Development and validation of a theory-based questionnaire to measure different types of cognitive load

Krieglstein, Felix ; Beege, Maik ; Rey, Günter Daniel ; Sanchez-Stockhammer, Christina ; Schneider, Sascha

Abstract: According to cognitive load theory, learning can only be successful when instructional materials and procedures are designed in accordance with human cognitive architecture. In this context, one of the biggest challenges is the accurate measurement of the different cognitive load types as these are associated with various activities during learning. Building on psychometric limitations of currently available questionnaires, a new instrument for measuring the three types of cognitive load—intrinsic, extraneous, and germane cognitive load—is developed and validated relying on a set of five empirical studies. In Study 1, a principal component analysis revealed a three-component model which was subsequently confirmed using a confirmatory factor analysis (Study 2). Finally, across three experiments (Studies 3–5), the questionnaire was shown to be sensitive to changes in cognitive load supporting its predictive validity. The quality of the cognitive load questionnaire was underlined by satisfactory internal consistencies across all studies. In sum, the proposed questionnaire can be used in experimental settings to measure the different types of cognitive load in a valid and reliable manner. The construction and validation process of the questionnaire has also shown that the construct germane cognitive load remains controversial concerning its measurement and theoretical embedding in cognitive load theory.

DOI: https://doi.org/10.1007/s10648-023-09738-0

Educational Psychology Review (2023) 35:9
https://doi.org/10.1007/s10648-023-09738-0

INTERVENTION STUDY

Development and Validation of a Theory-Based Questionnaire to Measure Different Types of Cognitive Load

Felix Krieglstein[1] · Maik Beege[2] · Günter Daniel Rey[1] · Christina Sanchez-Stockhammer[3] · Sascha Schneider[4]

Accepted: 7 December 2022 / Published online: 28 January 2023

Abstract

According to cognitive load theory, learning can only be successful when instructional materials and procedures are designed in accordance with human cognitive architecture. In this context, one of the biggest challenges is the accurate measurement of the different cognitive load types as these are associated with various activities during learning. Building on psychometric limitations of currently available questionnaires, a new instrument for measuring the three types of cognitive load—intrinsic, extraneous, and germane cognitive load—is developed and validated relying on a set of five empirical studies. In Study 1, a principal component analysis revealed a three-component model which was subsequently confirmed using a confirmatory factor analysis (Study 2). Finally, across three experiments (Studies 3–5), the questionnaire was shown to be sensitive to changes in cognitive load supporting its predictive validity. The quality of the cognitive load questionnaire was underlined by satisfactory internal consistencies across all studies. In sum, the proposed questionnaire can be used in experimental settings to measure the different types of cognitive load in a valid and reliable manner. The construction and validation process of the questionnaire has also shown that the construct germane cognitive load remains controversial concerning its measurement and theoretical embedding in cognitive load theory.

Keywords Cognitive load measurement · Scale development · Factor analysis · Subjective scales · Questionnaire

Introduction

Since its first full description in the late 1980s, *Cognitive Load Theory* (CLT) has become a prominent and an influential theory in instructional psychology (Paas and Sweller, 2021; Paas et al., 2004). The theoretical assumptions of this

✉ Felix Krieglstein
felix.krieglstein@phil.tu-chemnitz.de

Extended author information available on the last page of the article

framework are used to design learning materials and instructional procedures as conducive to learning as possible. In order to optimize instructional design, various CLT recommendations have been tested for their effectiveness using randomized, controlled experiments (Sweller, 2021). Such experiments often involve a cognitive load measurement, which is particularly challenging because this latent construct is not directly observable or physically measurable (Ayres, 2018; McNeish, 2018). In this context, the learners' perceived cognitive load including its types is usually measured with self-rating scales in which individuals reflect on past learning experiences (i.e., retrospective judgments; e.g., Möller, 2014). Hereby, a psychological scale or measurement must meet the requirements of reliability and validity to adequately measure a construct such as cognitive load (Korbach et al., 2018). A closer look at current available questionnaires reveals difficulties with item formulations as well as psychometric ambiguities (see section "*Psychometric limitations of current CLT Questionnaires*"). Considering these challenges, this work aimed to develop and validate a new questionnaire to measure the learners' perceived cognitive load during learning more validly and reliably.

Literature Review

Foundations of Cognitive Load Theory

CLT is based on our knowledge of human cognitive architecture and evolutionary educational psychology (Sweller, 2020, 2021). According to this theory, processing complex and novel information in a learning situation burden the learner's working memory capacity (Sweller et al., 1998, 2011). In this context, CLT aims to explain how to efficiently use the limited working memory for successful learning, defined as the construction and automation of knowledge in long-term memory (Sweller et al., 2019). Central to CLT is the interplay of working memory and long-term memory (Sweller, 2016). The term *working memory* refers to brain systems that are activated when (complex) mental tasks such as language comprehension, problem-solving, or reasoning must be carried out (Baddeley, 1992). However, when dealing with complex, novel information, working memory reaches its limits. Accordingly, it is assumed that people can process only a limited number of elements simultaneously (Cowan, 2010; Miller, 1956). Moreover, there is empirical evidence one can hold novel information in working memory for not more than 20 to 30 s until it is lost (Peterson and Peterson, 1959). These limitations protect learners from an overload of new information (Sweller, 2016). The role of long-term memory within human cognition was decisively influenced by findings from de Groot (1965). Using the example of playing chess, de Groot could show that players who could draw on many years of experience were able to recognize a large number of meaningful board configurations. It is assumed that long-term memory is unlimited in its capacity to store information. Accordingly,

the knowledge stored in long-term memory can be retrieved when needed (Unsworth et al., 2013). Within long-term memory, a huge amount of retrievable information is organized into schemata (Bartlett, 1932). Schemata are defined as cognitive structures which bundle several related pieces of information into one element. In this vein, learning arises when people construct and automate schemata in long-term memory (Kirschner, 2002). By this, the limitations of working memory can be overcome when schemata from the long-term memory are activated in a given learning situation. Novices and experts, therefore, differ in the amount of domain-specific knowledge stored in their long-term memory (Sweller, 2021). Recently, the foundations of CLT have been expanded by incorporating findings from evolutionary psychology (Geary, 2008). Following Geary (2005, 2008), information can be divided into two categories – biologically primary and secondary information. Primary information is essential for human functioning and is learned without explicit instruction (e.g., learning one's native language). In contrast, acquiring secondary knowledge is associated with more effort and is often supported by an external person. In this context, an instructor (e.g., a teacher) gives his or her knowledge to a person who lacks that knowledge. In this context, people obtain secondary knowledge from other people in different ways, for example by listening to them or reading their texts (*borrowing and reorganising principle*; Sweller, 2021). Knowledge and skills, such as mathematics or the ability to read as secondary information, are therefore taught in educational and training contexts causing a high cognitive load on the learners' working memory (Paas and Sweller, 2012). Recommendations derived from CLT are primarily focused on the acquisition of secondary knowledge. Accordingly, cognitive load is usually measured in learning settings in which such knowledge is to be learned.

Types of Cognitive Load

Traditionally, the load imposed on the learner's cognitive system is divided into three types, namely *intrinsic* (ICL), *extraneous* (ECL), and *germane cognitive load* (GCL; Sweller et al., 1998). On a conceptual level, it is assumed that ICL and ECL are additives (Sweller, 2010; Sweller et al., 2011). The resulting total cognitive load determines the amount of required working memory resources needed for learning. Ideally, cognitive resources are directed to handle the intrinsic load (*germane processing*; Paas and van Merriënboer, 2020). Once the demands of the learning task exceed the capacity of working memory, a cognitive overload occurs (Mayer and Moreno, 2003). The aim of CLT is thus to avoid a cognitive overload during learning by providing appropriate instructional materials (de Jong, 2010). For this, knowledge of the individual cognitive load types is necessary.

Intrinsic Cognitive Load

Intrinsic cognitive load (ICL) refers to the load imposed by the learning task's complexity. Accordingly, the complexity is reflected in the element interactivity

describing the learning task's inherent number of elements and their interrelations (Sweller, 2010). A learning task can thus be classified on a continuum between low and high element interactivity depending on how much information needs to be processed simultaneously (Sweller and Chandler, 1994). The intrinsic load imposed on working memory can be reduced when learners can resort to already formed schemata stored in long-term memory—that is domain-specific prior knowledge (Sweller et al., 2019). In summary, a high intrinsic load arises when a complex task containing high element interactivity must be handled by a person who has little or no prior knowledge of the subject matter. From the perspective of an instructional designer, the aim is not to reduce the amount of complexity but rather to help students to manage the intrinsic load (Mayer and Fiorella, 2021). In this context, it is possible to equip learners with relevant prior knowledge which should help to process the learning contents (*pre-training principle*; Mayer et al., 2002). It is assumed that people learn better when they are familiarized with key terms before the actual learning material is presented (Mayer, 2017).

