognitive load measurement as a means to advance cognitive load theory. Manuscript submitted for publication.
- Paas, F., & Van Merriënboer, J.J.G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, *35*, 737-743.
- Paas, F., & Van Merriënboer, J.J.G. (1994a). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*, 351-371.
- Paas, F., & Van Merriënboer, J.J.G. (1994b). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive load approach. *Journal of Educational Psychology*, *86*, 122-133.
- Paas, F., Van Merriënboer, J.J.G., & Adam, J.J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*, 419-430.
- Peavler, W.S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology*, *11*, 559-566.
- Penney, C.G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17*, 398-422.
- Perfect, T.J., & Maylor, E.A. (2000). Rejecting the dull hypothesis: The relation between method and theory in cognitive aging research. In T.J. Perfect & E.A. Maylor (Eds.), *Models of cognitive aging* (pp. 1-18). Oxford, UK: Oxford University Press.
- Pillay, H. (1994). Cognitive load and mental rotation: Structuring orthographic projection for learning and problem solving. *Instructional Science*, *22*, 91-113.
- Pirolli, P.L., & Anderson, J.R. (1985). The role of learning from examples in the acquisition of recursive programming skills. *Canadian Journal of Psychology*, *39*, 240-272.
- Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, *12*, 61-86.
- Rabbitt, P.M.A. (1993). Crystal quest: A search for the basis of maintenance of practiced skills into old age. In A.D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control: A tribute to Donald Broadbent* (pp.188-230). Oxford, UK: Clarendon Press/Oxford University Press.
- Ransopher, S.B., & Thompson, D.N. (1991). Einstellung rigidity, set induction, and problem solving in the elderly. *Educational Gerontology*, *17*, 219-227.
- Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology*, *23*, 90-108.
- Salthouse, T.A. (1993). Speed mediation of adult age differences in cognition. *Developmental Psychology*, *29*, 722-738.
- Salthouse, T.A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403-428.
- Salthouse, T.A. (2000). Steps toward the explanation of adult age differences in cognition. In T.J. Perfect & E.A. Maylor (Eds.), *Models of cognitive aging* (pp. 19-49). Oxford, UK: Oxford University Press.

References

- Salthouse, T.A., & Babcock, R.L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*, 763-776.
- Salthouse, T.A., Fristoe, N.M., Lineweaver, T.T., & Coon, V.E. (1995). Aging and attention: Does the ability to divide decline? *Memory & Cognition, 23*, 59-71.
- ScreenCam [Computer software]. (1997). Cambridge, MA: Lotus.
- Sekuler, R., & Owsley, C. (1983). Visual manifestations of biological aging. *Experimental Aging Research, 9*, 253-255.
- Snijders, J.Th., & Verhage, F. (1962). *Voorlopige handleiding bij de Groninger Intelligentie Test* [Provisional manual to the Groningen Intelligence Test]. Amsterdam: Swets & Zeitlinger.
- Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652-654.
- Sternberg, S. (1967). Retrieval of contextual information from memory. *Psychonomic Science, 8*, 55-56.
- Stoltzfus, E.R., Hasher, L., & Zacks, R.T. (1996). Working memory and aging: Current status of the inhibitory view. In J.T.E. Richardson, R.W. Engle, L. Hasher, R.H. Logie, E.R. Stoltzfus, & R.T. Zacks (Eds.), *Working memory and human cognition* (pp. 66-88). New York: Oxford University Press.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.
- Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology, 81*, 457-466.
- Sweller, J. (1993). Some cognitive processes and their consequences for the organisation and presentation of information. *Australian Journal of Psychology, 45*, 1-8.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295-312.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185-233.
- Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General, 119*, 176-192.
- Sweller, J., & Cooper, G.A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*, 59-89.
- Sweller, J., & Levine, M. (1982). Effects of goal specificity on means-ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 463-474.
- Sweller, J., Van Merriënboer, J.J.G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.
- Tabbers, H.K., Martens, R.L., & Van Merriënboer, J.J.G. (2001). The modality effect in multimedia instructions. In J.D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third annual conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*, 257-287.
- Van Gerven, P.W.M., Paas, F., & Van Merriënboer, J.J.G. (1998, April). *Instructional efficiency: Relations between cognitive load, task complexity, and aging*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), San Diego, CA.
- Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., Hendriks, M., & Schmidt, H.G. (2002). The efficiency of multimedia training into old age. Manuscript submitted for publication.
- Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., & Schmidt, H.G. (2000). Cogni-

- tive load theory and the acquisition of complex cognitive skills in the elderly: Towards an integrative framework. *Educational Gerontology*, *26*, 503-521.
- Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., & Schmidt, H.G. (2002a). Cognitive load theory and aging: Effects of worked examples on training efficiency. *Learning and Instruction*, *12*, 87-105.
- Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., & Schmidt, H.G. (2002b). Memory load and task-evoked pupillary responses in aging. Manuscript submitted for publication.
- Van Merriënboer, J.J.G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.
- Van Merriënboer, J.J.G., De Croock, M.B.M., & Jelsma, O. (1997). The transfer paradox: Effects of contextual interference on retention and transfer performance of a complex cognitive skill. *Perceptual and Motor Skills*, *84*, 784-786.
- Van Merriënboer, J.J.G., & Paas, F. (1990). Automation and schema acquisition in learning elementary computer programming: Implications for the design of practice. *Computers in Human Behavior*, *6*, 273-289.
- Van Merriënboer, J.J.G., Schuurman, J.G., De Croock, M.B.M., & Paas, F. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance, and training efficiency. *Learning and Instruction*, *12*, 11-37.
- Verney, P., Granholm, E., & Dionisio, D.P. (2001). Pupillary responses and processing resources on the visual backward masking task. *Psychophysiology*, *38*, 76-83.
- Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*, 1-39.
- Welford, A.T. (1985). Changes of performance with age: An overview. In N. Charness (Ed.), *Aging and human performance* (pp. 333-365). New York: Wiley.
- Wierwille, W.W., Rahimi, M., & Casali, J.G. (1985). Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. *Human Factors*, *27*, 489-502.
- Wingfield, A., Stine, E.A.L., Lahar, C.J., & Aberdeen, J.S. (1988). Does the capacity of working memory change with age? *Experimental Aging Research*, *14*, 103-107.
- Winn, B., Whitaker, D., Elliott, D.B., & Phillips, N.J. (1994). Factors affecting light-adapted pupil size in normal human subjects. *Investigative Ophthalmology & Visual Science*, *35*, 1132-1137.
- Zhu, X., & Simon, H.A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction*, *4*, 137-166.

