



Pergamon

Learning and Instruction 12 (2002) 87–105

Learning and
Instruction

www.elsevier.com/locate/learninstruc

Cognitive load theory and aging: effects of worked examples on training efficiency

P.W.M. Van Gerven ^{a,*}, F.G.W.C. Paas ^b,
J.J.G. Van Merriënboer ^b, H.G. Schmidt ^c

^a Department of Psychology, Maastricht University, P.O. Box 616, 6200 MD Maastricht,
The Netherlands

^b Educational Technology Expertise Center (OTEC), Open University of The Netherlands, Heerlen,
The Netherlands

^c Department of Psychology, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam,
The Netherlands

Abstract

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 article 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. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Cognitive load; Cognitive aging; Complex skills; Instructional design

* Corresponding author.

E-mail address: p.vangerven@psychology.unimaas.nl (P.W.M. Van Gerven).

1. Introduction

Cognitive aging entails both growth and decline. Growth concerns the 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 the working-memory capacity (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 (Allen, Madden, Groth, & Crozier, 1992; Hasher & Zacks, 1988), which is likely to withhold older 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, depending 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 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 (CLT) offers important tools for this purpose. 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 working-memory capacity as possible to relevant operations and is not forced to waste resources on operations that are not relevant or even detrimental to the learning process.

2. 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 (Just & Carpenter, 1992), it can be argued that working memory should be merely occupied by task-relevant oper-

ations, especially when dealing with complex material. CLT 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* as possible. 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 here is 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.

3. 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 (Czaja & Sharit, 1993; Gilinsky & Judd, 1994; Lorschach & Simpson, 1988). In other words, the cognitive declines associated with aging are particularly apparent 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 poor 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).

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 is made available for

learning. Second, the 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.

4. 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, wherein the goal state of the problem can be attained only 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 (Paas, 1992; Paas & Van Merriënboer, 1994b; Ward & Sweller, 1990) as well as with children (Pillay, 1994).

The central 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 caused by 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 efficient 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 (1993) developed a mathematical procedure involving the standardized scores of mental effort and performance (Section 6.2). 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.

5. Method

5.1. Participants

A first group of participants consisted of 30 psychology students (18 females, 12 males) of Maastricht University, ranging in age from 18 to 30 years (median 19.50 years). A second group consisted of 24 elderly participants (12 females, 12 males) ranging in age from 61 to 76 years (median 66.00 years). The elderly participants were selected from a subject 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 \$5) payment per hour and a refund of their travel expenses.

5.2. Materials

5.2.1. 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).

5.2.2. 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', etc. 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.

5.2.3. 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 Section 5.2.5 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 *averagely 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 ($\alpha=.82$).

5.2.4. 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 common sense rather than to their abstract abilities. To prevent a bottom 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. Luch-

ins' (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 & Polson, 1976; Atwood, Masson, & Polson, 1980). The task has mainly been used for studying the effect of *Einstellung* (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. Perceiving jugs 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 Fig. 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 content of a jug is poured into a smaller jug, a residual remains in the larger jug as soon as the smaller one 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

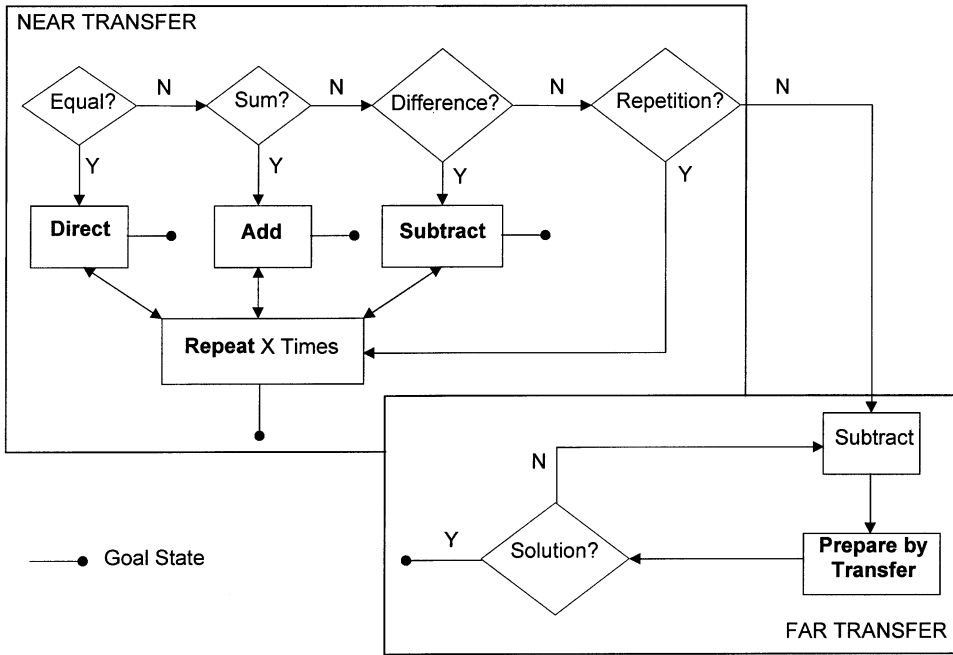


Fig. 1. A general processing model for the water-jug problem. Note that this is not a cognitive model, but merely a clarification of the difference between near- and far- transfer problems.

with a capacity of 7 units (like in Fig. 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 Fig. 1).

