Cognitive load theory, educational research, and instructional design: some food for thought

Ton de Jong

Abstract Cognitive load is a theoretical notion with an increasingly central role in the educational research literature. The basic idea of cognitive load theory is that cognitive capacity in working memory is limited, so that if a learning task requires too much capacity, learning will be hampered. The recommended remedy is to design instructional systems that optimize the use of working memory capacity and avoid cognitive overload. Cognitive load theory has advanced educational research considerably and has been used to explain a large set of experimental findings. This article sets out to explore the open questions and the boundaries of cognitive load theory by identifying a number of problematic conceptual, methodological and application-related issues. It concludes by presenting a research agenda for future studies of cognitive load.

Keyword Cognitive load theory

Introduction

This article discusses *cognitive load theory*, a theory relating *working memory characteristics* and the design of *instructional systems*. Theories of the architecture of human memory make a distinction between long term memory and short term memory. Long term memory is that part of memory where *large* amounts of information are stored *(semi-) permanently*, whereas short term memory is the memory system where *small* amounts of information are stored (Cowan 2001; Miller 1956) for a *very short duration* (Dosher 2003). The original term "short term memory" has since been replaced by "working memory" to emphasize that this component of memory is responsible for the processing of information. More recent and advanced theories distinguish two subsystems within working memory: one for holding visuospatial information (e.g., written text, diagrams) and one for phonological information (e.g., narration; Baddeley and Hitch 1974). The significance of

T. de Jong (✉)
Faculty of Behavioral Sciences, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
e-mail: a.j.m.dejong@utwente.nl

working memory capacity for cognitive functioning is evident. A series of studies have found that individual working memory performance correlates with cognitive abilities and academic achievement (for an overview see Yuan et al. 2006).

The characteristics of working memory have informed the design of artifacts for human functioning such as car operation (Forlines et al. 2005) and the processing of financial data (Rose et al. 2004). Another area in which the implications of working memory characteristics are studied is instructional design. This research field finds its roots in work by Sweller and colleagues in the late 1980s and early 1990s (Chandler and Sweller 1991; Sweller 1988, 1989; Sweller et al. 1990). Their cognitive load theory has subsequently had a great impact on researchers and designers in the field of education. Basically, cognitive load theory asserts that learning is hampered when working memory capacity is exceeded in a learning task. Cognitive load theory distinguishes three different types of contributions to total cognitive load. *Intrinsic cognitive load* relates to inherent characteristics of the content to be learned, *extraneous cognitive load* is the load that is caused by the instructional material used to present the content, and finally, *germane cognitive load* refers to the load imposed by learning processes.

Cognitive load theory has recently been the subject of criticism regarding its conceptual clarity (Schnotz and Kürschner 2007) and methodological approaches (Gerjets et al. 2009a). The current article follows this line of thinking. It explores the boundaries of cognitive load theory by presenting a number of questions concerning its foundations, by discussing a number of methodological issues, and by examining the consequences for instructional design. The core of the literature reviewed consists of the 35 most frequently cited articles, with "cognitive load" as a descriptor, taken from the Web of Science areas "educational psychology" and "education and educational research". (The virtual H-Index for "cognitive load" in these categories was 35.) These have been supplemented with a selection of recent articles on cognitive load from major journals. The current article concludes by suggesting a role for cognitive load theory in educational theory, research, and design.

The three types of cognitive load revisited

Intrinsic cognitive load

Intrinsic cognitive load relates to the difficulty of the subject matter (Cooper 1998; Sweller and Chandler 1994). More specifically, material that contains a large number of interactive elements is regarded as more difficult than material with a smaller number of elements and/or with a low interactivity. Low interactivity material consists of single, simple, elements that can be learned in isolation, whereas in high interactivity material individual elements can only be well understood in relation to other elements (Sweller 1994; Sweller et al. 1998). Pollock et al. (2002) give the example of a vocabulary where individual words can be learned independently of each other as an instance of low interactivity material, and grammatical syntax or the functioning of an electrical circuit as examples of high interactivity material. This implies that some (high interactivity) content by its nature consumes more of the available cognitive resources than other (low interactivity) material. However, intrinsic load is not merely a function of qualities of the subject matter but also of the prior knowledge a learner brings to the task (Bannert 2002; Sweller et al. 1998). An important premise regarding intrinsic load is that it cannot be changed by instructional treatments.

The first premise is that intrinsic load (which can be expressed as experienced difficulty of the subject matter) depends on the number of domain elements and their interactivity. This, however, is not the complete story and intrinsic load also seems to depend on other characteristics of the material. First, some types of content seem to be intrinsically more difficult than others, despite having the same number of elements and the same interactivity. Klahr and Robinson (1981), for example, found that children had considerably more difficulty solving Tower of Hanoi problems in which they had to end up with all pegs occupied compared to the traditional Tower of Hanoi problem where all disks must end up on one peg. This effect occurred despite the number of moves being the same in both problems. Second, learning difficulty increases independently of element interactivity when learners must change ontological categories. Chi (1992) gives a number of examples of learners who attribute the wrong ontological category to a concept (e.g., see "force" as a material substance) and find it very difficult to change to the correct ontological category. Chi (2005) adds to this that certain concepts (e.g., equilibrium) have emergent ontological characteristics, making them even more difficult for students to understand. Training students in ontological categories helps to improve their learning (Slotta and Chi 2006). Third, specific characteristics of relations are also seen as being related to difficulty. Campbell (1988) mentions in an overview study aspects of the material that contribute to difficulty such as "negative" relations between elements (this is conflicting information) and ambiguity or uncertainty of relations.

The second premise concerning intrinsic cognitive load is that it cannot be changed by instructional treatments. Ayres (2006b), for example, describes intrinsic load as "fixed and innate to the task…" (p. 489). As a consequence, intrinsic load is unaffected by external influences. "It [intrinsic load] cannot be directly influenced by instructional designers although… it certainly needs to be considered" (Sweller et al. 1998, p. 262). Or as Paas et al. (2003a, p. 1) state: "Different materials differ in their levels of element interactivity and thus intrinsic cognitive load, and they cannot be altered by instructional manipulations…". Also in more recent work it explicitly stated that intrinsic load cannot be changed (e.g., Hasler et al. 2007; Wouters et al. in press). A somewhat different stance is taken in other work. Sequencing the material in a simple-to-complex order so that learners do not experience its full complexity at the outset is a way to control intrinsic load (van Merriënboer et al. 2003). Pollock et al. (2002) introduced a similar approach in which isolated elements were introduced before the integrated task (see also, Ayres 2006a). Gerjets et al. (2004) also argue that instructional approaches can change intrinsic load. Along with the simple-to-complex sequencing, they suggest part-whole sequencing (partial tasks are separately trained) and "modular presentation of solution procedures". In the latter approach, learners are confronted with the complete task but without references to 'molar' concepts such as problem categories. Van Merriënboer et al. (2006) also suggest reducing intrinsic load by a whole-part approach in which the material is presented in its full complexity from the start, but learners' attention is focused on subsets of interacting elements. Van Merriënboer and Sweller (2005) see these approaches as all fully in line with cognitive load theory, because both simple-to-complex and whole-part approaches start with few elements and gradually build up complexity.

Intrinsic load as defined within cognitive load theory is an interesting concept that helps explain why some types of material are more difficult than others and how this may influence the load on memory. However, the analysis above also shows that difficulty (and thus memory load) is not determined solely by number and interaction of elements and that there are techniques that may help to control intrinsic load.

Extraneous cognitive load

Extraneous cognitive load is cognitive load that is evoked by the instructional material *and* that does not directly contribute to learning (schema construction). As van Merriënboer and Sweller (2005, p. 150) write: "Extranous cognitive load, in contrast, is load that is not necessary for learning (i.e., schema construction and automation) and that can be altered by instructional interventions". Following this definition, extraneous load is imposed by the material but could have been avoided with a different design. A number of general sources of extraneous cognitive load are mentioned in the literature. The "split-attention" effect, for instance, refers to the separate presentation of domain elements that require simultaneous processing. In this case, learners must keep one domain element in memory while searching for another element in order to relate it to the first. Split attention may refer to spatially separated elements (as in visual presentations) or temporal separation of two elements, as in multi-media presentations (Ayres and Sweller 2005; Lowe 1999). This can be remedied by presenting material in an integrated way (e.g., Cerpa et al. 1996; Chandler and Sweller 1992; Sweller and Chandler 1991). A second identified source of extraneous cognitive load is when students must solve problems for which they have no schema-based knowledge; in general, this refers to conventional practice problems (Sweller 1993). In this situation students may use means-ends-analysis as a solution procedure (Paas and van Merriënboer 1994a). Though this is an effective way of solving problems, it also requires keeping many elements (start goal, end goal, intermediate goals, operators) in working memory. To remedy this, students can be offered "goal free problems", "worked out problems" or "completion problems" instead of traditional problems (Atkinson et al. 2000; Ayres 1993; Renkl et al. 2009, 1998; Rourke and Sweller 2009; Sweller et al. 1998; Wirth et al. 2009). A third source of extraneous load may arise when the instructional design uses only one of the subsystems of working memory. More capacity can be used when both the visual and auditory parts of working memory are addressed. This "modality principle" implies that material is more efficiently presented as a combination of visual and auditory material (see amongst others Low and Sweller 2005; Sweller et al. 1998; Tindall-Ford et al. 1997). A fourth source of unnecessary load occurs when learners must coordinate materials having the same information. Cognitive resources can be freed by including just one of the two (or more) sources of information. This is called the "redundancy principle" (Craig et al. 2002; Diao and Sweller 2007; Sweller 2005; Sweller et al. 1998).