Summary

The mainspring behind the present thesis is the broadly recognized notion of a fast-developing technological society, in which a large and growing part of the population is of advanced age. Research has shown that this development implies a growing interest in retraining, both in the professional and non-professional realm. Designing instructional material for elderly learners — that is, people aged sixty and over — requires special care, in particular when the goal is to master complex subject matter. Instructional strategies aimed at acquiring complex skills have already been developed and tested with children and relatively young adults. A theory promoting such strategies is cognitive load theory (CLT). In numerous studies, instructional designs springing from cognitive load theory have proven to be superior relative to conventional instructional designs. Elderly learners have never been a target group in these studies, however. This is considered a gap within CLT research. To fill the gap, it was attempted in this thesis to merge cognitive load theory with knowledge emerging from research on cognitive aging.

Cognitive load theory is an instructional theory, in which fundamental knowledge of cognitive psychology is used to devise instructional designs that stimulate an optimal use of the available cognitive resources. One such fundamental notion is that human processing capacity is limited. More specifically, human working memory is able to process about seven information units at the same time. Another important feature of working memory is that it interacts with a long-term store, in which information is not so much stored as a pile of independent information elements, but rather as clusters of elements that are strongly interrelated. An advantage of these clusters is that they are considered as a single element in working memory. In a manner of speaking, they increase the processing capacity of working memory. A final feature of working memory is that it consists of two independent, modality-specific ‘slave’ units — a visual and an auditory unit — which are mediated by a central executive. The central executive is responsible for the conscious processing of information from the modality-specific stores as well as from the long-term store.

These basic ideas about working memory were adopted by cognitive load theory by putting them in an educational context. Within cognitive load theory, complex skills are believed to consist of information elements that are highly interdependent. This inbuilt characteristic causes a high intrinsic cognitive load. In order to grasp a complex skill, it has to be stored in long-term memory as whole. For that purpose, however, it must

pass the narrow bridge of working memory. According to cognitive load theory, this can be achieved efficiently if working memory is not engaged in unnecessary activities. Unnecessary activities cause an extraneous cognitive load, which can obstruct a proper transfer of complex information from working memory to long-term memory. Instead, working memory has to engage in activities that assemble interdependent information elements into relevant clusters or schemata. Cognitive schemata are organized in such a way that they enable a person to solve problems of a particular category. The quality of a schema determines the range of problems a person can solve. That is, the richer a schema, the broader the range of problems a person can handle. Thus, an elaborate schema enables a person to tackle problems that deviate substantially from the problems he or she has seen during training. This phenomenon is called transfer. Moreover, cognitive schemata can get automated. The application of automated schemata require little conscious processing, which is a relieve for working memory. The load associated with the construction of such schemata is called germane cognitive load.

In a nutshell, cognitive load theory states that when dealing with complex material — which is material that imposes a high intrinsic cognitive load — it is essential to keep the extraneous level of load as low as possible and make the germane portion of load as high as possible, that is within the limits of working-memory capacity. In other words, cognitive load theory promotes instructional designs that prevent learners from getting involved in unnecessary mental activity and at the same time stimulate the learner to engage in mental activity that contributes to the construction of high-quality cognitive schemata.

In Chapter 1, this idea was transferred to the domain of cognitive aging. In this thesis, cognitive aging refers to changes of working memory in elderly people. In this light, the literature describes three phenomena. First, the processing capacity of working memory seems to decrease. Second, there is a general decrease of cognitive processing speed. Third, older people have difficulty inhibiting irrelevant information, which can impede the proper processing of task-relevant information. These changes limit the rate at which older people are able to learn new material. It appears that the narrow bridge of working memory has become even narrower in the elderly. In the first chapter, it was argued that cognitive load theory offers the strategies to effectively cope with the learning difficulties associated with these working-memory constraints. That is, because cognitive load theory is concerned with mental limitations and these limitations are especially observable in the elderly, it goes without saying that CLT-based instructional formats should be particularly helpful for this group of learners. After all, lowering the level of extraneous cognitive load produces a proportionally greater effect in an older person than in a young person.

This idea was tested in Chapters 2, 3, and 5, applying different CLT-based training designs.

In Chapter 2, the cognitive efficiency of training with worked examples was tested against the efficiency of training with conventional problems. Within cognitive load theory, worked examples can be considered the basic format of efficient complex skill training. In conventional instructional formats, studying worked examples takes a minor position relative to solving conventional training problems, which arises from the common-and-garden philosophy that “practice makes perfect.” Indeed, practice — in the conventional sense — eventually makes perfect, but it appears that this does not automatically hold for an early stage of learning. In an early stage of learning, appropriate cognitive schemata have not yet developed. Thus, when confronted with a new complex skill, learners have little information at their disposal that could help them tackle the problems in this novel area. As a result, they have to process all the details of the problem consciously, without being aided by existing schemata. In this situation, learners are inclined to apply a so-called means-ends strategy. A means-ends strategy involves both backward and forward processing. In the backward-processing phase, a ‘pyramid’ of goals and subgoals is constructed, starting at the top of the pyramid, which represents the main goal of the problem, and subsequently working through the subordinate layers, which represent the subgoals. This phase halts at a layer containing manageable subgoals. From that point on, the forward-processing phase begins. In this phase, the learner solves the subgoals in the lower layers and works forward via the higher-level subgoals to the main goal. The combination of backward and forward processing makes a means-ends strategy a capacity-demanding exercise. More importantly, only half of the exercise — the forward-processing phase — can be considered necessary. In the forward phase, problem states are associated with the operators that are necessary to advance into the next problem state and eventually into the goal state. This is exactly what novices should learn. By studying worked examples, learners are stimulated to exclusively expend their cognitive resources to forward processing. Thus, extraneous load (i.e., backward processing) is reduced and germane load (i.e., forward processing) is optimized. Therefore, studying worked examples should lead to better results in terms of cognitive load, schema construction, and transfer performance than solving conventional problems.

