Bridging Cognitive Load and Self-Regulated Learning Research

CrossMark

Bridging Cognitive Load and Self-Regulated Learning Research: A complementary approach to contemporary issues in educational research

Anique B.H. de Bruin[*], Jeroen J.G. van Merriënboer

Maastricht University, The Netherlands

ABSTRACT

The aim of this Introduction to the Special Issue 'Bridging Cognitive Load and Self-Regulated Learning Research' is to explore how cognitive load theory, which is particularly relevant for how learners deal with complex information, and self-regulated learning theory, which is particularly relevant for how learners use information to monitor and control their learning, can be combined into one joint research paradigm that is relevant for contemporary and future developments in education. The first two sections introduce cognitive load theory and self-regulated learning theory. The third section discusses the main similarities and differences between the theories, and describes how the *cue-utilization framework* can be used as the basis for a joint research paradigm. The main idea postulated is that new instructional methods should help learners identify diagnostic *cues* in available information that provide an accurate indication of where learners stand in relation to criterion task performance. Use of these diagnostic cues when monitoring learning will lead to better regulation of learning activities and of mental resources allocated, and thus to more efficient learning and higher learning outcomes.

In the fourth section, the six studies and two commentaries presented in this special issue are positioned within this paradigm. In the fifth and final section, a common research agenda based on the joint CLT-SRL paradigm is sketched and its relevance for future developments is explained. The studies presented in this special issue and the two commentaries, which complete the Special Issue, should be seen as a very first step in executing this research agenda.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Learning in the 21st century is in many ways identical to learning in any other era. That is, only when elaborate processing of information and deliberate practicing of skills take place, either individually or in a community, will learners develop competencies transferable to other tasks at later times. Learning conditions have changed considerably, however, and these 21st century differences can be characterized by at least two phenomena. First, the amount of available information has increased and continues to increase dramatically. According to estimations, more information has been produced in the last 30 years than in the 5000 preceding years

(Jungwirth, 2002). Therefore, the necessity of adequate literacy skills is becoming ever more obvious. Over the last decade, this information explosion has come to include personal information (e.g., movement, studying, sleeping, emotions; Zhu, Satizabal, Blanke, Perez-Uribe, & Troster, 2016, and even EEG; Dikker et al., 2016). This change in learning conditions changes the goals of education with less emphasis on transmission of information, and more emphasis on development of domain-general skills, such as literacy skills and self-directed learning skills.

This shift in 21st century learning conditions and goals raises novel educational questions related to how teachers and educational technologies can aid in this process. Specific questions posed include: How does the information explosion affect learning and performance? And, how should feedback in a digital learning environment be designed to enable the learner to act upon it? Given the multifaceted nature of these questions, they ask for integration of research paradigms and theoretical frameworks. Specifically, these issues relate to models of instructional design

* Corresponding author. School of Health Professions Education, Department of Educational Development and Research, P.O. Box 616, 6200 MD Maastricht, The Netherlands.
E-mail address: Anique.debruin@maastrichtuniversity.nl (A.B.H. de Bruin).

and cognitive load, which deal with learning in complex and information-rich environments (Sweller, van Merriënboer, & Paas, 1998), and models of Self-Regulated Learning (SRL), which deal with students' monitoring and control of their learning processes (Bjork, Dunlosky, & Kornell, 2013; Winne & Hadwin, 1998). In the current special issue, these two research pillars are combined to present work on their interface and to discuss how aligning them can provide a novel theoretical ground for contemporary issues in educational science.

Consider a 13-year old high school student who is motivated to teach herself to program a website for her volleyball team. To do so, she needs to determine how to develop her skill acquisition (e.g., asking a skilled friend or teacher, selecting an online course, watching YouTube videos). She will likely need to examine several types of information; textual, visual, and auditory, and continuously monitor her progress in relation to her end goal; a fully functional website. In this editorial, we build upon the example of self-teaching programming skills, because it provides a prototypical example of a contemporary skill and because it relates closely to both the instructional design issues that CLT research tackles, and monitoring and regulation issues pertinent in SRL research. On the one hand, the student must be supported in *making sense* of the wealth of verbal and visual information she is confronted with. Relevant cognitive-load principles then include the multimedia principle (i.e, use a combination of text and pictures rather than only text or only pictures), the split attention principle (i.e., integrate necessary information sources in space or time), and the signaling principle (i.e., highlight pieces of information the learner should pay attention to; for an overview of principles, see Mayer, 2014). On the other hand, the student must be supported in her *use of* selected information for improving own learning and performance. Effectively teaching yourself to program thus requires the ability to monitor own learning and the ability to control or regulate own behavior on the basis of rich and complex information. Educational interventions for improving this process should also study motivation for being involved in these complex cognitive and metacognitive processes (Whittaker et al., 2012).

Bringing together two fields of research can only be accomplished when there is sufficient common ground between them, and when they can profit from each other's strengths. Both educational research pillars have a history of over 25 years of dense scientific output, which has led to insights into the learning mind and how education can adapt to it. Both are based in the psychology of learning and memory, both use subjective estimates of learners about study behavior (i.e., invested mental effort, judgments of learning), and both focus on the self-directed learning situation. Both have their unique strengths: the field of instructional design in studying how cognitive resources relate to learning problem solving skills, and the field of self-regulated learning in studying how subjective judgments relate to objective performance. It has cost 25 years to bring CLT and SRL research to where they are now. Development of a combined theoretical approach and research paradigm and generation of robust insights will probably take at least a similar amount of time.

Below, we will first describe the fundamentals of Cognitive Load Theory and Self-Regulated Learning Theory. Afterwards, we analyze how these two theoretical frameworks can be combined. We then describe the topics of the papers in this special issue and what research questions could be answered by combining CL and SRL research. We end this editorial by applying these insights to the issue of educational research on self-teaching programming skills.