The majority of studies in the traditional cognitive load approach have focused on extraneous cognitive load, with the aim of reducing this type of load (see for example Mayer and Chandler 2001; Mayer and Moreno 2002, 2003). This all seems quite logical; learners should not spend time and resources on processes that are not relevant for learning. However, there are still a few issues that merit attention. First, designs that seem to elicit extraneous processes may, at the same time, stimulate germane processes. For example, when there are two representations with essentially the same information, such as a graph and a formula, these can be considered as providing redundant information leading to higher extraneous load (Mayer et al. 2001); but relating the two representations and making translations from one representation to the other (including processes of abstraction) can equally well be regarded as a process of acquiring deep knowledge (Ainsworth 2006). Reducing extraneous load in these cases may therefore also remove the affordances for germane processes. Paas et al. (2004, pp. 3–4) provide an example of this when they write: "In some learning environments, extraneous load can be inextricably bound with germane load. Consequently, the goal to reduce extraneous load and increase germane load may pose problems for instructional designers. For instance, in nonlinear hypertext-based

learning environments, efforts to reduce high extraneous load by using linear formats may at the same time reduce germane cognitive load by disrupting the example comparison and elaboration processes." Second, there seem to be limits to the reduction of extraneous load; an interesting question is whether extraneous load can be zero. Can environments be designed that have no extraneous cognitive load? In practice there are limits to what can be designed. Representations should be integrated to reduce the effects of split attention, but in many cases this would lead to "unreadable" representations. Having information in different places is then preferable to cluttering one place with all the information. Third, like intrinsic load, extraneous load is not independent of the prior experience of the learner. A learner's awareness of specific conventions governing the construction of learning material assists with processing and thus reduces extraneous cognitive load. Some researchers even use a practice phase to "… avoid any extraneous load caused by an unfamiliar interface" (Schnotz and Rasch 2005, p. 50). Fourth, as recent research results indicate, some characteristics of instructional material that have always been regarded as extraneous may not hinder learning, if the material is well designed. Mayer and Johnson (2008), for example, found that redundant information is advantageous as long as it is short and placed near the information to which it refers.

Eliminating characteristics of learning material that are not necessary for learning will help students to focus on the learning processes that matter. This has been one of the important lessons for instructional design from cognitive load theory. However, the above analysis also shows that it is not always evident which characteristics of material can be regarded as being extraneous.

Germane cognitive load

Cognitive load theory sees the construction and subsequent automation of schemas as the main goal of learning (see e.g., Sweller et al. 1998). The construction of schemas involves processes such as interpreting, exemplifying, classifying, inferring, differentiating, and organizing (Mayer 2002). The load that is imposed by these processes is denominated germane cognitive load. Instructional designs should, of course, try to stimulate and guide students to engage in schema construction and automation and in this way increase germane cognitive load. Sweller et al. (1998) give the example of presenting practice problems under a high variability schedule (different surface stories) as compared to a low variability schedule (same surface stories). Sweller et al. (1998) describe the finding that students under a high variability schedule report higher cognitive load and also achieve better scores on transfer tests. They explain this by stating that the higher cognitive load must have been germane in this case. Apart from the fact that this is a "post-hoc" explanation there seem to be no grounds for asserting that processes that lead to (correct) schema acquisition will impose a higher cognitive load than learning processes that do not lead to (correct) schemas. It could even be argued that poorly performed germane processes lead to a higher cognitive load than smoothly performed germane processes. The fragility of Sweller et al.'s conclusion is further illustrated by work in which the effect mentioned by Sweller et al. did not occur. In a study in which learners had to acquire troubleshooting skills in a distillery system, participants in a condition in which practice problems were offered under a high variability schedule showed better performance on transfer problems than a low variability group, but, as measured with a subjective rating scale, did not show higher cognitive load during learning (de Croock et al. 1998).

An interesting question concerning germane load is whether germane load can be too high. Cognitive load theory focuses on the "bad" effects of intrinsic and extraneous load,

but since memory capacity is limited, even "good" processes may overload working memory. What would happen, for example, if an inexperienced learner were asked to abstract over a number of concrete cases? Could this task cause a memory overload for this learner? Some of the cognitive load articles suggest that this could indeed be the case. Van Merriënboer et al. (2002), for example, write that they intend to "increase, within the limits of total available cognitive capacity, germane cognitive load which is directly relevant to learning" (p. 12). This suggests both that too much germane load can be invoked, and also that the location of the borderline might be known. In another study, Stull and Mayer (2007) found that students who used author-generated graphic organizers outperformed students who created these organizers themselves. Stull and Mayer concluded that students in the self-generated organizer condition obviously had experienced more extraneous load (again as a post-hoc explanation), but it also might have been the case that these students had to perform germane processes that were too demanding. Kalyuga (2007, p. 527) tried to solve this conceptual issue in the following way: "If specific techniques for engaging learners into additional cognitive activities designed to enhance germane cognitive load (e.g., explicitly self-explaining or imagining content of worked examples) cause overall cognitive load to exceed learner working memory limitations, the germane load could effectively become a form of extraneous load and inhibit learning." This quote suggests that germane load can only be "good" and that when it taxes working memory capacity too much, it should be regarded as extraneous load. This, however, seems to be at odds with the accepted definition of extraneous load that was presented earlier.

Conceptual issues

Can the different types of cognitive load be distinguished?

The general assertion of cognitive load theory that the capacity of working memory is limited is not disputed. What may be questioned is another pillar upholding the theory, namely, that a clear distinction can be made between intrinsic, extraneous, and germane cognitive load. This section discusses differences between the different types of load at a conceptual level; a later section discusses whether the different forms of load can be distinguished empirically.

The first distinction to be discussed is that between intrinsic and germane load. The essential problem here is that these concepts are defined as being of a different (onto-logical) character. Intrinsic load refers to "objects" (the material itself) and germane load refers to "processes" (what goes on in learning). The first question is how intrinsic load, defined as a characteristic of the material, can contribute to the cognitive load of the learner. A contribution to cognitive load cannot come from the material as such, but can only take place when the learner starts processing the material. Without any action on the side of the learner, there cannot be cognitive load. This means that cognitive load only starts to exist when the learner relates elements, makes abstractions, creates short cuts, etc. Mayer and Moreno (2003), for example, use the term "essential processing" to denote processes of selecting words, selecting images, organizing words, organizing images, and integrating. Although Mayer and Moreno (2003) associate "essential processing" with germane load, the processes involved seem to be close to making sense of the material itself without relating it to prior knowledge and without yet engaging in schema con-struction, and as such to be characteristic of intrinsic load. A slightly alternative way to look upon the contribution of material characteristics to cognitive load would be to see

these characteristics as mediating towards germane processes, meaning that some germane processes are harder to realize for some material than for others, or that different germane processes may be necessary for different types of material. But that means that intrinsic load is then defined in terms of opportunities for germane processing. If intrinsic load and germane load are defined in terms of relatively similar learning processes, the difference between the two seems to be very much a matter of degree, and possibly non-existent. In this respect it is interesting to note that the very early publications on cognitive load do not distinguish between intrinsic and germane load but only between extraneous load and load exerted by learning processes (Chandler and Sweller 1991; Sweller 1988).

The second distinction to be addressed is that between extraneous and germane load. Paas et al. (2004, p. 3) present the following distinction: "If the load is imposed by mental activities that interfere with the construction or automation of schemas, that is, ineffective or extraneous load, then it will have negative effects on learning. If the load is imposed by relevant mental activities, i.e., effective or germane load, then it will have positive effects on learning." The first issue apparent in this definition is its quite tautological character (what is good is good, what is bad is bad). Second, this definition seems to extend the definition of extraneous load as originating from unnecessary processes that are a result of poorly designed instructional formats (which was the original definition; see above) to all processes that do not lead to schema construction (for a critical comment see also Schnotz and Kürschner 2007). Does this extension, therefore, broaden the idea of extraneous load related to mental activities that could be avoided but occupy mental space to include wrongly performed "germane" activities as well? Or, stated differently, do instructional designs that lead to "mistakes" in students' schemas (but still aim at schema construction processes) contribute to extraneous load? The difference between extraneous and germane load is additionally problematic because, as was the case with intrinsic load, this distinction seems to depend on the expertise level of the learner. Paas et al. (2004, pp. 2–3) write: "A cognitive load that is germane for a novice may be extraneous for an expert. In other words, information that is relevant to the process of schema construction for a beginning learner may hinder this process for a more advanced learner". This notion is related to the so-called "expertise-reversal effect", which means that what is good for a novice might be detrimental for an expert. In a similar vein, Kalyuga (2007, p. 515) states: "Similarly, instructional methods for enhancing levels of germane load may produce cognitive overload for less experienced learners, thus effectively converting germane load for experts into extraneous load for novice learners." Taken together, this implies that it is not the nature of the processes that counts but rather the way they function. A further problem is that germane processes can be considered to be extraneous depending on the learning goal (see also, Schnotz and Kürschner 2007). Gerjets and Scheiter (2003), for example, report on a study in which students in a surface-emphasizing approach condition outperformed students who had followed a structure-emphasizing approach in solving isomorphic problems. The group with the structure-emphasizing approach showed longer learning times. This is interpreted by the authors as time devoted to schema construction by use of abstraction processes. This, however, was not very useful for solving isomorphic problems and should thus be judged as extraneous in this case, according to Gerjets and Scheiter. In a similar vein, Scott and Schwartz (2007) studied the function of navigational maps for learning from hypertexts and found that depending on the use of the maps for understanding or navigation, the cognitive load could be regarded as either germane or extraneous.

The distinction between intrinsic and extraneous load may also be troublesome. Intrinsic load is defined as the load stemming from the material itself; therefore intrinsic load in principle cannot be changed if the underlying material stays the same. But papers

by influential authors present different views. For example, in a recent study, DeLeeuw and Mayer (2008) manipulated intrinsic load by changing complexity of sentences in a text while leaving the content the same. This manipulation seems to affect extraneous load more than intrinsic load.

Can the different types of cognitive load simply be added?

An important premise of cognitive load theory is that the different types of cognitive load can be added. Sweller et al. (1998, p. 263) are very clear on the additivity of intrinsic and extraneous load: "Intrinsic cognitive load due to element interactivity and extraneous cognitive load due to instructional design are additive". Also in later work, it is very clearly stated that the different types of load are additive (Paas et al. 2003a). There are, however, indications that the total load experienced cannot simply be regarded as the sum of the three different types of load. This is, for example illustrated in the relation between intrinsic and germane load, where there are two different possible interpretations. If intrinsic and germane load are seen as members of two distinct ontological categories ("material" and "cognitive processes", respectively), there are principled objections to adding the two together. However, if they are regarded as members of the same ontological category (namely "cognitive processes") the two may interact. For instance, if learners engage in processes to understand the material as such (related to intrinsic load) this will help to perform germane processes as well.

Is there a difference between load and effort?

Merriam-Webster's on-line dictionary defines effort as "the total work done to achieve a particular end" and load as "the demand on the operating resources of a system". Clearly effort and load are related concepts; one important difference is that effort is a unit that is exercised by a system itself whereas load is a factor that is experienced by a system. The terms effort and load are often used as synonyms in cognitive load theory, but sometimes a distinction is being made.

