

Cognitive Load Measurement as a Means to Advance Cognitive Load Theory

Fred Paas

*Educational Technology Expertise Center
Open University of The Netherlands, Heerlen*

Juhani E. Tuovinen

*School of Education
Charles Sturt University, Australia*

Huib Tabbers

*Educational Technology Expertise Center
Open University of the Netherlands, Heerlen*

Pascal W. M. Van Gerven

*Institute of Psychology
Erasmus University Rotterdam, the Netherlands*

In this article, we discuss cognitive load measurement techniques with regard to their contribution to cognitive load theory (CLT). CLT is concerned with the design of instructional methods that efficiently use people's limited cognitive processing capacity to apply acquired knowledge and skills to new situations (i.e., transfer). CLT is based on a cognitive architecture that consists of a limited working memory with partly independent processing units for visual and auditory information, which interacts with an unlimited long-term memory. These structures and functions of human cognitive architecture have been used to design a variety of novel efficient instructional methods. The associated research has shown that measures of cognitive load can reveal important information for CLT that is not necessarily reflected by traditional performance-based measures. Particularly, the combination of performance and cognitive load measures has been identified to constitute a reliable estimate of the mental efficiency of instructional methods. The discussion of previously used cognitive load measurement techniques and their role in the advancement of CLT is followed by a discussion of aspects of CLT that may benefit by measurement of cognitive load. Within the cognitive load framework, we also discuss some promising new techniques.

Cognitive load theory (CLT; Paas & van Merriënboer, 1994a; Sweller, van Merriënboer, & Paas, 1998) is concerned with the development of instructional methods that efficiently use people's limited cognitive processing capacity to stimulate their ability to apply acquired knowledge and skills to new situations (i.e., transfer). CLT is based on a cognitive architec-

ture that consists of a limited working memory, with partly independent processing units for visual/spatial and auditory/verbal information, which interacts with a comparatively unlimited long-term memory. Central to CLT is the notion that working memory architecture and its limitations should be a major consideration when designing instruction. The most important learning processes for developing the ability to transfer acquired knowledge and skills are schema construction and automation. According to CLT, multiple elements of information can be chunked as single elements in cognitive schemas, which can be automated to a

Requests for reprints should be sent to Fred Paas, Educational Technology Expertise Center, Open University of the Netherlands, P.O. Box 2960, 6401 DL Heerlen, The Netherlands. E-mail: fred.paas@ou.nl

large extent. Then, they can bypass working memory during mental processing thereby circumventing the limitations of working memory. Consequently, the prime goals of instruction are the construction and automation of schemas. However, before information can be stored in schematic form in long-term memory, it must be extracted and manipulated in working memory. Work within the cognitive load framework has concentrated on the design of innovative instructional methods that efficiently use the working memory capacity.

These instructional methods have proven successful for children, young adults (for an overview, see Sweller et al., 1998), and older adults (e.g., Paas, Camp, & Rikers, 2001; Van Gerven, Paas, van Merriënboer, & Schmidt, 2002a) in a wide variety of domains, such as physics, mathematics, statistics, computer programming, electrical engineering, paper folding, understanding empirical reports, and learning to use computer programs. Compared with conventional instructional tasks, CLT-based tasks have been found to be more efficient because they require less training time and less mental effort to attain the same or better learning and transfer performance. In this article we argue that the measurement of cognitive load has contributed to the success of CLT and can be considered indispensable to its further development.

It is understood that CLT recognizes the concept of cognitive load as a crucial factor in the learning of complex cognitive tasks. In fact, the instructional control of cognitive load to attain transfer can be considered the essence of the theory. CLT incorporates specific claims concerning the role of cognitive load within an instructional context. That is, cognitive load is not simply considered as a by-product of the learning process but as the major factor that determines the success of an instructional intervention. These theoretical claims require an empirical account of exactly how cognitive load relates to performance. We already know that the amount of working memory resources devoted to a particular task greatly affects how much is learned and the complexity of what is learned. Furthermore, performance has been shown to degrade by either underload or overload. Failures of learning and performing complex cognitive tasks, which are the focus of CLT, can normally be attributed to the task demands that exceed the available cognitive capacity, the inadequate allocation of cognitive resources, or both. The basis for an empirical account lies in the proper definition of the construct cognitive load and the instrumentation for measuring cognitive load. There is a clear need for tools to assess and predict cognitive load. By comparing prior analytical predictions of cognitive load to empirical assessments of cognitive load after an instructional manipulation, the CLT research community can provide tools that enable instructional designers to predict the level of cognitive load in an early design phase.