An experiment was conducted to test this hypothesis. Both young and elderly participants were trained in solving a computerized version of so-called water-jug problems. The goal of water-jug problems is to acquire a certain amount of water in a target jug by strategically pouring jugs of different sizes into each other. There was a handful of restrictions. First, there was a tap by means of which working jugs could be filled to the brim without overflowing. Second, if a jug was poured into a smaller one, pour-

ing stopped as soon as the latter was full; a residual remained in the former jug. The same held for a donating jug containing a residual of water that exceeded the spare capacity of a receiving jug. The residuals often were the basis for a solution. One half of both the young and the older participants were trained by solving four water-jug problems (i.e., the conventional condition). That is, they were confronted with the initial state of these problems and had to work their way to the goal states. The other half of the participants was confronted with a worked-out version of these same training problems. That is, they not only saw the initial state of the problems, but also the intermediate states that led to the goal state. The problem states were accompanied by explanatory text. After training, all participants were presented with a transfer test, containing both near- and far-transfer problems. The near-transfer problems were variations of the training problems. The far-transfer problems substantially deviated from the training problems and required additional reasoning. The results showed that the elderly experienced substantially less cognitive load in the worked-examples condition relative to the conventional condition. No performance differences were found between the training conditions. In other words, while the elderly in the worked-examples condition experienced less cognitive load, they achieved an equal performance level as the elderly participants in the conventional condition. It was concluded, therefore, that studying worked-examples is a more efficient means of training the elderly than solving conventional problems.

The efficiency of worked examples was further explored in Chapter 3. The starting point of this chapter was the notion that there are numerous ways of designing worked examples. One way is the use of multimedia, which enables an audiovisual presentation of the training material. An advantage of audiovisual presentation is that text can be removed from the visual material. Instead, textual material can be presented auditorily. In this format, the learner is not inclined to split his or her attention between text and the images. Thus, extraneous visual search is reduced to a minimum. Moreover, audiovisual material addresses both the visual and auditory components of working memory. Distribution of information over both components reduces the chances of overload in either component. The beneficial ratio between extraneous and germane cognitive load in multimedia-based worked examples should lead to lower levels of experienced cognitive load and advanced schema construction compared to both visual worked examples and conventional problems. Both young and older adults were trained with either conventional problems, purely visual worked examples, or multimedia-based worked examples. The results showed that multi-media-based worked examples are more efficient than conventional problems in that a lower level of training load leads to comparable transfer performance. However, no differences were found between either the conventional-problems condition and the visual worked-examples condition or

the visual worked-examples and multimedia-based worked-examples condition. Moreover, these effects were the same for the young and the elderly participants.

Chapter 5 describes an experiment that addresses some problems arising from the procedure of the experiment discussed in Chapter 3. First, in the previous experiment not only the number of modalities was changed (from one — only visual — to two: both visual and auditory), but also the pace of the worked examples. Where in Chapter 3 the visual worked examples were user-paced, the multimedia-based worked examples were system-paced. A second problem concerns the presentation format of the worked examples. Whereas in Chapter 3 the visual worked examples consisted of static images, the multimedia-based worked examples were presented in an animated format. A third problem arising the previous experiment concerned the cognitive-load measure. In the previous experiments, cognitive load was measured by means of a subjective rating scale, on which the participants were required to indicate their experienced level of cognitive load. Self ratings are known to be susceptible to all kinds of response biases, such as social desirability. This introduces a final problem, which concerns the rather low levels of cognitive load that were reported on these scales in either condition. If this phenomenon was not due to a response bias, the level of germane cognitive load was probably too low.

The experiment described in Chapter 5 addressed these hitches. First, the problem of pace control was resolved by making both the visual and bimodal worked examples system-paced. That is, the speed of presentation was the same in both conditions. Second, the problem of presentation format was resolved by presenting a series of non-dynamic images in both conditions. In the unimodal (i.e., visual) condition, explanatory text was imposed on the images. In the bimodal (i.e., audiovisual) condition, the same text was presented as narration. Third, the problem of self ratings was sidestepped by introducing an additional cognitive-load measure, which was a secondary-task paradigm. In this paradigm, the primary task (i.e., studying worked-out water-jug problems) is accompanied by a secondary task, which has a lower priority. The secondary task was the detection of a series of light signals, which were presented at random time intervals. As soon as the participants detected a signal, they had to push a response button. The idea behind this technique is that the response latencies of the secondary task give an indication of the load imposed by the primary task. That is, the more a participant is occupied by the primary task, the longer the response times. The problem of germane cognitive load, finally, was addressed by increasing the variability of the training problems. Increasing training variability promotes deeper processing. That is, if training problems are not clustered into groups of the same category, the learner is compelled to re-evaluate every training problem. This stimulates a more elaborate processing of every single problem and thus an in-

crease of cognitive load, which in its turn leads to better schema construction. Therefore, training variability was introduced as an extra factor. In the low-variability condition, eight training problems were clustered into four categories and presented as such. In the high-variability condition, training problems were presented in random sequences. The results revealed that bimodal training leads to lower levels of perceived cognitive load than unimodal training in both age groups. This effect was not reflected by the reaction times on the secondary task, however. Furthermore, a nearly-significant positive effect of training variability on performance was found for both age groups. However, there was no combined effect of training format and training variability in either age group. Again, the levels of experienced cognitive load were rather low. It was concluded that, although no effects were found that were exclusively advantageous for the elderly, the application of bimodal worked examples and the manipulation of training variability deserve further exploration, especially in less contrived problem domains.