Paas (1992, p. 429) presents an early attempt to differentiate the two concepts taking the above distinction into account: "Cognitive load is a multidimensional concept in which two components—mental load and mental effort—can be distinguished. Mental load is imposed by instructional parameters (e.g., task structure, sequence of information), and mental effort refers to the amount of capacity that is allocated to the instructional demands." Paas introduces here a distinction between cognitive load, mental load, and mental effort. In discussing the concept of cognitive load, Paas and van Merriënboer (1994a) distinguish causal factors, those that influence cognitive load (including the task, learner characteristics, and interactions between these two), and assessment factors. For the assessment factors, mental load is again seen as a component of cognitive load. Along with this, they distinguish mental effort and task performance as other components. According to Paas and van Merriënboer, mental load is related to the task, and mental effort and performance reflect all three causal factors. This means that mental load is determined by the task only, and that mental effort and performance are determined by the task, subject characteristics and their interactions. Paas and van Merriënboer are very clear with respect to the (non-existent) role of subject characteristics with regard to mental load: "With regard to the assessment factors, *mental load* is imposed by the task or environmental demands. This task-centered dimension, considered independent of subject characteristics,

is constant for a given task in a given environment" (Paas and van Merriënboer 1994a, p. 354). However, Paas et al. (2003b, p. 64) write: "*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." So, they present a different view of the role of participant characteristics in relation to mental load even when quoting the original paper.

The role of performance is often interpreted differently as well. In the paper by Paas and van Merriënboer, performance is regarded as one of the three assessments (along with mental load and mental effort) that is a reflection of (among other things) cognitive load. However, Kirschner (2002) gives a somewhat different place to performance when he writes: "Mental load is the portion of CL that is imposed exclusively by the task and environmental demands. Mental effort refers to the cognitive capacity actually allocated to the task. The subject's performance, finally, is a reflection of mental load, mental effort, and the aforementioned causal factors." (p. 4). According to Kirschner, these causal factors can be "characteristics of the subject (e.g., cognitive abilities), the task (e.g., task complexity), the environment (e.g., noise), and their mutual relations" (p. 3). In Kirschner's definition, performance is not one of the three measures reflecting (an aspect of) cognitive load, but it is determined by the other two assessment factors.

The implications of the definitions of load and effort from Merriam-Webster's on-line dictionary are very straightforward: load is something experienced, whereas effort is something exerted. Following this approach, one might say that intrinsic and extraneous cognitive load concern cognitive activities that must unavoidably be performed, so they fall under cognitive load; germane cognitive load is the space that is left over that the learner can decide how to use, so this can be labelled as cognitive effort. This seems to be the tenor of the message of Sweller et al. (1998, p. 266) when they write: "Mental load refers to the load that is imposed by task (environmental) demands. These demands may pertain to task-intrinsic aspects, such as element interactivity, which are relatively immune to instructional manipulations and to task-extraneous aspects associated with instructional design. Mental effort refers to the amount of cognitive capacity or resources that is actually allocated to accommodate the task demands." Kirschner's (2002) definition given above also suggests that the term "load" can be reserved for intrinsic and extraneous load, whereas "effort" is more associated with germane load. However, in this same paper, Kirschner (2002, p. 4) uses the terms again as synonyms when load is defined in terms of effort "… extraneous CL is the effort required to process poorly designed instruction, whereas germane CL is the effort that contributes, as stated to the construction of schemas". Another complication is that some authors state that processing extraneous characteristics is under learners' control. Gerjets and Scheiter (2003) report on a study in which a strong reduction of study time did not impair learning, whereas cognitive load theory would predict a drop in performance. They explain this by assuming that learners make strategic decisions under time pressure and that they may decide to increase germane load and ignore distracting information, that is, to decrease extraneous load.

A related conceptual problem is that it is not clear whether or not cognitive load is defined as relative to capacity in cognitive load theory. Working memory is assumed to consist of two distinct parts (visual and phonological). If both parts are used the capacity of working memory increases compared to the use of only one. As Sweller et al. (1998, pp. 281–282) write: "Although less than purely additive, there seems to be an appreciable increase in capacity available by the use of both, rather than a single, processor." It is not clear, however, if cognitive load is seen as relative to the total capacity or not. If the

material is presented in two modalities instead of one, will load decrease (if taken as relative to capacity) or will load stay the same (in that basically the same material is offered and only the capacity has changed)?

The above discussion shows that cognitive load is not only a complex concept but is also often not well defined. Although there are some indications given, none of the studies makes it very clear how mental load and mental effort relate to the central concepts of intrinsic, extraneous, and germane cognitive load. This terminological uncertainty has direct consequences for how studies on influencing cognitive load are designed as well as for how cognitive load is measured.

Research issues

The measurement of cognitive load

In many studies there is no direct measurement of cognitive load; the level of cognitive load is induced from results on knowledge post-tests. It is then argued that when results on knowledge post-test are low(er) cognitive load (obviously) has been (too) high(er) (see for example, Mayer et al. 2005). Recently DeLeeuw and Mayer (2008, p. 225) made this more explicit by stating that "one way to examine differences in germane processing… is to compare students who score high on a subsequent test of problem-solving transfer with those who score low". Similarly, Stull and Mayer (2007, p. 808) stated: "Although we do not have direct measures of generative and extraneous processing during learning in these studies, we use transfer test performance as an indirect measure. In short, higher transfer test performance is an indication of less extraneous processing and more generative processing during learning". Of course, there is much uncertainty in this type of reasoning and therefore many authors have expressed the need for a direct measurement of cognitive load. Mayer et al. (2002, p. 180), for example, state: "Admittedly, our argument for cognitive load would have been more compelling if we had included direct measures of cognitive load…". Three different groups of techniques are used to measure cognitive load: self-ratings through questionnaires, physiological measures (e.g., heart rate variability, fMRI), and secondary tasks (for an overview see, Paas et al. 2003b).

Measuring cognitive load through self-reporting

One of the most frequently used methods for measuring cognitive load is self-reporting, as becomes clear from the overview by Paas et al. (2003b). The most frequently used self-report scale in educational science was introduced by Paas (1992). This questionnaire consists of one item in which learners indicate their "perceived amount of mental effort" on a 9-point rating scale (Paas 1992, p. 430). In research that uses this measure, reported effort is seen as an index of cognitive load (see also, Paas et al. 1994, p. 420). Though used very frequently, questionnaires based on the work by Paas (1992) have no standard format. Differences are seen in the number of units used for the scale(s), the labels used as scale ends, and the timing (during the learning process or after).

Following Paas' original work, a scale with nine points is used most frequently (example studies are Kester et al. 2006a; Paas 1992; Paas et al. 2007; Paas and van Merriënboer 1993; van Gerven et al. 2002) but also a 7-point scale is used quite often (see e.g., Kablan and Erden 2008; Kalyuga et al. 1999; Moreno and Valdez 2005; Ngu et al. 2009; Pollock et al. 2002). Still others use a 5-point scale (Camp et al. 2001; Huk 2006;

Salden et al. 2004), a 10-point scale (Moreno 2004), a 100-point scale (Gerjets et al. 2006), or a continuous (electronic) scale with or without numerical values (de Jong et al. 1999; Swaak and de Jong 2001; van Gerven et al. 2002).

Questionnaires also differ in what are used as anchor terms at the extremes of the scale(s). In the original questionnaire by Paas, participants were asked for their mental effort, with the extremes on the scale being "very, very low mental effort" and "very very high mental effort". Moreno and Valdez (2005, p. 37) asked learners "How difficult was it for you to learn about the process of lightning". Ayres (2006a) also used a scale measuring difficulty and having the extremes "extremely easy" and "extremely difficult", but still called this mental effort. This was also the case in a study by Yeung et al. (1997), who used the extremes "very, very easy" and "very very difficult". Pollock et al. (2002) combined difficulty and understanding in one question by asking students "How easy or difficult was it to learn and understand the electrical tests from the instructions you were given?" (p. 68). In some studies the two issues (effort and difficulty) are queried separately and are then combined into one metric (Moreno 2007; Zheng et al. 2009). Still others (e.g., Moreno 2004) combine different aspects by asking participants "how helpful and difficult (mental effort)" the program was (p. 104). Clearly, questions differ in asking for "effort", "difficulty" and related concepts and differ in whether they ask for these concepts related to the material, the learning process, or the resulting knowledge (understanding). An empirical distinction between these questions is reported by Kester et al. (2006a), who used a 9-point scale to gauge the mental effort experienced during learning, but asked separately for the mental effort required to understand the subject matter. While their results showed an effect of interventions on experienced mental effort during learning, no effect was found on mental effort for understanding. In their conclusion on the use of efficiency measures in cognitive load research, van Gog and Paas (2008, p. 23) even state that "… the outcomes of the effort and difficulty questions in the efficiency formula are completely opposite". This means that the outcomes of studies may differ significantly depending on the specifics of the question asked.

The timing of the questionnaire also differs across studies. Most studies present the questionnaire only after learning has taken place (for example, Ayres 2006a; Hasler et al. 2007; Kalyuga et al. 1999; Pociask and Morrison 2008; Tindall-Ford et al. 1997), whereas other studies repeat the same questionnaire several times during (and sometimes also after) the learning process (for example, Kester et al. 2006a; Paas et al. 2007; Stark et al. 2002; Tabbers et al. 2004; van Gog et al. 2008; van Merriënboer et al. 2002). The more often cognitive load is measured the more accurate the view of the actual cognitive load is, certainly if it is assumed that cognitive load may vary during the learning process (Paas et al. 2003b). It is also not clear if learners are able to estimate an average themselves, which implies that for the researcher to measure cognitive load a few times and calculate an average may result in a different value than letting the learner decide on an average cognitive load by posing one question at the end. It is also questionable whether an average load over the whole process is the type of measure that does justice to principles of cognitive load theory, because it does not measure (instantaneous) cognitive overload (see also later).

Along with, and probably related to, diversity in use of the one-item questionnaire is inconsistency in the outcomes of studies that use this questionnaire. First, the obtained values have no agreed upon meaning and second, findings using the questionnaire are not very well connected to theoretical predictions. A problematic issue with the one-item questionnaire is that values on the scales can be interpreted differently between studies. Levels of effort or difficulty reported as beneficial in one study are associated with the

poorest scoring condition in other work. Pollock et al. (2002, Experiment 1), for example, report cognitive load scores of around 3 (43% on a scale of 7) for their "best" condition, whereas Kablan and Erden (2008) found a value for cognitive load of 2.5 (36% on a scale of 7) for their "poorest" (separated information) condition. The highest performing group for Pollock et al. thus reports a higher cognitive load than the poorest performing group for Kablan and Erden (and in both cases a low cognitive load score was seen as being profitable for learning). The literature shows a wide variety of scores on the one-item questionnaire, with a specific score sometimes associated with "good" and sometimes with "poor" performance. Therefore, it seems as if there is no consistency in what can be called a high (let alone a too high) cognitive load score or a low score. Nonetheless, some authors take the scores as a kind of absolute measure. For example, Rikers, in a critical overview of a number of cognitive load studies, writes when discussing a study by Seufert and Brünken (2006): "As a result of this complexity, intrinsic cognitive load was increased to such an extent that hyperlinks could only be used superficially. This explanation is very plausible and substantiated by the students' subjective cognitive load that ranged between $M = 4.38$ (63%) and 5.63 (80%)" (2006, p. 361). Apart from the fact that the different types of load were not measured separately in this study and that therefore there is no reason to assume that the scores were determined mainly by intrinsic load, it is also not clear why some scores are seen as high and others as low.

