MEASUREMENT OF COGNITIVE LOAD
IN INSTRUCTIONAL RESEARCH [1]

FRED G. W. C. PAAS, JEROEN J. G. VAN MERRIËNBOER

Department of Instructional Technology
University of Twente

JOS J. ADAM

Department of Movement Science
University of Limburg

Summary.—The results of two of our recent empirical studies were considered to assess the usefulness of subjective ratings and cardiovascular measures of mental effort in instructional research. Based on its reliability and sensitivity, the subjective rating-scale technique met the requirements to be useful in instructional research whereas the cardiovascular technique did not. It was concluded that the usefulness of both measurement techniques in instructional research needs to be investigated further.

As cognitive research has made us aware of the limits of the processing capacity of the human mind, cognitive load has become an important issue in instructional research. Sweller and collaborators have performed studies that provide insight into the relation between instruction and cognitive load (e.g., Chandler & Sweller, 1991, 1992; Sweller, 1988, 1989; Sweller, Chandler, Tierney, & Cooper, 1990). Despite the availability of several kinds of measurement of cognitive load, instructional research has exclusively concerned performance-based measures to estimate cognitive load (e.g., Sweller, 1988). Consequently, information on the usefulness of measurement in instructional research is limited. The goal of this paper was to assess the usefulness of two measurements of cognitive load.

The structure of this discourse is as follows. In the next section, the concept of cognitive load, its dimensions, its role in instructional research, and its measurement are discussed. In the section which follows, the basic requirements for measurement of cognitive load in instructional research are presented. In order, the criteria of reliability and sensitivity are described. In the next section, an evaluation of the usefulness of the rating-scale technique is made using two empirical studies (Paas, 1992; Paas & Van Merriënboer, 1994). The next section contains an evaluation of the usefulness of a cardiovascular technique, based on one empirical study (Paas & Van Merriënboer, 1994). Finally, in a concluding section, a discussion is offered on the usefulness of the two cognitive-load measurement techniques in instructional research.

THE CONCEPT OF COGNITIVE LOAD

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

With regard to the assessment, cognitive load can be conceptualized with respect to the dimensions of mental load and effort and performance. *Mental load* is imposed by the task or environmental demands. This task-centered dimension, which is considered independent of subjects' characteristics, is assumed constant for a given task. The human-centered dimension, *mental effort*, refers to the amount of resources actually allocated to accommodate the task demands. Mental effort reflects the amount of controlled processing in which the individual is engaged. The amount of invested mental effort comprises all three causal factors (task's environmental characteristics, subjects' characteristics, and interactions between these). Finally, the level of *performance* achieved is an indication of the third measurement dimension; it also reflects all three causal factors.

The intensity of effort expended by students is often considered the essence of cognitive load (Hamilton, 1979; Paas, 1992). Therefore, mental effort can be used as an index of cognitive load. Mental effort can be measured with "subjective" (rating scales) and "objective" techniques (physiological parameters). With performance measures, mental effort can provide information on the cognitive costs at which the performance is attained and on the relative efficiency of instructional conditions (see Paas & Van Merriënboer, 1993).

Cognitive Load and Instructional Research

Especially for novices learning a complex cognitive task, cognitive capacity is a limiting factor. The adverse effects of high or excessive cognitive load on learning have been reported in a diversity of relatively complex cognitive domains such as mathematics (Paas, 1992; Sweller, 1989; Tarmizi & Sweller, 1988), numerically controlled computer programming (Chandler & Sweller, 1992; Paas & Van Merriënboer, 1994; Sweller, *et al.*, 1990), and electrical engineering (Chandler & Sweller, 1991). Problems in these domains

are frequently characterized by an hierarchical goal structure, that is, the top goal can only be accomplished by successfully attaining all subgoals. Novices often cannot cope initially with these problems because they are overwhelmed by the amount of information and because they adopt cognitive capacity-demanding, weak problem-solving strategies, e.g., means-ends analysis, hill climbing, to solve them (Sweller, 1988).

