Journal of Educational Psychology
1994, Vol. 86, No. 1, 122–133

0022-0663/94/$3.00

Variability of Worked Examples and Transfer of Geometrical Problem-Solving Skills: A Cognitive-Load Approach

Fred G. W. C. Paas and Jeroen J. G. Van Merriënboer

Four computer-based training strategies for geometrical problem solving in the domain of computer numerically controlled machinery programming were studied with regard to their effects on training performance, transfer performance, and cognitive load. A low- and a high-variability conventional condition, in which conventional practice problems had to be solved (followed by worked examples), were compared with a low- and a high-variability worked condition, in which worked examples had to be studied. Results showed that students who studied worked examples gained most from high-variability examples, invested less time and mental effort in practice, and attained better and less effort-demanding transfer performance than students who first attempted to solve conventional problems and then studied work examples.

In complex cognitive domains such as mathematics, physics, or computer programming, problem solutions can often be characterized by a hierarchical goal structure. The goal of these solutions can be attained only by successfully attaining all subgoals. Learning and performance of complex cognitive tasks are typically constrained by limited processing capacity. The more complex a task, that is, the more subgoals (that can be performed in alternative ways) it contains, the higher the processing demands are, and the more likely it is to exceed the concurrent processing and response capabilities of novices. Failure to learn these tasks may be attributed to the inadequate allocation of attention and the related high or excessive "cognitive load" (Sweller, 1988).

Cognitive load can be considered to be a multidimensional construct that represents the load that performing a particular task imposes on the cognitive system of a particular learner (Paas & Van Merriënboer, in press-b). The construct can be conceived to consist of causal factors and assessment factors corresponding to factors that affect cognitive load and factors that are affected by cognitive load, respectively. The causal factors include the task environment characteristics, subject characteristics, and the interactions between task environment and subject characteristics. Task characteristics include such factors as task structure, task novelty, type of reward system, and time pressure. Task environment demands concern such factors as noise and temperature. Subject characteristics pertain to relatively stable factors, that is, factors that are not likely to experience sudden changes as a result of the task (environment), such as subjects' cognitive capabilities, cognitive style, and prior knowledge. Finally, the subject–task (environment) interactions can affect cognitive load through relatively unstable factors such as internal criteria of optimal performance, motivation, or state of arousal.

With regard to the assessment factors, cognitive load can be conceptualized with respect to the dimensions of mental load, mental effort, and performance. *Mental load* is imposed by the task or environmental demands. This task-centered dimension, which is considered to be independent of subject characteristics, is constant for a given task. The human-centered dimension, *mental effort*, refers to the amount of capacity or resources that is actually allocated to accommodate the task demands. Mental effort is usually said to reflect the amount of controlled processing in which the individual is engaged (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). The amount of invested mental effort comprises all three causal factors (task environment characteristics, subject characteristics, and interactions between both). Finally, the level of *performance* achieved is an indication of the third measurement dimension; it also reflects all three causal factors. Figure 1 shows a schematic representation of the cognitive load construct and its causal factors (Figure 1, left side) and assessment factors (Figure, right side).

The intensity of effort being expended by students is often considered to constitute the essence of cognitive load (Hamilton, 1979; Paas, 1992; Sanders, 1979). Therefore, mental effort can be used as an index of cognitive load. Mental effort can be measured with subjective techniques, which use rating scales, and with objective techniques, which use physiological parameters. It is said that together with performance measures, mental effort can provide information on (a) the cognitive costs at which the performance is attained and (b) the relative efficiency of instructional conditions (see Paas & Van Merriënboer, in press-a).

Fred. G. W. C. Paas and Jeroen J. G. Van Merriënboer, Division of Instructional Technology, University of Twente, Enschede, The Netherlands.

We wish to thank the r.k. (rooms katholieke) Middelbaar Technische School in Enschede, The Netherlands, in particular W. Hommel, A. Klein Goldewijk, and F. Konst for granting permission to carry out and for cooperating with the research in the school. We also thank Sanne Dijkstra, Ton de Jong, and two anonymous reviewers for valuable comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Fred G. W. C. Paas, who is now at the Department of Movement Sciences, University of Limburg, P. O. Box 616, 6200 MD Maastricht, The Netherlands. Electronic mail may be sent to fred.paas@bw.rulimburg.nl.

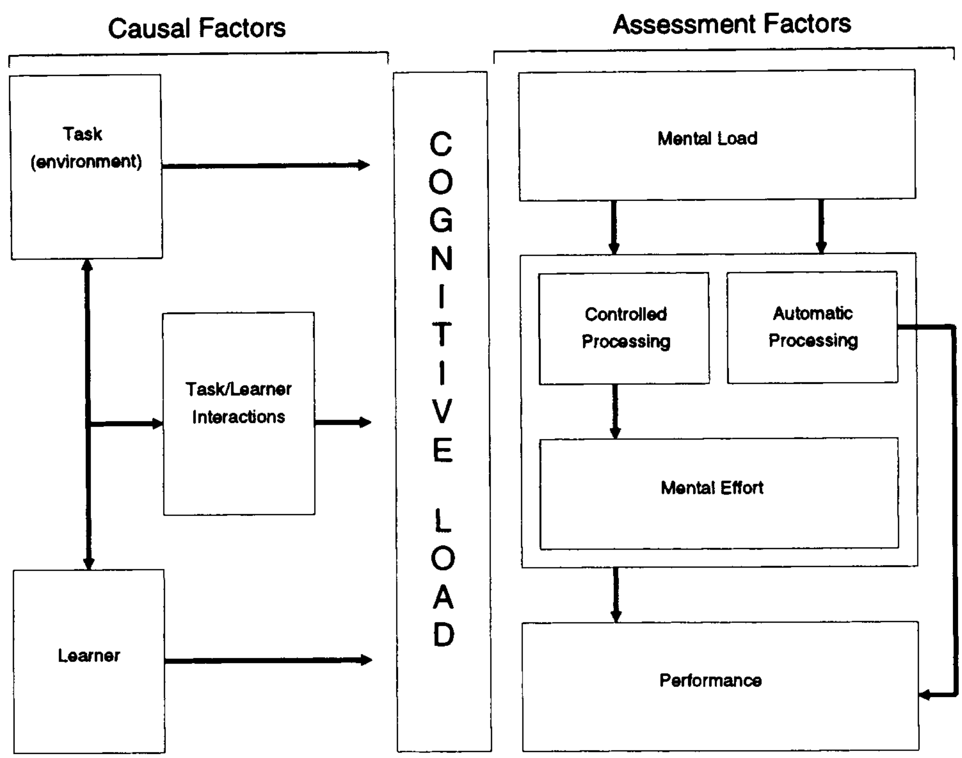

Figure 1. Schematic representation of the cognitive load construct.