Extraneous Cognitive Load

Extraneous cognitive load (ECL) results from processes that are not relevant to learning. In general, ECL is determined by the presentation format of the learning material (Sweller et al., 2019). An inappropriate design of the learning material represents an unnecessary (or unproductive) load on the learners' working memory (Kalyuga and Singh, 2016). Instructional designers should therefore focus their attention on reducing the ECL to free up enough cognitive resources for learning-relevant activities. In this vein, several recommendations were derived from CLT (Paas et al., 2003). For example, it is recommended to avoid search processes within the learning material by physically and temporally integrating related, learning-relevant information (e.g., Schroeder and Cenkci, 2020). From a cognitive load perspective, spatially distant formats generate extraneous load because learners are forced to hold information in working memory while searching for referential information (*split-attention effect*; Sweller et al., 2011). Instructional designers should therefore follow the principles of spatial and temporal contiguity when presenting learning contents (for a meta-analysis, see; Schroeder and Cenkci, 2018).

Germane Cognitive Load

While the position of both the ICL and ECL within CLT framework is relatively indisputable, the role of the third type – *germane cognitive load* (GCL) – is still controversial (Jiang and Kalyuga, 2020; Kalyuga, 2011). GCL was added to the CLT framework at a later stage of development (Sweller et al., 1998) because it has become increasingly clear that cognitive load is also a necessary prerequisite for learning. To transfer information to long-term memory (i.e., schemata construction and automation), learners must actively invest cognitive resources (Moreno and Park, 2010). Consequently, the GCL as a learning-relevant (or *productive*) load should be as high as possible, as this is interpreted as a sign of engaged learners devoting their cognitive resources to learning. Following these assumptions, GCL

rather holds an allocation role in distributing working memory resources to learning-relevant activities (i.e., dealing with the intrinsic load; Sweller, 2010). As indicated above, ICL and ECL additively form cognitive load (Sweller, 2010; Sweller et al., 2011). Thus, learning materials should be designed to reduce ECL so that working memory resources are freed to deal with ICL (i.e., complexity) by mental effort investment. This allows sufficient cognitive resources to be devoted to ICL what is described as *germane processing* (Paas and van Merriënboer, 2020). Ideally, working memory resources are used for the construction and automation of schemata (Kirschner, 2002). These remarks suggest that ICL and GCL are closely interrelated (Kalyuga, 2011). Instructional designers apply various design principles to foster GCL while learning. One way is asking students to mentally form a picture of the key material described in a scientific text or an auditory explanation (*imagination principle*; Leopold and Mayer, 2015). *Seeing with the mind's eye* is assumed to support learning as learners actively stimulate learning elements and their relations resulting in coherent mental representations. In line with the *generative learning theory*, translating information depicted in a text to a mental image initiates the learner to perform learning-relevant activities such as selecting relevant information, organizing information into meaningful mental models, and integrating these models with prior knowledge (Fiorella and Mayer, 2016).

Measuring Cognitive Load

Measuring the different types of cognitive load is of high importance for researchers as this is a valid way to examine why some learning materials are more difficult to learn than others. However, valid and reliable measurement is an ongoing challenge in CLT research (Ayres, 2018; de Jong, 2010; Moreno, 2010). In this vein, measuring instruments must be able to differentiate between the types of cognitive load to better assess the effectiveness of specific instructional designs and principles. According to Kirschner et al., (2011, p. 104), developing instruments that meet the conceptual types of cognitive load, "has become the holy grail of CLT research."

In general, the most common approaches to measuring cognitive load can be classified into dual task measures, physiological parameters, and self-rating measures (Paas et al., 2003). Since dual-task methods and physiological measurements are outside the scope of this work, they should be explained only briefly for the sake of completeness. In the application of the dual-task method, a second task is added to the learning scenario (Brünken et al., 2002). Learners are therefore required to work on two tasks at the same time, with learning performances measured in both tasks. It is assumed that performance in the secondary task drops when the primary task consumes too many cognitive resources and therefore represents a high cognitive load. Furthermore, physiological parameters have been increasingly used in recent years to infer cognitive load. It is assumed that parameters such as pupil dilation (Sibley et al., 2011) or electroencephalography measures (Antonenko et al., 2010) are related to processes described in CLT. In general, collecting physiological data requires a greater effort and is often hardly economical (Klepsch et al., 2017). As with dual-task methods, physiological parameters are not able to differentiate

between types of cognitive load. While self-rating measures are highly subjective, dual-task measures and physiological parameters are rather objective measures of cognitive load. All of these attempts have their strengths and weaknesses whereby subjective scales of cognitive load are the most used type of measurement in educational psychology and beyond (e.g., Schmeck et al., 2015).

Measuring Cognitive Load with Unidimensional Scales

One of the first attempts to measure cognitive load in research and practice with subjective scales was made by Hart and Staveland (1988), who developed the *NASA Task Load Index (TLX)* focusing on perceived workload. It is used, for example, to measure the workload of nurses in intensive care units (Tubbs-Cooley et al., 2018) or pilots during simulated flight tasks (Mansikka et al., 2019) demonstrating the important role of this construct in human factors research. Accordingly, the perceived workload is defined and measured as a broader construct which makes it rather useless for use in educational scenarios. For educational psychology research, probably the most popular and widely used subjective instrument was developed by Paas (1992). This single-item asks learners to rate their perceived mental effort on a 9-point scale ranging from very, very low mental effort to very, very high mental effort. Consequently, the perceived mental effort is an indicator of the cognitive load caused by the learning task. A high level of invested mental effort can be understood as an indication of a complex learning task. Although the scale from Paas (1992) can be used in a less intrusive way in educational contexts, it does not differentiate between types of cognitive load. Moreover, using single items is connected with psychometric problems in terms of reliability as, for example, reliability indices need to have more than one item to be calculated. Besides the approach to measure the overall cognitive load, there have been attempts in recent years to measure the different types of cognitive load. However, one-item scales were often used in these contexts. For instance, Ayres (2006) used a rating scale ranging from extremely easy to extremely difficult to assess the learners' perceived intrinsic load in terms of a learning task. Cierniak et al. (2009) measured extraneous load with the question of how difficult it was to learn with the material. As pointed out by Leppink et al. (2013), the use of these scales is problematic concerning different scale points and labels. Moreover, some of these scales have not been validated.

Measuring Cognitive Load with Multidimensional Scales

An often-used multidimensional questionnaire for the differentiated measurement of cognitive load in complex knowledge domains was developed by Leppink et al. (2013). The scale consists of ten items measuring the individual types of cognitive load. Concerning the formulation, the ICL items focus on the perceived complexity of formulas, concepts, and definitions associated with the learning task. The higher the complexity, the higher the ICL should be assessed. In contrast, the three ECL items refer to the instruction's clearness and effectiveness. For example, unclear language would therefore increase extraneous load. The four GCL items focus on the learners' understanding of the formulas, concepts, and definitions. The questionnaire

was validated in the area of statistics, mainly because this area of knowledge is considered challenging for students. Leppink et al. (2013) found promising results supporting the three-factor model of cognitive load and replicated the questionnaire validation in another study (Leppink et al., 2014), but within a different learning domain (language learning). For this purpose, the item formulations were adapted to the new learning setting, whereby the three-factor solution was confirmed as well. However, because of problems with the GCL scale (lack of correlation with learning performance), one GCL-related item was added in a further study to explicitly capture the effort required to cope with cognitive load. Accordingly, the new cognitive load instrument consists of 13 items. However, in recent years, the 10-item measurement instrument has become established in experimental practice (e.g., Beege et al., 2019; Chung & Cheon, 2020; Schneider et al., 2021; Wang et al., 2021). Nevertheless, it seems conceptually logical that learners need to exert mental effort to cope with both intrinsic and extrinsic loads (Klepsch and Seufert, 2021).

Another attempt to measure the different types of cognitive load was made by Klepsch et al. (2017). Similar to the scale from Leppink et al., (2013, 2014), it consists of multiple items directly measuring the different cognitive load types. Two items were designed to measure ICL referring to the task's complexity and the number of elements of information that must be kept in mind during learning. ECL is measured with three items focusing on the difficulty to link crucial information, the design of the task, and finding important information. Depending on whether the learning material contains prompts that may elicit GCL, GCL is measured with either two or three items. The items concentrate on the effort to understand the overall context as well as the ambition to understand everything correctly. The third item is only useful when the GCL is intentionally varied, e.g., with prompts, and asks whether the learning task consisted of elements that supported task comprehension. Another limitation refers to the item formulations. Some of the items contain the designation "task" which can lead to potential misunderstandings. Since it is common in experimental studies in the field of educational psychology to conduct a learning test after learning, the items could lead to the learning test rather than the learning intervention being assessed in terms of cognitive load. To avoid this, appropriate instruction is necessary. Interesting to note is the fact that learners completed the Klepsch et al. (2017) scale either with or without prior knowledge of CLT. Students assigned to the informed rating group received an introduction to CLT and its types. This should enable the learners to be able to better differentiate between cognitive load types with the awareness of how the types are also related to each other. In contrast, learners in the naïve rating group received no introduction to CLT. As assumed, providing learners with knowledge leads to a more valid measurement of cognitive load. Participants were able to assess the cognitive load as postulated by theoretical assumptions. However, besides being a promising way to ensure a valid cognitive load measurement, providing learners with an introduction to CLT is not always possible.