Failure of learning and performance of complex tasks may be attributed to the task demands which exceed the available cognitive capacity (cognitive overload), the inadequate allocation of attention, or both. Frequently occurring situations of cognitive overload may also lead to reduced motivation and consequently to failure.

The measurement of cognitive load is of crucial importance for instructional research. Whereas cognitive load is considered an important theoretical construct in models of instructional design for complex cognitive skills (e.g., Sweller, 1989; Van Merriënboer, Jelsma, & Paas, 1992), it is not common to measure cognitive load while conducting research on instruction. The few efforts in instructional research to measure cognitive load are almost exclusively concerned with performance measures. For example, to find evidence for cognitive load theory, Sweller (1988) used a computational model of cognitive load in which the cognitive processing load was based on the number of productions (if-then relations) underlying task performance, acquisition time, number of errors, and test performance.

Performance measures often cannot be used as indicators of potential performance problems if additional demands are imposed. Performance measures may show no deficits with increasing task or instructional demands because subjects are able to invest more mental effort to compensate for increasing cognitive load (Tulga & Sheridan, 1980). Measurement of mental effort, therefore, may yield important information that is not necessarily reflected in performance-based measures. Moreover, combining performance and mental effort measures allows calculation of an index of mental efficiency (Paas & Van Merriënboer, 1993, 1994).

Available Measurement Techniques

Three major classes of mental effort measurement techniques can be distinguished (Eggemeier, 1988). These include subjective, psychophysiological, and task- and performance-based indices. Subjective techniques use rating scales to report the expenditure of experienced effort or capacity. These techniques are based on the assumption that subjects are able to introspect on their cognitive processes and to report the amount of mental effort expended. Psychophysiological techniques use physiological indices, such as pupillary diameter, heart-rate variability, and event-related brain potentials, to obtain information on mental effort. These techniques are based on the assumption that changes in cognitive functioning are reflected in physiological

functioning. Task- and performance-based techniques use objective task characteristics and performance levels to obtain information on mental effort.

In the present paper a subjective technique is evaluated with regard to its usefulness in instructional research. This evaluation is based on two empirical studies (Paas, 1992; Paas & Van Merriënboer, 1994). Both studies were part of a larger project designed to assess the effects of alternative training strategies on transfer in the domains of statistics and geometry. Thus, the studies were not specifically designed to assess the usefulness of measurements of mental effort in instructional research. However, the results from the applied techniques can be analyzed and interpreted from this perspective.

In both studies a subjective rating scale was used; on this, numerical translations of the perceived amount of mental effort had to be given. In addition, in Study 2 (Paas & Van Merriënboer, 1994) a psychophysiological technique was used in which spectral analyses of heart-rate variability were used to indicate the intensity of mental effort.

Rating scale technique.—The rating scale that was used in the present research was a modified version of Bratfisch, Borg, and Dornic's scale (1972) for measuring perceived task difficulty (9-point, ranging from 1-very, very easy to 9-very, very difficult). Subjects had to report the amount of mental effort invested on a 9-point symmetrical category scale by translating the perceived amount of invested mental effort into a numerical value. The numerical values and labels assigned to the categories ranged from "1" to "9," corresponding to "very, very low mental effort" to "very, very high mental effort."

Spectral analysis of heart-rate variability.—A physiological method for measuring the intensity of mental effort is to subject the variability in heart rate to spectral analysis, a mathematical method for investigating whether a signal contains periodic components. The time between successive heart beats seems to be determined by three different feedback mechanisms, connected with respiration, blood pressure, and body-temperature regulation. Controlled processing is related to a specific cardiovascular state that is manifest in the heart-rate variability power spectrum band, which is related to the blood-pressure regulation, the so-called midfrequency band from 0.07 to 0.14 Hz. Intensity of effort is directly related to controlled processing, which in turn causes a change in this power spectrum. This blood-pressure-related component has been found to decrease with increasing mental effort (e.g., Aasman, Mulder, & Mulder, 1987; L. J. M. Mulder, 1988, 1992).