High cognitive load seems to be especially imposed in traditional instructional strategies in which conventional problems are emphasized. To solve such problems, the problem specifications of which consist of the description of an initial problem state and a goal state, novices typically use means–ends analyses. Among others, Owen and Sweller (1985) and Sweller (1988) have demonstrated that this weak problem-solving method consumes a large amount of the learner's limited cognitive capacity, partly for processes that are not directly relevant for learning; these processes are associated with so-called *extraneous* cognitive load. Consequently, traditional instructional strategies may interfere with learning.

Recent studies, which provide guidelines for the training of complex cognitive tasks, are in agreement on the belief that to perform the task fluently, one must consider the cognitive processes of schema acquisition and rule automation. Moreover, these processes are usually seen as essential to reach transfer of problem-solving skills, that is, the ability to apply acquired knowledge and skills to problems that differ from the problems in which one is trained (Jelsma, Van Merriënboer, & Bijlstra, 1990; Sweller, 1989; Sweller, Chandler, Tierney, & Cooper, 1990; Van Merriënboer, Jelsma, & Paas, 1992; Van Merriënboer & Paas, 1990). Cognitive schemata can be conceptualized as cognitive structures that enable problem solvers to recognize problems as belonging to particular categories requiring particular operations to reach a solution. Schemata can provide analogies in new problem-solving situations and can be used by mapping processes to reach solutions for unfamiliar aspects of the problem-solving task.

An *automated rule* is a task-specific procedure that can be used without conscious processing and that can directly control problem-solving behavior. Automated rules can provide identical elements that may help to solve new problems, and they can free processing resources that may be devoted to controlled processes. Thus, in solving a transfer problem, familiar aspects can be performed by automated rules, and new aspects can be solved by the use of schemata.

In view of the limited processing capacity of humans, instructional methods that are directed to attaining transfer should prevent novices from using time and cognitive-capacity demanding problem-solving methods by redirecting their attention to aspects of the task that facilitate schema acquisition (e.g., stereotyped solutions for particular [sub-]goals and relations between task components). Type and variability of practice are important determinants of schema acquisition and transfer. Practice types that have proved to be successful in the training for transfer of complex cognitive skills are training with (a) worked examples with a study assignment (Cooper & Sweller, 1987; Paas, 1992), (b) partly worked examples with a completion assignment (Van Merriënboer, 1990a, 1990b; Van Merriënboer & De Croock, 1992), and (c) expertlike problem analyses with a performance assignment (Dufresne, Gerace, Hardiman, & Mestre, 1992). Especially, the reported time- and capacity-saving characteristics of training with (emphasis on) worked examples provide the possibility to facilitate schema acquisition by adding quality (e.g., increasing variability) to training, to facilitate rule automation by adding quantity (e.g., increasing the number of examples) to training, or both.

Ranzijn (1991) and Shapiro and Schmidt (1982) pointed out that increased variability of practice along the task dimensions is beneficial to schema acquisition and hence to transfer of acquired skills because it increases the chances that similar features can be identified and that relevant features can be distinguished from irrelevant ones. Thus, the confrontation with a wide range of different problems and solutions of these problems is important to give inductive processes the opportunity to extend or restrict the range of applicability. However, because practice-problem variability is positively related to cognitive load, for complex cognitive domains in which novices' cognitive load is often high, increased variability may also be expected to hinder learning. From this cognitive-load viewpoint, increased variability is expected to make conventional practice even less effective in the acquisition of appropriate cognitive schemata. Instead, worked-examples practice may be expected to provide the possibility to add variability to training and to enhance schema acquisition and transfer.

For geometrical problems in the domain of computer numerically controlled (CNC) machinery programming, we compared the effects of four computer-based training strategies on transfer. A low- and a high-variability conventional condition, in which conventional practice problems had to be solved, were compared with a low- and a high-variability worked condition, in which worked examples had to be studied. In contrast to previous studies (Cooper & Sweller, 1987; Paas, 1992; Sweller et al., 1990) on worked examples, in which the worked condition also contained a number of conventional problems, in this study a pure worked condition, in which students only had to study worked examples, was used. Consistent with the format of a previous study (Paas, 1992), the subjects in the conventional conditions of the present study saw the worked examples after they had attempted to solve the problems. Thus, attempting to solve the problems was an additional step for them. Despite the opportunity to study the worked examples, in the previous study it was found that the conventional condition yielded lower transfer performance than did the worked-examples condition. A possible explanation for that finding is that inferior transfer performance in the conventional condition could have been caused by the incorporation of the failed solutions in the schemata.

The main hypothesis of this study was that students in the worked conditions would show more effective and efficient transfer performance, that is, higher performance reached with less time and less mental effort for training, than would students in the conventional conditions. In addition, it was expected that training with worked examples would cost less time and less mental effort than training with conventional problems. With regard to practice-problem variability, we hypothesized that subjects in the worked condition would gain more from high practice variability than would subjects in the conventional condition.

Furthermore, data were collected on the number of incorrect solutions generated during instruction. Whereas no hypotheses were formulated with regard to the incorrect solutions variable, it can be helpful for interpreting possible differences in transfer performance between the conventional and worked conditions.

Method

Subjects

The participants were 60 students (58 men and 2 women), 19–23 years of age. They were recruited from four 4th-year classes of a secondary technical school (Middelbaar Technische Scholen[1]) in The Netherlands. They participated as a part of a CNC-programming course. Best performance on the transfer test was rewarded with 100 Dutch guilders (approximately $50).

Materials

The materials were related to geometrical problems that might emerge in the programming of CNC machinery, such as drilling, turning, and milling machines. The following topics were treated successively: the $x - y$ coordinate system, basic CNC-programming code, Pythagorean theory, right-angled triangles, and the trigonometric functions sine, cosine, and tangent. After a general instruction, the students should be able to encode and translate graphical or textual information into CNC-programming code and vice versa, to use geometrical theories and principles to construct right-angled triangles in a two-dimensional space, and to recognize and apply trigonometric ratios and Pythagorean principles to these triangles.

Four computer-based training programs were developed on an Olivetti M240 computer using the programming languages AUTOMATOR mi (Direct Technology Limited, 1987) and UNICOMAL (Unicomal A/S, 1987). The experiment was performed by using an Olivetti M240 computer.

General instruction. In the general instruction, the CNC programming-related geometrical theory was explained and illustrated with four worked examples.

Specific instruction. In this phase, a set of six problems was presented. In the worked conditions, all problems were presented with their solutions; in the conventional conditions, only the problems were presented. The first, third, and fifth problems required the students to calculate the length of a line, to calculate the distance between two points, and to transform absolute coordinates to relative coordinates, respectively. These problems were identical for the low- and high-variability conditions. The other problems (second, fourth, and sixth) differed from these problems as a function of practice-problem variability, such that in the low-variability conditions these problems had only different values, and in the high-variability conditions both values and problem formats (i.e., problem goal and localization in the cross of axes) were different. Figure 2A shows the identical first problems from the low- and high-variability conditions in a conventional and in a worked format, and Figure 2B shows the second problem from the low- and high-variability conventional condition.