The most recent approach to develop and validate a new CLT questionnaire was made by Dönmez et al. (2022). Similar to the mentioned instruments, the questionnaire measured the cognitive load in a differentiated way. Three of the ICL items focus on the learner's prior knowledge while one item is related to the learning topic

and asks respondents to decide whether the topic was quite strange to them. From a conceptual perspective, this wording is not fully aligned with CLT since it does not measure the complexity of the learning content. Five ECL items refer to the clearness and adequacy of the language and the instructions and explanations. The four GCL items of the questionnaire from Dönmez et al. (2022) are also disputable concerning the theoretical fit to CLT. Wordings like pleasure (item GCL2) and interesting (item GCL4) while learning are not considered in this cognitively oriented theory and express more affective feelings during learning. In addition, the active use of cognitive resources is not addressed in the items.

Psychometric Limitations of Current CLT Questionnaires

In recent years, the questionnaires by Leppink et al. (2013) and Klepsch et al. (2017) have become established in CLT research. In this vein, these instruments have been used in a variety of empirical studies to measure cognitive load. Nevertheless, a closer look reveals psychometric ambiguities. The questionnaire developed by Leppink et al. (2013) was verified with an exploratory and confirmatory factor analysis that is in line with common recommendations (e.g., DeVellis and Thorpe, 2021; Worthington and Whittaker, 2006) and has been done in other psychology-related scale developments (e.g., Alisat and Riemer, 2015; Exline et al., 2014; Shea et al., 2019). In addition, a randomized experiment was added to examine the effects of two different formats of worked examples (familiar vs. unfamiliar) on cognitive load types and learning outcomes. Here, theory-consistent results could be found. For instance, learners with a higher level of prior knowledge reported a lower ICL and the order of the examples had a significant impact on ECL perceptions. However, no separate experiments were performed to verify whether the questionnaire could explicitly differentiate between types of cognitive load, which would be an important criterion for assessing validity. Similarly, the paper from Dönmez et al. (2022) did not report any experimental studies in which the types of cognitive load were manipulated separately. In the validation paper of their questionnaire, Klepsch et al. (2017) have not reported any exploratory factor analysis. For example, it is not clear whether items had to be removed from the questionnaire because of factor loadings. Another critical point is related to the number of items per cognitive load type. Cognitive load types are measured with two (ICL) or three items (ECL, GCL). Measuring an abstract construct such as cognitive load with a small number of items is problematic (Carpenter, 2018). In this context, various methodologists recommend using at least four or five items per factor or construct (e.g., Costello and Osborne, 2005; Reise et al., 2000). Furthermore, the questionnaire by Klepsch et al. (2017) was developed and validated in the German language. To make the questionnaire applicable to the broad scientific public, an English translation was given. However, the paper did not explain in more detail how the translation was done—whether it was done, for example, by a native speaker or a linguist. The questionnaire from Dönmez et al. (2022) was developed with one exploratory factor analysis complemented by two confirmatory factor analyses which is in line with methodological recommendations (e.g., Worthington and Whittaker, 2006). However, from a psychometric view, the mixture of reversed

items (i.e., positive and negative items) could be problematic (e.g., Barnette, 2000). In this context, a study by Sonderen et al. (2013) showed that reverse-worded items may lead to inattention when responding to a questionnaire. Similarly, considering findings from Swain et al. (2008), respondents seem to make errors when the items do not reflect their experiences, i.e., claim the opposite. In sum, it is recommended to formulate all items in the same direction (Sonderen et al., 2013).

Construction of the Questionnaire

Process of the Questionnaire Development

The proposed questionnaire was developed following recommendations from various methodologists (e.g., Carpenter, 2018; DeVellis and Thorpe, 2021; Worthington and Whittaker, 2006). In a first step, the authors jointly developed a catalog of items of 17 items (Table 1), each measuring one of the three types of cognitive load (Greco et al., 2011). This was preceded by an extensive literature research to determine what the cognitive load items should measure (see section "*Types of Cognitive Load*"). In this context, the authors resorted to both foundational literature (e.g., Sweller, 1988, 1994; Sweller et al., 1998) as well as newer considerations (e.g., Sweller, 2020, 2021; Sweller et al., 2019) to formulate the items as precisely as possible. As suggested by Klepsch et al. (2017) as well as Leppink et al., (2013, 2014), the questionnaire developed in this work aims to measure the learners' perceived cognitive load by measuring all types of cognitive load separately with multiple items (see Table 1). For example, the items representing ICL should reflect the conceptual components of this cognitive load type (perceived task complexity as well as prior knowledge). Since ICL focuses on the learning content (along with the information it contains), the designation "learning content" was used in the items. Thus, it should be easier for learners to assess exactly what makes ICL unique. In contrast to previous questionnaires, one ICL item explicitly focuses on prior knowledge as an essential part of our understanding of intrinsic load (e.g., Sweller et al., 2019). Items intended to measure ECL were oriented to the design of the learning material and not its complexity. In this context, the items were formulated to fit a variety of sources of extraneous load (for a discussion see Krieglstein et al., 2022b). Since there is growing evidence that learners seem to have problems differentiating between ICL and ECL, the strict use of the terms "learning content" (ICL) and "design or structure of the learning material" (ECL) should make it easier for learners to separate the two loads in the assessment of learning experience. In addition, cognitive load types can be assigned to either active or passive load types (Klepsch and Seufert, 2021). The complexity (ICL) and design (ECL) of the learning intervention are experienced passively by the learner, while the learner must invest cognitive resources to cope with the passively experienced loads (see also Krell, 2017). In line with these assumptions, the items intended to measure both the ICL and ECL were formulated passively (e.g., by assessing the complexity of the learning content). In contrast, the items designed to measure GCL were formulated from a first-person perspective ("I"). This is intended to make it easier for learners to assess the

Table 1 Item catalog designed for the principal component analysis (PCA)

Cognitive load	Proposed item (in German)	Proposed item (in English)
ICL		
ICL1 (Item 1)	Die Lerninhalte waren schwer zu verstehen	The learning content was difficult to understand
ICL2 (Item 2)	Die Erklärungen des Lerninhalts waren schwer nachvollziehbar	The explanations of the learning content were difficult to understand
ICL3 (Item 3)	Die Lerninhalte waren komplex	The learning contents were complex
ICL4 (Item 4)	Die Lerninhalte enthielten viele komplexe Informationen	The learning content included much complex information
ICL5 (Item 5)	Ohne Vorwissen waren die Informationen nicht verständlich	Without prior knowledge, the information was not understandable
ECL		
ECL1 (Item 6)	Es war schwierig, einen Überblick über den Aufbau des Lernmaterials zu erlangen	It was difficult to gain an overview of the structure of the learning material
ECL2 (Item 7)	Die Gestaltung des Lernmaterials machte es schwer, die Zusammenhänge zwischen den einzelnen Informationen herzustellen	The design of the learning material made it difficult to recognise links between individual information units
ECL3 (Item 8)	Das Lernmaterial war ungünstig gestaltet	The design of the learning material was inconvenient
ECL4 (Item 9)	Die Gestaltung des Lernmaterials machte es schwierig, wichtige Informationen zügig zu finden	The design of the learning material made it difficult to find relevant information quickly
ECL5 (Item 10)	Aufgrund der Gestaltung des Lernmaterials hatte ich das Gefühl, mich nicht auf die Lerninhalte konzentrieren zu können	Because of the design of the learning material, I had the impression that I could not concentrate on the learning content
GCL		
GCL1 (Item 11)	Ich habe aktiv über die Lerninhalte nachgedacht	I actively reflected upon the learning content
GCL2 (Item 12)	Das Lernmaterial hat mich angeregt, aktiv über die Lerninhalte nachzudenken	The learning material encouraged me to actively think about the learning content
GCL3 (Item 13)	Ich habe mich bemüht, die Lerninhalte zu verstehen	I made an effort to understand the learning content
GCL4 (Item 14)	Ich war in der Lage, die einzelnen Informationen in einen Zusammenhang zu bringen	I was able to put the individual pieces of information into context
GCL5 (Item 15)	Ich habe ein umfassendes Verständnis der Lerninhalte erlangt	I achieved a comprehensive understanding of the learning content
GCL6 (Item 16)	Ich konnte mein bestehendes Wissen mit den Lerninhalten erweitern	I was able to expand my prior knowledge with the learning content
GCL7 (Item 17)	Das Wissen, das ich durch das Lernmaterial erworben habe, kann ich schnell und sicher anwenden	I can apply the knowledge that I acquired through the learning material quickly and accurately

active investment of their cognitive resources for learning. An exception is item 10 (ECL) in which learners are asked to estimate whether the design of the learning material caused them to lose concentration on the learning contents. Here, the first-person perspective was also used. This formulation was intended to help respondents to assess more accurately the extent to which the design of the learning material results in a negative effect on learning. In detail, the items intended to measure GCL refer to *germane processing* as the underlying definition assumes that working memory resources should be devoted to intrinsic load (Sweller et al., 2011). The GCL items consider this definition by focusing both on mental effort (Paas and van Merriënboer, 2020) as well as schema construction and automation, which is defined as the overall goal of learning (Moreno and Park, 2010). Thus, the items asked whether learners actively reflected upon the learning content and to what extent they made an effort to understand the learning content. To include the schema construction and automatization processes, items refer to the extent to which the learning content was comprehensively understood and the extent to which the existing prior knowledge could be expanded with the learned information. To record the aspect of a successful learning process, learners were asked whether they can apply the acquired knowledge quickly and accurately. It becomes clear that the five items intended to measure GCL follow a chronological order along the learning process. Learners must make a mental effort (i.e. actively understand the learning content). By this, a schema can be constructed and automated. The term "learning content" was consequently used to help learners to better assess the extent to which they devote cognitive resources to intrinsic load.