Taking the variations above into account, it may not be surprising to find that the relation of cognitive load measures to instructional formats is not very consistent. In their overview of cognitive load measures, Paas et al. (2003b) claim that measures of cognitive load are related to differences in instructional formats in the majority of studies. This conclusion seems not to be completely valid. Not all of the studies they mention have shown such results. For example, in their overview Paas et al. (2003b, p. 69) state that in the original study by Paas (1992) it was found that worked-out examples and completion problems were superior in terms of extraneous load, but *extraneous* load during learning was *not* measured in this study; beyond that, *no* differences between conditions were found on invested mental effort *during* the learning process. As Paas (1992) himself writes: "Apparently the processes required to work during specific instruction demanded the same amount of mental effort in all conditions,…" (p. 433). And despite the fact that many studies have indeed found differences in cognitive load, it is also interesting to consider studies in which differences in resulting performance were found without detecting any associated differences in cognitive load, such as in the original study by Paas (1992), Stark et al. (2002), Tabbers et al. (2004), Lusk and Atkinson (2007), Hummel et al. (2006), de Westelinck et al. (2005), Clarebout and Elen (2007), Seufert et al. (2007, study 2), Fischer et al. (2008), Beers et al. (2008), Wouters et al. (2009), Amadieu et al. (2009), Kester and Kirschner (2009), and de Koning et al. (in press). Conversely, there are also studies in which differences in experienced cognitive load were found but not differences on performance tests (Paas et al. 2007; Seufert et al. 2007). Often these unexpected outcomes are explained by pointing to external factors (e.g., motivation), but the validity of the test itself is never questioned. Seufert et al. (2007) found different effects on cognitive load as measured with a one-item questionnaire in different studies with the same set-up. Their speculative interpretation is that in one study participants interpreted the question in terms of extraneous cognitive load, whereas in the other study participants seemed to have used an overall impression of the load.

The overview above shows that there are many variations in how the technique of self-reporting is applied, that there are questions about what is really measured, and that results cannot always be interpreted unequivocally. Paas et al. (2003b, p. 66), however, claim that

the uni-dimensional scale is reliable, sensitive to small differences in cognitive load, and valid (and non-intrusive). The claims by Paas et al. (2003b) are based on work by Paas et al. (1994) who presented an analysis of the reliability and sensitivity of the subjective rating method. Their data, again, come from two other studies, namely Paas (1992) and Paas and van Merriënboer (1994b). Students in these two studies had to solve problems, both in a training and in a test phase. The subjective rating of perceived mental effort was performed after solving each problem. The studies had several conditions but only the problems that were the same in all conditions (separately for both studies) were used in the analysis. Concerning *reliability*, Paas et al. report Cronbach's α's of .90 and .82 for two different studies. This tells us that there is a high correlation between the effort reported for the different problems. However, that would characterize cognitive load, measured as mental effort, as a student bound characteristic, while according to the theory, mental effort is influenced by the task, participants' characteristics, and the interaction between those two factors (Paas et al. 1994). The interaction component in particular could mean that some problems elicit more effort from some students and other problems elicit more effort from other students. In that case the theory would not predict a high reliability. Concerning *sensitivity*, Paas et al. (1994) claim that the one-item questionnaire is sensitive to experimental interventions. Their supporting evidence comes from the Paas (1992) study in which several expected differences were found in the subjective ratings of test items based on the experimental conditions. However, they fail to say that, in that same study, no differences on the practice problems were found between conditions, despite the fact that these differences were expected to occur. Above, quite a number of other studies were also reported in which no differences were found on the subjective scale administered although they had been expected. Of course, those differences could be non-existent, but it could also mean that the measure is not as sensitive as is stated. In addition, in the Paas (1992) study, the questionnaire was given after each problem, while most of the later studies used only one question after the entire learning session. The latter procedure may also negatively influence the sensitivity of the measure. Paas et al. (2003) also claim that the Paas et al. (1994) study determined the *validity* of the construct as measured by the subjective questionnaire, but no data can be found on this in the quoted work. There were no comparisons made with other measures to find any of the central aspects of validity (concurrent, discriminant, predictive, convergent, and criterion validity). Some articles (e.g., Kalyuga et al. 2001a, b; Lusk and Atkinson 2007) claim that older work by Moray (1982) validated the subjective measure of cognitive load with objective measures. However, Moray only looked at the concept of "perceived difficulty" within the domain of cognitive tasks and, in fact, Moray warned that there had been no work relating (subjective) difficulty and mental workload.

Physiological measures as indications for cognitive load

A second set of measures of cognitive load are physiological measures. One of these measures is heart rate variability. Paas and van Merriënboer (1994b) classified this measure as invalid and insensitive (and quite intrusive); in addition, Nickel and Nachreiner (2003) concluded that heart rate variability could be used as an indicator for time pressure or emotional strain, but not for mental workload. Recent work, however, gives more support to this measure (Lahtinen et al. 2007) or to related measures that combine heart rate and blood pressure (Fredericks et al. 2005). Pupillary reactions are also used and are regarded as sensitive for cognitive load variations (Paas et al. 2003b; van Gog et al. 2009), but have only been used in a limited number of studies (e.g., Schultheis and Jameson 2004). There

are, however, indications that the sensitivity of this measure to workload changes diminishes with age of the participants (van Gerven et al. 2004). Cognitive load can also be assessed by using neuro-imaging techniques. Though this is a promising field (see e.g., Gerlic and Jausovec 1999; Murata 2005; Smith and Jonides 1997; Volke et al. 1999) there are still questions on how to relate brain activations precisely to cognitive load. Overall, physiological measures show promise for assessing cognitive load but (still) have the disadvantage of being quite intrusive.

Dual tasks for estimating cognitive load

A third way of measuring cognitive load is through the dual-task or secondary-task approach (Brünken et al. 2003). In this approach, borrowed from psychological research on tasks such as car driving (Verwey and Veltman 1996), a secondary task is introduced alongside the main task (in this case learning). A good secondary task could be a simple monitoring task such as reacting as quickly as possible to a color change (Brünken et al. 2003) or a change of the screen background color (DeLeeuw and Mayer 2008), or detecting a simple auditory stimulus (Brünken et al. 2004) or a visual stimulus (a colored letter on the screen) (Cierniak et al. 2009). A more complicated secondary task could be remembering and encoding single letters when cued (Chandler and Sweller 1996) or seven letters (Ayres 2001). The principle of this approach is that slower or more inaccurate performance on the secondary task indicates more consumption of cognitive resources by the primary task. A dual task approach has an advantage over, for example, a questionnaire, in that it is a concurrent measure of cognitive load as it occurs. Though Brünken et al. (2003) have empirically shown the suitability of the dual task approach for measuring cognitive load, this technique is used relatively rarely (for exceptions see, Ayres 2001; Brünken et al. 2002; Chandler and Sweller 1996; Ignacio Madrid et al. 2009; Marcus et al. 1996; Renkl et al. 2003; Sweller 1988).

Overall problems with cognitive load measures

There are three main problems remaining with cognitive load measures. The first is that cognitive load measures are always presented as relative. The second is that an overall rating of cognitive load, as is often utilized, does not provide much help with interpreting results in terms of cognitive load theory, because the contributions to learning of the different kinds of cognitive load are different. And the third is that the most frequently used measures are not sensitive to variations over time.

The basic principle of cognitive load is that learning is hindered when cognitive overload occurs, in other words when working memory capacity is exceeded (see e.g., Khalil et al. 2005). However, the measures that are discussed here cannot be used to measure overload; the critical level indicating overload is unknown. Instead, many studies compare the level of cognitive load for different learning conditions and conclude that when a certain condition has a lower reported load, this condition is the one best suited for learning (e.g., van Merriënboer et al. 2002). However, what should count in cognitive load theory is not the relative level of effort or difficulty reported, but the absolute level. Overload occurring in one condition is independent of what happens in another condition. In fact, all conditions in a study could show overload.

Studies that measure only one overall concept of cognitive load do not do justice to its multidimensional character (Ayres 2006b). Though more popular in other science domains, the use of multi-dimensional rating scales for workload such as the NASA-TLX is

exceptional within educational science (for examples see, Fischer et al. 2008; Gerjets et al. 2006; Kester et al. 2006b). When cognitive load is measured as one concept, there is no distinction between intrinsic load, extraneous load, and germane load. This means that if one instructional intervention shows a lower level of cognitive load than another, this may be due to a lower level of extraneous load, a lower level of germane load or both (assuming that the intrinsic load is constant over interventions). Attributing better performance on a post-test to a reduction of extraneous load, as most of these studies do, is a post-hoc explanation without direct evidence, despite the measurement of cognitive load. This issue is clearly phrased by Wouters et al. (in press), who found differences between conditions in performance but no differences in mental effort, and who explain this in the following way:

> The mental effort measure that was used did not differentiate between mental effort due to the perceived difficulty of the subject matter, the presentation of the instructional material or engaging in relevant learning activities. It is possible that the effects in the conditions with or without illusion of control on mental effort have neutralized each other. In other words, the illusion of control conditions may have imposed rather high extraneous cognitive load and low germane cognitive load, whereas the no illusion of control may have imposed rather low extraneous cognitive load and high germane cognitive load.

And in a related study these same authors (Wouters et al. 2009, p. 7) write:

The mental effort measure used did not differentiate between mental effort due to perceived difficulty of the subject matter, presentation of the instructional material, or being engaged in relevant learning activities. It is possible that the effect on the perceived mental effort of the varying design guidelines, that is, modality and reflection prompts, have neutralized each other.