In this article, the current state of cognitive load measurement in the CLT context is presented. Specifically, the different measurement techniques are described with respect to their contribution to the theory. First, cognitive load and related concepts are defined. Then, the instrumen-

tation for the measurement of cognitive load is explained, and an overview is presented of the different measurement techniques that have been used in cognitive load research. Special attention is given to a computational approach for visualizing the relative mental efficiency of instructional conditions based on mental effort and performance measures. Finally, the role of cognitive load measurement in the advancement of the CLT is discussed.

THE CONSTRUCT OF COGNITIVE LOAD

Cognitive load can be defined as a multidimensional construct representing the load that performing a particular task imposes on the learner's cognitive system (Paas & van Merriënboer, 1994a). According to the general model presented by Paas and van Merriënboer (1994a), the construct has a causal dimension reflecting the interaction between task and learner characteristics and an assessment dimension reflecting the measurable concepts of mental load, mental effort, and performance. Task characteristics that have been identified in CLT research are task format, task complexity, use of multimedia, time pressure, and pacing of instruction. Relevant learner characteristics comprise expertise level, age, and spatial ability. Some interactions that have been found relate to age and task format, indicating that instructions involving goal-specific or goal-free tasks disproportionately compromise or enhance elderly people's learning and transfer performance, respectively (Paas et al., 2001); to expertise level and task format, indicating that the positive effects on learning found for novices disappear or reverse with increasing expertise (see Kalyuga, Ayres, Chandler, & Sweller, 2003); and to spatial ability and use of multimedia, indicating that only high-spatial learners are able to take advantage of contiguous presentation of visual and verbal materials (see Mayer, Bove, Bryman, Mars, & Tapangco, 1996; Mayer & Moreno, 2003).

Mental load is the aspect of cognitive load that originates from the interaction between task and subject characteristics. According to Paas and van Merriënboer's (1994a) model, mental load can be determined on the basis of our current knowledge about task and subject characteristics. As such, it provides an indication of the expected cognitive capacity demands and can be considered an a priori estimate of the cognitive load. *Mental effort* is the aspect of cognitive load that refers to the cognitive capacity that is actually allocated to accommodate the demands imposed by the task; thus, it can be considered to reflect the actual cognitive load. Mental effort is measured while participants are working on a task. *Performance*, also an aspect of cognitive load, can be defined in terms of learner's achievements, such as the number of correct test items, number of errors, and time on task. It can be determined while people are working on a task or thereafter.

According to Paas and van Merriënboer (1993, 1994a), the intensity of effort being expended by learners can be consid-

ered the essence to get a reliable estimate of cognitive load. It is believed that estimates of mental effort may yield important information that is not necessarily reflected in mental load and performance measures. For example, instructional manipulations to change the mental load will only be effective if people are motivated and actually invest mental effort in them. Also, it is quite feasible for two people to attain the same performance levels; one person needs to work laboriously through a very effortful process to arrive at the correct answers, whereas the other person reaches the same answers with a minimum of effort.

Figure 1 presents the attributes of cognitive load and a framework of cognitive load definitions. CLT distinguishes between three types of cognitive load: *intrinsic load*, *extraneous* or *ineffective load*, and *germane* or *effective load*. Intrinsic cognitive load through element interactivity is determined by an interaction between the nature of the material being learned and the expertise of the learners. It cannot be directly influenced by instructional designers. Extraneous cognitive load is the extra load beyond the intrinsic cognitive load resulting from mainly poorly designed instruction, whereas germane cognitive load is the load related to processes that contribute to the construction and automation of schemas. Both extraneous and germane load are under the direct control of instructional designers. The basic assumption is that an instructional design that results

in unused working memory capacity because of low extraneous cognitive load due to appropriate instructional procedures may be further improved by encouraging learners to engage in conscious cognitive processing that is directly relevant to the construction and automation of schemas. Because intrinsic load, extraneous load, and germane load are additive, from a cognitive load perspective, it is important to realize that the total cognitive load associated with an instructional design, or the sum of intrinsic cognitive load, extraneous cognitive load, and germane cognitive load, should stay within working memory limits.

Xie and Salvendy (2000) presented a detailed conceptual framework that is useful for understanding the construct and estimating cognitive load. They distinguished between *instantaneous load*, *peak load*, *accumulated load*, *average load*, and *overall load* (see Figure 1). Instantaneous load represents the dynamics of cognitive load, which fluctuates each moment someone works on a task. Peak load is the maximum value of instantaneous load while working on the task. Accumulated load is the total amount of load that the learner experiences during a task. Mathematically, it can be defined as the integration of instantaneous workload for the time interval that was spent on the task (i.e., the area below the instantaneous load curve). Average load represents the mean intensity of load during the performance of a task. According to Xie and Salvendy (2000), it is the average