In the conventional conditions, the initial problem states and the accompanying goal states were presented. After the subjects had

[1] Middelbaar Technische Scholen are Dutch schools that prepare students for applied technical occupations.

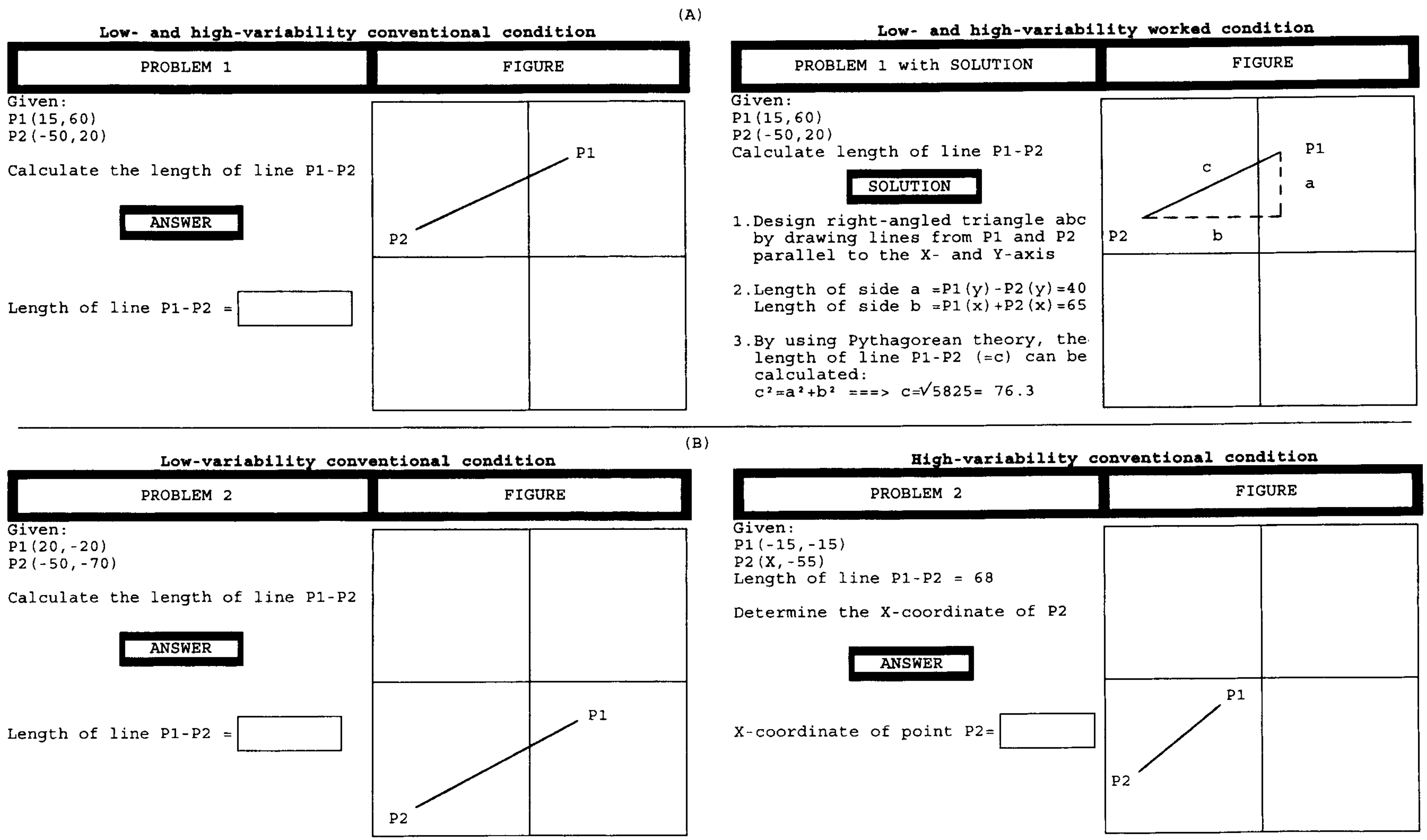

Figure 2. (A) First practice problem of the low- and high-variability condition in a conventional and worked format. (B) The second practice problem of the low- and high-variability conventional condition. (These problems are translated from the Dutch language.)

attempted to solve the problems, they could study the solution. In the worked conditions the subjects could immediately study the problems with their solutions.

Cognitive-load measurement. In the present study a rating-scale technique, in which numerical translations of the perceived amount of mental effort had to be given on a rating scale, and a spectral-analysis technique of the heart-rate variability, in which the energy in the heart-rate variability power-spectrum bands was calculated, were used to estimate the intensity of mental effort.

With the rating-scale technique, subjects reported their perceived amount of invested mental effort on a 9-point symmetrical category scale by translating their perception into a numerical value. The rating scale was a modified version of Bratfisch, Borg, and Dornic's (1972) scale for measuring perceived task difficulty. The numerical values and labels assigned to the categories ranged from *very, very low mental effort* (1) to *very, very high mental effort* (9). The scale was provided to, explained to, and illustrated for the students just before the beginning of the experiment and again during the general instruction.

The rating-scale technique is based on the assumptions that subjects are able to introspect on their cognitive processes and can report the amount of mental effort expenditure. Among others, Gopher and Braune (1984) found that subjects can introspect on their cognitive processes and have no difficulty in assigning numerical values to the imposed mental load (i.e., the invested mental effort). Furthermore, these subjective measures are easy to obtain, nonintrusive, easy to analyze, and have very high face validity (O'Donnel & Eggemeier, 1986).

In comparing the effects of three training strategies on transfer, Paas (1992) obtained a coefficient of internal consistency (Cronbach's coefficient alpha) of .90 with the modified scale. In the present study the internal consistency of the scale was again estimated with Cronbach's coefficient alpha. For the test problems, a coefficient of .82 was obtained.

The spectral analysis of heart-rate variability offers a physiological measure for the intensity of mental effort. Spectral analysis is a method for investigating whether a signal contains periodic components. The spectral-analysis technique of heart-rate variability is based on the assumption that changes in cognitive functioning are reflected in physiological functioning. Among others, Aasman, Mulder, and Mulder (1987), G. Mulder (1980), and L. J. M. Mulder (1988) have validated this technique with several cognitive tasks (e.g., multidimensional classification and sentence comprehension). In the present study spectral analyses on heart-rate variability were performed with the cardiovascular spectral analysis (CARSPAN) program (Van der Meulen & Mulder, 1990). The algorithm used for spectral computations is called the *sparse discrete Fourier transform*. For the spectral computation of the heart-rate variability, a relative variability measure, the squared *modulation index* (*MI*), was used. As a result of the use of the MI (relative variability), the calculated power values are dimensionless. The unit $(micro\text{-}MI)^2$ is used for the spectral-energy values (see L. J. M. Mulder, 1992).