There are developments, mostly recent, that now aim to measure the different types of cognitive load separately. In de Jong et al. (1999), participants were asked, through an on-line pop-up questionnaire, to use sliders to answer three questions about (1) the perceived difficulty of the subject matter, (2) the perceived difficulty of interacting with the environment itself, and (3) the perceived helpfulness of the instructional measures (the functionality of the tools) that they had used. Gerjets et al. (2006, see pp. 110–111) asked participants three questions concerning cognitive load: 'task demands' (how much mental and physical activity was required to accomplish the learning task, e.g., thinking, deciding, calculating, remembering, looking, searching, etc.), 'effort' (how hard the participant had to work to understand the contents of the learning environment), and 'navigational demands' (how much effort the participant had to invest to navigate the learning environment). These represented, according to Gerjets et al. intrinsic, germane, and extraneous load, respectively. In another recent study, Corbalan et al. (2008) used two measures for assessing cognitive load. Participants had to rate their "effort to perform the task" (p. 742) on a 7-point scale and in addition, after each task, participants were also asked to indicate, on a one-item 7-point scale, their "effort invested in gaining understanding of the relationships dealt with in the simulator and the task" (p. 744). Corbalan et al. call the first a measure of task load and the second a measure of germane load. Cierniak et al. (2009) asked students three different questions. The intrinsic load question was "How difficult was the learning content for you?". The extraneous load question was: "How difficult was it for you to learn with the material?". The question asking for germane load was: "How much did you concentrate during learning?". Gerjets et al. (2009b) used five questions to measure cognitive load: one item for intrinsic load, three for extraneous load, and one for germane load. Overall, results from these studies give inconsistent results, raising doubts whether students can themselves distinguish between different types of load.

A third issue is that measuring cognitive load is rarely time-related. Paas et al. (2003b), citing Xie and Salvendy (2000), distinguish between different types of load in relation to time: instantaneous load, peak load, accumulated load, average load, and overall load. Instantaneous load is the load at a specific point in time, and peak load is the maximum experienced instantaneous load. Average load is the mean instantaneous load experienced during a specified part of the task, and overall load refers to average load over the whole period of learning. In addition, there is accumulated load, which is the total load experienced. Paas et al. (2003b) state that both instantaneous load and overall load are important for cognitive load theory, but since the theory only refers to addition of types of load and to overload; only instantaneous load should be considered for these purposes. According to Paas et al. (2003b) physiological measures are suited for capturing instantaneous load because they can monitor the load constantly, and overall load is often measured with questionnaire based measures. However, physiological measures are very rarely used in cognitive load research, whereas questionnaire measures of overall load prevail. A related problematic issue is that it is not clear if learners who use a questionnaire have overall load or accumulated load in mind when answering the question. In this respect, Paas et al. (2003b, p. 69) state: "It is not clear whether participants took the time spent on the task into account when they rated cognitive load. That means it is unknown 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." These considerations show that the time aspect is not well considered when measuring cognitive load.

Cognitive load and learning efficiency measures

Many articles on cognitive load present an instructional efficiency measure that expresses in a metric the (relative) efficiency of an instructional approach. This measure was introduced by Paas and van Merriënboer (1993), who used the formula $E = |R - P| / \sqrt{2}$ where P is the standardized performance score (on a post-test) and R is the standardized mental effort score (as measured in the test phase). Paas and van Merriënboer explain the formula by stating that if $R - P < 0$, then E is positive, and if $R - P > 0$, then E is negative. (Many later authors have used the more simple expression $E = (P - R)/\sqrt{2}$.) In this formula the mental effort is the effort reported by participants in relation to the post-test. Therefore, the metric basically expresses the level of automatization of the students' resulting knowledge. For this reason it expresses the (relative) *effectiveness* of the instructional intervention, rather than, as Pass and van Merriënboer state, its *efficiency*. In part for similar reasons, Tuovinen and Paas (2004) state that this measure is more related to transfer. The measure, as introduced by Paas and van Merriënboer, has led to some confusion in the literature, at both a more technical and a conceptual level. As an example of the first type of confusion, Tindall-Ford et al. (1997), write the formula as $E = (M - P)/\sqrt{2}$ (M = mental effort; P = performance), thus reversing the original formula in their presentation. These authors write "E can be either positive or negative. The greater the value of E, the more efficient the instructional condition" (p. 272) which would mean that higher mental effort and lower performance implies better efficiency. A second example involves Marcus et al. (1996), who use the original formula by Paas and van Merriënboer in which the absolute value of $M - P$ was taken and who nonetheless state that this value can be negative or positive. A third example can be found in the article by Kalyuga et al. (1998, p. 7) who display the formula as $E = (P - R) \div |\sqrt{2}|$, and thus erroneously take the absolute value of $\sqrt{2}$.

Along with these more technical issues, there have also been conceptual misunder-
standings regarding the efficiency measure. As indicated above, the original measure is
actually a measure of the quality of the resulting knowledge and not of the instructional
condition. This has led many authors to use mental effort related to learning as the mental
effort factor in the equation in order to express a quality of learning. A recent overview by
van Gog and Paas (2008) shows indeed that the majority of studies have used a cognitive
load rating associated with the learning phase; very few studies have used a cognitive load
rating associated with the testing phase. Van Gog and Paas conclude that the original
measure (using cognitive load during the test phase) is more suitable when instructional
measures aim to increase germane load, and the measure using cognitive load in the learning
phase is more suited for situations in which the intention is to reduce extraneous load. This
does not seem to be justified. Cognitive load theory focuses on what happens during
knowledge acquisition, so that measuring cognitive load during learning is essential.
Measuring cognitive load during the performance after learning may add interesting
information on the quality of the participants' knowledge, and in this respect it is valuable
information. That would be the case in any study that concerns learning and performance,
not just studies within the cognitive load tradition. But there are more conceptual difficulties
with the measure. A low performance associated with a low cognitive load would result in a
similar efficiency value as a high performance with a high cognitive load, which leads to
unclear experimental situations. Moreno and Valdez (2005, p. 37), for example, studying
the difference between students who were presented with a set of frames in the right order
(the NI group) and students who had to order those frames themselves (the I group) write:

> However, predictions about differences in the relative instructional efficiency of I
> and NI conditions are unclear. The predicted combination of relatively higher per-
> formance and higher cognitive load for Group I and relatively lower performance
> and lower cognitive load for Group NI may lead to equivalent intermediate efficiency
> conditions.

Studies from cognitive load theory sometimes use plain cognitive load measures to
compare instructional conditions (e.g., de Koning et al. in press) and sometimes use
efficiency measures (e.g., Paas et al. 2007) even when, as in these two examples, the same
phenomenon is studied (animations) on the same topic (the cardiovascular system). Since it
is known that the two measures may lead to different results, shifting between measures
does not assist in drawing overall interpretations of experimental outcomes. Cognitive load
theory research would certainly benefit from standardization in this respect.

Individual differences

Individual differences may influence the outcomes of studies conducted within the cog-
nitive load tradition. One such aspect that is taken up by cognitive load research and
frequently reported nowadays is the fact that instructional treatments that should reduce
extraneous load work differently for individuals with low versus high expertise. This is
called the expertise-reversal effect (Kalyuga 2007; Kalyuga et al. 2003). This phenomenon
implies that instructional designs that follow cognitive load recommendations are only
beneficial for "… learners with very limited experience" (Kalyuga et al. 2003, p. 23).
Examples of studies in which this effect was found are Yeung et al. (1997), Kalyuga et al.
(1998), Tuovinen and Sweller (1999), Cooper et al. (2001), Clarke et al. (2005), and Ayres
(2006a). The idea of the expertise reversal effect is that intrinsic load decreases with
increasing expertise. This means that treatments that reduce extraneous load are more

effective for less knowledgeable students than for more experienced students and experts. It also means that instruction can gradually be adapted to a learner's developing expertise. Renkl and Atkinson (2003), for example, state that intrinsic load can be lower in later stages of knowledge acquisition so that "fading" can take place. For instance, a transition in later stages from worked-out problems to conventional problems might be superior for learning. An extensive discussion of the expertise reversal effect and an overview of related studies is given in Kalyuga (2007).

Another individual characteristic that has been found to interact with effects of cognitive load related instructional treatments is spatial ability. Several studies report differential effects for high and low spatial ability students. For example, Mayer and Sims (1994) performed a study in which participants had to learn about the functioning of a bicycle tire pump. One group saw an animation with concurrent narration explaining the working of the pump; another group saw the animation before the narration. The concurrent group outperformed the successive group overall, as could be expected on the basis of the contiguity (or split attention) effect, but this effect, although strong for high spatial ability students, was not prominent for those with low spatial ability. Others studies that report differential cognitive load effects in relation to spatial ability are Plass et al. (2003), Moreno and Mayer (1999b), and Huk (2006).

An individual characteristic that is by nature very much related to cognitive load theory is working memory capacity. It is clear from many studies that next to intra-individual differences in working memory capacity (see e.g., Sandberg et al. 1996) there are also inter-individual differences in working memory capacity (see e.g., Miyake 2001). Different tests for measuring working memory capacity exist (for an overview see, Yuan et al. 2006). The most well known of these tasks is the operation span task (OSPAN) introduced by Turner and Engle (1989); different operationalizations of this test exist (see e.g., Pardo-Vazquez and Fernandez-Rey 2008). Several research fields use working memory capacity tests, and have found relations between scores on these test and performance. Examples include studies of reading ability (de Jonge and de Jong 1996) and mathematical reasoning (Passolunghi and Siegel 2004; Wilson and Swanson 2001). But the influence of working memory capacity in learning is also repeatedly found (see e.g., Dutke and Rinck 2006; Perlow et al. 1997; Reber and Kotovsky 1997). In cognitive load research, however, working memory capacity is hardly ever measured. Exceptions can be found in work by van Gerven et al. (2002, 2004), who used a computation span test to control for working memory capacity between experimental groups. Another recent example is a study by Lusk et al. (2009), who measured participants' individual working memory capacity with the OSPAN test and assessed the effects of segmentation of multimedia material in relation to working memory capacity. They found that students with high working memory capacity recalled more than students with low working memory capacity and generated more valid interpretations of the material presented. They also found a positive effect of segmentation on post-test scores for both recall and interpretation, and an interaction in which lack of segmentation was especially detrimental for students with a low working memory capacity. A further very recent example is a study by Seufert et al. (2009), who measured working memory capacity with a "numerical updating test" in which participants had to memorize additions and subtractions they had to perform in an evolving matrix of numbers. This study addressed the modality effect, which appeared to hold only for low working memory capacity learners, while combining audio and visual information was less beneficial for learners with a high working memory capacity than presenting just visual information. Finally, Berends and van Lieshout (2009) measured participants' memory capacity with a digit span test and found that adding illustrations to algebra word problems decreased

students' performance overall, but this effect was less obvious for students with a higher working memory capacity.