In *Study 1*, the item catalog was examined with the help of a principal component analysis (PCA) to determine the number of components (or factors). After the number of factors (cognitive load types) was determined, the item structure was further confirmed using confirmatory factor analysis (CFA) in *Study 2* (based on a new sample). This approach is generally accepted when developing questionnaires and is therefore used in educational psychology (Leppink et al., 2013) and beyond (e.g., Gim Chung et al., 2004; Lecerf and Canivez, 2018; Meichsner et al., 2016).

Procedure of the Measurement Validation

After these factor-analytical considerations, the proposed questionnaire should be tested in experimental settings to find out whether intentionally manipulated instructional designs are reflected in the cognitive load scales. For this purpose, three experimental studies are conducted, each varying one cognitive load type (e.g., Klepsch and Seufert, 2020), to determine whether the instrument can differentiate between the certain cognitive load types as would be expected based on theoretical assumptions as well as the intentional manipulation of the instructional design. For this purpose, each cognitive load type was manipulated separately (*Study 3*: ICL; *Study 4*: ECL; *Study 5*: GCL) to further validate the questionnaire. To make the proposed cognitive load questionnaire applicable to the broad scientific public in educational psychology research and beyond, an additional translation of the questionnaire

into English is described. For this reason, the construction process was assisted by a professor of English and digital linguistics.

Study 1: Principal Component Analysis

Because of the rather small sample size, a principal component analysis (PCA) was preferred over an exploratory factor analysis (EFA) as it is less stringent in terms of assumptions (e.g., Leppink et al., 2013). This multivariate statistical method aims to reduce the dimensionality of data into a few components (Ringnér, 2008).

Method

Participants For the first study, data from 69 participants (78.3% female) was collected. The sample consisted of students from *Chemnitz University of Technology* in Germany ($M_{age} = 23.28$; $SD_{age} = 3.28$). They were enrolled in the study courses media communication (55.1%), media and instructional psychology (42.0%), and others (2.9%). Most of the students were in the first (40.6%), third semester (26.1%), or fifth semester (23.2%) of their studies.

Instructional Material The learning material used in the first study dealt with the processes in a wastewater treatment plant. Therefore, the participants received a static picture including written explanations of the processes explaining how wastewater is purified in various basins before being discharged into neighboring rivers or streams. While the schematic picture of the processes was shown on the left-hand side, the corresponding explanations were presented in text form on the right-hand side (see Fig. 1).

Procedure The study was conducted online via the open-source survey software *LimeSurvey*. Participants, who were invited to participate via email mailing lists, were able to work on the study on their own. They were instructed in writing to study the learning material carefully, as they would then be asked to complete questionnaires. Afterward, participants completed the proposed cognitive load questionnaire. The 17 items were assessed on a 9-point Likert-type scale. This scale width was chosen because a meta-analysis by Krieglstein et al. (2022a) showed that nine scale points used in the existing cognitive load questionnaire were associated with higher reliability values. Participants were asked to fill in the questionnaire based on the following instruction: "Your task is to evaluate the previous learning intervention. Please answer the following statements on a Likert-type rating scale from 1 (not at all applicable) to 9 (fully applicable)." Since only the endpoints of the scale were labeled, the scale can be considered interval-equivalent (Wildt and Mazis, 1978). Such numerical scales are the common way to measure constructs like cognitive load (Ouwehand et al., 2021).

Analysis Plan

PCA was performed using *IBM SPSS Statistics* (IBM Corp., 2021). The catalog consisting of 17 items was factor-analyzed using oblique (i.e., oblimin) rotation for analytic rotation. Similarly, Costello and Osborne (2005) recommend using oblique rotation in the field of social sciences. As the responses were made on an interval-equivalent scale, the correlation matrix was analyzed. The number of components to be extracted was determined based on parallel analysis (O'Connor, 2000). This analysis is based on the parallel comparison between eigenvalues extracted from random data sets and the actual data (Buja and Eyuboglu, 1992). After calculating the analysis with subsequent verification of sampling adequacy (Dziuban and Shirkey, 1974), the next step was to retain and delete items based on previously defined criteria (Schreiber, 2021; Worthington and Whittaker, 2006). As pointed out by Carpenter (2018), it is common in scale development to remove items. Besides, it is important to meet the requirement of an optimal scale length to ensure participants' motivation (Carpenter, 2018). The decision whether an item should be retained or deleted should be made based on previously defined criteria. In general, it is assumed that an item can be deleted because of (1) a poor factor loading (<0.05; Mertler and Vannatta, 2001), and/or (2) cross-loadings (>0.30; Worthington and Whittaker, 2006). In addition, several methodologists recommend to measure a construct (or factor) with at least four to five items (e.g., Costello and Osborne, 2005; Fabrigar et al., 1999). In line with Worthington and Whittaker (2006) as well as Schreiber (2021), the PCA must be rerun with the revised version of the instrument. To ensure internal consistency of the components, McDonald's omega (ω; McDonald, 1999) was calculated for the revised cognitive load questionnaire.

Results

The adequacy of the sample was verified with the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test of sphericity (Bartlett, 1950; Kaiser, 1974). Results revealed that the sampling adequacy can be accepted with a middling value of KMO = 0.794 (Kaiser, 1974) indicating low partial correlations between variables. Furthermore, Bartlett's test of sphericity (Bartlett's $\chi^2(136) = 746.59$, $p < 0.001$) revealed that the data are suitable for factor analysis. The means with corresponding standard deviations, skewness, kurtosis, and component loadings are displayed in Table 2. Parallel analysis revealed that three components can be extracted from the data. After the principal component analysis has been performed, the number of items was reduced as described in the *Analysis plan* section. In general, two items had to be removed. The items intended to measure ICL and ECL could be maintained completely because of high factor loadings (0.77–0.90) and small cross-loadings (0.01–0.21) following recommendations from Worthington and Whittaker (2006). However, two items intended to measure GCL had to be removed for subsequent analysis steps. Specifically, item 14 was removed because it had the lowest factor loading (-0.49) and cross-loadings with the two other components (-0.14; -0.24). Item 12 was also removed from the questionnaire since it had the second lowest factor loading (-0.70) within the component. The cross-loadings of item

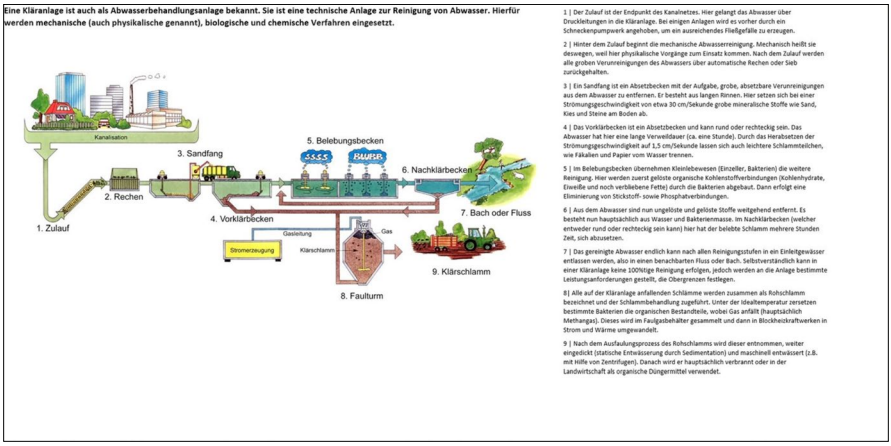

Fig. 1 Learning material used in Study 1 (presented in German)

12 (0.17; −0.15) also led to exclusion of this item. In addition, the inter-item correlation of the items was calculated because of its importance for item selection (McDonald, 1999). All items showed values above 0.50 which can be considered a good inter-item correlation (Kelava and Moosbrugger, 2020). Another important criterion for exclusion was the number of items. By removing item 12 it was possible to ensure that all components contained the same number of items resulting in a set of 15 items. Subsequently, the PCA was repeated with the revised instrument. The item reduction increased the KMO to 0.835 which can be interpreted as a meritorious value (Kaiser, 1974). Again, the Bartlett test of sphericity (Bartlett's $\chi^2(105)=726.39$, $p<0.001$) revealed data fit. As can be seen in Table 2, the item reduction did not lead to any noteworthy changes in the factor loadings and cross-loadings for the components. The final solution consisted of 15 items underlying three components that explained 74.3% of the total variance which is above the recommended value of 50% (Henson and Roberts, 2006). Interestingly, the signs of the factor loadings of the items intended to measure the GCL reversed (negative to positive). Internal consistencies were satisfactory across the cognitive load types ($\omega_{ICL}=0.93$; $\omega_{ECL}=0.91$; $\omega_{GCL}=0.88$). Unsurprisingly, the overall reliability of $\omega=0.61$ was rather low, which is reasonable when considering that the individual items are intended to measure different types of cognitive load. Furthermore, correlations between the three components were calculated. ICL and ECL were positively correlated ($r=0.29$; $p=0.015$). In contrast, ICL and GCL ($r=-0.36$; $p=0.002$) as well as ECL and GCL ($r=-0.38$; $p=0.002$) were negatively correlated.