2. Cognitive Load Theory

Cognitive Load Theory (CLT) was initially developed in the 1980s

(Sweller, 1988). It aims to develop instructional design guidelines based on a model of human cognitive architecture that is broadly supported by research in cognitive psychology and which can be explained from an evolutionary perspective (Sweller, 2008). CLT assumes that the human cognitive system has a severely limited working memory for "the retention of a small amount of information in a readily accessible form" (Cowan, 2014, p. 197). The capacity of working memory is limited by storage (i.e., only a very small number of information elements can be simultaneously active), duration (i.e., information elements quickly decay unless they are refreshed by rehearsal), and possibly other cognitive mechanisms (Shipstead, Lindsey, Marshall, & Engle, 2014). The theory emphasizes that these working memory limitations only apply to novel information obtained through sensory memory. Working memory has no known limitations when dealing with information retrieved from long-term memory. Long-term memory holds cognitive schemas that vary in their degree of complexity and automation. Human expertise comes from knowledge organized by these schemas, *not* from an ability to engage in reasoning with many elements that have not been organized in long-term memory - human working memory simply is not able to process many elements.

Expertise develops as learners combine simple schemas into more complex ones. These schemas organize knowledge but also heavily reduce working memory load because even a highly complex schema can be dealt with as *one* element in working memory. Schemas can be constructed by the learner by bringing elements together (i.e., chunking), by incorporating new elements in schemas already available in long-term memory (i.e., elaboration) or, more commonly, by obtaining already schematized information from other people. Schemas can then be treated as one single element in working memory and thus heavily decrease cognitive load. Constructed schemas may become fully automated if they are repeatedly applied and yield desired results. As is the case for schema construction, automation can free working memory capacity for other activities because an automated schema, acting as a central executive, directly steers behavior without the need to be processed in working memory. Because automation requires a great deal of practice, automated schemas only develop for those aspects of performance that are consistent across task situations. Thus, from an instructional design perspective, well-designed instruction should not only encourage schema construction but also schema automation for those aspects that are consistent across tasks (Van Merriënboer & Kirschner, 2013).

The ease with which information may be processed in working memory is a focus of CLT. Working memory load is affected by different processes yielding intrinsic, extraneous, or germane cognitive load (Sweller, van Merriënboer, & Paas, 1998; Van Merriënboer & Sweller, 2005). *Intrinsic load* is a direct function of the complexity of the performed task and the expertise of the learner; it cannot be altered without altering the task to be learned (e.g., by simplification) or by the act of learning itself. It depends on the extent of element interactivity of the materials or tasks that must be learned. The only way to reduce intrinsic cognitive load is to develop schemas that incorporate the interacting elements. *Extraneous load* is a result of superfluous processes that do *not* contribute to learning. It may be imposed, for example, when learners must integrate information sources that are distributed in place or time. *Germane load*, finally, is caused by effortful learning processes that deal with intrinsic cognitive load by adding elements to working memory that are relevant for learning, either from long-term memory (i.e., elaboration) or from the learning context (i.e., generalization from multiple cases).

CLT assumes that the different types of cognitive load are additive. Whereas CLT may not be relevant to teaching simple tasks, it

is critical when complex learning tasks are used because they pose a high load on the learner's cognitive system (Van Merriënboer, Kester, & Paas, 2006). Then, extraneous load and intrinsic load must be lowered to free up processing resources necessary for learning. The more extraneous load is reduced, the more working memory resources can be devoted to intrinsic load and so the easier it becomes to induce germane load for learning. Extraneous load can be reduced, for example, by the use of goal-free tasks (Ayres, 1993), worked examples (Renkl, 2002) and completion tasks (Van Merrienboer, Schuurman, de Croock, & Paas, 2002), by integrating different sources of information (Ginns, 2006), by using multiple modalities (Ginns, 2005), and by reducing redundancy (Diao & Sweller, 2007). Intrinsic load can be managed, for example, by simple-to-complex ordering of learning tasks (Pollock, Chandler, & Sweller, 2002). Germane load can be optimized, for example, by increasing variability over tasks (Paas and van Merriënboer, 1994) and evoking self-explanation (Gerjets, Scheiter, & Catrambone, 2006). Finally, it is important to note that guidelines for managing cognitive load are dependent on the expertise of the learner; due to the so-called expertise reversal effect (Kalyuga, Rikers, & Paas, 2012), instruction for novice learners differs from instruction for more experienced learners.

3. Self-Regulated Learning Theory

In virtually all levels of formal education, students encounter learning environments in which they possess some level of autonomy of their own learning decisions. Students greatly vary how effective they are in steering their own learning, but generally, students have difficulty monitoring their learning and making effective study decisions. Accurate monitoring and regulation of learning is paramount to optimizing learning and performance (Dunlosky & Rawson, 2012), but formal training is scarce (Bjork et al., 2013). SRL theories attempt to outline the underlying processes when students make decisions about their own learning, and research typically focuses on (1) delineating characteristics and behavior of effective self-regulated learners, ineffective self-regulated learners, or both, and (2) designing and testing interventions to improve students' self-regulated learning.

In contrast to CLT, there is no uniform SRL theory, but several theories have been formulated in the literature. Theories, however, show considerable overlap (e.g., Pintrich, Wolters, & Baxter, 2000; Winne & Hadwin, 2008) and in all of them monitoring and control (regulation) of learning play a central role. This is depicted most straightforward in the model by Nelson and Narens (1990) (See Fig. 1).

When a student reads, e.g., a textbook chapter (the object level), she may experience less confidence in her understanding of some of the information (the meta-level). This can lead her to decide to restudy that information so as to improve her knowledge (control).
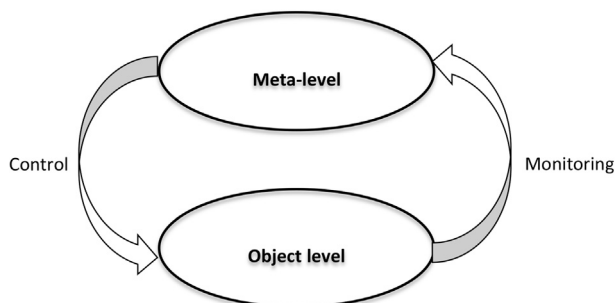

Fig. 1. The intricate relation between monitoring and control (Nelson & Narens, 1990).