5.2.5. Apparatus and interface

The experiment was run on an IBM-compatible computer. The software controlling the experiment was 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 Fig. 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

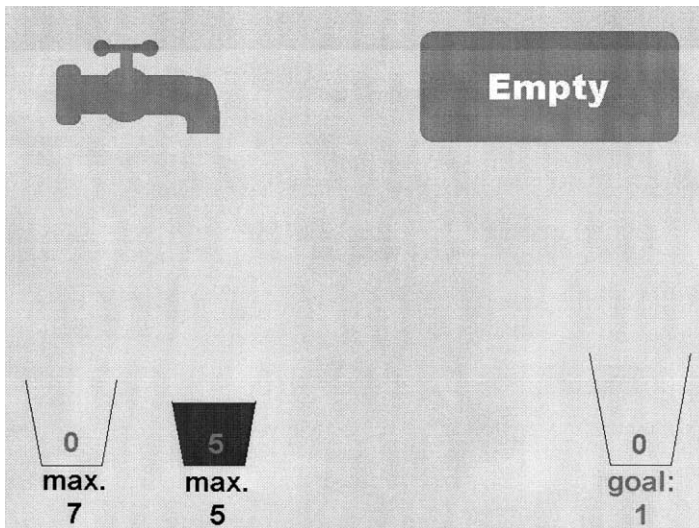


Fig. 2. The interface of the computerized version of Luchins' (1942) water-jug problem. Depicted is the initial phase of the solution of a far-transfer problem. The second working jug has just been filled. (Labels were originally in the Dutch language.)

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.

5.2.6. 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 Fig. 3). That is, for each solution step to appear, they had to push a 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.

5.2.7. 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 12–20; see Fig. 1), or they had to obtain a certain distribution of water over the jugs (problems 20–23; after Atwood & Polson, 1976).







<p>1. Goal: 1 liter. Solution: If the first working jug is filled and poured into the second working jug twice, 1 liter remains in the first one.</p> 	<p>2. Thus, fill the first working jug and...</p> 
<p>3. ... pour its content into the second one.</p> 	<p>4. Refill the first working jug...</p> 
<p>5. ... and again pour its content into the second one. 1 Liter remains in the first working jug.</p> 	<p>6. Pour this residual into the target jug and the problem is solved!</p> 

Fig. 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. (Texts and labels were originally in the Dutch language.)

5.3. 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 Examples 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)×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, she/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, she/he was informed about the remaining time. After the maximum time had passed, the participant was asked to stop with the test.

6. 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)×2 (Training Format)×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 1. Outcomes of the analyses are summarized in Table 2.

Table 1
Means and standard deviations of the dependent variables

		Conventional problems		Worked examples	
		M	SD	M	SD
Young (<i>n</i> =30)	Training efficiency	0.86	0.46	0.55	0.61
	Near-transfer training efficiency	0.74	0.35	0.32	0.99
	Far-transfer training 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 training efficiency	-1.18	1.40	-0.15	1.13
	Far-transfer training 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

Table 2

F values, significance levels, degrees of freedom, and mean square errors (note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. AG=age group; TF=training format; TD=transfer distance)

	AG	TF	AG×TF	AG×TF×TD	df	MSE
Training efficiency	36.27***	1.35	5.45*	0.10 ^{a,b}	1,50	0.92
Near-transfer training efficiency	18.77***	1.20	6.91*	–	1,50	1.02
Far-transfer training 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

^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.

6.1. 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 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 ($M=119.47$, $SD=8.70$), $F(1, 50)=13.66$, $MSE=82.60$, $p=0.001$.

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, $F(1, 50)=28.40$, $MSE=1.45$, $p<0.001$. 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.

6.2. 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, $F(1, 50)=6.94$, $MSE=2.23$, $p<0.05$, 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, $F(1, 50)=1.66$, $p>0.05$. Not surprisingly, there is a huge performance difference between the age groups. This holds for both Near-Transfer Performance, $F(1, 50)=10.98$, $MSE=.00$, $p<0.01$, and Far-Transfer Performance, $F(1, 49)=44.09$, $MSE=.05$, $p<0.001$. However, there is neither an effect of Training Format nor an interaction (see Table 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 Fig. 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.