Chapter 4 was a special chapter in that it explored the usefulness of pupillometry as an 'objective' measure of cognitive load. Since the 1950s it is known that the pupil dilates as a function of mental load. However, little research has been done on the cognitive pupillary response in elderly people. Therefore, the study described in this chapter explicitly addressed the question whether pupillometry is a useful measure of cognitive load in both young and elderly people. The level of cognitive load was varied by means of a memory-search task. In this task, participants were required to memorize sets of one to six digits. Subsequently, they were presented with a series of single-digit probes. For every probe, they had to decide as quickly as possible whether it was in the memory set or not. Thus, the critical events in this task were the presentation of the memory set, which initiated an encoding process, and the presentation of the probe, which initiated a search process. Three so-called task-evoked pupillary responses (TEPRs) were measured: mean dilation, peak dilation, and latency to the peak. Mean pupil dilation is the average dilation during a particular time window following a critical event. Peak dilation represents the largest pupil dilation during this period. Latency to the peak is the time that has elapsed between the onset of the critical event and the peak. These TEPRs typically intensify as a function of cognitive load. In this experiment, it was found that encoding larger memory sets leads to a larger mean pupil dilation in both age groups. The same effect was found for the search process, but only in the young participants. The absence of an effect in the elderly could have been due either to age-related cognitive resource limits or physiological constraints. Finally, no effects with respect to peak dilation and peak latency were found in either age group. It was concluded from this study that mean pupil dilation is a useful TEPR for measuring cognitive load in both young and elderly people.

In Chapter 6, finally, some general conclusions were drawn from the previous chapters. A first conclusion was that training formats based on cognitive load theory usually impose less cognitive load on the learner than conventional training formats. A second conclusion was that in spite of the difference in training load, these two approaches lead to a comparable level of performance. The favorable ratio between training load and performance makes CLT-based training formats the better choice in that they make a cost-effective use of the available mental resources. The central question to be answered was whether CLT-based training formats are particularly beneficial for elderly learners. It was concluded that the disproportional ‘gain’ in terms of experienced cognitive load found in Chapters 2 and 3 definitely make CLT-based formats a highly recommendable choice for this particular group. Moreover, the low levels of overall load suggest that germane cognitive load can still be raised in order to acquire performance effects. The study described in Chapter 5 suggested an effect in this direction. However, this finding needs further exploration.

Chapter 6 also provided a critical reflection on cognitive load theory and its associated research methods. The core issue of this chapter was that cognitive load theory incorporates specific claims concerning the role of cognitive load within an instructional context. That is, cognitive load is not simply considered as a by-product of the learning process, but as the major factor determining the success of an instructional intervention. It was argued that these theoretical claims require an empirical account of exactly how cognitive load relates to performance. It was asserted that this account has not yet been provided by CLT-research, including the research reported in this thesis. Suggestions and recommendations are given to consolidate the theoretical basis of cognitive load theory. An essential step in this direction is the application of two measures of cognitive load, which cancel out each other’s drawbacks. For example, one could combine the sensitivity of a subjective rating scale with the objectivity of pupillometry. Furthermore, a thorough method is needed to disentangle intrinsic from extraneous, and extraneous from germane cognitive load. Two possibilities are task and verbal protocol analysis. It was stated that only a combination of empirical (i.e., measurement) and analytical methods could reveal the true mechanism behind the success of cognitive load theory.

Finally, the foregoing was put in perspective by conceding that one could uphold a more pragmatic approach. That is, one could be satisfied with the result that a CLT-based training format imposes less cognitive load on the learner, leads to advanced performance, and is thus more efficient than a conventional format without attaching importance to the underlying cognitive mechanism. It is concluded that choosing the pragmatic approach should definitely entail the pursuit of making the elderly catch up with the young.

Samenvatting

Het is allang geen nieuws meer dat de westerse maatschappij steunt op een hoogstaande en zich snel ontwikkelende technologie. Het is ook niet nieuw dat de westerse bevolking in hoog tempo aan het vergrijzen is. Onderzoek heeft aangetoond dat deze ontwikkelingen zowel in de professionele als in de niet-professionele sfeer een groeiende behoefte aan bijscholing met zich meebrengen. Deze constatering vormt de drijfveer achter dit proefschrift. Het ontwerpen van cursusmateriaal voor ouderen — dat wil zeggen mensen vanaf zestig jaar — vereist namelijk speciale aandacht, met name als het gaat om het leren van complexe vaardigheden, zoals het omgaan met een computer. Richtlijnen voor het ontwerpen van instructiemateriaal gericht op het leren van complexe vaardigheden zijn reeds ontwikkeld en getest op jongvolwassenen en kinderen. De basis voor deze richtlijnen is gelegd door theorieën als de cognitieve belastingtheorie (CBT). Talloze studies hebben aangetoond dat instructies, die zijn gebaseerd op de cognitieve belastingtheorie, tot betere leerresultaten leiden dan conventionele instructies. Ouderen vormden echter nooit een doelgroep binnen dit soort studies. Om dit gemis teniet te doen, is geprobeerd de cognitieve belastingtheorie te combineren met kennis die voortvloeit uit cognitief verouderingsonderzoek.

De cognitieve belastingtheorie is gebaseerd op fundamentele kennis uit de cognitieve psychologie. De theorie wordt toegepast bij het ontwerpen van instructiemateriaal dat de cognitieve capaciteit van de lerende optimaal benut. Eén van de belangrijkste gegevens binnen de cognitieve psychologie is de beperking van de mentale verwerkingscapaciteit. Om precies te zijn: het menselijk werkgeheugen kan ongeveer zeven informatie-eenheden tegelijkertijd verwerken. Een ander belangrijk kenmerk van het werkgeheugen is dat het in verbinding staat met het lange-termijngeheugen. Hierin ligt informatie niet zozeer opgeslagen als een hoop losse elementen, maar veeleer als klwens van elementen die een sterke relatie met elkaar hebben. Een voordeel van deze klwens is dat ze door het werkgeheugen worden beschouwd als een geheel, dus als één enkel element. Ze besparen dus werkgeheugencapaciteit. Een laatste kenmerk van het werkgeheugen is dat het bestaat uit twee onafhankelijke opslagcomponenten — een visuele en een auditieve — die worden gecoördineerd door een centrale verwerkingseenheid. De centrale verwerkingseenheid is verantwoordelijk voor het bewust verwerken van informatie uit de twee opslagcomponenten en uit het lange-termijngeheugen.