Monitoring thus acts to control or regulate learning, in order to optimize the acquisition of knowledge and skills. But control also influences monitoring: when a student experiences improved understanding of the chapter when restudying it, this leads her to updates in her monitoring. When students' monitoring is poor, they will terminate study prematurely or study concepts that do not need restudying. This will jeopardize their learning outcomes.

Research aiming to improve self-regulated learning either studies how learners can move through this cycle more effectively (e.g., Sonnenberg & Bannert, 2015) or examines how separate parts of the model can improve (e.g., Hartwig & Dunlosky, 2014). Monitoring and regulation research aims to explain how students come to metacognitive judgments, why these are often inaccurate, and what should be done to improve them.

An influential framework in relation to monitoring and regulation accuracy is the cue utilization framework (Koriat, 1997). Within this framework, monitoring and regulation judgments are described as inferential in nature, depending on the cues students infer from available information when making metacognitive judgments. That is, students do not have direct access to the quality of their cognitive states, but need to infer cues informative of their learning and performance (Koriat, 1993). One pervasive cue is the fluency, or ease, with which information is processed (Hertzog, Dunlosky, Robinson, & Kidder, 2003). For example, if a text is easily read, a student typically judges her understanding of the text as high. The accuracy of her monitoring judgment then depends on the diagnosticity of reading ease for her actual understanding of the text. If better understanding of the text correlates with reading speed, reading speed is a diagnostic cue for monitoring understanding. Unfortunately, the cues that students use are often not diagnostic of their actual learning. A second threat to monitoring and regulation accuracy is a selection problem; when diagnostic cues are available but not *used* by students. This happens, for example, when students generate a summary of a text immediately after studying the text. The details of the text that are still in short-term memory distract students from the diagnostic cues of the gist of their summary (Thiede & Anderson, 2003). Cue prompts are therefore needed to improve use of diagnostic cues when judging monitoring and regulation of learning (e.g., Thiede, Anderson, & Therriault, 2003; Van Loon, De Bruin, Van Gog, Van Merriënboer, & Dunlosky, 2014). Effective cue prompts when learning texts are generating keywords (Thiede et al., 2003), providing summaries (Thiede & Anderson, 2003), and completing empty diagrams of the texts (Van Loon et al., 2014) prior to monitoring learning of the texts. These cue prompts can also improve regulation of learning and lead to improved learning outcomes (Thiede et al., 2003). Cue diagnosticity and cue utilization are thus two important concepts to improve self-regulated learning.

4. Similarities and differences between CLT and SRL: bridging the gap

Both CLT and SRL research have generated robust insights and combining them would allow for covering a broad range of learning tasks, relating to issues of self-regulation and design of instruction simultaneously. A starting point for analysing commonalities and differences is provided by the measures that are typically used in CLT and SRL research (See Table 1).

An elemental independent variable in both CLT and SRL research is the subjective estimate students generate about their learning. In CLT research, estimates mainly relate to ratings of invested mental effort as a proxy for overall cognitive load. The most popular measurement instrument is the Paas scale, where learners rate their invested mental effort on a unidimensional 9-point rating scale ranging from (1) very very low mental effort, to (9) very, very

Table 1
Measurements in cognitive load and self-regulated learning research.

	Process measures (Subjective estimates of learning and effort)	Outcome measures	Computed measures
Cognitive load theory	Mental effort/overall cognitive load - Intrinsic load - Extraneous load - Germane load Regulation decisions - Task selection - Allocation of cognitive resources	Retention performance Transfer performance	Efficiency of learning Task selection efficiency Task involvement
Self-regulated learning theories	Monitoring judgments - Judgments of Learning - Judgments of comprehension - Judgments of performance Regulation decisions - (Re)study decisions - Allocation of time	Recall Text compre-hension Problem solving accuracy	Monitoring accuracy (relative and absolute) Regulation accuracy

high mental effort. Mental effort is then defined as the cognitive capacity that is allocated to accommodate the demands imposed by the task (Paas, 1992). As cognitive load theory is mainly involved with complex problem solving, the main outcome measures are performance on test problems that are structurally identical to the learning problems (i.e., retention performance), and performance on test problems that are structurally different (i.e., transfer performance, either near or fars).

Popular computed measures are efficiency and task involvement. Efficiency is defined to be high when learners show relatively high performance in combination with low invested mental effort; efficiency is low when learners show relatively low performance and high invested mental effort (See Paas and van Merrienboer (1993), for how efficiency is calculated and represented graphically in relation to effort and performance). In a similar way, task involvement is defined to be high when learners show relatively high performance combined with high invested mental effort ("no pain, no gain"); it is low when learners show relatively low performance in combination with low invested mental effort (Paas, Tuovinen, van Merrienboer, & Darabi, 2005). In SRL research, efficiency is also a relevant outcome measure, typically depicted by calculating the relation between invested time and learning outcomes (Ariel, Dunlosky, & Bailey, 2009).