Transfer test. The transfer test consisted of the same six problems for all conditions, which had to be solved by the students. The transfer problems were designed by the experimenter in cooperation with two teachers in CNC programming. The transfer problems were designed in such a way that they differed from the examples in both the high- and the low-variability conditions. The appropriate combination of strategies for solving the transfer problems was less obvious than with the training problems. The transfer problems typically had a hierarchical goal structure; that is, the solutions of the problems could only be accomplished by setting subgoals (more than during practice), which could be executed in alternative orders. Figure 3 shows an example of a transfer problem.

Design and Procedure

Subjects were randomly assigned to one of the four instructional conditions so that each condition contained 15 subjects. Each subject was tested individually. Before the experiment started, the students received an explanation sheet, which provided general procedural information and scrap paper, and a pocket calculator, and they were told about the procedure of the experiment.

After this verbal and written instruction the student was wired to an electrocardiogram (ECG) R-wave toptrigger (ECT4), for heart-rate data acquisition. The ECG signal was delivered by three disposable, pregelled, silver–silver chloride ECG electrodes (Medtronic). Two electrodes were placed at the level near the second intercostal space along the midclavicular positions, and the other electrode was placed just above the left processus iliaca anterior superior. An event-data multiplexer personal computer

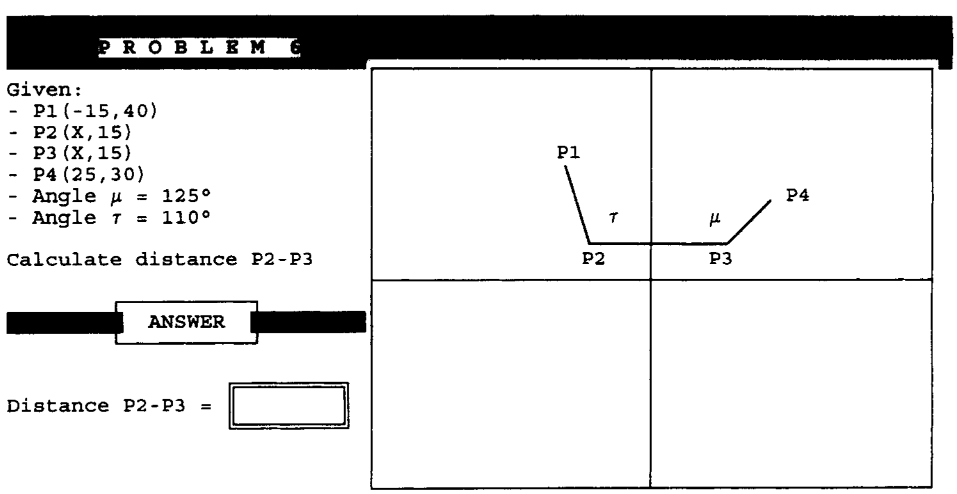

Figure 3. Transfer problem. (The problems in the experiment were presented in the Dutch language.)

(EDM-PC) for stimulus–response collection was connected to a computer that collected the heart-rate data and to the computer on which the training programs were presented (both were Olivetti M240s). The time of an R-top event and the synchronizing signal, which was generated by the presentation computer each time the student started or finished predefined parts in the program, were collected and transmitted by the EDM-PC to the data-collection computer. The ECG was recorded from the subjects during the whole experiment, that is, from 4 min before the general instruction to 4 min after the transfer test.

After connecting the student to the measurement device, the computer-based training program started with a 4-min rest phase in which a baseline measurement of heart rate was determined. To obtain a reliable baseline, we ensured that these rest phases appeared at regular intervals: after the general instruction, after the specific instruction, and after the transfer test. During the experiment, subject characteristics, mental-effort ratings, problem solutions, and study and problem-solving times were automatically registered by the computer. Students were allowed to use a pocket calculator. Figure 4 shows the consecutive stages of the experiment.

In the general instruction stage students had the opportunity to study procedural information, basic CNC-programming theory, and geometrical theory related to CNC-programming, between a program-controlled minimum and maximum time. One worked example accompanied each of the following theoretical parts: representation of basic CNC-programming code in the $x - y$ coordinate system, Pythagorean theory, right-angled triangles, and trigonometric functions.

In the specific instruction stage, after each of the six problems the students had to rate the perceived amount of mental effort that they invested in the problem. Students in the conventional conditions were allowed to work on each problem for a minimum of 0.5 min and a maximum of 10 min. On every problem the student had two attempts. After typing in one wrong answer, the student was told by the computer that the answer was not correct and that he or she had one other attempt to solve the problem. The program automatically continued when the students had typed in the right answer, after two wrong answers, or when the maximum time had passed. Then, students were presented with the worked example, which they could study for a maximum of 5 min. The worked examples presented in the conventional conditions were identical to those in the worked conditions. Students in the worked conditions had 6 min to study a worked example.

Finally, the transfer test, which also consisted of six problems, was administered. Students were allowed to work 8 min on each problem. Students could proceed with the next problem at any time, irrespective of the quality of the solution. However, they were told that no feedback would be provided and that it was not permissible to return to a previous problem. As was the case during instruction, ratings of the perceived amount of mental effort, heart rate, problem solutions, and solution times were automatically registered by the computer.

Results

The data were analyzed with 2 (type of practice: conventional vs. worked) × 2 (variability of practice: low vs. high) analyses of variance (ANOVAs). The results for the training and test phases are described separately. For the training phase the analyses were conducted on the following dependent variables: (a) time on general instruction, (b) time on specific instruction divided into time on solving problems and time on studying problem solutions, (c) the perceived amount of mental effort invested during specific instruction, and (d) the spectral energy of the heart-rate variability during specific instruction. Additional tests were performed on the mean total time spent on problem solving during conventional practice, the mean time spent on studying a worked problem in the conventional conditions, and the number of initially generated incorrect solutions during practice in the conventional conditions.

Training Phase

Table 1 shows, in order, the means and standard deviations for time (in seconds) on general instruction, time (in seconds) on specific instruction, perceived amount of mental effort (1–9) during specific instruction, and spectral energy in the midfrequency band of the heart-rate variability, $(micro\text{-}MI)^2$, during the specific instruction, as a function of type of practice and variability of practice.