Study 2: Confirmatory Factor Analysis

In the next step, the retained items were examined using confirmatory factor analysis—a more restrictive form of factor analysis (CFA; Lance and Vandenberg, 2002). It is a deductive approach intended for hypothesis testing focusing on the examination of relations among latent constructs (in this case, cognitive load; Jackson et al.,

Table 2 Descriptive values and component loadings of the items of the principal component analysis (Study 1)

Component/item	Mean (*SD*)	Skewness	Kurtosis	Inter-item correlation	Component loadings before and (after item reduction)		
					CL1	CL2	CL3
Component 1 (ICL)							
Item 1	2.70 (1.65)	1.62	2.69	0.80	.86 (.88)	−.11 (.01)	−.03 (.03)
Item 2	2.75 (1.79)	1.44	1.53	0.84	.86 (.86)	.04 (.03)	.10 (−.10)
Item 3	3.30 (1.87)	1.15	0.78	0.83	.89 (.89)	.02 (.00)	−.03 (.01)
Item 4	3.41 (1.90)	1.30	1.29	0.84	.89 (.90)	.08 (.08)	−.09 (.08)
Item 5	2.86 (1.90)	1.30	1.28	0.77	.81 (.82)	−.07 (−.06)	.12 (−.12)
Component 2 (ECL)							
Item 6	3.32 (2.19)	0.87	−0.12	0.74	.21 (.23)	.77 (.76)	.01 (−.01)
Item 7	4.06 (2.29)	0.32	−0.99	0.77	.08 (.07)	.82 (.82)	.01 (−.03)
Item 8	4.38 (2.46)	0.41	−1.02	0.82	−.18 (−.18)	.88 (.91)	.11 (−.11)
Item 9	4.49 (2.23)	0.21	−0.98	0.77	−.06 (−.04)	.90 (.91)	−.10 (.10)
Item 10	3.93 (2.45)	0.47	−0.81	0.77	.02 (.03)	.82 (.83)	.05 (−.04)
Component 3 (GCL)							
Item 11	6.17 (1.77)	−0.52	−0.39	0.80	.14 (.16)	.04 (−.02)	−.94 (.95)
*Item 12**	*5.29 (2.27)*	*−0.18*	*−1.04*	*0.58*	*.17*	*−.15*	*−.70*
Item 13	7.20 (1.65)	−1.37	1.91	0.64	.03 (.04)	.01 (−.05)	−.76 (.77)
*Item 14**	*6.72 (1.54)*	*−0.62*	*0.02*	*0.53*	*−.14*	*−.24*	*−.49*
Item 15	5.41 (1.68)	−0.33	−0.38	0.74	−.07 (−.05)	.01 (−.04)	−.82 (.82)
Item 16	6.14 (2.13)	−0.63	−0.44	0.70	−.24 (−.22)	.03 (.01)	−.73 (.74)
Item 17	4.74 (1.93)	−0.14	−0.63	0.67	−.08 (−.05)	.07 (.04)	−.77 (.78)

*Item 12 and Item 14 were removed from the questionnaire and are therefore shown in italics

2009). Since the previous PCA resulted in a three-component solution and previous cognitive load questionnaires also relied on a three-factor model (Dönmez et al., 2022; Klepsch et al., 2017; Leppink et al., 2013), the a priori hypothesis explicitly assumed that the observed ratings of the items loaded on three latent variables (factors). Besides the evaluation of the proposed cognitive load questionnaire with its latent structure, the second objective was to test for construct validity defined as the embedding of the measure in a nomological network with other, theoretically aligned variables (Cronbach and Meehl, 1955). Specifically, it should be checked whether the developed questionnaire meets the psychometric requirement of convergent validity. Convergent validity is present when a proposed questionnaire is related to other scales that measure the same construct (Krabbe, 2017). In this context, the questionnaire by Klepsch et al. (2017) was used in addition to the proposed instrument. To ensure convergent validity, both instruments (or rather the respective

cognitive load types) should correlate. For example, our proposed ICL items should correlate with the ICL items proposed by Klepsch et al. (2017). If this case occurs, it can be assumed that our instrument measures something similar to the cognitive load questionnaire of Klepsch et al. (2017).

Method

Participants For the second study, participants were recruited via *Prolific* (https://www.prolific.co/), a platform for recruiting participants for online experiments or surveys with selected target groups. There is empirical evidence proving the high data quality of *Prolific* for behavioral research (e.g., Eyal et al., 2021). In line with recommendations from various methodologists (e.g., Bentler and Chou, 1987), a ratio of 10:1 (participants to items) was used to determine appropriate sample size. Overall, 158 participants ($M_{age} = 32.08$; $SD_{age} = 10.81$) took part in the second study. 44.3% of the participants were female. The majority of the participants were employees (45.6%) or students (32.9%). All participants spoke German as their native language to ensure that both the learning material as well as the cognitive load questionnaire could be understood adequately.

Instructional Material In accordance with the first study, the learning material of the second study consisted of a static picture accompanied by textual explanations. The static picture illustrated the schematic structure of a nerve cell. Hereby, the picture and eight corresponding labels were presented spatially separated. In addition to the name of the respective component (e.g., axon, soma, and myelin sheath), information about the function was also given (see Fig. 2).

Procedure Again, the study was conducted online via the open-source survey software *LimeSurvey*. After participants received the invitation to participate from *Prolific*, they could start the study independently. Participants were instructed in writing to study the learning material carefully, as they would then be asked to complete questionnaires. After engaging with the learning material, participants completed two cognitive load questionnaires (the proposed questionnaire and the questionnaire by Klepsch et al., 2017). In order to avoid potential confounding effects from specific item orders, the items were presented randomly. The items were assessed by the participants based on a 9-point Likert-type scale (1 = not at all applicable; 9 = fully applicable).

Analysis Plan

A confirmatory factor analysis was performed using the robust maximum likelihood (robust ML) estimation. This estimator was used since the assumption of multivariate normal distribution was violated and robust ML is less dependent on the assumption of multivariate normality (Li, 2016). The 15 remaining items from *Study 1* were factor analyzed to check whether the items load on the three proposed cognitive load types. Analysis was performed using *RStudio* (RStudio Team, 2022). Oriented to Hu

and Bentler (1998), the model fit was checked relying on several indices, namely the comparative fit index (CFI), the root-mean-square error of approximation (RMSEA), the goodness-of-fit index (GFI), and the standardized root mean squared residual (SRMR). Values greater than 0.95 are usually interpreted as an acceptable fit for the CFI (Bentler, 1990), while values greater than 0.90 indicate an acceptable fit for the GFI (Marsh and Grayson, 1995). Concerning the RMSEA, the value should not exceed the cut-off 0.10 (Browne and Cudeck, 1992). With respect to the SRMR, the value should be close to 0.08 or below (Hu and Bentler, 1998).

Results

Bartlett's test of sphericity (Bartlett's $\chi^2(87) = 210.98$, $p < 0.001$) confirmed sample fit for factor analysis. The initial model showed an acceptable fit of the data (CFI = 0.92; RMSEA = 0.10; GFI = 0.92; SRMR = 0.07). Factor loadings revealed that all items load on the intended factor ($p < 0.001$). A closer look shows that the items intended to measure ICL and ECL have high factor loadings ranging from 0.73 to 0.93. For the GCL, however, the factor loadings were lower, ranging from 0.35 to 0.91. The items 11 (0.41) and 13 (0.35) showed lower factor loadings. These items were not deleted to maintain the theoretical convergence of the construct (e.g., Carpenter, 2018). Table 3 illustrates descriptive values, squared multiple correlations as an indicator of item reliability, and factor loadings. In addition, the internal consistency of the three factors (respectively cognitive load types) can be assessed as good to very good ($\omega_{ICL} = 0.93$; $\omega_{ECL} = 0.93$; $\omega_{GCL} = 0.80$). To check for convergent validity, correlations between the proposed cognitive load items and the items by Klepsch et al. (2017) were calculated with Pearson's correlation coefficient r (Table 4). In sum, evidence for moderate convergent validity could be found. With regard to ECL, the correlation showed satisfactory convergent validity considering the recommendations of Carlson and Herdman (2012). The correlation between the ICL items was lower, but still in the acceptable range. For the GCL items, the lowest correlation was found. The problems concerning this cognitive load type will be taken up again in the "*General Discussion*" section.

Study 3: Varying Intrinsic Cognitive Load

Lastly, three experiments were added to verify the sensitivity of the proposed cognitive load questionnaire. All three studies follow the identical methodological approach by manipulating one type of cognitive load resulting in similarities between the studies in terms of procedure and measures. They are all between-subjects designed by randomly assigning participants to either the experimental or control group.