Nowadays, more instruments become available to measure the three different types of cognitive load instead of an overall estimate. An example is the questionnaire developed by Leppink, Paas, van Gog, van der Vleuten, and van Merriënboer (2014), where learners answer 13 questions on a scale from 0 to 10; 4 items measure intrinsic cognitive load (e.g., "I invested a very high mental effort in the complexity of this activity"); 4 items measure extraneous cognitive load (e.g., "I invested a very high mental effort in unclear and ineffective explanations and instructions in this activity"), and 5 items measure germane cognitive load (e.g., "I invested a very high mental effort during this activity in enhancing my knowledge and understanding"). First, this refinement enables the formulation of more specific hypotheses with regard to effects of interventions on cognitive load. It should discourage researchers from classifying the type of induced cognitive load only after the research has been conducted, and so help them to predict and explain their findings But second, distinguishing the different types of load also complicates interpreting the relations with outcome measures and the computation of efficiency and task involvement. An obvious example is provided by the relation between extraneous cognitive load and germane cognitive load with performance: High extraneous load will typically hamper learning because effort expended that is irrelevant to learning will have a detrimental effect on learning outcomes, but the opposite is true for

high germane load, because the more mental effort is put into schema construction, the higher learning outcomes will be. For an overview of possibilities for relating cognitive load to performance and for computing efficiency and task involvement measures, see Van Gog and Paas (2008).

Analysis of students' task selection is an important indicator of regulation accuracy in CLT research. When students are asked to select a next task for study, measures of accuracy, effort, and efficiency can be computed. In terms of regulation accuracy, the selected task can be compared with a teacher-selected task (both retention and transfer tasks) or with performance on a similar task. Mental effort put into selection of the task can be rated by students, and efficiency of the task selection can be measured by relating effort put into selection to accuracy of the task selection. The latter has, to our knowledge, not been empirically studied yet.

In SRL research, subjective estimates relate to both the experienced level or quality of learning (e.g., monitoring judgments) and to the next steps that learners consider needed in the learning process (e.g., control or regulation judgments). In experimental settings, these judgments are typically made explicitly on a numerical or pictorial judgment scale. The quality of students' monitoring judgments is usually termed 'monitoring accuracy' and is measured by correlating monitoring judgments and actual learning/performance (i.e., relative monitoring accuracy) and/or by calculating the absolute difference between monitoring judgments and actual learning/performance (i.e., absolute accuracy). With regard to relative accuracy, a high, positive correlation provides an indication of higher discrimination between the better and lesser known items. With regard to absolute accuracy, a non-significant difference from zero indicates a well-calibrated judgment, and deviance from zero is interpreted as over or underconfidence. Calculating absolute monitoring accuracy is only possible when judgments and actual learning or performance are made on the same rating scale.

Regulation accuracy is determined by correlating study choices with actual test performance. The more negative this relation, the higher regulation accuracy is, assuming that accurate regulation entails selecting those learning tasks for future study that were less learned. Alternatively, comparison between students' monitoring judgments and their regulation judgments provides a measure of regulation accuracy. This measure indicates to what extent students choose to dedicate study time to learning tasks they *think* they perform poorly on (although, under certain circumstances, students may decide to spend most time on medium difficulty tasks, see Undorf & Ackerman, 2017). Also, as a measure of regulation accuracy control sensitivity can be calculated by asking students to volunteer or withhold a response (i.e., 'betting' on an item) and

relating those decisions to monitoring judgments (Koriat & Goldsmith, 1996). These numerical indicators of metacognitive monitoring and regulation allow for comparison of the effect of interventions between different conditions or groups of learners, and across separate studies. For more details on metacognitive judgments, see De Bruin and van Gog (2012).

One of the main differences between the theoretical frameworks becomes clear in the second column of Table 1, specifying the learning outcomes under study. Specific to CLT research is its focus on learning to solve problems and the acquisition of complex cognitive skills (e.g., mathematics, science, engineering). By comparison, SRL research typically studies conceptual knowledge acquisition, such as vocabulary learning, idiom learning or text comprehension. To a lesser extent, actual effects on learning outcomes have been shown, which are essential to validate the effect of interventions (e.g., Koriat & Bjork, 2006; Lauterman & Ackerman, 2014; Thiede et al., 2003).

How can we further strengthen the similarities and reduce the differences between the two frameworks, in order to work towards one integrated framework? We argue that the cue-utilization framework described above is a particularly relevant bridging aspect. By means of Fig. 2, we will explain how the cue-utilization framework can bridge CLT and SRL research. SRL research attempts to unravel the basis of metacognitive judgments and test interventions ('cue prompts') that improve monitoring accuracy. When studying expository texts, cue prompts known to improve monitoring accuracy are, for example, generating keywords or summaries about the texts (Thiede & Anderson, 2003; Thiede et al., 2003). Van Loon et al. (2014) show how a diagram cue prompt improved cue diagnosticity, cue utilization, and monitoring accuracy after studying cause-and-effect texts. Students mostly relied on the number of non-completed (i.e., blank) boxes and the number of correct boxes in the diagram. Since those cues were diagnostic of ultimate test performance, the cue prompt was effective. The top of Fig. 2 illustrates the relation between monitoring judgments and actual learning or performance (monitoring accuracy) and the relation between cues and actual learning or performance (diagnosticity of cues). Finally, cue utilization is determined by students' actual *use* of cues when monitoring their learning, measured by relating cues to monitoring judgments. For further explanation, see De Bruin, Dunlosky, & Cavalcanti, 2017.