The mean time spent on the general instruction did not differ across conditions. There were no significant main effects of practice type, $F(1, 56) = 2.19$, $p > .10$, $MS_e = 5{,}405.58$; practice variability, $F(1, 56) < 1.0$; or Practice Type × Practice Variability interaction, $F(1, 56) < 1.0$. For the

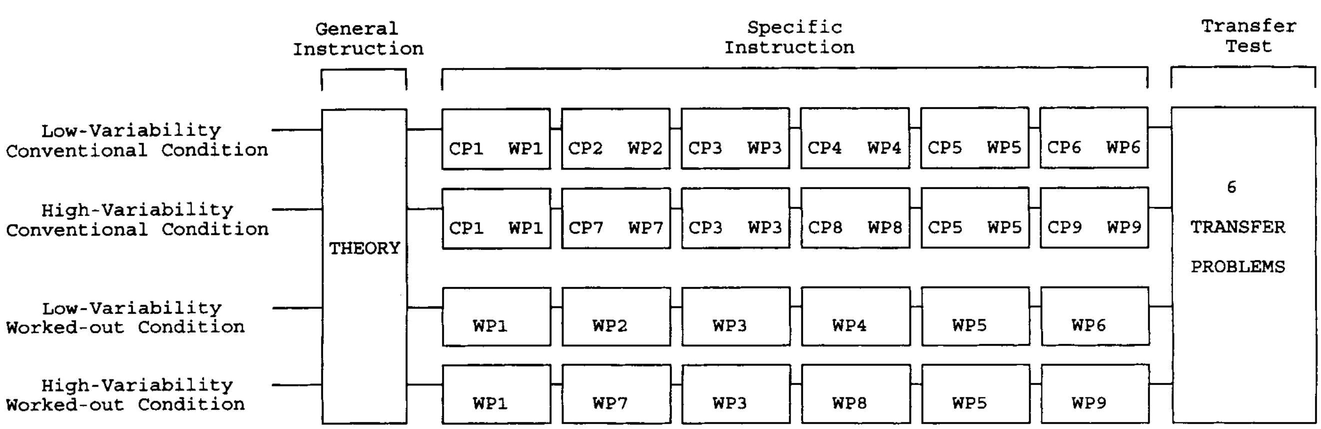

Figure 4. Schematic representation of the procedure of the experiment. CP = conventional problem; WP = worked out problem.

Table 1
Dependent Variables During the Training Phase as a Function of Condition

Dependent variable	Condition			
	LVC	HVC	LVW	HVW
Time on task (in seconds)				
General instruction				
M	888	892	920	916
SD	42	61	90	89
Specific instruction				
M	1,230	1,406	561	625
SD	280	332	170	182
Perceived mental effort (1–9)[a]				
Specific instruction				
M	4.20	4.50	3.20	3.30
SD	1.08	1.05	1.10	1.14
Spectral energy, $(\mu MI)^2$				
Specific instruction				
M	2,434	2,280	2,410	2,311
SD	1,013	807	1,786	1,390

Note. LVC = low-variability conventional; HVC = high-variability conventional; LVW = low-variability worked; HVW = high-variability worked.
[a] The subjects rated perceived mental effort on a scale ranging from *very, very low mental effort* (1) to *very, very high mental effort* (9).

dependent variable mean total time spent on practice (specific instruction), there was a significant main effect for type of practice, such that the students who studied worked examples needed about 45% of the time that was needed by the students in the conventional conditions to solve conventional problems and study their solutions, $F(1, 56) = 125.46$, $p < .0001$, $MS_e = 62{,}901.93$. The main effect for variability of practice was not significant at the .05 level, although the differences between the low- and high-variability conditions were in the expected direction of high-variability practice lasting longer than low-variability practice, $F(1, 56) = 3.44$ $p < .10$. There was no interaction between type of practice and variability of practice, $F(1, 56) < 1.0$.

The time on practice in the conventional conditions (see Table 1) was separated into mean total time spent solving the problem ($M = 978$, $SD = 225$, for the low-variability condition; $M = 1{,}146$, $SD = 309$, for he high-variability condition) and mean total time spent studying the solution ($M = 252$, $SD = 76.7$, for the low-variability condition; $M = 260$, $SD = 78.6$, for the high-variability condition). A *t* test performed on the problem-solving data revealed no difference in mean total time between the low- and high-variability conventional conditions, $t(28) = 1.70, p > .05$. An ANOVA performed on the mean total problem-solution study times yielded a significant main effect for practice-problem type, $F(1, 56) = 92.9$, $p < .0001$, $MS_e = 18{,}273.5$, which indicates that the students in the conventional conditions spent less time on studying the problem solutions than did the students in the worked conditions. Neither the main effect for problem variability, $F(1, 56) = 1.01$, $p > .10$, nor the interaction between problem type and variability, $F(1, 56) < 1.0$, were significant.

To ascertain whether students in the conventional conditions took more time to study the solution after solving a problem incorrectly than after solving a problem correctly, we examined the study times of two groups of students. One group consisted of students with three or more problems incorrect (incorrect group; $n = 10$, $M = 46$ s, $SD = 10.7$) and the other consisted of students who solved at most only one problem wrong (correct group; $n = 13$, $M = 36$ s, $SD = 1.6$).[2] A Mann-Whitney *U* test, which was conducted because the homogeneity assumption was not met, revealed that the correct group invested significantly shorter time in studying the solution than did the incorrect group, $U = 26.5$, $p < .05$.

No reliable conclusions could be drawn on error types. In spite of regular instructions from the experimenter to write down all the solution steps, the students did not systematically do so. Students were allowed to use a pocket calculator to decrease the chance of calculation errors. With some reservations, it may thus be assumed that most errors were due to methodological or procedural misinterpretations. Students in the worked conditions could not make errors during practice because they only had to study worked examples. To determine whether the number of errors, operationalized as the mean number of initially generated incorrect solutions during practice, differed across the conventional conditions, we performed an additional *t* test. The results indicated that students in both conditions did not differ significantly in the average number of incorrect solutions generated (out of 12 possible), $t(28) = 1.41$, $p > .10$ ($M = 4.53$, $SD = 2.49$, for the low-variability conventional condition; $M = 5.66$, $SD = 2.75$, for the high-variability conventional condition).

The ANOVA performed on the mean amounts of perceived mental effort during the specific instruction showed that the main effect for type of practice was significant, $F(1, 56) = 15.36$, $p < .001$, $MS_e = 1.20$, with students in the worked conditions perceiving less mental effort than students in the conventional conditions. The main effect for variability of practice and the interaction between practice type and practice variability were not significant, both $Fs(1, 56) < 1.0$.

The results of the ANOVA on the mean spectral energy in the midfrequency band of the heart-rate variability (0.07–0.14 Hz) during the specific instruction revealed no main effects and no interactions, all $Fs(1, 56) < 1.0$.

Test Phase

For the test phase the analyses were conducted on the following dependent variables: time on transfer test (in seconds), transfer test performance (percentage correct), perceived amount of invested mental effort (1–9) during the

[2] Note that a problem was considered incorrect if a student had generated two incorrect solutions, and it was considered correct if a student had generated at most only one incorrect solution to the problem as well as the correct solution.