Participants

For the last three studies, 54 students from *Chemnitz University of Technology* in Germany were recruited who participated in each study. The sample consisted of

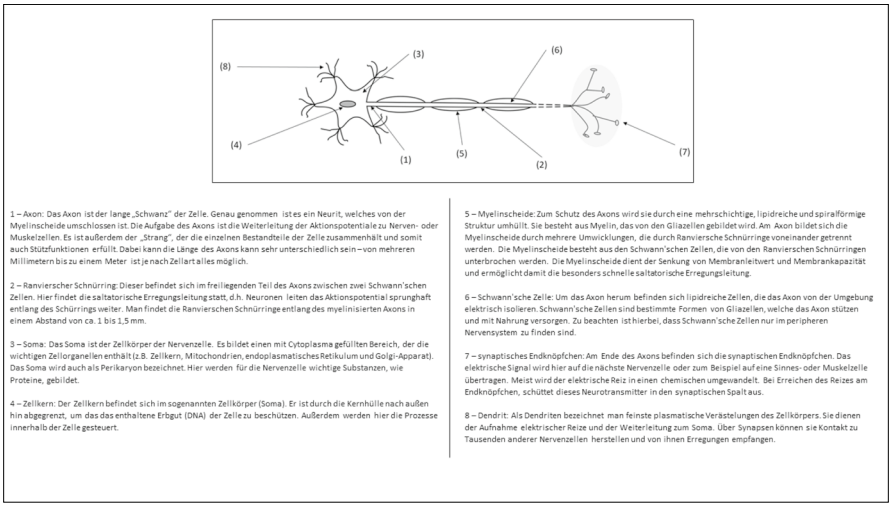

Fig. 2 Learning material used in Study 2 (presented in German)

bachelor and master students who were studying media communication, media, and instructional psychology, computer science and communication studies, and other courses. Table 5 provides an overview of the distribution of demographical data for each study.

Procedure

The procedure of the three experimental studies is almost identical (possible deviations from the following explanations are indicated in the respective studies). All studies were conducted online via the open-source web conferencing system *BigBlueButton*. Participants were instructed to focus only on the learning content as they would then answer questions about it. In this context, participants were asked to share their screens to ensure that participants continuously worked with the learning material and did not check other websites. No personal data was viewed or recorded. Taking notes was also forbidden. At the beginning of each study, participants signed an informed consent form and were instructed about the procedure of the study. As a next step, participants' prior knowledge was assessed with some open-answer questions since it influences cognitive load perceptions and consequently learning performance (e.g., Zambrano et al., 2019). Participants were then randomly assigned to the experimental or control group and worked on the learning material. Directly after the learning intervention, the cognitive load questionnaire had to be filled in. Each cognitive load type was captured with five items. All items had to be rated on a 9-point Likert-type scale ranging from 1 (not at all applicable) to 9 (fully applicable). Finally, the participants worked on the learning test. In the last part of each study, the participants had to work on the knowledge test followed by the request for demographic data. Since the three studies included different learning topics, prior knowledge and learning tests were created separately for each study. All parts of the

Table 3 Descriptive values and factor loadings of the items of the confirmatory factor analysis (Study 2)

Factor/item	Mean (*SD*)	Skewness	Kurtosis	Inter-item correlation	Factor loading	Squared multiple correlation (R^2)
Factor 1 (ICL)						
Item 1	5.37 (2.26)	−0.26	−0.88	.89	.93	.86
Item 2	4.54 (2.15)	0.08	−0.93	.85	.88	.78
Item 3	6.15 (1.99)	−0.71	−0.20	.82	.87	.75
Item 4	6.49 (2.06)	−0.69	−0.27	.83	.88	.77
Item 5	4.90 (2.22)	0.02	−1.04	.70	.73	.54
Factor 2 (ECL)						
Item 6	4.30 (2.24)	0.16	−1.13	.75	.78	.61
Item 7	4.70 (2.43)	0.08	−1.22	.83	.88	.78
Item 8	4.70 (2.41)	0.02	−1.19	.87	.91	.82
Item 9	4.98 (2.52)	−0.09	−1.26	.83	.87	.76
Item 10	4.31 (2.37)	0.29	−1.01	.78	.82	.67
Factor 3 (GCL)						
Item 11	6.65 (1.85)	−0.92	0.56	.54	.41	.17
Item 13	7.84 (1.33)	−1.96	6.29	.44	.35	.12
Item 15	4.63 (2.06)	−0.03	−0.86	.66	.86	.74
Item 16	6.24 (2.23)	−0.80	−0.16	.56	.55	.30
Item 17	4.21 (2.05)	0.20	−0.63	.67	.91	.83

Table 4 Correlations between the cognitive load types and the cognitive load items by Klepsch et al. (2017) in Study 2

Variables	1	2	3	4	5	6
1. ICL	–					
2. ECL	.55***	–				
3. GCL	−.45***	−.35***	–			
4. $ICL_{Klepsch}$	.56***	.29***	−.06	–		
5. $ECL_{Klepsch}$	.68***	.83***	−.44***	.47***	–	
6. $GCL_{Klepsch}$	−.11	−.20*	.44***	.23**	−.13	–

*$p < .05$, **$p < .01$, ***$p < .001$

studies (i.e., prior knowledge, learning material, CLT questionnaire, and learning test) were connected via hyperlinks. Each study lasted about 20 to 25 min. A maximum of four participants took part in the survey concurrently.

Analysis Plan

Each of the three studies was analyzed in the same way using *IBM SPSS Statistics* (IBM Corp., 2021). Prior to the analyses, it was checked whether the two

Table 5 List of experimental studies and corresponding demographical data (Studies 3–5)

Study	Learning material	Experimental groups	n	M_{age} (SD_{age})	% female
3 – ICL manipulation	Pulley system	Pre-training	26	23.81 (3.64)	84.0%
		No pre-training	28	23.14 (3.00)	82.1%
4 – ECL manipulation	Cellular respiration	Integrated format	26	22.88 (2.67)	72.0%
		Separated format	28	24.00 (3.78)	92.9%
5 – GCL manipulation	Human respiratory system	Imagination	27	23.85 (3.42)	81.5%
		No imagination	27	23.07 (3.22)	84.6%

groups to be compared were similar regarding several control variables. In this context, possible differences between the two groups about prior knowledge, age, gender, subject of study, and the distribution of bachelor and master students were assessed by resorting to independent *t*-tests and chi-square tests. By this, comparable groups as a result of randomization should be ensured (Suresh, 2011). Since the learners' domain-specific prior knowledge was assessed with open-answer questions across all three studies, two independent raters evaluated the answers based on a list of correct answers. Their agreement was calculated with Cohen's kappa (κ; McHugh, 2012) – an indicator for interrater reliability. In case of disagreement between the two raters, the average of the two ratings was calculated. Independent *t*-tests were conducted to check whether the experimental and control group differed in terms of the cognitive load types ICL, ECL, and GCL, as well as learning performance. Based on the means and standard deviations, the effect size of Cohen's *d* was calculated. For interpretation, the benchmarks proposed by Cohen (1988) were followed assuming that 0.20 is a small, 0.50 is a medium, and 0.80 is a large effect size. Since the *t*-test is a parametric statistic assuming a specific distribution, the homogeneity of variance was checked with Levene's test (Glass, 1966). It was omitted to check for normal distribution because commonly used tests used for verification (e.g., Kolmogorov–Smirnov test) do not have sufficient statistical power for small sample sizes and the *t*-test reacts robustly to a violation of normality (Lumley et al., 2002). Internal consistencies, measured with McDonald's ω, are presented for all cognitive load scales in all three studies in Table 6. Moreover, correlations between prior knowledge, cognitive load types, and learning performance were calculated separately for each study (Table 7). The results of the three studies will be discussed in the "*General Discussion*" section.

Instructional Material and Measures

In *Study* 3, ICL was varied by applying the *pre-training principle* (Mayer et al., 2002). Based on Mayer et al. (2002), the experimental group received a pre-training before learning, whereas the control group received no pre-training. Since learners are equipped with prior knowledge resulting from the

Table 6 Internal consistencies for each cognitive load type in Studies 3 to 5

Study	McDonald's ω		
	ICL	ECL	GCL
3 – ICL manipulation	.91	.93	.85
4 – ECL manipulation	.88	.95	.60
5 – GCL manipulation	.93	.96	.81

intervention, they should report a lower ICL than the control group (e.g., Mayer and Moreno, 2003). This difference should consequently be reflected in the questionnaire. Concerning ECL and GCL, there should be no differences between the experimental and control group because these types of cognitive load were not intentionally manipulated. Overall, 54 students participated in this study, whereby 26 students were assigned to the experimental group (with pre-training) and 28 students were assigned to the control group (no pre-training). The two groups did not differ significantly in terms of the participants' age ($p = 0.233$) and prior knowledge ($p = 0.344$). Furthermore, chi-square tests revealed no differences concerning gender distribution ($p = 0.857$), subject of study ($p = 0.477$), and the distribution of bachelor and master students ($p = 0.835$). The learning material consisted of an instructional text explaining the spatial structure and functioning of a pulley system (208 words). The text was taken from Eitel et al. (2013). Prior to the learning intervention, the experimental group received a schematic illustration of a pulley system that was taken from Hegarty (2005). In contrast, participants in the control group received no pre-training and started directly with the learning material. The learner's domain-specific prior knowledge was assessed with two open-answer questions in which the participants had to explain how a pulley system works and what purpose it serves. The two raters showed a strong agreement in the evaluation of the answers (question 1: $\kappa = 0.83$; question 2: $\kappa = 0.88$). Overall, the students showed rather low prior knowledge ($M = 1.04$; $SD = 1.03$; maximum of five points). Learning performance was measured with eleven decision questions in which statements (e.g., "If the rope is deflected over two pulleys, you only have to pull on the rope with half the force") had to be assessed as either true or false.