The power density spectrum, which results from the spectral analysis of the heart-rate variability, provides the important advantage of separating the effects of respiratory rate (high-frequency band from 0.15 to 0.40 Hz) and thermoregulation (low-frequency band from 0.02 to 0.06 Hz) from the men-

tal-effort-related blood-pressure component. The spectrum shows the contribution of oscillations within a range of frequencies to the total variance in the signal. So, the spectral decomposition of the variance provides information about both spontaneous and task-related fluctuations in heart rate. Among others, Aasman, et al. (1987), G. Mulder (1980), and L. J. M. Mulder (1988) have validated this technique with several cognitive tasks, e.g., multidimensional classification, sentence comprehension, and continuous working memory.

REQUIREMENTS IN MEASUREMENT OF MENTAL EFFORT IN INSTRUCTIONAL RESEARCH

The properties considered in evaluating the usefulness of the measurement of mental effort in this paper are reliability and sensitivity. It should be noted that other important properties of measurement such as validity and intrusiveness could not be assessed from the results of the present empirical studies.

Criterion of Reliability

Information on the *reliability* of the subjective rating-scale technique was obtained by performing analyses of the internal consistency of the ratings. These analyses yielded Cronbach coefficients alpha of reliability. With regard to the heart-rate variability-based technique, the reliability was estimated by calculating Pearson's product-moment correlation coefficients for comparable measurement periods.

Criterion of Sensitivity

Sensitivity refers to the capability of the measurement technique to reflect differences in the expenditure of processing capacity associated with training and transfer conditions. The sensitivity was obtained by assessing whether the techniques discriminate between differences in training conditions (manipulated by varying the type and variability of training problems), the results of the training conditions that are reflected in expenditure of mental effort during the transfer test and in task complexity. Expectations regarding the sensitivity of the rating scale and the spectral analysis were based on the assumptions that training conditions affect the efficiency of mental performance, that there is a positive relationship between task complexity and expenditure of cognitive capacity, and that increased expenditure of cognitive capacity will be accompanied by physiological changes and feelings of effort reflected in appropriate indices (Eggemeier, 1988).

EVALUATION OF THE RATING SCALE

Method

The usefulness of the rating scale was evaluated on the basis of two studies.

Study 1.—Paas (1992) studied the effects of three computer-based train-
ing strategies on training performance, transfer performance, and cognitive
load in the domain of statistics. The *conventional, worked,* and *completion*
strategies emphasized, in order, the solving of conventional problems, the
study of worked-out problems, and the completion of partly worked-out
problems. All students ($N = 42$) were offered identical formal instruction and
the same set of 12 statistical problems. For the conventional training strate-
gy, all problems in the set were goal-specific open problems (conventional
problems). For the worked strategy, the first two problems of each subset of
three consisted of problems together with worked-out problem solutions,
while each third problem in the subset of three was a conventional one. For
the completion strategy, the first two problems in each subset of three were
problems together with partly worked-out solutions that had to be complet-
ed, while each third problem was, again, a conventional problem. After the
instruction the subjects had to perform a 12-problem near-transfer test and a
12-problem far-transfer test. Except for specific values, near-transfer prob-
lems resembled the conventional training problems. The far-transfer prob-
lems were characterized by presentation formats that differed from the prob-
lems during instruction, involved unstructured data presentations, and re-
quired application of different combinations of problem-solving strategies (for
examples, see Paas, 1992).
    The rating scale was provided, explained, and illustrated just before the
beginning of the experiment and during the general instruction. After solv-
ing a conventional problem, studying a worked-out problem, or completing a
completion problem, the students had to score the amount of mental effort
invested in the preceding problem.
    Study 2.—Paas and Van Merriënboer (1994) studied the effects of four
computer-based training strategies, in the context of geometrical problem
solving in Computerized Numerically Controlled programming, on training
performance, transfer performance, and cognitive load. A low- and high-vari-
ability *conventional* condition were compared with a low- and high-variability
worked condition. During instruction in the conventional conditions a set of
six problems was presented which had to be solved. During instruction in
the worked conditions the same set of six problems was presented with their
written-out solutions which had to be studied. 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 the absolute coordinates
to relative coordinates, respectively. These problems were identical for the
low- and high-variability conditions. The other problems (second, fourth,
sixth) differed from these problems as a function of practice-problem vari-
ability such that in the low-variability conditions these problems had only
different values and in the high-variability conditions both values and prob-

lem formats, i.e., problem goal and localization in the cross of axes, were different. After the instruction the subjects had to perform a six-problem transfer test (for examples, see Paas & Van Merriënboer, 1994).