Table 2
Dependent Variables During the Test Phase as a Function of Condition

Dependent variable	LVC	HVC	LVW	HVW
Time on task (in seconds)				
Transfer test				
M	1,602	1,506	1,483	1,605
SD	308	319	439	361
Performance (% correct)				
Transfer test				
M	28.90	27.80	47.80	62.20
SD	11.70	16.60	13.90	16.00
Perceived mental effort (1–9)[a]				
Transfer test				
M	5.90	6.10	5.50	5.20
SD	1.13	0.66	1.11	1.13
Spectral energy, $(\mu MI)^2$				
Transfer test				
M	2,460	2,409	2,157	2,141
SD	938	929	1,355	1,103

Note. LVC = low-variability conventional; HVC = high-variability conventional; LVW = low-variability worked; HVW = high-variability worked.
[a] The subjects rated perceived mental effort on a scale ranging from *very, very low mental effort* (1) to *very, very high mental effort* (9).

transfer test, and spectral energy of the heart-rate variability, $(micro\text{-}MI)^2$, during the transfer test. An additional *t* test was performed on the differences between spectral energy in mentally active (instruction and transfer test) and mentally inactive (baseline) periods.[3]

Table 2 shows the means and standard deviations of time on transfer test, transfer test performance, perceived amount of mental effort invested in the transfer problems, and the spectral energy in the midfrequency band of the heart-rate variability during the transfer test, respectively, as a function of type of practice and variability of practice.

The ANOVA performed on the mean time needed for solving the transfer problems revealed that the times to solve problems did not differ across conditions, for type of practice and variability of practice, both $Fs(1, 56) < 1.0$, and for the interaction, $F(1, 56) = 1.37, p > .10$.

Transfer test performance was expressed as the mean percentage of all problems answered correctly. The ANOVA on transfer test performance revealed a significant main effect of type of practice, such that the students in the worked condition outperformed the students in the conventional condition, $F(1, 56) = 50.09, p < .0001, MS_e = 212.96$. The main effect of variability of practice did not yield significant differences between conditions, $F(1, 56) = 3.13, p < .10$. As predicted, the interaction between Practice Type × Practice Variability was significant, $F(1, 56) = 4.26, p < .05$. To determine the nature of this interaction, we conducted a post hoc analysis of simple effects using Tukey's method of multiple comparisons (honestly significant difference [HSD] = 14.10, $\alpha = .05$). This analysis indicated that subjects in the worked conditions exhibited significantly higher transfer performance than subjects in the conventional conditions under both low- and high-variability practice. Furthermore, with regard to transfer performance in the worked conditions, there was a significant advantage of high-variability practice. Finally, variability had no effect on transfer performance in the conventional conditions.

With regard to the mean amount of perceived mental effort invested in the transfer problems, the main effect for practice type was significant, $F(1, 56) = 6.50, p < .025, MS_e = 1.06$, thus indicating that the students who had practice with worked examples perceived that they invested less mental effort in solving the transfer problems than did the students who had practice with conventional problems. The main effects for practice variability and the interaction of Practice Type × Practice Variability were not significant, both $Fs(1, 56) < 1.0$.

Mean transfer performance scores and mean mental effort scores were also combined according to the *efficiency approach* introduced by Paas and Van Merriënboer (in press-a). The efficiency approach is based on the conversion of raw mental effort data and raw performance data to *z* scores, which can be displayed in an *M* (mental effort) – *P* (performance) cross of axes. It is argued that the combined effects on mental effort and performance of an experimental instructional condition can be deduced from the position of its related point in the cross of axes in relation to points that represent other instructional conditions. Relative condition efficiency is defined as the observed relation between mental effort and performance in a particular condition in relation to a (hypothetical) baseline condition in which each unit of invested mental effort equals one unit of performance. Relative condition efficiency can be calculated according to the following formula:[4]

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

The sign for relative condition efficiency *E* is dependent on *M* and *P*, according to the following rules:

If $M - P < 0$, then E is positive;

if $M - P > 0$, then E is negative.

[3] To obtain a baseline for the spectral-analysis technique, we calculated spectral energy for four regularly spaced periods in which the subjects were asked to refrain from mental activity. In order, the first, second, third, and fourth periods were obtained (a) before the start of the general instruction, (b) between the general and the specific instruction, (c) between the specific instruction and the transfer test, and (d) after the transfer test.

[4] Square root 2 in the formula for relative condition efficiency comes from the general formula for calculation of the distance of a point $p(x,y)$ to a line $ax+by+c=0$, distance$=(|ax+by+c|)/(a^2+b^2)^{1/2}$. In this case we have a *M* (mental effort) – *P* (performance) coordinate system where we want to know the distance of point $p(M, P)$ to the line $M - P = 0$. (For further details, see Paas & Van Merriënboer, in press.)

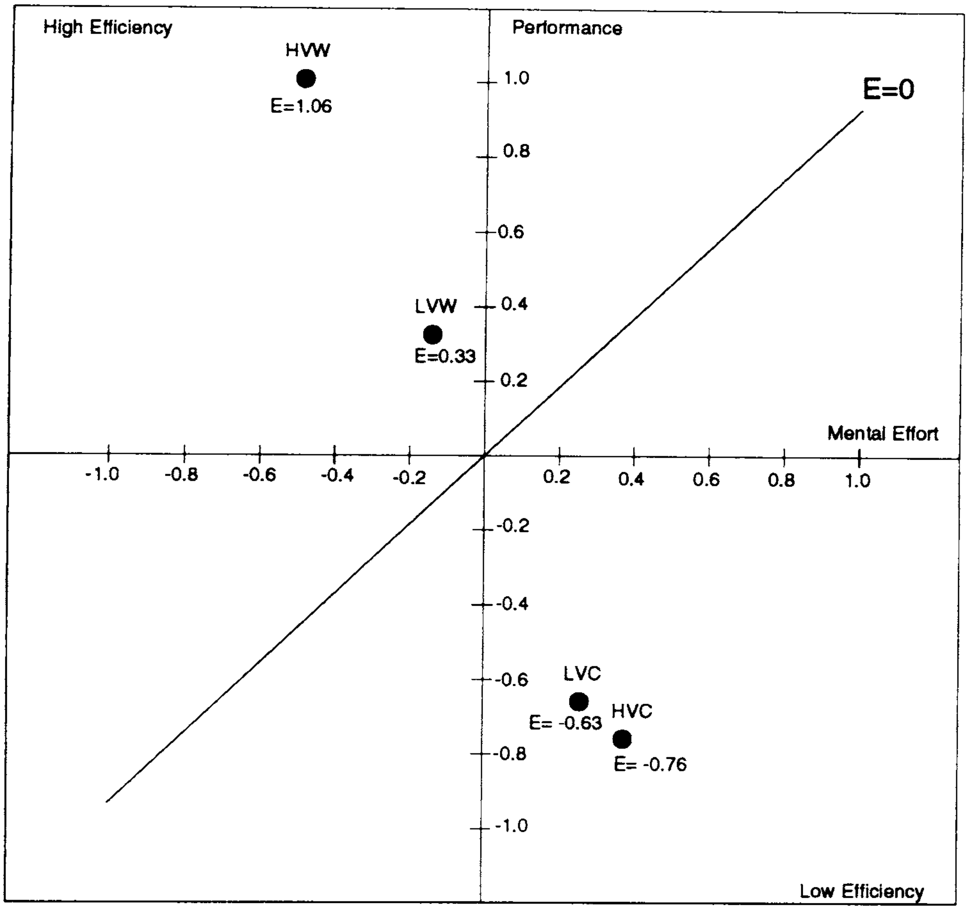

Figure 5. Relative condition efficiency as a function of practice type and practice variability. LVC = low-variability conventional condition; HVC = high-variability conventional condition; LVW = low-variability worked condition; HVW = high-variability worked condition. E = efficiency.