In parallel, the bottom of Fig. 2 shows how relating mental effort judgments to actual learning and performance provides an estimate of the (im)proper use of cognitive resources (i.e., efficiency and task involvement). Relating mental effort ratings to actual mental effort indicates mental effort rating accuracy. The bottom of Fig. 2 illustrates how the relation between cues and actual learning or performance also indicates cue diagnosticity. For example, a learner who is solving a problem using trial-and-error may use the experienced high extraneous load as an invalid cue for learning ("this costs me a lot of effort so I must have learned a lot"). A cue prompt asking the learner to explain the problem to a peer may help to recognize the low *germane* load as a more diagnostic cue ("I cannot explain what I did so I did not invest enough effort in learning"). Thus, in the example, the cue inferred from the ability to explain the problem to someone else would yield a radically different mental effort rating than a cue based on trial-and-error processing. Furthermore, the use of proper cues might also help learners to distinguish between intrinsic, extraneous and germane load; for instance, cues that help learners monitor attention control may be diagnostic for extraneous load, while cues that help learners monitor their understanding may be more diagnostic for germane load. This model emphasizes the need for unravelling cue use and designing effective cue prompts to improve cue use.
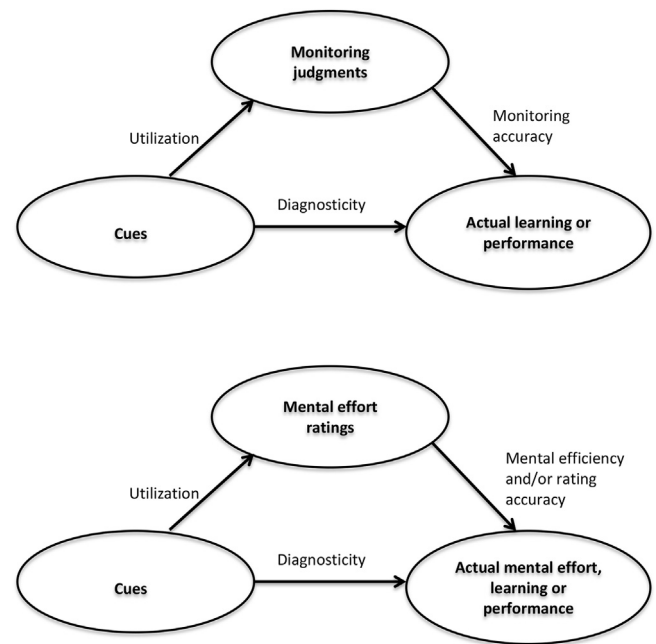

Fig. 2. Monitoring judgments, cues, learning and performance determine monitoring accuracy, cue diagnosticity, and cue utilization. Mental effort ratings, cues, learning and performance determine mental efficiency, mental effort rating accuracy, cue diagnosticity, and cue utilization.

5. The studies in this special issue

Having outlined the background, similarities, and differences of CLT and SRL research, and after presenting the cue-utilization framework as a first step toward integration, we now turn to the studies presented in this special issue. We focus particularly on those elements that we consider relevant for the development of a combined research paradigm that will accommodate educational research into future themes, such as the cue-utilization framework. Detailed discussion of the findings is offered in the commentaries by Boekaerts and by Sweller and Paas at the end of this special issue. An overview of the central measures studied in the papers can be found in Table 2.

Raaijmakers, Baars, Schaap, Paas, & Van Gog (1) and Van Loon, Destan, Spiess, De Bruin and Roebers (2) show two novel and relevant steps in CL and SRL research: interventions to examine cue use for CL ratings, and the effect of feedback on metacognitive judgments, respectively. Where cue use interventions are common in SRL research and feedback manipulations are often used in CLT research, the reverse of this as presented in these two manuscripts is uncommon and novel. They use tasks of relatively low ecological validity, but apt for answering the research questions. The first study by Raaijmakers et al. (1) investigated the effect of feedback valence (positive, negative, absent) after a learning task on mental effort ratings. That is, they examined how the perceived level of learning is used as a cue to judge the effort expended in the learning task. Because the 'Day of the week' problem they used is difficult to self-assess, they were able to manipulate the feedback valence independent of actual performance. The results overall revealed that negative feedback led to higher ratings of mental effort than positive feedback, despite the lack of differences in performance. This is one of the first studies that examines the cues learners rely on when judging mental effort (for research on how timing and frequency affect mental effort ratings, see e.g., Schmeck, Opfermann, van Gog, Paas, & Leutner, 2015).

Table 2
Overview of the central measures studied in the papers, specifically with regard to cue use, CLT, and SRL aspects.

	Cues studied	Monitoring judgments/ Mental effort ratings	Learning/Performance outcomes	(1) What cues influence CL and SRL ratings?	(2) How is cue use in CL and SRL related?	(3) How to design predictive cue prompts?
Raaijmakers et al. (2017)	Feedback valence	CL ratings	Problem solving accuracy	X		
Van Loon et al. (2017)	Intrinsic load (Item difficulty) Cues generated through feedback	Confidence judgments	Recall	X		X
Schleinschok et al. (2017)	Cues generated through drawing	CL ratings, JOLs, restudy decisions	Text comprehension	X	X	X
Glogger-Frey et al. (2017)	Cues generated through self-regulated learning or worked example study	Extraneous load Perceived knowledge gaps Self-efficacy	Recall of surface and deep structure Procedural knowledge task Transfer task	X	X	X
Sidi et al. (2017)	Contextual cues (through screen or paper)	Confidence judgments Cognitive load (through time pressure)	Problem solving success rate Problem solving efficiency	X	X	X
Maranges et al. (2017)	Cues available because of ego depletion	Self-regulation depletion through reading tone instruction (Exp 1), writing essay without letters A, N, or I, O (Exp 2, 3) Cognitive load manipulation through dual task (Exp 1, 2, 3)	Time in cold water (Exp 1), affect (Exp 2), visual memory (Exp 2), lexical judgment task (Exp 3)	X	X	

In the second article, Van Loon et al. (2) study the development of children's monitoring of own performance and, especially, how this is affected by feelings of difficulty, for which the intrinsic cognitive load imposed by the task may serve as a cue, and how this is influenced by self-protection biases. They also study the effects of feedback on performance, because this is known to increase the quality of monitoring processes. The main dependent measure in their study is monitoring accuracy. Although it is known that monitoring accuracy increases with age, it is unknown whether this is due to a greater reliance on feelings of difficulty, less self-protective biases, or a combination of both. The reported study approaches this question by comparing two age groups, and also studies how monitoring accuracy in these age groups is affected by performance feedback. Their results showed that both age groups improved self-evaluations after performance feedback, although the older children better discriminated between incorrect and correct responses and the younger children continued to provide inappropriately high self-rewards after feedback.