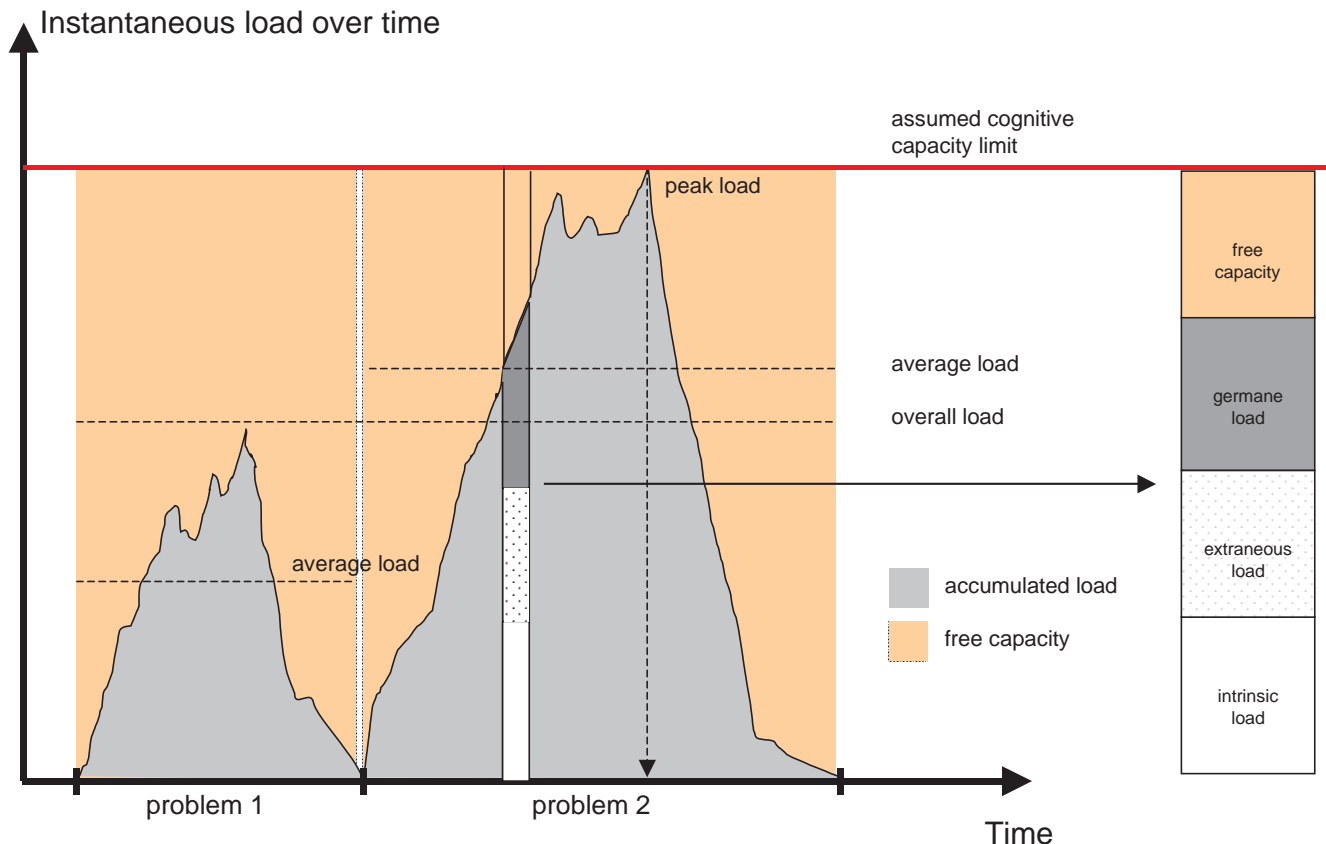


FIGURE 1 Attributes of cognitive load and a framework of cognitive load definitions.

value of instantaneous load and equals the accumulated load per time unit. Finally, overall load is the experienced load based on the whole working procedure or the mapping of instantaneous load or accumulated and average load in the learner's brain. From the CLT perspective, both instantaneous load and its derived measures and overall load can be useful to get a detailed view of the dynamics of the cognitive load within periods of task performance and an overall view of cognitive load across periods of task performance, respectively. Whereas the instantaneous load curve can be considered a visual representation of physiological processes, the overall load measure has a clear psychological basis. The implications of this distinction in terms of cognitive load measurement is discussed in the next section

MEASUREMENT OF COGNITIVE LOAD

The question of how to measure the multidimensional construct of cognitive load has proven difficult for researchers. Following Paas and van Merriënboer's (1994a) model, it is clear that cognitive load can be assessed by measuring mental load, mental effort, and performance. Other researchers have used analytical and empirical methods to classify cognitive load measurement (Linton, Plamondon, & Dick, 1989; Xie & Salvendy, 2000). Analytical methods are directed at estimating the mental load and collect subjective data with techniques such as expert opinion and analytical data with techniques such as mathematical models and task analysis. Empirical methods, which are directed at estimating the mental effort and the performance, gather subjective data using rating scales, performance data using primary and secondary task techniques, and psychophysiological data using psychophysiological techniques. Table 1 shows that whereas empirical techniques for measuring mental effort have received a lot of attention from CLT researchers, analytical techniques have been used only in one study (Sweller, 1988). In particular, rating scale, psychophysiological, and secondary task techniques have been used to determine the cognitive load in cognitive load research.