The mean relative condition efficiency data are pictured in Figure 5. The ANOVA performed on these data revealed a significant effect of practice type, $F(1, 56) = 5.13, p < .025$, $MS_e = 0.86$, thus indicating that the conditions in which students had to study worked examples were more efficient than the conditions in which students had to solve conventional problems before studying the solutions. The main effect of practice variability and the Practice Type × Practice Variability interaction were not significant, both *F*s (1, 56) < 1.0.

With regard to the spectral-energy data in the midfrequency band of the heart-rate variability, the results of the ANOVA on the transfer test revealed no main effects and no interactions, all $Fs(1, 56) < 1.0$. Additional paired sample *t* tests were conducted to test the differences between the baseline spectral energy in the midfrequency band (mentally inactive) and the energy during the phases in which the students had to be mentally active. These *t* tests indicated that this method was sensitive for relatively large differences in mental activity. Students invested significantly more mental energy during practice, $t(58) = 8.27, p < .0001$, and transfer, $t(58) = 8.91, p < .0001$, than they did during the mentally inactive periods.

Discussion

This study examined the effects of practice-problem type, variability of practice, and combinations of these variables with regard to their effects on training performance, transfer performance, and cognitive load. The results show that training with worked examples requires less time and is perceived as demanding less mental-effort than training with conventional problems. In addition, worked-example training leads to better transfer performance and is perceived as demanding less effort than training with conventional problems. Practice-problem variability only had an influence in the worked conditions, such that high-variability practice resulted in better transfer performance than did low-variability practice.

The interaction that was found with regard to transfer performanee between problem type and problem variability indicates that the timesaving and cognitive-capacity-saving character of worked examples enabled subjects in the worked condition to use the high variability in practice problems to their advantage. Although the variability of problem type during training increases cognitive load, it is important to schema acquisition and should be included. The interaction provides evidence for our claim that the positive effects of variability of problem type during training on schema acquisition will be manifested only if extraneous cognitive load is reduced.

The results regarding the measures of cognitive load that are based on mental effort revealed that the physiological technique was only sensitive to differences between mentally inactive (baseline) and mentally active periods (instruction and transfer test), whereas the rating-scale technique could successfully differentiate between different instructions conditions, which are assumed to involve different levels of mental activity. It must be concluded that the spectral-energy measure is not sensitive to the differences in cognitive load demonstrated in the tasks presented here. The findings concerning the rating-scale technique are in agreement with the results of Paas (1992), thus indicating that subjective reports of the amount of invested mental effort are sensitive to relatively small differences in task parameters that are expected to have an influence on cognitive load. Although these findings suggest that the rating-scale technique can be considered a valuable research tool for estimation of cognitive load in instructional research (see Paas, 1992, 1993), further research is needed to establish the exact relationship between perceived mental effort and cognitive load.

The students in the conventional conditions reported investing more mental effort during practice than did students in the worked conditions. In addition, students in the conventional conditions spent about twice as much time on practice. These results indicate that the processes that were required in the training phase of the conventional conditions (namely, solving problems, generating incorrect solutions, and studying correct solutions) imposed higher cognitive load than did the process of studying correct solutions in the worked conditions. Note that the differences in training time and perceived mental effort were inversely reflected in the quality of transfer performance. Better and less effort-demanding transfer performance in the worked conditions suggests that a considerable part of the mental effort in the conventional conditions was invested in processes that were not relevant for learning (i.e., extraneous cognitive load; see also Van Merriënboer & De Croock, 1992). Furthermore, it shows that the level of perceived mental effort during instruction is not invariably associated with the quality of transfer performance. Whether differences in perceived mental effort during instruction result in differences in transfer performance seems to depend on the relevance of the processes required to work with the instruction.

The superior transfer results that occurred with the worked examples are similar to those obtained by Paas (1992), Reed and Actor Bolstad (1991), and Sweller (1988). The results indicate that the presentation of six worked examples was a particularly effective instructional method for obtaining schema acquisition. Lower transfer performance in the conventional conditions might be due to the interference between schema acquisition and the problem-solving methods used as well as the frequent generation of incorrect solutions. Both factors are cognitive-capacity demanding and—as a result of the limited cognitive capacity—may not leave enough spare cognitive capacity for the acquisition of schemata.

In addition, the generation of incorrect solutions might have interfered with the organization and accessibility of knowledge in schemata (Bassok & Holoyoak, 1989). Although the evidence is indirect, the negative effects of initially constructing incorrect solutions on the quality of the schemata seem to exceed the positive influence of subsequently studying the correct solutions. It is possible that schema acquisition was impaired by the concurrent processing of the initial construction of incorrect solutions and the subsequent opportunity to study the correct solution. This hypothesis seems to be supported by research findings of Ross (1987, 1989) and Ross and Kennedy (1990). A main finding of Ross and his colleagues' research was that early experiences have a major impact on later performance because learners try to remind themselves of earlier similar experiences, irrespective of the quality of the examples remembered.

The effects of interactions between type of practice and failed solutions on schema acquisition and transfer are important topics for additional research. For instance, it would be interesting to investigate whether the cognitive system exhibits a priority hierarchy in which more attention is given to the earliest practical experiences than is given to later mental experiences.

The experimental conditions in this study differed from previous studies (e.g., Cooper & Sweller, 1987; Paas, 1992; Sweller et al., 1990) in that in the worked conditions only worked examples had to be studied. By alternating worked examples with conventional problems, previous studies on worked examples have left open the possibility that the sequence in which the students proceeded through problem-study and problem-solving cycles might have caused the superiority of the worked conditions. However, in this study the worked condition only contained worked examples that had to be studied, thereby excluding the aforementioned alternative explanation.

Practice-problem type seems to play a major role in the effect of the instructional design on transfer of geometrical problem-solving skills. The positive results of the numerous studies on the effects of instruction with worked examples on transfer are overwhelming. The fact that most studies dealt with rather complex cognitive problem domains, such as mathematics, physics, and computer programming, seems to be no coincidence. In these domains problem solving requires a combination of rule-based and knowledge-based behavior, which implies a necessity to transfer acquired knowledge and skills. It is exactly these domains that are characterized by high processing load, in which the largest effects of worked examples could be expected. The superior transfer performance in the high-variability worked condition (this study) fulfills this expectation.