Schleinschok, Eitel, and Scheiter (3), Glogger-Frey, Gaus and Renkl (4) and Sidi, Shpigelman, Zalmanov, and Ackerman (5) present studies in which interventions are designed and tested that aim to improve metacognitive monitoring, regulation, and performance. They use and combine CL and SRL measurements in varying ways, and aim at ecologically valid learning tasks. In the third article, Schleinschok et al. (3) investigate how students' generating of a drawing after text reading influences cognitive load, monitoring accuracy, regulation, and ultimate learning from the text. Also, they analyze the relation between cognitive load ratings and judgments of learning. Contrary to most previous work in this area, they look at how the generative drawing task affects restudy of the text, to study how regulation behavior affects actual learning outcomes. They showed that students who generated drawings improved both their monitoring and regulation of learning more than those who did not generate drawings.

In the fourth article, Glogger-Frey et al. (4) compare the effects of self-regulated versus guided preparatory activities, which require students to compare and contrast several cases, on learning from subsequent direct instruction. As usual in CLT research, their main measure is transfer performance. On the one hand, self-regulated preparatory activities may yield higher extraneous cognitive load and thus hamper transfer of learning; on the other hand, such self-regulated activities may help learners to use *cues* that signify their understanding of structural relations and may thus increase transfer of learning. Their study provides interesting insights in the balance between both tendencies. Their findings show that self-regulated activities led to higher performance on transfer problems than worked examples.

In the fifth article, Sidi et al. (5) investigate how effort regulation, confidence, and problem solving (i.e., logic problems) are influenced by on screen versus on paper processing. Effort regulation is an interesting concept on the brink of CLT and SRL research. It underlines the importance of students' ability to regulate their depth of processing depending on the task and context. Specifically, they study how effort regulation is affected by increased cognitive load (due to time pressure) or by framing the task as of low importance, and how this differs for on screen versus on paper processing. They are interested in examining how the contextual cues differing between paper and screen affect monitoring and regulation of learning. Their results reveal that high cognitive load (through time pressure) or under low perceived importance of the task, working on screen but not on paper compromised both cognitive and metacognitive processes.

In the sixth and final article, Maranges, Schmeigel, and Baumeister (6) present a study comparing the effects of self-regulatory resource depletion (i.e., ego depletion) and cognitive load on pain tolerance and negative feelings. Ego depletion may undermine top-down control over negative feelings, thus, when people are depleted they become more sensitive to cues eliciting negative feelings that interfere with learning, whereas these feelings would otherwise be suppressed. Cognitive load may also affect negative feelings, but not necessarily in the same way as ego depletion: When load is high, attention may be distracted away from negative physical and emotional feelings. This hypothesis is tested in three experiments with findings that have interesting implications for learning and instruction. This is one of the few studies examining effect of self-regulation and cognitive load on emotions, and how this affects learning. In short, their findings show that cognitive load can reduce the experience of physical and emotional feelings.

Finally, the two commentaries summarize and critically discuss

the findings of the separate papers in the Special Issue. The commentary by Boekaerts (2017) emphasizes the need for more objective and immediate measurements of invested mental effort, such as through EEG, pupil size and eye movement data. The use of more objective measures of − the different types of − cognitive load would strengthen cognitive load theory because subjective ratings will not always be a valid measure of cognitive load (see Korbach, Bruenken, & Park, 2017). It also underlines the potential of application of the cue utilization framework to CLT, to gain insight into the basis for subjective mental effort ratings. Sweller and Paas' commentary (2017) discuss the difficulty that arises when wanting to combine SRL and CLT research due to its distinct knowledge bases. That is, they consider self-regulation to be a biologically primary skill and therefore unteachable, whereas cognitive load theory stresses biologically secondary, domain-specific knowledge that requires extensive and explicit instruction.

6. Towards a joint research agenda

As the studies in this special issue demonstrate, metacognitive cues play an essential role in optimizing the quality of both monitoring judgments and mental effort ratings. Contrary to what Sweller and Paas (2017) argue, we see evidence that SRL skills are trainable and improvable through practice (e.g., Thiede et al., 2003; Van Loon et al., 2014). Even if a child learns how to walk without instruction, extensive practice and explicit instruction is needed to make it an athlete. Similarly, even though children's ability to regulate behavior develops with age, monitoring and regulating learning in academic settings requires practice and instruction. SRL research shows that interventions to improve monitoring and regulation are domain-specific and dependent on the criterion learning task. As such, we argue that SRL and CLT are sufficiently similar to warrant exploration of a joined research paradigm. Based on the studies in this special issue and the theoretical underpinnings of CLT and SRL, we come to the following research questions central to a combined CLT-SRL paradigm:

(1) What cues (both diagnostic and non-diagnostic) influence CL and SRL ratings? (all papers in this special issue)
(2) How is cue use in CL and SRL related? (Papers 3, 4, 5, 6)
   a. How and when are cues used for CL ratings and SRL judgments similar and/or different?
   b. How does cognitive load influence SRL judgments and how does self-regulation influence CL ratings?
(3) How can we design effective cue prompts that accurately predict actual learning and actual mental effort? (Papers 2, 3, 4, 5)

How can these general research questions be translated to contemporary issues in educational research? We will revert to the example of self-teaching programming skills as a contemporary and exemplary issue in education and thus as subject of educational research. As explained, self-teaching programming skills contains strong elements of self-regulated learning (e.g., it asks for continuous monitoring of goal achievement and regulation of learning activities) and of cognitive load (e.g., it requires individuals to manage their cognitive load effectively and efficiently; it is done on devices using multimedia). These two elements are intricately related and therefore should be studied simultaneously, as well as looking at the interaction between them.

A first step (related to RQ 1) then would be to (further) unearth the cues that influence monitoring judgments, regulation judgments, and mental effort ratings. This research question calls for both fundamental and applied research approaches to gain in-depth understanding of the basis of these subjective judgments.