Studies in the cognitive load field predominantly use a between-subjects design. This means, however, that cognitive load differences between conditions are also influenced by "… individual differences, such as abilities, interest, or prior knowledge…" (Brünken et al. 2003, p. 57). A good recommendation, based also on the results above, is to include these individual characteristics as control variables in experimental set-ups so that differences between groups cannot be attributed to differences in these relevant student characteristics. Working memory capacity in particular should be included as a measure even in studies at an individual level. If we accept the definition of load presented above as "the demand on the operating resources of a system", it is not enough to know the demand (which could mean the effort applied); the system's capacity must also be known in order to determine (over)load.

The external validity of research results

Cognitive load theory is used in presenting guidelines for instructional design. Overviews of instructional designs based on cognitive load theory are given, for example, by Sweller et al. (1998), Mayer (2001), and Mayer and Moreno (2003). This presupposes that results from cognitive load research are applicable in realistic teaching and learning situations. The next few sections explore some questions concerning the external validity of cognitive load research.

When does "overload" occur in realistic situations?

Two generally accepted premises of cognitive load theory are that overload may occur and that overload is harmful for learning. Overload means that *at some point in time* the requested memory capacity is higher than what is available (the so-called instantaneous load). In realistic learning situations, however, learners have means to avoid instantaneous overload. First, when there is no time pressure persons who perform a task can perform different parts of it sequentially, thus avoiding overload at a particular moment. Second, in realistic situations learners will use devices (e.g., a notepad) to offload memory. People rarely perform a complex calculation mentally without noting down intermediate results. This is hardly ever allowed in cognitive load research. A study by Bodemer and Faust (2006) found that when students were asked to integrate separated pieces of information, learning was better when learners could do this physically, compared to when they had to do this mentally. And this is indeed what a normal learner in a normal learning situation would do. Cognitive load theory *does* apply to learning highly demanding, complex, time critical tasks, such as flying a fighter plane, for which tasks learners must use all available resources to make the right decisions in a very short time and they cannot swap, replay, or make annotations. Nearly all of the studies in the cognitive load tradition are using tasks where the limited capacity of working memory is not actually at stake. However, these studies are often designed in such a way as to prevent swapping or off-loading, thus creating a situation that is artificially time-critical.

Participants and study time

A substantial portion of the research on cognitive load theory includes participants who have no specific interest in learning the domain involved and who are also given a very

short study time. Mayer and Johnson (2008), for example, gave students a PowerPoint slideshow of about 2.5 min. In Moreno and Valdez (2005) learners had 3 min to study a sequence of multimedia frames. In Hasler et al. (2007) participants saw a system-paced video of almost 4 min. In Stull and Mayer (2007), psychology students studied a text about a biology topic for 5 min. DeLeeuw and Mayer (2008) presented participants with a 6 min long multimedia lesson. In many other studies participants worked a little longer, but the time period was still too short for a realistic learning task. Also, in this field of research, it is common for the research subjects to be psychology students who are given the task of learning material chosen by the researcher (see, e.g., Ignacio Madrid et al. 2009; Moreno 2004). Research with short study times and with students who have no direct engagement with the domain may very well be used to test the basic cognitive mechanisms of cognitive load theory but raise problems when these results are translated into practical recommendations. This has been acknowledged within cognitive load research itself by van Merriënboer and Ayres (2005) who recommend studying students who are working on realistic tasks and using realistic study times. A recent example of this type of work can be found in Kissane et al. (2008).

Study conditions

There are other aspects of the experimental conditions set up by researchers that undermine the relevance of study outcomes to educational practice. We can take as an example one of the central phenomena in cognitive load theory, the *modality principle* (Mayer 1997, 2001). This principle says that if information (graphs, diagrams, or animations) is presented over different modalities (visual and auditory) more effective learning will result than when the information is only presented visually. The modality principle has been observed in a large set of studies, as can be gathered from several overviews on this principle (Ginns 2005; Mayer 2001; Moreno 2006). A few recent studies, however, could not find support for the modality principle (Clarebout and Elen 2007; Dunsworth and Atkinson 2007; Elen et al. 2008; Opfermann et al. 2005; Tabbers et al. 2004; Wouters et al. 2009). The main difference between the studies in which the modality effect could be demonstrated and the ones that failed to find evidence for the effect seems to be the amount of learner control. In studies in which the modality effect was found (e.g., Kalyuga et al. 1999, 2000; Mayer and Moreno 1998; Moreno and Mayer 1999a; Mousavi et al. 1995) information presentation was system-paced and generally very rapid, so that even the information presented in the complete visual condition could hardly be processed. A relevant quote in this respect comes from a paper by Moreno and Mayer (2007) who discuss the modality principle and write: "By the time that the learner selects relevant words and images from one segment of the presentation, the next segment begins, thereby cutting short the time needed for deeper processing" (p. 319). The primary explanation Tabbers et al. (2004) give for what they call the "reverse modality effect" is that students in this study could pace their own instruction and move forward and backward through the material. Opfermann et al. (2005) explain their results from the fact that students in their study had sufficient time to study the material, comparable to a realistic study situation. Students also had the opportunity to replay material. Elen et al. (2008) point as well to the importance of learner control versus program control for achieving learning gains. Wouters et al. (2009) conclude, "Apparently, there are conditions (self-pacing, prompting attention) under which the modality effect does not hold true anymore". Mayer et al. (2003) found support for the modality effect in a condition where there was learner control, but in addition they found that interactivity, the possibility of pacing and controlling the order of

the presentation, enhanced learning outcomes. Or, as the authors write, "… interactivity reduces cognitive load by allowing learners to digest and integrate one segment of the explanation before moving on to the next." (Mayer et al. 2003, p. 810). Harskamp et al. (2007) studied the modality effect in a realistic situation (although still with a relatively short study time of 6–11 min) and no possibility of taking notes. The domain was biology (animal behavior) and the material consisted of the presentation of pictures with paper-based or narrated text. The data confirmed the modality effect, showing that there are cases in which the effect is also found under the condition of learner control. However, a further experiment by Harskamp et al. (2007) showed that this was only the case for the fast learners, who more or less acted as if there was system control. For learners who decided on their own to take more time to learn, the modality effect was not found.

These results could indicate that the modality effect holds mainly in situations where there are (very) short learning times and there is system control. It could also mean that the learning conditions in these studies are such that in the system-paced conditions, students who see all the information on screen simply do not have time to see everything in the short time the system allows (see Bannert 2002). This explanation also means that the limitations of short term memory (that indeed become evident in the type of conditions mentioned) can be overcome by allowing students to use more time and to switch back and forth between the information, as is the case in normal learning situations.

Conclusions

Cognitive load theory has without doubt inspired many researchers and has given great momentum to educational research. In this paper, however, some critical questions were posed concerning the conceptual clarity, the methodological rigor, and the external generalizability of this work. In so doing, no attempt was made to give an exhaustive and structured meta-analysis of the work. Instead, by highlighting a number of issues and illustrating these with considerations and results from this field of research, including the 35 most influential ISI publications, some fundamental problems in the area have been highlighted. Here a short synthesis of these issues is presented together with a number of topics for a research agenda.

An important point to note is that cognitive load theory is constructed in such a way that it is hard or even impossible to falsify (see also, Gerjets et al. 2009a). In particular, the fact that cognitive load is composed of three different elements that are "good" (germane), "bad" (extraneous), or just there (intrinsic) means that every outcome fits within the theory post-hoc. If learners perform better, a higher cognitive load must have been composed of germane processes; if they perform poorly, a higher cognitive load must have been extraneous. In addition, the fact that processes can be regarded as germane in one case and the same processes as extraneous in another case means that the theory can account for nearly every situation. Many studies in cognitive load theory make rather speculative interpretations of what happened with cognitive load during learning on the basis of learning performances. Of course, what would help to make these interpretations valid is a suitable measure of cognitive load.

So far cognitive load research has lacked such a measure. The measure that is most often used is the one-item questionnaire as introduced by Paas (1992). Apart from the fact that this measure is used in many different forms and operationalizations throughout cognitive load research, it has two essential drawbacks in that it does *not give a concurrent measure* of cognitive load and it does not measure an essential concept in cognitive load

theory, namely *cognitive overload*. New ways to measure cognitive load based on neu-roscientific techniques might be able to overcome these two problems. Recently, Jaeggi et al. (2007) observed specific patterns of brain activity in cases of high cognitive load. It would be another step forward if techniques were available to distinguish between the three different types of cognitive load: intrinsic, extraneous and germane. This could also help to clear up a number of the conceptual difficulties in cognitive load theory that have been identified in this paper. It would be especially interesting to see if the distinction between intrinsic and germane load stands firm with a more precise assessment. This distinction entered cognitive load theory only at a later stage. Intrinsic and germane load belong to different ontological categories, and if learning processes are examined in both categories quite a few similarities turn up. A return to distinguishing just between processes that do contribute to knowledge generation and processes that do not may help to avoid a number of problems with the current version of the theory.

What has cognitive load theory brought to the field of educational design? The three main recommendations that come from cognitive load theory are: present material that aligns with the prior knowledge of the learner (intrinsic load), avoid non-essential and confusing information (extraneous load), and stimulate processes that lead to conceptually rich and deep knowledge (germane load). These design principles have been around in educational design for a long time (see e.g., Dick and Carey 1990; Gagné et al. 1988; Reigeluth 1983). Work in cognitive load theory often denies the existence of this earlier research, as illustrated in the following quote by Ayres (2006a, p. 288): "Whereas strat-egies to lower extraneous load are well documented… methods to lower intrinsic load have only more recently been investigated" (p. 288). In his study, Ayres introduces part-tasks as one of the initial approaches to lower cognitive load. Describing this as a "recent" approach denies much of the history of instructional design. The value of cognitive load theory until now has certainly been in directing extra and detailed attention to character-istics of instructional designs that may not contribute to learning. If intrinsic and extra-neous load are high, it is clear that learning may be hampered, but having these two forms of load set low doesn't guarantee that learning will take place (see also, Schnotz and Kürschner 2007). This situation has also been recognized in cognitive load research itself, and a number of recent publications point to the fact that cognitive load research has shifted its attention to stimulating germane processes (see e.g., Ayres and van Gog 2009; Kirschner 2002; van Merriënboer and Sweller 2005). What cognitive load research should do now is examine germane processes and estimate which germane processes are most suited for which learners so that experienced cognitive load is optimized and cognitive overload is avoided. The great challenge will be to find load-reducing approaches for intensive knowledge producing mechanisms such as learning from multiple representations (e.g., Ainsworth 2006; Brenner et al. 1997), self-explanations (Chi et al. 1994), inquiry learning (Linn et al. 2006), collaborative learning (Lou et al. 2001), or game-based learning (Nelson and Erlandson 2008). Combining these approaches, which strongly stimulate germane processes, with enough structure to avoid cognitive overload will most probably be one of the leading research themes in the near future (de Jong 2006; Kirschner et al. 2006; Mayer 2004).