Results

Means and standard deviations of the prior knowledge and dependent variables are presented in Table 8. As expected, the group with pre-training reported a significantly lower ICL than the group without pre-training, $t(52) = 1.91$, $p = 0.031$, $d = 0.52$. In contrast, no significant difference between the groups was found for ECL, $t(52) = 1.56$, $p = 0.062$, $d = 0.43$. Rather unexpectedly, it was found that the pre-training group reported a significantly higher GCL than the no pre-training group, $t(52) = 1.91$; $p = 0.031$; $d = 0.52$. Considering learning performance, no significant difference between the pre-training and no pre-training group could be found, $t(52) = 0.40$, $p = 0.345$, $d = 0.11$.

Table 7 Correlations between prior knowledge, cognitive load types, and learning performance in Studies 3 to 5

	1	2	3	4
Study 3 (ICL manipulation)				
1. Prior knowledge	–			
2. ICL	− .08	–		
3. ECL	− .22	.81***	–	
4. GCL	.42**	− .46***	− .45***	–
5. Learning performance	.31*	− .13	− .19	.15
Study 4 (ECL manipulation)				
1. Prior knowledge	–			
2. ICL	− .53***	–		
3. ECL	− .36**	.79***	–	
4. GCL	.44***	− .58***	− .50***	–
5. Learning performance	.31*	− .41**	− .34*	.52***
Study 5 (GCL manipulation)				
1. Prior knowledge	–			
2. ICL	− .43***	–		
3. ECL	− .33***	.69***	–	
4. GCL	.24	− .51***	− .47***	–
5. learning performance	.34*	− .24*	− .27*	.44***

*$p < .05$, **$p < .01$, ***$p < .001$

Study 4: Varying Extraneous Cognitive Load

Instructional Material and Measures

In *Study* 4, ECL was manipulated with the help of the *split-attention effect* (Ayres and Sweller, 2021). The experimental group received the learning material in a separated format in which corresponding elements were presented spatially separated from each other (e.g., Cierniak et al., 2009). In the control group, related elements were presented in an integrated format, that is, without spatial distance. Drawing on theoretical assumptions (Sweller et al., 2011) as well as empirical findings (Pouw et al., 2019), the spatially separated format should result in higher ECL. Consequently, the experimental group should report higher ECL than the control group. Concerning ICL and GCL, no differences were expected. The 54 participating students were assigned to either the integrated format (control group: $n = 26$) or the separated format (experimental group: $n = 28$). Both groups did not differ in terms of their age ($p = 0.110$) as well as prior knowledge ($p = 0.360$). Chi-square tests revealed that the two groups differed in terms of gender distribution ($p = 0.044$), but not to the subject of study ($p = 0.575$) as well as the distribution of bachelor and master students ($p = 0.358$). Since gender is not a relevant variable for explaining

Table 8 Means and standard deviations of the prior knowledge and all dependent variables (Study 3)

	Pre-training ($N=26$)		No pre-training ($N=28$)	
	M	*SD*	*M*	*SD*
Prior knowledge (0–5)	1.10	1.02	0.98	1.06
ICL (1–9)	3.76	1.91	4.77	1.98
ECL (1–9)	5.18	2.33	6.13	2.15
GCL (1–9)	6.60	1.28	5.92	1.33
Learning performance (0–11)	7.96	1.59	8.14	1.72

cognitive load effects, it was not included as a covariate in the following analyses.[1] The learning material consisted of a schematic illustration accompanied by text labels displaying partial steps and material transformations of cellular respiration. In detail, the four phases of glycolysis, oxidative decarboxylation, citrate cycle, and respiratory chain were illustrated to explain the metabolic pathway in which glucose is broken down and adenosine triphosphate is produced. For the control condition (integrated format), corresponding labels were displayed close to the picture. In contrast, for the experimental condition (separated format), the picture and corresponding labels were presented separately. The text labels were located under the picture to generate a spatial distance between corresponding learning elements (e.g., de Koning et al., 2020). Learners' domain-specific prior knowledge was assessed with three open-answer questions (e.g., "What is the difference between aerobic and anaerobic respiration?"). The two raters showed a strong agreement in the evaluation of the answers (question 1: $\kappa=0.91$; question 2: $\kappa=0.91$; question 3: $\kappa=0.95$). Overall, the students showed rather low prior knowledge ($M=1.32$; $SD=1.55$; maximum of six points). After the participants engaged with the learning material, the cognitive load questionnaire was presented. Lastly, participants worked on the knowledge test consisting of five multiple-choice questions (e.g., "Where do the subprocesses of cellular respiration take place?" given with the answer options cytosol, nucleus, mitochondrion, cell membrane, and cell wall) and five short open-answer questions (e.g., "How much adenosine triphosphate is released during the respiratory chain?"), resulting in a total of 29 points.

Results

Means and standard deviations of the prior knowledge and dependent variables are presented in Table 9. Because Levene's test showed that variance homogeneity was violated for the dependent variables ICL and ECL, non-parametric tests were conducted. A Mann–Whitney *U* test revealed that the separated format resulted in a

[1] Multivariate analyses of covariance (MANCOVAs) with gender as covariate indicated that gender had no significant influence on ICL, ECL, GCL, as well as learning performance.

significantly higher ICL than the integrated format, $U = 228.50$; $Z = 2.35$; $p = 0.019$; $d = 0.68$. As expected, students in the separated format reported a significantly higher ECL than students in the integrated format, $U = 158.00$; $Z = 3.57$; $p < 0.001$; $d = 1.12$. With respect to GCL, no significant difference occurred between the integrated and separated format, $t(52) = 0.98$; $p = 0.165$; $d = 0.27$. In terms of learning, it could be shown that the integrated format group achieved significantly higher learning performance than the separated format group, $t(52) = 2.05$; $p = 0.023$; $d = 0.56$.

Study 5: Varying Germane Cognitive Load

Instructional Material and Measures

In *Study* 5, GCL was varied by applying the *imagination principle* (Leopold, 2021). Both the experimental (imagination) as well as the control group (no imagination) received the same instructional text. However, while the experimental group received a specific imagery instruction, the control group did not receive such an instruction (e.g., Leopold and Mayer, 2015; Leopold et al., 2019). The 54 participating students were assigned to either the imagination ($n = 27$) or the no imagination condition ($n = 27$). The groups did not differ in terms of their age ($p = 0.197$) as well as their prior knowledge ($p = 0.287$). Chi-square tests revealed that the two groups were similar in terms of gender distribution ($p = 0.761$), the subject of study ($p = 0.470$) as well as the distribution of bachelor and master students ($p = 0.776$).

The learning material consisted of nine slides (787 words) explaining how the human respiratory system works and was adapted from Leopold and Mayer (2015). On each slide, one separate paragraph was presented. In detail, the structure of the human respiratory system was first explained, and then the processes of inhalation, gas exchange, and exhalation were illustrated. The explanations were all text-based, and no pictures were implemented. Within the learning environment, participants could click on the next button to go to the next slide, but going back to the previous slide was not possible. To realize the GCL manipulation, an imagination instruction was added under the paragraph, for example, "Please imagine the steps in the thoracic cavity and airways when the diaphragm and rib muscles receive a signal to inhale." In contrast, the control group received no imagination instruction. The participants' prior knowledge was assessed with four open-answer questions (e.g., "What is the function of the diaphragm in human respiration?"). Again, two independent raters evaluated the answers based on a list of correct answers. They showed a strong agreement in the evaluation of the answers (question 1: $\kappa = 0.92$; question 2: $\kappa = 0.88$; question 3: $\kappa = 0.91$; question 4: $\kappa = 0.87$). Learning performance was assessed with five short open-answer questions. For example, learners were asked to name the components of the respiratory system. In this context, participants could achieve a maximum of 16 points.

Table 9 Means and standard deviations of the prior knowledge and all dependent variables (Study 4)

	Integrated format ($N=26$)		Separated format ($N=28$)	
	M	*SD*	*M*	*SD*
Prior knowledge (0–6)	1.40	1.59	1.25	1.55
ICL (1–9)	5.98	1.87	7.16	1.23
ECL (1–9)	4.62	2.40	7.07	1.75
GCL (1–9)	5.59	1.41	5.24	1.26
Learning performance (0–29)	19.04	4.00	17.04	3.16

Results

Means and standard deviations of the prior knowledge and dependent variables are presented in Table 10. In line with our assumptions, independent *t*-tests revealed that the imagination and no imagination group did not differ significantly with respect to ICL, $t(52)=0.32$; $p=0.376$; $d=0.09$, and ECL, $t(52)=0.79$; $p=0.218$; $d=0.21$. In contrast, the imagination group reported a significantly higher GCL than the no imagination group, $t(52)=1.73$; $p=0.044$; $d=0.47$. It could be also confirmed that the imagination outperformed the no imagination group with regard to learning performance, $t(52)=2.02$; $p=0.024$; $d=0.55$.