The mental-effort rating scale was provided, explained, and illustrated just before the beginning of the experiment and again during the general instruction. After each problem the students had to rate the amount of mental effort invested in the preceding problem.

Results and Discussion

Interpretations regarding the rating-scale data of both studies should be considered under the assumptions that individual differences in scaling were equally divided across experimental conditions and that there was no interaction between individual differences in scaling and conditions.

Reliability of the rating scale.—In comparing the effects of the training strategies on transfer, Study 1's coefficient of reliability (Cronbach alpha) was 0.90 for the rating scale. The calculation of the coefficient is based on 28 problems which were the same in each experimental condition, that is, four conventional training problems and 24 transfer-test problems.

For Study 2, a coefficient of reliability (Cronbach α) of 0.82 was obtained in comparing the effects of the training strategies on transfer. The calculation was based on the six transfer-test problems which were the only problems that were the same in each condition.

Sensitivity of the rating scale.—In Study 1, the perceived amount of mental effort invested during the near- and far-transfer tests differed across conditions (for the Near transfer test, Conventional $M = 4.4$, Worked $M = 3.8$, Completion $M = 3.1$; $F_{2,39} = 9.14$, $p < .001$, $MS_e = 1.01$; for the Far-transfer test, Conventional $M = 5.6$, Worked $M = 4.1$, Completion $M = 3.6$; $F_{2,39} = 14.97$, $p < .001$, $MS_e = 1.00$). *Post hoc* tests using Fisher's significant-difference procedure $(p < .05)$ showed that the perceived amount of mental effort invested in the conventional condition was higher than those in the worked and completion conditions. These results indicate that the mental-effort rating scale was sensitive to the differences in cognitive structure that resulted from the different training strategies.

To examine the sensitivity of the rating scale for differences in task complexity, mean mental-effort scores of all subjects were collapsed across practice, near-transfer, and far-transfer conditions. Although the classification of practice, near-transfer, and far-transfer problems as problems low, medium, and high in complexity is very rough, it is the most obvious classification in this study.

A one-way analysis of variance with repeated measurements on the factor of condition indicated that the mean mental-effort scores differed across the practice $(M = 2.4)$, near-transfer $(M = 3.5)$, and far-transfer $(M = 4.4)$ conditions $(F_{2,82} = 56.39, p < .0001)$. Multiple comparisons showed that all means

differed from each other ($p < .01$). Subjective ratings of mental effort increased monotonically with task complexity. So, the relative simple practice problems were rated lower than the more complex near-transfer problems, and these problems were rated lower than the most complex far-transfer problems. The differences in task complexity were reflected in the differences in scores on the mental-effort rating scale and the scale seems to be sensitive to task complexity.

With regard to the mean amount of mental effort invested in the transfer problems in Study 2, the main effect for practice type was significant ($F_{1,56} = 6.5$, $p < .025$, $MS_e = 1.06$), indicating that the students who had practice with worked examples rated mental effort to solve the transfer problems lower (Low variability, $M = 5.5$, High variability, $M = 5.2$) than the students who had practice with conventional problems (Low variability, $M = 5.9$, High variability, $M = 6.1$). The main effect for practice variability and the interaction of practice type × practice variability were not significant ($Fs_{1,56} < 1.0$).

To examine the sensitivity of the rating scale for differences in task complexity mean mental-effort ratings of all subjects were collapsed across practice and transfer conditions. The practice problems were classified as low in complexity and the transfer problems as high in complexity. A one-way analysis of variance with repeated measurements on the condition factor indicated that the mean mental-effort ratings differed across the practice ($M = 3.8$) and transfer ($M = 5.7$) conditions ($F_{1,59} = 181.08$, $p < .0001$). In conclusion, the differences in task complexity were reflected in the differences in scores on the mental-effort rating scale.