The great achievement of cognitive load theory is that it has created unity in a diverse set of instructional design principles and that it has described a cognitive basis underlying these principles. It should not, however, remain at the level of confirming these general principles. Rather, its role is now to move forward and try to determine (1) which instructional treatments lead to which cognitive processes (and how), (2) what the corre-sponding effects are on memory workload and potential overload, (3) what characteristics

of the learning material and the student mediate these effects and (4) how best to measure effects on working memory load in a theory-related manner. This will give a firmer foundation to principles of instructional design theory so that they can be applied in a more dedicated, flexible, and adaptive way.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*, 183–198.

Amadieu, F., Tricot, A., & Mariné, C. (2009). Prior knowledge in learning from a non-linear electronic document: Disorientation and coherence of the reading sequences. *Computers in Human Behavior, 25*, 381–388.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*, 181–214.

Ayres, P. (1993). Why goal free problems can facilitate learning. *Contemporary Educational Psychology, 18*, 376–381.

Ayres, P. (2001). Systematic mathematical errors and cognitive load. *Contemporary Educational Psychology, 26*, 227–248.

Ayres, P. (2006a). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology, 20*, 287–298.

Ayres, P. (2006b). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction, 16*, 389–400.

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 135–147). Cambridge, UK: Cambridge University Press.

Ayres, P., & van Gog, T. (2009). State of the art research into cognitive load theory. *Computers in Human Behavior, 25*, 253–257.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47–89). New York: Academic Press.

Bannert, M. (2002). Managing cognitive load—recent trends in cognitive load theory. *Learning and Instruction, 12*, 139–146.

Beers, P., Boshuizen, H., Kirschner, P. A., Gijselaers, W., & Westendorp, J. (2008). Cognitive load measurements and stimulated recall interviews for studying the effects of information and communications technology. *Educational Technology Research and Development, 56*, 309–328.

Berends, I. E., & van Lieshout, E. C. D. M. (2009). The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load. *Learning and Instruction, 19*, 345–353.

Bodemer, D., & Faust, U. (2006). External and mental referencing of multiple representations. *Computers in Human Behavior, 22*, 27–42.

Brenner, M. E., Mayer, R. E., Moseley, B., Brar, T., Duran, R., Reed, B., et al. (1997). Learning by understanding: The role of multiple representations in learning algebra. *American Educational Research Journal, 34*, 663–689.

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*, 53–62.

Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science, 32*, 115–132.

Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology, 49*, 1–12.

Camp, G., Paas, F., Rikers, R. M. J. P., & 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.

Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of Management Review, 13*, 40–52.

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. (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.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*, 151–170.

Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. N. Giere (Ed.), *Cognitive models of science* (Vol. 15, pp. 129–186). Minneapolis, MN: University of Minnesota Press.

Chi, M. T. H. (2005). Common sense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences, 14*, 161–199.

Chi, M. T. H., Deleeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439–477.

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior, 25*, 315–324.

Clarebout, G., & Elen, J. (2007). In search of pedagogical agents' modality and dialogue effects in open learning environments. *e-Journal of Instructional Science and Technology*, 10 [Electronic Version].

Clarke, T., Ayres, P., & Sweller, J. (2005). The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications. *Educational Technology Research & Development, 53*, 15–24.

Cooper, G. (1998). Research into cognitive load theory and instructional design at UNSW. From http://education.arts.unsw.edu.au/CLT_NET_Aug_97.HTML.

Cooper, G., Tindall-Ford, S., Chandler, P., & Sweller, J. (2001). Learning by imagining. *Journal of Experimental Psychology: Applied, 7*, 68–82.

Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2008). Selecting learning tasks: Effects of adaptation and shared control on learning efficiency and task involvement. *Contemporary Educational Psychology, 33*, 733–756.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences, 24*, 87–114.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features and redundancy. *Journal of Educational Psychology, 94*, 428–434.

de Croock, M. B. M., van Merriënboer, J. J. G., & Paas, F. (1998). High vs. low contextual interference in simulation-based training of troubleshooting skills: Effects on transfer performance and invested mental effort. *Computers in Human Behavior, 14*, 249–267.

de Jong, T. (2006). Computer simulations—technological advances in inquiry learning. *Science, 312*, 532–533.

de Jong, T., Martin, E., Zamarro, J.-M., Esquembre, F., Swaak, J., & van Joolingen, W. R. (1999). The integration of computer simulation and learning support; an example from the physics domain of collisions. *Journal of Research in Science Teaching, 36*, 597–615.

de Jonge, P., & de Jong, P. F. (1996). Working memory, intelligence and reading ability in children. *Personality and Individual Differences, 21*, 1007–1020.

de Koning, B. B., Tabbers, H., Rikers, R. M. J. P., & Paas, F. (in press). Attention guidance in learning from a complex animation: Seeing is understanding. *Learning and Instruction*.

de Westelinck, K., Valcke, M., de Craene, B., & Kirschner, P. A. (2005). Multimedia learning in social sciences: Limitations of external graphical representations. *Computers in Human Behavior, 21*, 555–573.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*, 223–234.

Diao, Y., & Sweller, J. (2007). Redundancy in foreign language reading comprehension instruction: Concurrent written and spoken presentations. *Learning and Instruction, 17*, 78–88.

Dick, W., & Carey, L. (1990). *The systematic design of instruction* (3rd ed.). New York: Harper Collins College Publishers.

Dosher, B. (2003). Working memory. In *Encyclopedia of cognitive science* (Vol. 4, pp. 569–577). New York: Wiley.

Dunsworth, Q., & Atkinson, R. K. (2007). Fostering multimedia learning of science: Exploring the role of an animated agent's image. *Computers & Education, 49*, 677–690.

Dutke, S., & Rinck, M. (2006). Multimedia learning: Working memory and the learning of word and picture diagrams. *Learning and Instruction, 16*, 526–537.

Elen, J., van Gorp, E., & Kempen, K. H. (2008). The effects of multimedia design features on primary school learning materials. *International Journal of Instructional Media, 35*, 7–15.

Fischer, S., Lowe, R. K., & Schwan, S. (2008). Effects of presentation speed of a dynamic visualization on the understanding of a mechanical system. *Applied Cognitive Psychology, 22*, 1126–1141.

Forlines, C., Schmidt-Nielsen, B., Raj, B., Wittenburg, K., & Wolf, P. (2005). A comparison between spoken queries and menu-based interfaces for in-car digital music selection. *Lecture Notes in Computer Science, 3585*, 536–549.

Fredericks, T. K., Choi, S. D., Hart, J., Butt, S. E., & Mital, A. (2005). An investigation of myocardial aerobic capacity as a measure of both physical and cognitive workloads. *International Journal of Industrial Ergonomics, 35*, 1097–1107.

Gagné, R. M., Briggs, L. J., & Wager, W. W. (1988). *Principles of instructional design* (3rd ed.). Fort Worth, TX: Harcourt Brace Jovanovich.

Gerjets, P., & Scheiter, K. (2003). Goal configurations and processing strategies as moderators between instructional design and cognitive load: Evidence from hypertext-based instruction. *Educational Psychologist, 38*, 33–42.

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science, 32*, 33–58.

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction, 16*, 104–121.

Gerjets, P., Scheiter, K., & Cierniak, G. (2009a). The scientific value of cognitive load theory: A research agenda based on the structuralist view of theories. *Educational Psychology Review, 21*, 43–54.

Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., & Eysink, T. H. S. (2009b). Learning with hypermedia: The influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior, 25*, 360–367.

Gerlic, I., & Jausovec, N. (1999). Multimedia: Differences in cognitive processes observed with EEG. *Educational Technology Research and Development, 47*, 5–14.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*, 313–331.

Harskamp, E. G., Mayer, R. E., & Suhre, C. (2007). Does the modality principle for multimedia learning apply to science classrooms? *Learning and Instruction, 17*, 465–477.

Hasler, B. S., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology, 21*, 713–729.

Huk, T. (2006). Who benefits from learning with 3D models? The case of spatial ability. *Journal of Computer Assisted Learning, 22*, 392–404.

Hummel, H. G. K., Paas, F., & Koper, E. J. R. (2006). Timing of cueing in complex problem-solving tasks: Learner versus system control. *Computers in Human Behavior, 22*, 191–205.

Ignacio Madrid, R., van Oostendorp, H., & Puerta Melguizo, M. C. (2009). The effects of the number of links and navigation support on cognitive load and learning with hypertext: The mediating role of reading order. *Computers in Human Behavior, 25*, 66–75.

Jaeggi, S. M., Buschkuehl, M., Etienne, A., Ozdoba, C., Perrig, W. J., & Nirkko, A. C. (2007). On how high performers keep cool brains in situations of cognitive overload. *Cognitive Affective & Behavioral Neuroscience, 7*, 75–89.

Kablan, Z., & Erden, M. (2008). Instructional efficiency of integrated and separated text with animated presentations in computer-based science instruction. *Computers & Education, 51*, 660–668.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23–32.

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. (2001a). Learner experience and efficiency of instructional guidance. *Educational Psychology Review, 21*, 5–23.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001b). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579–588.

Kester, L., & Kirschner, P. A. (2009). Effects of fading support on hypertext navigation and performance in student-centered e-learning environments. *Interactive Learning Environments, 17*, 165–179.

Kester, L., Kirschner, P. A., & van Merriënboer, J. J. G. (2006a). Just-in-time information presentation: Improving learning a troubleshooting skill. *Contemporary Educational Psychology, 31*, 167–185.

Kester, L., Lehnen, C., van Gerven, P. W. M., & Kirschner, P. A. (2006b). Just-in-time, schematic supportive information presentation during cognitive skill acquisition. *Computers in Human Behavior, 22*, 93–112.

Khalil, M. K., Paas, F., Johnson, T. E., & Payer, A. F. (2005). Design of interactive and dynamic anatomical visualizations: The implication of cognitive load theory. *The Anatomical Record (New Anat.), 286B*, 15–20.

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*, 1–10.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimally guided instruction does not work. *Educational Psychologist, 41*, 75–86.

Kissane, M., Kalyuga, S., Chandler, P., & Sweller, J. (2008). The consequences of fading instructional guidance on delayed performance: The case of financial services training. *Educational Psychology, 28*, 809–822.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem-solving and planning processes in preschool children. *Cognitive Psychology, 13*, 113–148.