Deze basale opvattingen over het werkgeheugen vormen het uitgangspunt van de cognitieve belastingtheorie. Binnen de cognitieve belastingtheorie wordt aangenomen dat een complexe vaardigheid bestaat uit elementen die sterk met elkaar zijn verbonden. Deze structuur zorgt voor een hoge zogenaamde “intrinsieke belasting.” Om een complexe vaardigheid te leren, moet deze als een geheel in het lange-termijn-geheugen worden opgeslagen. Daarvoor moet de informatie wel eerst langs het nauwe pad van het werkgeheugen. Volgens de cognitieve belastingtheorie kan dit op een efficiënte manier gebeuren door het werkgeheugen niet te belasten onnodige activiteiten. Dit soort activiteiten zorgt namelijk voor een zinloze belasting, waardoor de overdracht van informatie van het werkgeheugen naar het lange-termijn-geheugen kan worden belemmerd. In plaats daarvan zou het werkgeheugen zich louter moeten bezighouden met activiteiten die ervoor zorgen dat onderling afhankelijke informatie-eenheden worden samengevoegd in de vorm van clusters of schema’s. Informatie die is opgeslagen in de vorm van schema’s stelt iemand in staat om problemen van een bepaalde categorie op te lossen. De kwaliteit van een schema bepaald het scala aan problemen dat iemand kan oplossen. Hoe uitgebreider een schema, des te breder is de waaier aan problemen die een persoon kan oplossen. Een uitgebreid schema kan een persoon zelfs in staat stellen om problemen op te lossen die hij of zij nog nooit eerder is tegenkomen. Dit fenomeen wordt ook wel “transfer” genoemd. Daarnaast kunnen cognitieve schema’s worden geautomatiseerd. Het gebruik van geautomatiseerde schema’s vergt slechts weinig bewuste verwerking en zorgt dus voor een verdere verlichting van het werkgeheugen. De belasting die geassocieerd wordt met het construeren van cognitieve schema’s wordt ook wel “zinvolle belasting” genoemd.

Logischerwijs beweert de cognitieve belastingtheorie dat bij het leren van complex materiaal — materiaal dat een hoge intrinsieke belasting met zich meebrengt — de zinloze belasting zo klein mogelijk moet zijn en de zinvolle belasting zo hoog mogelijk. Daarbij moeten de capaciteitsgrenzen van de werkgeheugen uiteraard in acht worden genomen. Met andere woorden, de cognitieve belastingtheorie staat voor instructiemateriaal dat enerzijds voorkomt dat de lerende zich bezighoudt met onnodige mentale activiteit en anderzijds stimuleert dat mentale activiteit wordt ontplooid die bijdraagt aan de constructie van cognitieve schema’s van hoge kwaliteit.

In Hoofdstuk 1 werd dit idee in verband gebracht met cognitieve veroudering. In dit proefschrift verwijst cognitieve veroudering naar veranderingen van het werkgeheugen die zich voordoen bij ouderen. Daaromtrent worden in de literatuur drie verschijnselen beschreven. Ten eerste lijkt de verwerkingscapaciteit van het werkgeheugen kleiner te worden bij ouderen. Ten tweede is er sprake van een algehele achteruitgang van de cognitieve verwerkingsnelheid. En ten derde blijken ouderen moeite te hebben

met het onderdrukken van irrelevante informatie, wat ten koste gaat van het verwerken van relevante informatie. Deze veranderingen beperken het vermogen van ouderen om nieuwe vaardigheden te leren. Het lijkt erop dat het smalle pad van het werkgeheugen bij ouderen nog veel smaller is geworden. In het eerste hoofdstuk werd betoogd dat de cognitieve belastingtheorie strategieën oplevert om deze leerbeperkingen te omzeilen. Omdat de cognitieve belastingtheorie de beperkingen van het werkgeheugen probeert te compenseren en deze beperkingen vooral een rol spelen bij ouderen, spreekt het voor zich dat CBT-gebaseerde instructievormen des te meer nut hebben voor deze specifieke groep. Tenslotte heeft het verlagen van zinloze cognitieve belasting een proportioneel groter effect op een oud dan op een jong persoon. Dit idee werd met verschillende CBT-gebaseerde trainingsvormen getoetst in de Hoofdstukken 2, 3 en 5.

In Hoofdstuk 2 werd de cognitieve efficiëntie van het bestuderen van uitgewerkte voorbeelden getoetst tegen de efficiëntie van het oplossen conventionele oefenproblemen. Binnen de cognitieve belastingtheorie worden uitgewerkte voorbeelden beschouwd als de hoeksteen van efficiënt leren. In conventionele trainingsmethoden spelen uitgewerkte voorbeelden echter een ondergeschikte rol. Onder het motto “oefening baart kunst” ligt de nadruk doorgaans op het oplossen van lange reeksen oefenproblemen. Uiteraard baart oefening uiteindelijk kunst, maar dit gaat niet automatisch op in een vroeg leerstadium. In een vroeg leerstadium beschikt de lerende namelijk nog niet over de nodige cognitieve schema's. Als lerenden in dit stadium dus worden geconfronteerd met een nieuw domein, dan beschikken ze over weinig voorkennis om problemen binnen dit domein op te lossen. Het gevolg is dat ze ieder detail van een probleem bewust moeten verwerken zonder daarbij geholpen te worden door bestaande cognitieve schema's. In deze situatie is de lerende gedwongen om een zogenaamde doel-middel-strategie toe te passen. Een doel-middel-strategie bestaat uit een achterwaartse en een voorwaartse fase. In de achterwaartse fase wordt een soort piramide van doelen en subdoelen geconstrueerd, beginnend bij de top. De top wordt gerepresenteerd door het hoofddoel van het probleem en de onderliggende lagen door de subdoelen. Deze fase stopt als een laag wordt bereikt met vervulbare subdoelen. Vanaf dat punt begint de voorwaartse fase van de doel-middel-strategie. In deze fase worden eerst de subdoelen in de onderste lagen opgelost, vervolgens de subdoelen in de hogere lagen en tenslotte het hoofddoel aan de top van de piramide. De combinatie van een voorwaartse en een achterwaartse fase maken de doel-middel-analyse een cognitief belastend proces. Daarbij komt dat slechts de helft van dit proces — de voorwaartse fase — nuttig is. In de voorwaartse fase wordt een probleemtoestand geassocieerd met de middelen die ervoor zorgen dat nieuwe probleemtoestanden en uiteindelijk het einddoel wordt bereikt. Dit is precies wat iemand moet leren bij het zich eigen maken van een nieuwe vaardigheid. Door uitgewerkte voorbeelden

te bestuderen, wordt de lerende gestimuleerd om zijn of haar mentale capaciteit uitsluitend aan het voorwaartse proces te besteden. Hierdoor wordt de zinloze belasting dus verminderd en de zinvolle belasting geoptimaliseerd. Het bestuderen van uitgewerkte voorbeelden zou daarom dus tot betere resultaten moeten leiden in termen van cognitieve belasting, schemaconstructie en transfer dan het oplossen van reeksen conventionele problemen.