EVALUATION OF HEART-RATE VARIABILITY

Method

The method of Study 2 has already been described above. In this study, an electrocardiogram (ECG) was recorded from each sitting subject throughout the experimental session, that is, from 4 min. before the start of the computer-based training program to 4 min. after the end of the program. Disposable, pregelled, silver/silver chloride ECG electrodes, manufactured by Medtronic, were used to collect the heart-rate signal. The ECG signal was delivered by three surface electrodes to an ECG-R-wave toptrigger (ECT-4). 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 processus iliaca anterior superior. An Event-Data Multiplexer-PC (EDM-PC) for stimulus-response collection was connected to a personal computer for data collection and presentation (both Olivetti M240) and to the ECT-4. 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 specific parts in the program, were collected and transmitted by the

EDM-PC to the data-collection computer on which these data were collected and displayed. The program for collection of the R-top time series and the synchronizing signals was written in the programming language Turbo Pascal.

To obtain a reliable baseline, spectral analyses were conducted on four regularly planned rest phases, in which subjects were asked to refrain from mental activity, and which lasted four minutes.

After artefact correction according to L. J. M. Mulder (1988), spectral analyses on the R-top time series were conducted using the program "CARdiovascular SPectral ANalysis," CARSPAN (van der Meulen & L. J. M. 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 (squared modulation index) was used. This means all signal fluctuations were considered as relative variations with respect to the mean value in the time segment under consideration.

Results and Discussion

Reliability of heart-rate variability.—To obtain information on the reliability of heart-rate variability, Pearson's product-moment correlation coefficients were calculated for the four regularly spaced comparable periods (4 min.) during which a baseline measurement of the heart rate was performed. To be reliable, the collected heart-rate measures of the different periods should correlate. The correlation matrix in Table 1 shows that, except for the correlation between Periods 1 and 3, all periods were significantly correlated. Although significant, the correlations were so low it might be concluded that this technique is not reliable.

TABLE 1
CORRELATION MATRIX WITH PEARSON'S PRODUCT-MOMENT CORRELATION COEFFICIENTS
OF FOUR BASELINE MEASUREMENT PERIODS OF HEART-RATE VARIABILITY

Period	Period		
	Baseline 2	Baseline 3	Baseline 4
Baseline 1	0.354*	0.218	0.437*
Baseline 2		0.448*	0.348*
Baseline 3			0.439*

*$p < .01$.

Sensitivity of heart-rate variability.—The results of the analyses of variance on the mean spectral energy in the midfrequency band in the specific instruction and in the transfer test gave no main effects and no interactions ($Fs_{1,56} < 1.0$).

Paired-sample t tests, conducted to test the differences between the baseline spectral energy in the midfrequency band (mentally inactive periods) and the energy during the phases in which the students had to be mentally active, indicated that students invested significantly more mental energy during

practice ($M = 2358$; $t_{58} = 8.27$, $p < .0001$) and transfer ($M = 2292$; $t_{58} = 8.91$, $p < .0001$) than during the mentally inactive periods ($M = 3408$). To test the sensitivity of the spectral analysis technique with regard to the energy in the midfrequency band, for differences in task complexity, spectral energy during practice and transfer was analyzed using a one-way analysis of variance with repeated measurements on the factor of condition (practice vs transfer). The analysis yielded no significant differences between practice and transfer ($F_{1,59} <$ 1.00).

In conclusion, the spectral analysis of the heart-rate variability is not sensitive to differences in task complexity. However, it might be argued that mentally inactive periods (baseline) can be considered as extremely simple. This argument may result in the conclusion that the technique is only sensitive to relatively large differences in task complexity.

GENERAL DISCUSSION

The rating-scale technique of the perceived amount of mental-effort expenditure proved to be reliable and sensitive in both studies, that is, the differences in type of practice problem and variability in the training conditions, the effects of the training conditions on transfer performance, and two levels of task complexity were reflected in the subjective ratings of mental effort. Further, the rating scale was easy to implement in the classroom. Under the assumption that subjects are able to report accurately how much effort they invested in a task, it may be concluded that this technique is useful for obtaining information on the invested amount of mental effort in instructional research.