Lahtinen, T. M. M., Koskelo, J. P., Laitnen, T., & Leino, T. K. (2007). Heart rate and performance during combat missions in a flight simulator. *Aviation Space and Environmental Medicine, 78*, 387–391.

Linn, M. C., Lee, H.-S., Tinker, R., Husic, F., & Chiu, J. L. (2006). Teaching and assessing knowledge integration in science. *Science, 313*, 1049–1050.

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71*, 449–521.

Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 147–158). Cambridge, UK: Cambridge University Press.

Lowe, R. K. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education, 14*, 225–244.

Lusk, M. M., & Atkinson, R. K. (2007). Animated pedagogical agents: Does their degree of embodiment impact learning from static or animated worked examples? *Applied Cognitive Psychology, 21*, 747–764.

Lusk, D. L., Evans, A. D., Jeffrey, T. R., Palmer, K. R., Wikstrom, C. S., & Doolittle, O. E. (2009). Multimedia learning and individual differences: Mediating the effects of working memory capacity with segmentation. *British Journal of Educational Technology, 40*, 636–651.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 88*, 49–63.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*, 1–19.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into Practice, 41*, 226–232.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*, 14–19.

Mayer, R. E., & 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., Dow, G. T., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds. *Journal of Educational Psychology, 95*, 806–813.

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied, 11*, 256–265.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*, 187–198.

Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology, 100*, 380–386.

Mayer, R. E., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology, 94*, 171–185.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312–320.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*, 107–119.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43–52.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389–401.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review, 63*, 81–97.

Miyake, A. (2001). Individual differences in working memory: Introduction to the special section. *Journal of Experimental Psychology-General, 130*, 163–168.

Moray, N. (1982). Subjective mental workload. *Human Factors, 24*, 25–40.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*, 99–113.

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning, 22*, 149–158.

Moreno, R. (2007). Optimising learning from animations by minimising cognitive load: Cognitive and affective consequences of signalling and segmentation methods. *Applied Cognitive Psychology, 21*, 765–781.

Moreno, R., & Mayer, R. E. (1999a). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358–368.

Moreno, R., & Mayer, R. E. (1999b). Multimedia supported metaphors for meaning making in mathematics. *Cognition and Instruction, 17*, 215–248.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*, 309–326.

Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research and Development, 53*, 35–45.

Mousavi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*, 319–334.

Murata, A. (2005). An attempt to evaluate mental workload using wavelet transform of EEG. *Human Factors, 47*, 498–508.

Nelson, B., & Erlandson, B. (2008). Managing cognitive load in educational multi-user virtual environments: Reflection on design practice. *Educational Technology Research and Development, 56*, 619–641.

Ngu, B., Mit, E., Shahbodin, F., & Tuovinen, J. (2009). Chemistry problem solving instruction: A comparison of three computer-based formats for learning from hierarchical network problem representations. *Instructional Science, 37*, 21–42.

Nickel, P., & Nachreiner, F. (2003). Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload. *Human Factors, 45*(57), 5–590.

Opfermann, M., Gerjets, P., & Scheiter, K. (2005). *Exploration of hypermedia environments: Learner control and adaptive user strategies.* Paper presented at the European Association for Research in Learning and Instruction (EARLI), Nicosia, Cyprus. August 23–27.

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., Renkl, A., & Sweller, J. (2003a). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1–4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*, 1–8.

Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. M. (2003b). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63–72.

Paas, F., van Gerven, P. W. M., & Wouters, P. (2007). Instructional efficiency of animation: Effects of interactivity through mental reconstuction of static key frames. *Applied Cognitive Psychology, 21*, 783–793.

Paas, F., & van Merriënboer, J. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors, 35*, 737–743.

Paas, F., & van Merriënboer, J. J. G. (1994a). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*, 351–372.

Paas, F., & van Merriënboer, J. J. G. (1994b). Variability of worked examples and transfer of geometric 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.

Pardo-Vazquez, J. L., & Fernandez-Rey, J. (2008). External validation of the computerized, group administrable adaptation of the "operation span task". *Behavior Research Methods, 40*, 46–54.

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology, 88*, 348–367.

Perlow, R., Jattuso, M., & Moore, D. D. W. (1997). Role of verbal working memory in complex skill acquisition. *Human Performance, 10*, 283–302.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*, 221–243.

Pociask, F. D., & Morrison, G. R. (2008). Controlling split attention and redundancy in physical therapy instruction. *Educational Technology Research and Development, 56*, 379–399.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*, 61–86.

Reber, P. J., & Kotovsky, K. (1997). Implicit learning in problem solving: The role of working memory capacity. *Journal of Experimental Psychology-General, 126*, 178–203.

Reigeluth, C. M. (Ed.). (1983). *Instructional-design theories and models: An overview of their current status*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*, 15–22.

Renkl, A., Gruber, H., Weber, S., Lerche, T., & Schweizer, K. (2003). Cognitive load during learning from worked-out examples. *Zeitschrift Fur Pädagogische Psychologie, 17*, 93–101.

Renkl, A., Hilbert, T., & Schworm, S. (2009). Example-based learning in heuristic domains: A cognitive load theory account. *Educational Psychology Review, 21*, 67–78.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23*, 90–108.

Rikers, R. M. J. P. (2006). A critical reflection on emerging topics in cognitive load research. *Applied Cognitive Psychology, 20*, 359–364.

Rose, J. M., Roberts, F. D., & Rose, A. M. (2004). Affective responses to financial data and multimedia: The effects of information load and cognitive load. *International Journal of Accounting Information Systems, 5*, 5–24.

Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognise designers' styles. *Learning and Instruction, 19*, 185–199.

Salden, R. J. C. M., Paas, F., Broers, N. J., & van Merriënboer, J. J. G. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in air traffic control training. *Instructional Science, 32*, 153–172.

Sandberg, E. H., Huttenlocher, J., & Newcombe, N. (1996). The development of hierarchical representation of two-dimensional space. *Child Development, 67*, 721–739.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*, 469–508.

Schnotz, W., & Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning. *Educational Technology Research & Development, 53*, 47–58.

Schultheis, H., & Jameson, A. (2004). Assessing cognitive load in adaptive hypermedia systems: Physiological and behavioral methods. In P. DeBra & W. Nejdl (Eds.), *Adaptive hypermedia and adaptive web-based systems* (Vol. 3137, pp. 225–234). Berlin: Springer.

Scott, B. M., & Schwartz, N. H. (2007). Navigational spatial displays: The role of metacognition as cognitive load. *Learning and Instruction, 17*, 89–105.

Seufert, T., & Brünken, R. (2006). Cognitive load and the format of instructional aids for coherence formation. *Applied Cognitive Psychology, 20*, 321–331.

Seufert, T., Jänen, I., & Brünken, R. (2007). The impact of intrinsic cognitive load on the effectiveness of graphical help for coherence formation. *Computers in Human Behavior, 23*, 1055–1071.

Seufert, T., Schütze, M., & Brünken, R. (2009). Memory characteristics and modality in multimedia learning: An aptitude-treatment-interaction study. *Learning and Instruction, 19*, 28–42.

Slotta, J. D., & Chi, M. T. H. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction, 24*, 261–289.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology, 33*, 5–42.

Stark, R., Mandl, H., Gruber, H., & Renkl, A. (2002). Conditions and effects of example elaboration. *Learning and Instruction, 12*, 39–60.

Stull, A. T., & Mayer, R. E. (2007). Learning by doing versus learning by viewing: Three experimental comparisons of learner-generated versus author-provided graphic organizers. *Journal of Educational Psychology, 99*, 808–820.

Swaak, J., & de Jong, T. (2001). System vs. learner control in using on-line support for simulation-based discovery learning; effects on knowledge tests, interaction measures, and a subjective measure of cognitive load. *Learning Environments Research, 4*, 217–241.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem-solving in mathematics and science. *Journal of Educational Psychology, 81*, 457–466.

Sweller, J. (1993). Some cognitive-processes and their consequences for the organization and presentation of information. *Australian Journal of Psychology, 45*, 1–8.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312.

Sweller, J. (2005). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 159–167). Cambridge, UK: Cambridge University Press.

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction, 8*, 351–362.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185–233.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology, 119*, 176–192.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.

Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*, 71–81.

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. E., & Paas, F. (2004). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional Science, 32*, 133–152.

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.

Turner, M. L., & Engle, R. W. (1989). Is working memory task dependent. *Journal of Memory and Language, 28*, 127–154.

van Gerven, P. W. M., Paas, F., van Merriënboer, J. J. G., & Schmidt, H. G. (2002). 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. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology, 41*, 167–174.

van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior, 25*, 325–331.

van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*, 16–26.

van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction, 18*, 211–222.

van Merriënboer, J. J. G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research & Development, 53*, 5–13.

van Merriënboer, J. J. G., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: Balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology, 20*, 343–352.

van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*, 5–14.

van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance and training efficiency. *Learning and Instruction, 12*, 11–37.

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*, 147–177.

Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied, 2*, 270–285.

Volke, H. J., Dettmar, P., Richter, P., Rudolf, M., & Buhss, U. (1999). Evoked coherences of EEG in mental load: An investigation in chess players. *Zeitschrift Fur Psychologie, 207*, 233–262.

Wilson, K. M., & Swanson, H. L. (2001). Are mathematics disabilities due to a domain-general or a domain-specific working memory deficit? *Journal of Learning Disabilities, 34*, 237.

Wirth, J., Künsting, J., & Leutner, D. (2009). The impact of goal specificity and goal type on learning outcome and cognitive load. *Computers in Human Behavior, 25*, 299–305.

Wouters, P., Paas, F., & van Merriënboer, J. J. G. (2009). Observational learning from animated models: Effects of modality and reflection on transfer. *Contemporary Educational Psychology, 34*, 1–8.

Wouters, P., Paas, F., & van Merriënboer, J. J. G. (in press). Observational learning from animated models: Effects of studying-practicing alternation and illusion of control on transfer. *Instructional Science*.

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., 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.

Yuan, K., Steedle, J., Shavelson, R., Alonzo, A., & Oppezo, M. (2006). Working memory, fluid intelligence, and science learning. *Educational Research Review, 1*, 83–98.

Zheng, R., McAlack, M., Wilmes, B., Kohler-Evans, P., & Williamson, J. (2009). Effects of multimedia on cognitive load, self-efficacy, and multiple rule-based problem solving. *British Journal of Educational Technology, 40*, 790–803.