Om deze hypothese te toetsen werd een experiment uitgevoerd. In dit experiment werden jongere en oudere deelnemers getraind in het oplossen van een gecomputeriseerde versie van zogenaamde water-jug problemen. Bij een water-jug (waterbeker) probleem is het de bedoeling dat een bepaalde hoeveelheid water wordt verkregen door bekertjes van verschillende grootte en met verschillende hoeveelheden water strategisch in elkaar over te schenken. Daarbij moesten een paar regels in acht worden genomen. Zo konden bekertjes met behulp van een kraan alleen volledig worden gevuld. Als een bekertje vol water in een kleinere bekertje werd overgeschonken, dan stopte het overschenken zodra de kleinere bekertje vol was en bleef er een residu over in de grotere bekertje. Hetzelfde gold voor een bekertje met een residu water dat groter was dan de vrije ruimte van een ontvangende bekertje. De residuen vormden vaak de basis voor een oplossing. De helft van de proefpersonen werd getraind door ze vier problemen zelf te laten oplossen (de conventionele conditie). Dat wil zeggen ze kregen de beginfase van een probleem te zien en moesten vervolgens zelf tot het doel zien te komen. De andere helft van de proefpersonen werd getraind door ze een uitgewerkte versie van deze zelfde problemen te laten bestuderen. Dat wil zeggen ze kregen niet alleen de beginfase van een probleem te zien maar ook de stappen die tot de oplossing leidden. De afzonderlijke probleemtoestanden werden voorzien van een verhelderende tekst. Na de training werden alle deelnemers geconfronteerd met een transfertest. Deze test bevatte zowel 'nabije' als 'verre' transferproblemen. De nabije transferproblemen waren varianten van de trainingsproblemen. De verre transferproblemen weken aanzienlijk af van de trainingsproblemen en vereisten meer denkstappen. De resultaten lieten zien dat ouderen de uitgewerkte voorbeelden als aanzienlijk minder belastend ervoeren dan de conventionele problemen. Bij beide leeftijdsgroepen werden echter geen prestatieverschillen gevonden op de transfertest. Met andere woorden, ouderen ervoeren minder cognitieve belasting bij het bestuderen van uitgewerkte voorbeelden terwijl ze minstens even goed scoorden op de transfertest als ouderen in de conventionele conditie. Geconcludeerd werd dan ook dat het bestuderen van uitgewerkte voorbeelden een efficiëntere trainingsmethode is voor ouderen dan het oplossen van conventionele problemen.

De efficiëntie van uitgewerkte voorbeelden werd nader onderzocht in Hoofdstuk 3. Het uitgangspunt van dit hoofdstuk was dat er talloze manieren zijn om uitgewerkte voorbeelden te presenteren. Eén van die ma-

nieren is een audiovisuele presentatie met behulp van multimedia. Een voordeel van een audiovisuele presentatie is dat de begeleidende tekst uit het visuele materiaal kan worden weggehaald en in plaats daarvan verbaal kan worden gepresenteerd. Op deze manier wordt de lerende niet gedwongen zijn of haar aandacht te splitsen tussen een diagram enerzijds en de bijbehorende tekst anderzijds. Zinloos visueel zoeken wordt dus tot een minimum beperkt. Bovendien maakt een audiovisuele presentatie gebruik van zowel de visuele als de auditieve component van het werkgeheugen. Het verdelen van informatie over beide componenten voorkomt overbelasting van één van de componenten. De voordelige verhouding tussen zinloze en zinvolle belasting zou van audiovisuele voorbeelden een efficiëntere trainingsvorm moeten maken ten opzichte van puur visuele voorbeelden en conventionele problemen. Om dit te testen, werden zowel jongere als oudere deelnemers getraind met behulp van audiovisuele voorbeelden, puur visuele voorbeelden en conventionele problemen. Het probleemdoorniein betrof ook hier weer het water-jug probleem. Uit de resultaten bleek dat audiovisuele voorbeelden efficiënter zijn dan conventionele problemen, omdat een lagere cognitieve belasting leidde tot een vergelijkbaar prestatieniveau op de transfertest. Er werden echter geen verschillen gevonden tussen conventionele problemen en puur visuele voorbeelden en tussen puur visuele voorbeelden en audiovisuele voorbeelden. Bovendien waren deze resultaten hetzelfde voor de jonge en oudere proefpersonen.

In Hoofdstuk 5 werd inhaakt op enkele procedurele problemen die voortkwamen uit Hoofdstuk 3. Ten eerste werd in het vorige experiment niet alleen een auditieve presentatievorm toegevoegd, maar was er ook een verschil in presentatietempo tussen de condities. In Hoofdstuk 3 werd het tempo binnen de puur visuele conditie namelijk bepaald door de deelnemer zelf, terwijl het tempo in de audiovisuele conditie door de computer werd bepaald. Een tweede probleem heeft betrekking op de presentatievorm van de uitgewerkte voorbeelden. Terwijl in de puur visuele voorbeelden sprake was van bewegingloze afbeeldingen, werden de audiovisuele voorbeelden in geanimeerde vorm weergegeven. Een derde probleem betrof het instrument voor het meten van cognitieve belasting. In alle voorgaande experimenten werd gebruik gemaakt van een subjectieve 9-punts-schaal, waarop de deelnemers zelf hun ervaren mentale belasting moesten scoren. Van subjectieve maten is echter bekend dat ze gevoelig zijn voor storende effecten, zoals sociale wenselijkheid. Dit is mogelijk de oorzaak van de nogal lage inschatting van de ervaren mentale belasting, die in de Hoofdstukken 2 en 3 in alle condities naar voren kwam. Als de oorzaak van deze lage inschatting niet was toe te schrijven aan een storend effect als sociale wenselijkheid, dan was er waarschijnlijk sprake van een te lage zinvolle cognitieve belasting.