Rating scale techniques are based on the assumption that people are able to introspect on their cognitive processes and to report the amount of mental effort expended. Although self-ratings may appear questionable, it has been demonstrated that people are quite capable of giving a numerical indication of their perceived mental burden (Gopher & Braune, 1984). Paas (1992) was the first to demonstrate this finding in the context of CLT. With regard to the model presented in Figure 1, most subjective rating techniques use the psychologically oriented concept of overall load. Subjective techniques usually involve a questionnaire comprising one or multiple semantic differential scales on which the participant can indicate the experienced

level of cognitive load. Most subjective measures are multidimensional in that they assess groups of associated variables, such as mental effort, fatigue, and frustration, which are highly correlated (for an overview, see Nygren, 1991). Studies have shown, however, that reliable measures can also be obtained with unidimensional scales (e.g., Paas & van Merriënboer, 1994b). Moreover, it has been demonstrated that such scales are sensitive to relatively small differences in cognitive load and that they are valid, reliable, and unintrusive (e.g., Gimino, 2002; Paas, van Merriënboer, & Adam, 1994).

Physiological techniques are based on the assumption that changes in cognitive functioning are reflected by physiological variables. These techniques include measures of heart activity (e.g., heart rate variability), brain activity (e.g., task-evoked brain potentials), and eye activity (e.g., pupillary dilation, and blink rate). Psychophysiological measures can best be used to visualize the detailed trend and pattern of load (i.e., instantaneous, peak, average, and accumulated load). An example of a physiological method used within the cognitive load framework is presented in Paas and van Merriënboer's (1994b) study. They measured heart-rate variability to estimate the level of cognitive load, and they found this measure to be intrusive, invalid, and insensitive to subtle fluctuations in cognitive load. Unlike heart-rate variability and other physiological measures, the cognitive pupillary response seems a highly sensitive instrument for tracking fluctuating levels of cognitive load. Beatty and Lucero-Wagoner (2000) identified three useful task-evoked pupillary responses (TEPRs): mean pupil dilation, peak dilation, and latency to the peak. These TEPRs typically intensify as a function of cognitive load. In Van Gerven, Paas, van Merriënboer, and Schmidt's (2002b) study, these TEPRs were measured as a function of different levels of cognitive load in both young and old participants. They found that mean pupil dilation is a useful TEPR for measuring cognitive load, especially for young adults.

Task- and performance-based techniques include two subclasses: primary task measurement, which is based on task performance, and secondary task methodology, which is based on the performance of a task that is performed concurrently with the primary task. In this procedure, performance on a secondary task is supposed to reflect the level of cognitive load imposed by a primary task. Generally, the secondary task entails simple activities requiring sustained attention, such as detecting a visual or auditory signal. Typical performance variables are reaction time, accuracy, and error rate. Although secondary task performance is a highly sensitive and reliable technique, it has rarely been applied in research on CLT. The only exceptions can be found in the studies by Brünken, Plass, and Leutner (2003), Chandler and Sweller (1996), Marcus, Cooper, and Sweller (1996), Sweller (1988), and Van Gerven, Paas, van Merriënboer, and Schmidt (2002c). This may relate to an important drawback of secondary task performance: It can interfere considerably with the

TABLE 1
Studies That Measured Cognitive Load and Calculated Mental Efficiency and the Measurement Technique They Used

<i>Studies</i>	<i>Cognitive Load Measurement Technique</i>	<i>Mental Efficiency</i>
Sweller (1988)	PS, ST	
Paas (1992)	RS9	
Paas & van Merriënboer (1993)	RS9	ME
Paas & van Merriënboer (1994b)	RS9, HRV	ME
Cerpa, Chandler, & Sweller (1996)	RS9	ME
Chandler & Sweller (1996)	ST	
Marcus, Cooper, & Sweller (1996)	RS7, ST	ME
Tindall-Ford, Chandler, & Sweller (1997)	RS7	ME
Yeung, Jin, & Sweller (1997)	RS9	ME
de Croock, van Merriënboer, & Paas (1998)	RS9	
Kalyuga, Chandler, & Sweller (1998)	RS7	ME
Kalyuga, Chandler, & Sweller (1999)	RS7	ME
Tuovinen & Sweller (1999)	RS9	ME
Yeung (1999)	RS9	ME
Kalyuga, Chandler, & Sweller (2000)	RS7	ME
Kalyuga, Chandler, & Sweller (2001)	RS7	ME
Kalyuga, Chandler, Tuovinen, & Sweller (2001)	RS9	ME
Mayer & Chandler (2001)	RS7	
Pollock, Chandler, & Sweller (2002)	RS7	ME
Stark, Mandl, Gruber, & Renkl (2002)	RS9	
Tabbers, Martens, & van Merriënboer (2002)	RS9	
Tabbers, Martens, & van Merriënboer (in press)	RS9	
Van Gerven, Paas, van Merriënboer, Hendriks, & Schmidt (2002)	RS9	ME
Van Gerven, Paas, van Merriënboer, & Schmidt (2002a)	RS9	ME
Van Gerven, Paas, van Merriënboer, & Schmidt (2002b)	PR	
Van Gerven, Paas, van Merriënboer, & Schmidt (2002c)	RS9, ST	ME
van Merriënboer, Schuurman, de Croock, & Paas (2002)	RS9	ME