With regard to the spectral-analysis technique of the heart-rate variability, it was concluded that its reliability and sensitivity are low. The psychophysiological technique was only sensitive to relatively large differences in invested mental effort, that is, to differences between mentally active periods (instruction and transfer test) and mentally inactive periods (baseline). The technique was insensitive to relatively small differences in invested mental effort. That is, the obtained differences in training and transfer performance and differences in task complexity were not reflected in the spectral energy in the midfrequency band (0.07–0.14 Hz). These results are consistent with a statement of Aasman, et al. (1987) that there is much intrinsic variability in the signal, which is one of the sources of its low reliability and relative insensitivity to small differences in processing load between tasks.

The high intrinsic variability of the heart-rate variability signal may be caused by the fact that it is collected during a predefined period (at least 100 sec.) in which a subject is expected to be mentally active or inactive. However, a period in which a subject is expected to be mentally active may also contain irrelevant cognitive processes, such as moments of inactivity. With the spectral-analysis technique it is impossible to filter the heart-rate

signal to remove this noise. On the other hand, with the rating-scale technique students probably will only consider the relevant cognitive processes, that is, mental activity to deal with a problem, in their reflection on mental effort investment. So, it may be argued that the subjective ratings of mental effort were based on relevant cognitive processes, whereas there is a chance that the psychophysiological measure of mental effort was based on both relevant and irrelevant cognitive processes.

Since specification of the learner's performance levels is central to most instructional research, measures of the primary task would be applied to gather such information. Depending upon the objectives and practical constraints of the study, measurement of mental effort could be employed to provide additional information on cognitive load. From the presented studies it may be concluded that the rating-scale technique is an appropriate alternative for measuring the amount of invested mental effort in instructional research. This technique can represent an important supplement to primary task information, since equivalent levels of performance do not provide information on the cognitive costs at which the performance was attained.

From the results on the spectral-analysis technique of the heart-rate variability one may conclude that this technique is not appropriate for application in instructional research. However, because this study was, as far as is known, the only application of the cardiovascular and rating-scale techniques in instructional research, and because other important properties such as validity and intrusiveness could not be assessed from the present data, more research is needed to investigate the usefulness of both techniques in instructional research.

REFERENCES

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

BRATFISCH, O., BORG, G., & DORNIC, S. (1972) *Perceived item-difficulty in three tests of intellectual performance capacity*. Report 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.

CHANDLER, P., & SWELLER, J. (1992) The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62, 233-246.

EGGEMEIER, F. T. (1988) Properties of workload assessment techniques. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload*. Amsterdam, The Netherlands: North-Holland, Elsevier. Pp. 41-62.

HAMILTON, P. (1979) Process entropy and cognitive control: mental load in internalized thought processes. In N. Moray (Ed.), *Mental workload: its theory and measurement*. New York: Plenum. Pp. 289-298.

MULDER, G. (1980) The heart of mental effort: studies in the cardiovascular psychophysiology of mental work. Unpublished Ph.D. dissertation, Univer. of Groningen, Groningen, The Netherlands.

MULDER, L. J. M. (1988) Assessment of cardiovascular reactivity by means of spectral analysis. Unpublished Ph.D. dissertation, Univer. 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.

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. (1994) Variability of worked examples and transfer of geometrical problem-solving skill: a cognitive-load approach. *Journal of Educational Psychology*, 86, 122-133.

PAAS, F. G. W. C., & VAN MERRIËNBOER, J. J. G. (in press) Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*.

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.

TARMIZI, R. A., & SWELLER, J. (1988) Guidance during mathematical problem solving. *Journal of Educational Psychology*, 80, 424-436.

TULGA, M. K., & SHERIDAN, T. B. (1980) Dynamic decisions and workload in multitask supervisory control. *IEEE Transactions on Systems, Man, and Cybernetics*, 10, 217-232.

VAN DER MEULEN, P., & MULDER, L. J. M. (1990) *CARSPAN (Version 1.2) user's guide*. Groningen, The Netherlands: Univer. of Groningen, Institute for Experimental Psychology.

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.