Fundamental research could examine what cues students typically use or are persuaded to use through contextual influences on self-judgments during prototypical learning tasks, whereas applied research could study how this varies in more ecologically valid settings. The studies in the papers by Raaijmakers et al., 2017; Van Loon et al., 2017 and Schleinschok et al. (2017) provide examples of research into the basis of monitoring judgments and mental effort ratings. Further steps would be to use the cue diagnosticity and cue utilization framework (Fig. 2) for an in-depth analysis of what cues students actually use and how diagnostic these are. Work within the Meta-Reasoning Framework is worth mentioning here (Ackerman & Thompson, 2015; in press), as it shows that outside the traditional verbal and textual learning tasks, the concepts of cues, monitoring, and regulation apply similarly to reasoning and problem-solving tasks.

While these are questions studied in relation to actual learning tasks, we also see a need for fundamental research in controlled environments (related to RQ 2) focused at the timing and frequency aspects of simultaneously tapping monitoring judgments and mental effort ratings (i.e., do they influence each other inadvertently?). The study by Schleinschok et al. (2017) examines how JOLs, CL ratings, and performance correlate, although they do not manipulate factors that potentially contribute to that correlation such as timing and frequency. Testing hypotheses around this question will aid design of applied research. Applied research could then study how and why SRL versus CL ratings differ, and how this relates to diagnostic and non-diagnostic cue use. Ultimately, this will provide insight into how students can optimally benefit from developing accurate monitoring and regulation abilities and manage their mental effort and learning effectively. The strength of combining CLT and SRL research paradigms would lie in enabling learners to use their estimates of both mental effort and learning to better regulate their learning processes when engaging in a broad range of tasks (from vocabulary learning to complex skill acquisition). The study by Glogger-Frey et al. (2017) provides an interesting example of how self-regulated learning might increase extraneous cognitive load, but improve awareness of knowledge gaps. This paper shows how the two approaches can have opposing hypotheses. Moreover, a neglected question within both SRL and CLT research is how learners regulate the effort they exert in learning tasks. This question (related to RQ 2b) addresses the step in-between monitoring and regulation of learning, and addresses how learners translate insights from monitoring judgments to investing effort into a novel task or restudy of a task. Only when learners decide to invest mental effort when regulating learning, will learning outcomes improve. Research into regulation of effort will likely integrate insights from mental effort research (e.g., on eye movements and pupil dilation) into SRL paradigms (e.g., by studying how and when learners distribute mental effort when regulating learning).

The work by Maranges et al. (2017) is also relevant in relation to RQ 2b. It demonstrates how affect plays a role in issues of self-regulation and cognitive load, and how these topics can be combined to come to a more holistic approach to SRL-CLT research. Their approach shows that it is not only a matter of eliciting explicit monitoring/mental effort judgments, but when emotional processes are involved also a matter of directly influencing SRL and CL. In relation to self-teaching programming skills, affect measurements can be incorporated, to determine how affect influences SRL, CL and learning behavior. Such a line of research can complement the explicit judgment research.

Finally, insight into the diagnosticity and non-diagnosticity of cues as the basis for SRL and CL ratings, and the relation between SRL and CL ratings will provide the foundation for research on teaching students to use diagnostic cues and prevent non-

diagnostic cues when monitoring and regulating effort and learning (related to RQ 3). The study by Schleinschok et al. is a direct example of this approach (examining how a drawing tasks affects monitoring judgments and CL ratings), and the work by Sidi et al. is a more indirect example of this approach (examining how screen versus paper presentation affects monitoring, regulation, and cognitive load). Future studies could investigate how to teach students to choose whether to study on screen versus on paper depending on contextual cues. The programming skills example demonstrates that teaching students to use diagnostic cues is mostly an issue of helping students filter the diagnostic cues from the non-diagnostic ones, given that a broad array of cues (from feedback but also subjective cues) will be available at any point in time.

As described by Van Merriënboer (2016), the design of the cue prompt directly depends on the criterion task. For example, learning routine aspects of tasks will require prompts that generate cues about the fluency of these routine aspects (e.g, can they be performed effortless, fast, parallel to other tasks − performance accuracy as such is not a valid cue). In contrast, monitoring learning of conceptual information will be aided by cue prompts that generate information about understanding of the conceptual information (e.g., making summaries, keywords, diagrams − ease of processing is not a valid cue). A thorough analysis of both the criterion task and of the optimal way to generate cues that indicate where students stand in relation to the criterion task is needed to design effective cue prompts (De Bruin et al., 2017). The model presented in Fig. 2 can then be used to determine diagnosticity and utilization of the cues generated through the prompt.

Multimedia environments provide digital functionality to incorporate cue prompts, but also to provide feedback on the quality of the cues generated. If students, for example, at the end of a study day are asked to reflect on the quality of their learning, a mobile phone application can provide a summary of relevant data (e.g., length of study sessions, quality of summaries, and quiz performance) and analyze the discrepancy between students' reflection and actual learning. This can also be done across several days or weeks so as to provide information on development of SRL skills. In relation to CLT aspects, cue prompts can be used to teach students to accurately rate invested mental effort, and learn to distinguish between germane and intrinsic/germane load; a distinction crucial for learning. As indicated before, this also requires the further development of *objective* measures of − different types of − cognitive load, such as secondary task performance, eye-tracking and pupillometric analysis, as well as physiological data such as heart rate, neuronal activity, or electrodermal activity (Korbach et al., 2017). Such objective measures may help to identify biases in learners' subjective measures.

7. Conclusions

In this editorial to the Special Issue 'Bridging Cognitive Load and Self-Regulated Learning Research', we have attempted to outline and demonstrate the potential of development and execution of a joint research agenda. In a future where massive amounts of information will be continuously available providing learners with abundant opportunities to self-teach knowledge and skills, issues of self-regulated learning and cognitive load are omnipresent and will become inseparable. We describe how the diagnosticity and utilization of cues when monitoring and regulating learning, and when managing mental effort are essential to enable learners to cope and learn in such an environment. The papers described in this special issue can be seen as first pieces of this complex puzzle; a puzzle that is only just starting to take shape.