Note. Studies are listed in chronological order. PS = production system; ST = secondary task technique; RS = rating scale (9-point or 7-point scale); ME = mental efficiency; HRV = heart rate variability; PR = pupillary responses.

primary task, especially if the primary task is complex and if cognitive resources are limited, such as in the elderly (Van Gerven, Paas, van Merriënboer, & Schmidt, 2002c; but also see Brünken et al., 2003).

Considering CLT's distinction of intrinsic, extraneous, and germane load, it is important to note that researchers have measured the total cognitive load and have not been able to use one of the measurement techniques to differentiate between these three cognitive load components.

MENTAL EFFICIENCY OF INSTRUCTIONAL CONDITIONS

Although the individual measures of cognitive load can be considered important to determine the power of different instructional conditions, a meaningful interpretation of a certain level of cognitive load can only be given in the context of its associated performance level and vice versa. This was recognized by Paas and van Merriënboer (1993) who developed a computational approach to combine measures of mental effort with measures of the associated primary task performance to compare the mental efficiency of instructional conditions. Since then, a whole range of studies has successfully applied this method or an alternative method combining learning effort and test performance (see Table 1). Paas and van

Merriënboer (1993) argued that the complex relation between mental effort and performance can be used to compare the mental efficiency of instructional conditions in such a way that learners' behavior in a particular instructional condition is considered more efficient if their performance is higher than might be expected on the basis of their invested mental effort or equivalent if their invested mental effort is lower than might be expected on the basis of their performance.

Within the limits of their cognitive capacity, learners can compensate for an increase in mental load (e.g., increasing task complexity) by investing more mental effort, thereby maintaining performance at a constant level. Consequently, the cognitive costs associated with a certain performance level cannot be consistently inferred from performance-based measures. Instead, the combination of measures of mental effort and performance can reveal important information about cognitive load, which is not necessarily reflected by performance and mental load measures alone. Paas and van Merriënboer's (1993) approach provides a tool to relate mental effort to performance measures. In this approach, high-task performance associated with low effort is called *high-instructional efficiency*, whereas low-task performance with high effort is called *low-instructional efficiency*.

In the computational approach the student scores for mental effort and performance are standardized, yielding a *z* score

for mental effort and a z score for performance. Then, an instructional condition efficiency score (E) is computed for each student as the perpendicular distance between a dot (the z score for mental effort and the z score for performance) and the diagonal, $E = 0$, where mental effort and performance are in balance using the following formula. Note that square root 2 in this formula comes from the general formula for the calculation of the distance from a point, $p(x, y)$, to a line, $ax + by + c = 0$. Note

$$E = \frac{z_{\text{Performance}} - z_{\text{Mental Effort}}}{\sqrt{2}}$$

Subsequently, graphics are used to display the information on a Cartesian axis, performance (vertical) and mental effort (horizontal), helping to visualize the combined effects of the two measures. This method is depicted in Figure 2. For example, if the instructional condition efficiency for three groups (e.g., A, B, and C) is compared, the group means of the standardized mental effort and performance are plotted on the performance–effort axis, and the resulting point's distance from the $E = 0$ line gives a measure and direction of the instructional efficiency. In this illustrative example, the highest efficiency occurred in Condition A, in which high performance is obtained with relatively low-mental effort. The lowest efficiency occurred in Condition C, which combined low performance with high-mental effort expenditure. Condition B is an intermediate neutral efficiency condition.