Om de bovenstaande problemen het hoofd te bieden, werd een nieuw experiment uitgevoerd. Ten eerste werd het presentatietempo van de visu-

ele en de audiovisuele conditie aan elkaar gelijkgesteld door het tempo van beide presentatievormen door de computer te laten bepalen. Ten tweede werd de presentatievorm gelijkgesteld door in beide condities gebruik te maken van statische afbeeldingen. In de visuele conditie werden de afbeeldingen vergezeld door verhelderende tekst. In de audiovisuele conditie werd deze tekst verbaal gepresenteerd. Ten derde werd een nieuwe maat voor cognitieve belasting toegevoegd, namelijk de tweede-taak-methode. Bij deze methode wordt een primaire taak (in dit geval het bestuderen van uitgewerkte voorbeelden) vergezeld van een tweede taak met een lagere prioriteit. Deze tweede taak bestond uit het detecteren van een serie lichtsignalen, die op willekeurige tijdstippen werden gepresenteerd. Zodra een deelnemer een lichtsignaal ontwaarde, moest hij of zij zo snel mogelijk op een knop drukken. Het idee hierachter is dat de reactietijd (de tijd tussen het detecteren van een signaal en het drukken op de knop) een indicatie geeft van de ervaren cognitieve belasting. Dat wil zeggen hoe meer iemand in beslag wordt genomen door de primaire taak, des te langer wordt de reactietijd in de tweede taak. Tenslotte werd het probleem van de lage zinnvolle belasting opgelost door de variabiliteit van de trainingsproblemen op te voeren. Een grotere variabiliteit bevordert namelijk de diepgang van het leerproces. Als trainingsproblemen niet worden gegroepeerd in categorieën (lage variabiliteit), maar in willekeurige volgorde worden aangeboden (hoge variabiliteit), dan is de lerende gedwongen om elk trainingsprobleem opnieuw te analyseren. De kans op routine, en daarmee oppervlakkig leren, kan zo worden geminimaliseerd. Op deze manier zouden er betere cognitieve schema's moeten worden geconstrueerd. Variabiliteit werd daarom als extra factor aan het experiment toegevoegd. In een conditie met lage variabiliteit werden acht trainingsproblemen gegroepeerd in vier categorieën en als zodanig gepresenteerd. In een conditie met hoge variabiliteit werden dezelfde trainingsproblemen in willekeurige volgorde gepresenteerd. Uit de resultaten bleek dat audiovisuele voorbeelden bij beide leeftijdsgroepen tot een lagere cognitieve belasting leiden dan visuele problemen. Dit effect was echter niet terug te zien in de reactietijden. Verder werd een bijna significant effect gevonden van variabiliteit op de prestatie in beide leeftijdsgroepen: in de conditie met een hoge variabiliteit was de prestatie gemiddeld beter. In geen van de leeftijdsgroepen werd echter een gecombineerd effect gevonden van presentatievorm (visueel of audiovisueel) en variabiliteit. Bovendien was te gerapporteerde cognitieve belasting wederom bijzonder laag. Ondanks het ontbreken van een effect dat exclusief in het voordeel van de oudere deelnemers was, werd geconcludeerd dat het toepassen van audiovisuele presentatievormen en het manipuleren van trainingsvariabiliteit zeker nader onderzoek verdienen. Vooral in een minder kunstmatig probleem domein zou dit tot interessante resultaten kunnen leiden.

Hoofdstuk 4 was een speciaal hoofdstuk, dat gericht was op de vraag of het meten van pupilverwijding kan worden gebruikt als objectieve maat voor cognitieve belasting. Reeds sinds de jaren vijftig is bekend dat de pupil verwijdt als functie van mentale belasting. Er is echter nog maar weinig onderzoek gedaan naar deze cognitieve pupilrespons bij ouderen. Met een experiment werd daarom geprobeerd een antwoord te geven op de vraag of pupillometrie een bruikbare maat is voor cognitieve belasting bij ouderen. Cognitieve belasting werd gemanipuleerd door middel van een zogenaamde “geheugen-zoek-taak.” Bij deze taak was het allereerst de bedoeling dat deelnemers reeksen van één tot met zes cijfers uit hun hoofd leerden. Vervolgens kregen de deelnemers een serie losse cijfers te zien, waarvan ze achtereenvolgens zo snel mogelijk moesten beslissen of deze in de onthouden reeks voorkwamen of niet. De kritische momenten in deze taak waren enerzijds de presentatie van de cijferreeks, hetgeen een opslagproces inluidde, en anderzijds de presentatie van de losse cijfers, dat tot een zoekproces leidde. Tijdens deze taak werden drie zogenaamde taakgeïnduceerde pupilresponsen gemeten: de gemiddelde verwijding, de maximale verwijding en de tijd tot de maximale verwijding. De gemiddelde verwijding werd vastgesteld over een bepaalde periode volgend op een kritisch moment. De maximale verwijding was de grootste pupilverwijding gedurende deze periode. De tijd tot de maximale verwijding was de tijd tussen een kritisch moment en het optreden van het maximum. Onderzoek heeft aangetoond dat deze pupilresponsen groter worden bij een verhoogde cognitieve belasting. De resultaten van het experiment lieten zien dat de gemiddelde pupilverwijding bij zowel jongeren als ouderen toeneemt bij een verhoogde geheugenbelasting tijdens de opslagfase. Hetzelfde effect werd gevonden voor de zoekfase van de taak, maar in dit geval alleen bij de jongere deelnemers. De afwezigheid van een effect bij de oudere deelnemers kan te maken hebben gehad met de beperkingen van het werkgeheugen bij deze groep of met fysiologische beperkingen van de pupil. Verder werden nog bij de jongeren nog bij de ouderen effecten gevonden met betrekking tot de maximale pupilverwijding en de tijd tot het maximum. Geconcludeerd werd dat voor zowel jongeren als ouderen gemiddelde pupilverwijding een goede maat is voor cognitieve belasting.