References

Ackerman, R., & Thompson, V. (2015). Meta-Reasoning: What can we learn from meta-memory. In A. Feeney, & V. Thompson (Eds.), *Reasoning as memory* (pp. 164−178). Hove, UK: Psychology Press.

Ackerman, R., & Thompson, V. (in press). Meta-Reasoning: Shedding meta-cognitive light on reasoning research. L. Ball & V. Thompson (Eds.), International handbook of thinking & reasoning. Psychology Press.

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

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417−444.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review, 26*, 197−223.

De Bruin, A. B. H., Dunlosky, J., & Cavalcanti, R. B. (2017). Monitoring and regulation of learning in medical education: The need for predictive cues. *Medical Education, 51*, 575−584.

De Bruin, A. B. H., & van Gog, T. (2012). Improving self-mon-itoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction, 22*, 245−252.

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

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., McClintock, J., Oostrik, M., et al. (2016). *Brain-to-brain synchrony predicts student engagement and social dynamics in the classroom: A crowd-sourcing neuroscience study*. Poster presented at the Annual Meeting for the Cognitive Neuroscience Society, NY.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*, 271−280.

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.

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

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*, 511−525.

Glogger-Frey, I., Gaus, K., & Renkl, A. (2017). Learning from direct instruction: Best prepared by several self-regulated or guided invention activities? *Learning and Instruction, 51*, 26−35.

Hartwig, M., & Dunlosky, J. (2014). The contribution of judgment scale to the unskilled-and-unaware phenomenon: How evaluating others can exaggerate over- (and under-) confidence. *Memory & Cognition, 42*(1), 164−173.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29*, 22−34.

Jungwirth, B. (2002). Information overload: Threat or opportunity. *Journal of Adolescent & Adult Literacy, 45*(5), 90−91.

Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review, 24*, 313−337.

Korbach, A., Bruenken, R., & Park, B. (2017). Differentiating different types of cognitive load: A comparison of different measures. *Educational Psychology Review*. http://dx.doi.org/10.1007/s10648-017-9404-8. First Online.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609−639.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349−370.

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions. A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1133−1145.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490−517.

Lauterman, T., & Ackerman, R. (2014). Overcoming screen inferiority in learning and calibration. *Computers in Human Behavior, 35*, 455−463.

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32−42.

Maranges, H., Schmeigel, B., & Baumeister, R. F. (2017). Comparing cognitive load and self-regulatory depletion: Effects on emotions and cognitions. *Learning and Instruction, 51*, 74−84.

Mayer, R. E. (Ed.). (2014). *The Cambridge handbook of multimedia learning* (2nd ed.). New York: Cambridge University Press.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation, 26*, 125−141.

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., Tuovinen, J. E., van Merrienboer, J. J. G., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development, 53*(3), 25−34.

Paas, F., & van Merrienboer, J. J. G. (1993). The efficiency of instructional conditions:

An approach to combine mental effort and performance measures. *Human Factors, 35*, 737–743.

Paas, F., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*, 122–133.

Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self- regulated learning. In G. Schraw, & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43–97). Lincoln, Nebraska: Buros Institute of Mental Measurements.

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

Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., & Van Gog, T. (2017). Effects of performance feedback valence on perceptions of invested mental effort. *Learning and Instruction, 51*, 36–46.

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction, 12*, 529–556.

Schleinschok, K., Eitel, A., & Scheiter, K. (2017). Do drawing tasks improve monitoring and control when learning from texts? *Learning and Instruction*.

Schmeck, A., Opfermann, M., van Gog, T. A. J. M., Paas, G. W. C., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science, 43*, 93–114.

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language, 72*, 116–141.

Sidi, Y., Shpigelman, M., Zalmanov, H., & Ackerman, R. (2017). Understanding metacognitive inferiority on screen by exposing cues for depth of processing. *Learning and Instruction, 51*, 61–73.

Sonnenberg, C., & Bannert, M. (2015). Discovering the effects of metacognitive prompts on the sequential structure of SRL-processes using process mining techniques. *Journal of Learning Analytics, 2*, 72–100.

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

Sweller, J. (2008). Instructional implications of David C. Geary's evolutionary educational psychology. *Educational Psychologist, 43*, 214–216.

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

Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology., 28*, 129–160.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*, 66–73.

Undorf, M., & Ackerman, R. (2017). The puzzle of study time allocation for the most challenging items. *Psychonomic Bulletin & Review*. http://dx.doi.org/10.3758/s13423-017-1261-4.

Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*(1), 16–26.

Van Loon, M. H., De Bruin, A. B. H., Van Gog, T., Van Merriënboer, J. J., & Dunlosky, J. (2014). Can students evaluate their understanding of cause-and-effect relations? The effects of diagram completion on monitoring accuracy. *Acta Psychologica, 151*, 143–154.

Van Loon, M. H., Destan, N., Spiess, M., A.B.H., De Bruin, & Roebers, C. (2017). Developmental progression in performance evaluations: Effects of children's cue-utilization and self-protection. *Learning and Instruction, 51*, 47–60.

Van Merriënboer, J. J. G. (2016). How people learn. In N. Rushby, & D. W. Surry (Eds.), *The wiley handbook of educational technology*.

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. (2013). *Ten steps to complex learning* (2nd Rev. Ed.). New York: Routledge.

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.

Whittaker, S., Kalnikaitė, V., Petrelli, D., Sellen, A., Villar, N., Bergman, O., et al. (2012). Socio-technical lifelogging: Deriving design principles for a future proof digital past. *Human-Computer Interaction, 27*(1–2), 37–62.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 227–304). Mahwah, NJ: Erlbaum.

Zhu, Z., Satizabal, H. F., Blanke, U., Perez-Uribe, A., & Troster, G. (2016). Naturalistic recognition of activities and mood using wearable electronics. *IEEE Transactions on Affective Computing, 7*, 272–285.