OVERVIEW OF STUDIES

Table 1 shows a detailed overview of all studies related to CLT that have employed measurement of cognitive load. The first attempt was made by Sweller (1988) in his study on cognitive load during problem solving. Using an analytical approach, he developed a production system that modeled the difference in working memory load between means–ends

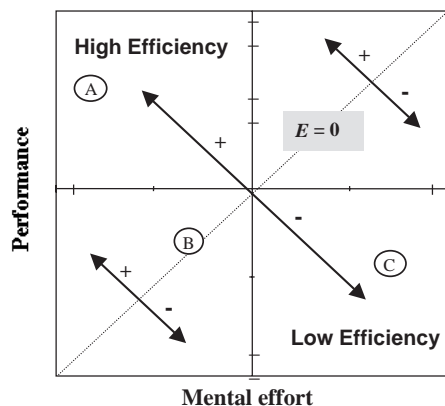


FIGURE 2 Graphic presentation used to visualize instructional efficiency.

analysis and a forward-working strategy, and he found empirical support in a subsequent experiment in which goal-free problem solving led to fewer errors on a secondary task compared with means–ends problem solving. The secondary task consisted of a memory task of the givens and the solution of a previously solved problem. The secondary task technique has also been used in studies by Brünken et al. (2003), Chandler and Sweller (1996; memory task of letters), Marcus et al. (1996; simple tone detection task and memory task of numbers), and Van Gerven, Paas, van Merriënboer, and Schmidt (2002c; simple visual signal-detection task). Whereas the secondary task technique for estimating relative differences in working memory load has only been applied in a few studies, most studies in cognitive load research have used a subjective rating scale. This scale was originally developed by Paas (1992), who based it on a measure of perceived task difficulty devised by Borg, Bratfisch, and Dornic (1971). In Paas's (1992) study, participants had to report their invested mental effort on a symmetrical scale ranging from 1 (*very, very low mental effort*) to 9 (*very, very high mental effort*) after each problem during training and testing. The scale's reliability and sensitivity (Paas et al., 1994) and moreover its ease of use have made this scale, and variants of it, the most widespread measure of working memory load within CLT research. A final category of measures that has been used in cognitive load research concerns physiological measures, such as heart-rate variability (Paas & van Merriënboer, 1994b) and pupil dilation (Van Gerven, Paas, van Merriënboer, & Schmidt, 2002b).

An interesting observation is that even though researchers are continuously trying to find or develop physiological and secondary task measures of cognitive load, subjective workload measurement techniques using rating scales remain popular, because they are easy to use; do not interfere with primary task performance; are inexpensive; can detect small variations in workload (i.e., sensitivity); are reliable; and provide decent convergent, construct, and discriminate validity (Gimino, 2002; Paas et al., 1994). However, we must stress that the internal consistency of these measures requires further studies.

DISCUSSION

It is clear that cognitive load measurement is of general theoretical interest to cognitive psychologists and relevant to applied problems in instructional technology, particularly for more complex tasks. But, how exactly has the measurement of cognitive load advanced CLT? The main contribution is of course that the measurement of cognitive load has given an empirical basis for the hypothetical effects of instructional interventions on cognitive load. First of all, differences in the intrinsic instructional load were reflected in measures of experienced memory load in experiments by Marcus et al. (1996), Van Gerven, Paas, van Merriënboer, and Schmidt

(2002c), and Pollock, Chandler, and Sweller (2002). Moreover, worked-out examples and completion problems have shown to be superior to conventional problems, not only in terms of test performance but also in terms of extraneous load (Paas, 1992; Paas & van Merriënboer, 1994b; Sweller, 1988; Van Gerven, Paas, van Merriënboer, Hendriks, & Schmidt, 2002; van Merriënboer, Schuurman, de Crook, & Paas, 2002). Studies on the split-attention effect and the modality effect have also demonstrated the difference in working memory load between integrated and nonintegrated presentation formats (Cerpa, Chandler, & Sweller, 1996; Chandler & Sweller, 1996; Kalyuga, Chandler, & Sweller, 1999; Tabbers, Martens, & van Merriënboer, 2002, in press; Tindall-Ford, Chandler, & Sweller, 1997; Van Gerven, Paas, van Merriënboer, & Schmidt, 2002c; Yeung, 1999). Furthermore, differential effects on memory load have been found in studies on expertise (Kalyuga, Chandler, & Sweller, 1998, 2000, 2001; Pollock et al., 2002; Tuovinen & Sweller, 1999) and on age differences (Van Gerven, Paas, van Merriënboer, Hendriks, et al., 2002; Van Gerven, Paas, van Merriënboer, & Schmidt, 2002a, 2002b, 2002c). Finally, in one of their experiments, van Merriënboer et al. (2002) found some empirical support for an increase in germane load. So, although some other studies failed to find hypothesized differences in working memory load (de Crook, van Merriënboer, & Paas, 1998; Kalyuga, Chandler, Tuovinen, & Sweller, 2001; Tabbers et al., 2002, Experiment 2; Yeung, Jin, & Sweller, 1997), the majority of the results support the main hypothesis of CLT: Differences in effectiveness between instructional formats are largely based on differences in memory load.