In Hoofdstuk 6 tenslotte werden een paar algemene conclusies getrokken. Een eerste conclusie was dat trainingsvormen die zijn gebaseerd op de cognitieve belastingtheorie over het algemeen een lagere cognitieve belasting met zich meebrengen dan conventionele trainingsvormen. Ten tweede werd geconcludeerd dat, ondanks de verschillen in cognitieve belasting, de twee trainingsvormen tot een vergelijkbaar prestatieniveau leiden. De gunstige verhouding tussen cognitieve belasting en prestatie maakt CBT-gebaseerde trainingsvormen de betere keus, omdat ze een efficiënt gebruik maken van de beschikbare mentale capaciteit. De hamvraag was echter of CBT-gebaseerde trainingsvormen vooral voor ouderen een

gunstig effect hebben. Geconcludeerd werd dat de grotere effecten in termen van ervaren cognitieve belasting bij ouderen (zie de Hoofdstukken 2 en 3) een sterk argument vormen voor het aanbevelen van CBT-gebaseerde trainingsvormen voor deze groep. Bovendien suggereren de lage cognitieve belastingniveaus dat de zinvolle belasting verder verhoogd kan worden om effecten in prestatie 'af te dwingen.' Het experiment van Hoofdstuk 5 lijkt hiervoor een aanwijzing te geven. Meer zekerheid hierover vereist echter nader onderzoek.

In Hoofdstuk 6 werd tevens een kritische blik geworpen op de cognitieve belastingtheorie en de bijbehorende onderzoeksmethodologie. Het belangrijkste discussiepunt betrof de specifieke uitspraken, die de cognitieve belastingtheorie doet over de rol van mentale belasting bij het trainen van complexe vaardigheden. Cognitieve belasting wordt namelijk niet beschouwd als een bijproduct van het leerproces, maar als de bepalende factor voor het succes van een trainingsvorm. Betoogd werd dat deze theoretische beweringen een empirisch beschrijving vereisen van de invloed van cognitieve belasting op de prestatie. Geconstateerd werd dat een dergelijke beschrijving nog niet is geleverd, hetgeen ook geldt voor dit proefschrift. Daarom werden suggesties en aanbevelingen gedaan om de theoretische basis van de cognitieve belastingtheorie empirisch te versterken. Als een belangrijke stap hierbij werd het toepassen van twee verschillende maten voor cognitieve belasting voorgesteld. Deze zouden elkaars nadelen kunnen opheffen en elkaars voordelen kunnen aanvullen. Men zou bijvoorbeeld de gevoeligheid van een subjectieve schaal kunnen combineren met de objectiviteit van pupillometrie. Daarnaast zou er een goede methode moeten worden gevonden om intrinsieke, zinloze en zinvolle cognitieve belasting van elkaar te kunnen onderscheiden. Twee mogelijke methoden daarvoor zijn taak- en verbale protocolanalyse. Op grond hiervan werd betoogd dat alleen een combinatie van empirische en analytische methodes het exacte mechanisme achter het succes van de cognitieve belastingtheorie kunnen achterhalen.

Men zou echter ook een pragmatisch standpunt kunnen innemen door tevreden te zijn met het resultaat dat een CBT-gebaseerde trainingsvorm tot een lagere mentale belasting, een betere prestatie en dus een grotere efficiëntie leidt dan een conventionele trainingsvorm. Het achterliggende cognitieve mechanisme zou op die manier naar de achtergrond geschoven worden. Geconcludeerd werd dat als voor deze pragmatische benadering gekozen wordt, men op zijn minst moet proberen om de verschillen in prestatie, cognitieve belasting en efficiëntie tussen jongeren en ouderen te slechten.

Curriculum Vitae

Pascal van Gerven was born on January 26, 1971, in Venlo, The Netherlands. He received secondary education (Atheneum) at the Rijksscholengemeenschap Den Hulster in his birthplace. After an academic digression, appraising chemistry and biology, he started studying psychology at the University of Nijmegen in September 1991. His graduation research concerned word production in bilinguals, on the basis of which he received his master's degree in experimental psychology in May 1997. Shortly after graduation, he began his PhD project at the Department of Psychology of Maastricht University. The results of this project are reported in the present thesis. Since September 2001 he works as a lecturer at the Institute of Psychology (Department of Social Sciences) of Erasmus University in Rotterdam.

Papers

- Paas, F., Tuovinen, J.E., Tabbers, H., & Van Gerven, P.W.M. (2002). Cognitive load measurement as a means to advance cognitive load theory. Manuscript submitted for publication.
- Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., Hendriks, M., & Schmidt, H.G. (2002). The efficiency of multimedia learning into old age. Manuscript submitted for publication.
- Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., & Schmidt, H.G. (2000). Cognitive load theory and the acquisition of complex cognitive skills in the elderly: Towards an integrative framework. *Educational Gerontology*, *26*, 503-521.
- Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., & Schmidt, H.G. (2002a). Cognitive load theory and aging: Effects of worked examples on training efficiency. *Learning and Instruction*, *12*, 87-105.
- Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., & Schmidt, H.G. (2002b). Memory load and task-evoked pupillary responses in aging. Manuscript submitted for publication.
- Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., & Schmidt, H.G. (2002c). Modality and variability as factors in training the elderly. Manuscript submitted for publication.

Presentations

- Van Gerven, P.W.M., Paas, F., & Van Merriënboer, J.J.G. (1998, April). Instructional efficiency: Relations between cognitive load, task complexity, and aging. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

- Van Gerven, P.W.M., Paas, F., & Van Merriënboer, J.J.G. (1999, July). Cognitive load theory and aging: Effects of worked examples. Presentation at the 6th European Congress of Psychology, Rome, Italy.
- Van Gerven, P.W.M., Paas, F., & Van Merriënboer, J.J.G. (1999, August). Implications of cognitive load theory for the elderly. Paper presented at the 8th European Conference for Learning and Instruction (EARLI), Göteborg, Sweden.
- Van Gerven, P.W.M., Paas, F., & Van Merriënboer, J.J.G. (2000, April). Worked examples versus conventional problems in the acquisition of complex skills in the elderly: A cognitive-load approach. Poster presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., Hendriks, M., & Schmidt, H.G. (2001, April). A multimedia strategy for efficient complex skill training into old age. Poster presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.
- Van Gerven, P.W.M., Paas, F., & Van Merriënboer, J.J.G., & Schmidt, H.G. (2000, May). Worked examples as an efficient means of complex skill training in the elderly. Paper presented at the Onderwijs Research Dagen (ORD), Leiden, The Netherlands.