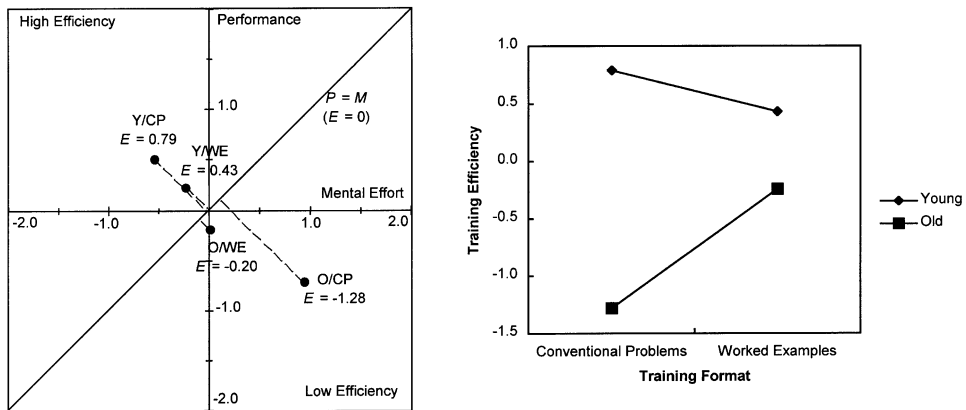


Fig. 4. Mean overall Training Efficiency per experimental group, based on Subjective Cognitive Load in the training phase and Performance in the test phase. The left diagram is a representation after Paas and Van Merriënboer (1993), where Y=Young, O=Old, CP=Conventional Problems, and WE=Worked Examples. The right diagram clearly demonstrates the obtained interaction.

As can be seen in Fig. 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, $F(1, 50)=36.27$, $MSE=0.92$, $p<0.001$, as well as an interaction between Age Group and Training Format, $F(1, 50)=5.45$, $p<0.05$, which is in accordance with Hypothesis 1. However, there is no interaction between Age Group, Training Format, and Transfer Distance, $F(1, 49)=0.10$, $MSE=0.19$, $p>0.05$, 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: $F(1, 50)=6.91$, $MSE=1.02$, $p<0.05$, against $F(1, 49)=8.20$, $MSE=0.81$, $p<0.01$, respectively.

6.3. Training time

Fig. 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, $F(1, 50)=65.73$, $MSE=2408.00$, $p<0.001$, and Training Format $F(1, 50)=28.72$, $p<0.001$. There is also a significant interaction between Age Group and Training Format, $F(1, 50)=15.73$, $p<0.001$. The interaction pattern, depicted in Fig. 5, suggests that the elderly participants are disproportionately favored when studying worked examples, which supports Hypothesis 1.

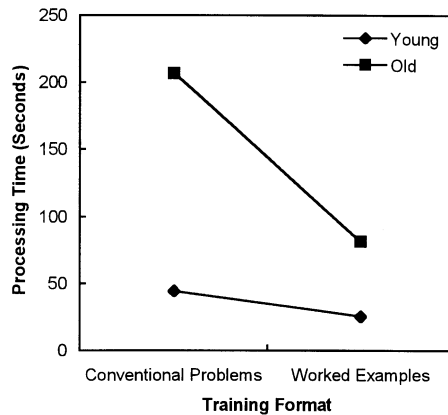


Fig. 5. Mean Training Time per conventional problem and worked example.

7. Discussion

The study reported in this article 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 poorer, and showed higher levels of cognitive load and training time than the young (see Tables 1 and 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; Lorschach & 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, 1994; 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 (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 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 that high that no differential effects were obtained.

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, such as described by Mayer and Moreno (2002), offers great opportunities (see also Mayer & Moreno, 1998; Tindall-Ford, Chandler, & Sweller, 1997). Such an approach would further reduce the amount of extraneous cognitive load caused by split attention or visual search 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.

Acknowledgements

The authors wish to thank Jelle Jolles, Peter Houx, and Astrid Quist of the Maastricht Aging Study (MAAS) for permitting us to use their subject pool. Furthermore, the Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged for funding this project.

References

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

- Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107–129.
- 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.
- 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.
- 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.
- 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.
- 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 and Zeitlinger.
- 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., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, 12, 107–119.
- 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.
- Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. *Journal of Educational Psychology*, 84, 429–434.
- Paas, F. G. W. C., & 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. G. W. C., & 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. G. W. C., & 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. G. W. C., Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, 79, 419–430.
- Pillay, H. (1994). Cognitive load and mental rotation: structuring orthographic projection for learning and problem solving. *Instructional Science*, 22, 91–113.
- 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.
- 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., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, 27, 763–776.

- Snijders, J. Th., & Verhage, F. (1962). *Voorlopige handleiding bij de Groninger intelligentie test* [Provisional manual to the Groningen intelligence test]. Amsterdam: Swets and Zeitlinger.
- Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science*, *12*, 257–285.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*, 295–312.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296.
- 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. G. W. C., 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 Merriënboer, J. J. G., & Paas, F. G. W. C. (1990). Automation and schema acquisition in learning elementary computer programming: implications for the design of practice. *Computers in Human Behavior*, *6*, 273–289.
- Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*, 1–39.
- 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.