An important issue for future research is raised by the results of the studies of Paas (1992) and Van Merriënboer (1990b), which showed that the further the required transfer was, the more subjects could gain from instruction with worked examples or partly worked examples with a completion assignment. A possible explanation for these results can be found in the timesaving and mental-effort-saving characteristics of worked examples that provide the possibility to increase the quantity of training, the quality (e.g., variability, in this study) of training, or both, and consequently to facilitate schema acquisition, rule automation, and transfer. With worked examples students' attention can be directed to the goal-relevant aspects of the task, thereby preventing them from irrelevant and capacity-demanding actions.

It is clear that the cognitive-load oriented claims that we make on the basis of the present results need further experimental confirmation. For example, one could claim that it was simply the appearance of incorrect solutions in the conventional conditions per se, rather than cognitive load, that caused the present transfer performance differences. Therefore, future research should include more process-based measures to verify that the learning processes used, such as weak problem-solving methods including the frequent generation of incorrect solutions, can be explained through the concept of cognitive load. Among others, verbal protocols could be used to investigate competing process-based explanations and to gain more insight into the cognitive processes that are deployed.

Furthermore, it could be argued that the present results are not only consistent with a cognitive-load approach but also with a capacity-constrained approach (e.g., Just & Carpenter, 1992). Whereas the former assumes a continuous relationship between cognitive load and learning, an important implication of capacity-constrained views is that performance differences among individuals emerge primarily when the task demands consume sufficient working-memory capacity to exhaust the resources of some subjects. The present research cannot provide a conclusive answer about the superiority of one view over the other for explaining the results that we found. Future research could be directed at contrasting cognitive-load views with capacity-constrained views.

In formulating instructional prescriptions from the present research, we need to consider some aspects carefully. First, Chandler and Sweller (1991), Sweller (1989), and Ward and Sweller (1990) have pointed out that worked examples have to be structured effectively. According to these researchers, students should be prevented from activities that impose extraneous cognitive load, such as mentally integrating mutually referring, disparate sources of information (e.g., text and diagram). These researchers suggested that the instructional designer should integrate the multiple sources of information in the worked examples.

Second, Chi, Bassok, Lewis, Reimann, and Glaser (1989) pointed out that good students use worked examples in a way that is different from the way that poor students use them. Chi et al. concluded from students' self-explanations that the students' ability level determines the way students make use of worked examples. During problem solving, good students used the examples for a specific reference, whereas poor students reread them to search for a solution. Furthermore, good students seemed to refer to the examples less frequently within each solution attempt.

Third, Van Merriënboer and Paas (1990) pointed out that the acquisition and use of cognitive schemata are controlled processes that require the voluntary investment of mental effort from the learner. Whereas in experimental studies learners are often highly motivated and inclined to invest mental effort (especially if high performance is rewarded), in typical school settings one often has to deliberately provoke mindful abstraction. It is possible that a pure worked instructional design, as it was used in the present study, will not have the same effects in school settings.

However, because wholesale advocacy of instruction with worked examples seems premature, we believe that when training for transfer in complex cognitive domains, effectively structured worked examples proved to be a crucial part of the instructional strategy.

References

Aasman, J., Mulder, G., & Mulder, L. J. M. (1987). Operator effort and the measurement of heart-rate variability. *Human Factors, 29,* 161–170.

Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics on algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 153–166.

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

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

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

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

Direct Technology Limited. (1987). *Automator mi* [Computer program]. Isleworth, Middlesex, England: Author.

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

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors, 26,* 519–532.

Hamilton, P. (1979). Process entropy and cognitive control: Mental load in internalized thought processes. In N. Moray (Ed.), *Mental workload: Its theory and measurement* (pp. 289–298). New York: Plenum Press.

Jelsma, O., Van Merriënboer, J. J. G., & Bijlstra, J. P. (1990). The ADAPT design model: Towards instructional control of transfer. *Instructional Science, 19,* 89–120.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

Mulder, G. (1980). *The heart of mental effort: Studies in the cardiovascular psychophysiology of mental work.* Unpublished doctoral dissertation, University of Groningen, Groningen, The Netherlands.

Mulder, L. J. M. (1988). *Assessment of cardiovascular reactivity by means of spectral analysis.* Unpublished doctoral dissertation, University of Groningen, Groningen, The Netherlands.

Mulder, L. J. M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology, 34,* 205–236.

O'Donnel, R. & Eggemeier, F. T. (1986). Workload assessment methodology. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (pp. 42.1–42.49). New York: Wiley.

Owen, E., & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Educational Psychology, 77,* 272–284.

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. (1993). *Cognitive-load measurement in instructional research: An evaluation of the usefulness of a rating scale and spectral analysis of the heart-rate variability.* Manuscript submitted for publication.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (in press-a). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors.*

Paas, F. G. W. C. & Van Merriënboer, J. J. G. (in press-b). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review.*

Ranzijn, F. J. A. (1991). The number of video examples and the dispersion of examples as instructional design variables in teaching concepts. *Journal of Experimental Education, 59,* 320–330.

Reed, S. K., & Actor Bolstad, C. (1991). Use of examples and procedures in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 753–766.

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 629–639.

Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 456–468.

Ross, B. H., & Kennedy, P. T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 42–55.

Sanders, A. F. (1979). Some remarks on mental load. In N. Moray (Ed.), *Mental workload: Its theory and measurement* (pp. 41–77). New York: Plenum Press.

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

Shapiro, D. C., & Schmidt, R. A. (1982). The schema theory: Recent evidence and developmental implications. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and coordination* (pp. 113–150). New York: Wiley.

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

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, 4,* 457–466.

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.

Unicomal A/S. (1987). *Unicomal* [Computer program]. Jels, Rodding, Denmark: Unicomal A/S & Unicomal Documentation Center.

Van der Meulen, P., & Mulder, L. J. M. (1990). *CARSPAN (version 1.2) user's guide.* Groningen, The Netherlands: University of Groningen, Institute for Experimental Psychology.

Van Merriënboer, J. J. G. (1990a). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of Educational Computing Research, 6,* 265–287.

Van Merriënboer, J. J. G. (1990b). *Teaching introductory computer programming: A perspective from instructional technology.* Enschede, The Netherlands: Bijlstra & Van Merriënboer.

Van Merriënboer, J. J. G., & De Croock, M. B. M. (1992). Strategies for computer-based programming instruction: Program completion vs. program generation. *Journal of Educational Computing Research, 8,* 365–394.

Van Merriënboer, J. J. G., Jelsma, O., & Paas, F. G. W. C. (1992). Training for reflective expertise: A four-component instructional design model for complex cognitive skills. *Educational Technology Research and Development, 40,* 23–43.

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.

Received November 25, 1992
Revision received August 26, 1993
Accepted September 8, 1993 ■