Another major advantage of cognitive load measures is that they enable us to calculate relative mental efficiency of instructional conditions (Paas & van Merriënboer, 1993). As we have seen, these scores combine the effort invested during learning or test performance with the performance on a test. Thus, instructional formats can be compared not only in terms of their effectiveness but also in terms of their efficiency. A number of experiments have demonstrated the added value of the efficiency measure by showing that differences in effectiveness are not always identical to differences in efficiency (Kalyuga et al., 1998, Experiment 1; 2001, Experiment 1; Kalyuga, Chandler, Tuovinen, et al., 2001, Experiment 1; Marcus et al., 1996; Paas & van Merriënboer, 1994b; Pollock et al., 2002, Experiment 1; Van Gerven, Paas, van Merriënboer, Hendriks, et al., 2002; Van Gerven, Paas, van Merriënboer, & Schmidt, 2002a, 2002c; van Merriënboer et al., 2002). In these cases, the calculation of mental efficiency scores has enriched our knowledge on the effect that different instructional formats have on the learning process.

Although the cognitive load rating scale has been found to be very useful, a point of concern seems to be justified. The reliability in terms of internal consistency (Cronbach's alpha) and the sensitivity to differences in task complexity of the original 9-point scale with labels relating to mental effort were found to be very good in some studies (Paas, 1992; Paas

& van Merriënboer, 1994b). Note that many of the subsequent studies using rating scales have referred to these reliability and sensitivity scores, but they have used different scales with fewer categories and different labels relating to task difficulty. The reliability and sensitivity data for these adapted scales were not reported. We believe that the psychometrical properties of each adapted scale need to be re-established.

Paas and van Merriënboer's (1993, 1994b) original calculation of relative instructional efficiency was based on the combination of test performance and mental effort invested to attain this test performance. Subsequent studies have combined mental effort spent during training with test performance to calculate efficiency. Whereas the original approach reflects the mental efficiency of test performance, which may be related to transfer, the latter approach may reflect the mental efficiency of training or learning efficiency. Although both approaches are important, an approach combining the three aspects of learning effort, test effort, and test performance incorporates the advantages of both approaches (see Tuovinen & Paas, 2002) and provides a more sensitive measure for comparing instructional conditions. When the effort during learning and the outcome performance measures are combined, the performance effort is neglected. However, two learners may reach the same performance after equal learning effort, but they still may require different amounts of effort to produce that performance standard. It appears that the student achieving the performance with less effort had learned the content better; thus, it is desirable to take both learning and performance effort into account when computing an instructional measure in conjunction with a performance measure.

A factor that has been neglected in the measurement of cognitive load and the calculation of mental efficiency is time on task. It is not clear whether participants took the time spent on the task into account when they rated cognitive load. That means we do not know whether a rating of 5 on a 9-point scale for a participant who worked for 5 min on a task is the same as a rating of 5 for a participant who worked for 10 min on a task. If the rating was based on the overall cognitive load, representing the experienced load based on the whole working procedure including the time spent, then there would be no need to consider time on task together with mental effort and performance. Otherwise, an efficiency measure incorporating mental effort, performance, and the time factor would be helpful.

An interesting topic for future research is related to CLT's heavy reliance on the distinction between intrinsic, extraneous, and germane load and the current inability to obtain differentiated knowledge about these load components. To determine the utility of CLT for generating instructional designs, it is important that the specification of what is intrinsic, extraneous, and germane can be done a priori. We argue that only the combination of analytical and empirical measures will enable us to determine intrinsic, extraneous, and germane cognitive load. If we can analytically identify the inter-

active elements of a task, the aspects of the task that interfere with schema construction and automation, and the aspects of the task that are beneficial to these processes, and if we can determine their cognitive consequences empirically, then it will be possible to determine empirically intrinsic, extraneous, and germane load, respectively.

In addition, future research could establish the usefulness of CLT for new psychological approaches to cognitive load measurement like subjective time estimation (e.g., Fink & Neubauer, 2001) and the opportunities provided by new neuroimaging techniques such as positron-emission tomography and functional magnetic resonance imaging. These techniques have been used to dissociate the verbal, spatial, and central executive units of working memory (e.g., Smith, Jonides, & Koeppe, 1996) and to identify the physiological capacity constraints in working memory (e.g., Callicot et al., 1999). Although the techniques currently are much too intrusive to be used in instructional research, they might enable us to determine the neural substrates underlying different types of cognitive load.

Finally, an interesting line for future research is to examine the possibility to use the combined mental effort and performance measures in intelligent interactive learning systems to construct personalized training. The knowledge that individual differences resulting from interactions between task and learner characteristics are important determinants for the level of cognitive load leads to the assumption that optimized efficiency can ultimately only be achieved when the assignment of learning tasks suits the learner's needs and capabilities. One possible way to achieve this goal is to monitor the learning process and use the cognitive state of the learner to select the appropriate learning tasks. Camp, Paas, Rikers, and van Merriënboer's (2001) study on personalized dynamic problem selection based on performance, mental effort, or mental efficiency can be considered a first step in this direction.

REFERENCES

- Beatty, J., Lucer-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). Cambridge, MA: Cambridge University Press.
- Borg, G., Bratfisch, O., & Dornic, S. (1971). On the problem of perceived difficulty. *Scandinavian Journal of Psychology*, 12, 249–260.
- Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38, 53–61.
- Callicot, J. H., Mattay, V. S., Bertolini, A., Finn, K., Coppola, R., Frank, J. A., et al. (1999). Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cerebral Cortex*, 9, 20–26.
- Camp, G., Paas, F., Rikers, R., & van Merriënboer J. J. G. (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort and mental efficiency. *Computers in Human Behavior*, 17, 575–595.
- Cerpa, N., Chandler, P., & Sweller, J. (1996). Some conditions under which integrated computer-based training software can facilitate learning. *Journal of Educational Computing Research*, 15, 345–367.
- Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*, 10, 151–170.
- de Croock, M. B. M., van Merriënboer, J. J. G., & Paas, F. (1998). High versus low contextual interference in simulation-based training of troubleshooting skills: Effects on transfer performance and invested mental effort. *Computers in Human Behavior*, 14, 249–267.
- Fink, A., & Neubauer, A. C. (2001). Speed of information processing, psychometric intelligence, and time estimation as an index of cognitive load. *Personality & Individual Differences*, 30, 1009–1021.
- Gimino, A. (2002, April). *Students' investment of mental effort*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors*, 26, 519–532.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38, 23–31.
- Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors*, 40, 1–17.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13, 351–371.
- Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, 92, 126–136.
- Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology*, 21, 5–23.
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, 93, 579–588.
- Linton, P. M., Plamondon, B. D., & Dick, A. O. (1989). Operator workload for military system acquisition. In G. R. McMillan, D. Beevis, E. Salas, M. H. Strub, R. Sutton, & L. van Breda (Eds.), *Applications of human performance models to system design* (pp. 21–46). New York: Plenum.
- Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology*, 88, 49–63.
- Mayer, R., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology*, 88, 64–73.
- Mayer, R., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology*, 93, 390–397.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38, 43–52.
- Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, 33, 17–33.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84, 429–434.
- Paas, F., Camp, G., & Rikers, R. (2001). Instructional compensation for age-related cognitive declines: Effects of goal specificity in maze learning. *Journal of Educational Psychology*, 93, 181–186.
- Paas, F., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, 35, 737–743.
- Paas, F., & van Merriënboer, J. J. G. (1994a). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6, 51–71.
- Paas, F., & van Merriënboer, J. J. G. (1994b). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive-load approach. *Journal of Educational Psychology*, 86, 122–133.
- Paas, F., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, 79, 419–430.

- Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, 12, 61–86.
- Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex*, 6, 11–20.
- Stark, R., Mandl, H., Gruber, H., & Renkl, A. (2002). Conditions and effects of example generation. *Learning and Instruction*, 12, 39–60.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296.
- Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (2002). *A closer look at the modality effect in multimedia learning*. Manuscript submitted for publication.
- Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (in press). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*.
- Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied*, 3, 257–287.
- Tuovinen, J., & Paas, F. (2002). *Measurement and computation of 2 and 3 factor instructional condition efficiency*. Manuscript in preparation.
- Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology*, 91, 334–341.
- Van Gerven, P. W. M., Paas, F., van Merriënboer, J. J. G., Hendriks, M., & Schmidt, H. G. (2002). *The efficiency of multimedia training into old age*. Manuscript submitted for publication.
- Van Gerven, P. W. M., Paas, F., van Merriënboer, J. J. G., & Schmidt, H. G. (2002a). Cognitive load theory and aging: Effects of worked examples on training efficiency. *Learning and Instruction*, 12, 87–105.
- Van Gerven, P. W. M., Paas, F., van Merriënboer, J. J. G., & Schmidt, H. G. (2002b). *Memory load and task-evoked pupillary responses in aging*. Manuscript submitted for publication.
- Van Gerven, P. W. M., Paas, F., van Merriënboer, J. J. G., & Schmidt, H. G. (2002c). *Modality and variability as factors in training the elderly*. Manuscript submitted for publication.
- van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance and training. *Learning and Instruction*, 12, 11–39.
- Xie, B., & Salvendy, G. (2000). Prediction of mental workload in single and multiple task environments. *International Journal of Cognitive Ergonomics*, 4, 213–242.
- Yeung, A. S. (1999). Cognitive load and learner expertise: Split-attention and redundancy effects in reading comprehension tasks with vocabulary definitions. *The Journal of Experimental Education*, 67, 197–217.
- Yeung, A. S., Jin, P., & Sweller, J. (1997). Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology*, 23, 1–21.