General Discussion

The aim of this work and the five studies involved was to develop and validate a new instrument to measure perceived cognitive load types with multiple items. Current available cognitive load questionnaires show methodological limitations in the process of development and validation upon closer examination. Taking up these points, this work aimed at depicting the process of developing and validating a theory-based cognitive load questionnaire. As recommended in the literature, an exploratory approach (PCA) was initially chosen to reduce the dimensionality of the data. Subsequently, the three-factor structure (representing the three cognitive load types) was verified using a deductive approach (CFA). In the last step, the proposed cognitive load questionnaire was evaluated in terms of predictive validity by testing whether it is sensitive to intentional manipulation of each cognitive load type. Across three experimental studies, empirically proven design principles from multimedia learning research were applied. The result of this work is a questionnaire consisting of 15 items that researchers can use to measure cognitive load in a differentiated way (Table 11).

In more detail, the PCA confirmed the assumed three-components model consisting of the three types proposed in CLT (Sweller et al., 1998). Items intended to measure ICL and ECL showed particularly high factor loadings and low cross-loadings so that none of the items had to be removed. For the items intended to

Table 10 Means and standard deviations of the prior knowledge and all dependent variables (Study 5)

	Imagination ($N=27$)		No imagination ($N=27$)	
	M	*SD*	*M*	*SD*
Prior knowledge (0–7)	1.44	1.26	1.67	1.61
ICL (1–9)	3.73	1.78	3.88	1.81
ECL (1–9)	4.28	2.25	4.79	2.46
GCL (1–9)	6.97	0.91	6.45	1.26
Learning performance (0–16)	8.44	2.83	6.96	2.55

measure GCL, the two items with the lowest factor loading were removed resulting in a questionnaire consisting of 15 items. As expected, the reduction did not result in any noteworthy changes in the factor loadings of the ICL and ECL items. Concerning the GCL items, the initially negative factor loadings turned positive. Thus, higher levels in the assessment of the items resulted in a higher GCL. In terms of item formulation, this result is important because the items are formulated in such a way that a higher value in the positive direction increases the respective type of cognitive load. The individual components showed high reliabilities (i.e., internal consistencies), whereas the internal consistency across all three components was rather low. This is an indication that they measure different facets of cognitive load which is in line with CLT's differentiated view on cognitive processes while learning (e.g., Moreno and Mayer, 2010). The CFA confirmed the proposed three-factor model. While the factor loadings of the ICL and ECL items were very satisfactory, factor loadings of the GCL items showed lower values but items are still related to the construct GCL (i.e., they measure it). It seems that GCL is at least partially related to the other types of cognitive load. This assumption has already been widely discussed in CLT research (e.g., Kalyuga, 2011; Sweller, 2010). In this context, it is argued that GCL and ICL are interrelated as germane resources must be expended to manage the ICL. Concerning ECL, it can be assumed that germane resources are reduced when the learning material is inappropriately designed (e.g., by unnecessary search processes). One possible explanation for the correlated factors can be found in method variance which is a potential problem in questionnaire research. In this context, part of the variance is attributable to the measurement method rather than to the measured constructs (e.g., Podsakoff et al., 2003). Variance in item responses is biased by the measurement method (in this case, self-reported items). Thus, correlations between two constructs could be artificially inflated when learners adjust their response tendencies (e.g., Lindell and Whitney, 2001). In order to reduce this bias, future research could implement a method factor including all items in the tested model (for an example from personality psychology, see Biderman et al., 2011). By this, it could be determined whether correlations between the three cognitive load types are reduced.

In the experimental validation, two challenges had to be mastered: (1) the intentional manipulation of the respective cognitive load type had to be successful, and

(2) the learners' responses in the questionnaire must correspond to what would be expected based on theoretical assumptions as well as the intentional manipulation of the instructional design. *Study 3* showed that the proposed ICL items are sensitive to complexity differences since the pre-training group reported a significantly lower ICL. However, it could be shown that the pre-training manipulation also resulted in significant GCL effects. In line with recent debates surrounding the relationship between ICL and GCL (e.g., Kalyuga, 2011; Sweller, 2010), this study also showed the theoretical closeness of the two constructs. When learning contents are perceived as complex (i.e., induce a high ICL), the learner must invest more cognitive resources resulting in higher GCL. In contrast to theoretical assumptions and meta-analytical findings (e.g., Krieglstein et al., 2022a), the ICL reduction caused by pre-training was not significantly associated with learning performance. Although the ECL was not significantly altered by the ICL manipulation, it seems that changes in complexity may influence the perception of the design. In *Study 4*, it could be confirmed that the ECL items are sensitive to changes in the format of the learning material. The separated format resulted in significantly higher ECL perceptions supporting both the *split-attention effect* (Ayres and Sweller, 2021) as well as the validity of the proposed questionnaire. The strength of this effect was supported by large effect size. Unfortunately, it was also shown that the separated format leads to a higher ICL. Learners may have problems making a distinction between the complexity (ICL) and presentation (ECL) of information (e.g., Sweller, 2010). This assumption is not new and represents a central issue in CLT research that has been widely discussed by Krieglstein et al. (2022a). When learners are confronted with very complex learning contents (in this case cellular respiration), it seems to be difficult to differentiate between the complexity and the presentation. Similarly, presenting complex learning material in a simple way is probably not quite possible. It became apparent that the ICL and ECL are less separable from each other on a measurement level than the theory describes. Following the literature (e.g., Sweller et al., 2019), lower ECL ratings came along with a significantly better learning performance demonstrating their validity. For the manipulation of the GCL, the questionnaire showed theory-consistent results (Leopold, 2021) in *Study 5*. The intentional manipulation using the imagination principle was reflected in the GCL items. Following the active–passive distinction proposed by Klepsch and Seufert (2021), the passively experienced cognitive load types ICL and ECL were not affected by the GCL manipulation showing a good differentiation of constructs. Consistent with Moreno and Mayer's (2010) recommendations for promoting *generative processing*, imagination instruction was associated with higher learning performance.

Another important point that needs discussion refers to correlations between prior knowledge, cognitive load types, and learning performance across the experimental studies. These correlations were mostly theory-consistent and thus provide further evidence for the validity of the questionnaire. In line with our understanding of ICL, it was negatively correlated with prior knowledge (Krieglstein et al., 2022a). When learners can draw on prior knowledge, the complexity of the learning content is lower as learners can resort to already learned schemata stored in long-term memory (Paas and van Merriënboer, 2020). However, it also became clear that learners have problems differentiating between ICL and ECL which is demonstrated

Table 11 Final cognitive load questionnaire in German and English language

Cognitive load type	Item – German	Item – English
ICL		
ICL1 (Item 1)	Die Lerninhalte waren schwer zu verstehen	The learning content was difficult to understand
ICL2 (Item 2)	Die Erklärungen des Lerninhalts waren schwer nachvollziehbar	The explanations of the learning content were difficult to understand
ICL3 (Item 3)	Die Lerninhalte waren komplex	The learning contents were complex
ICL4 (Item 4)	Die Lerninhalte enthielten viele komplexe Informationen	The learning content included much complex information
ICL5 (Item 5)	Ohne Vorwissen waren die Informationen nicht verständlich	Without prior knowledge, the information was not understandable
ECL		
ECL1 (Item 6)	Es war schwierig, einen Überblick über den Aufbau des Lernmaterials zu erlangen	It was difficult to gain an overview of the structure of the learning material
ECL2 (Item 7)	Die Gestaltung des Lernmaterials machte es schwer, die Zusammenhänge zwischen den einzelnen Informationen herzustellen	The design of the learning material made it difficult to recognise links between individual information units
ECL3 (Item 8)	Das Lernmaterial war ungünstig gestaltet	The design of the learning material was inconvenient
ECL4 (Item 9)	Die Gestaltung des Lernmaterials machte es schwierig, wichtige Informationen zügig zu finden	The design of the learning material made it difficult to find relevant information quickly
ECL5 (Item 10)	Aufgrund der Gestaltung des Lernmaterials hatte ich das Gefühl, mich nicht auf die Lerninhalte konzentrieren zu können	Because of the design of the learning material, I had the impression that I could not concentrate on the learning content
GCL		
GCL1 (Item 11)	Ich habe aktiv über die Lerninhalte nachgedacht	I actively reflected upon the learning content
GCL3 (Item 13)	Ich habe mich bemüht, die Lerninhalte zu verstehen	I made an effort to understand the learning content
GCL5 (Item 15)	Ich habe ein umfassendes Verständnis der Lerninhalte erlangt	I achieved a comprehensive understanding of the learning content
GCL6 (Item 16)	Ich konnte mein bestehendes Wissen mit den Lerninhalten erweitern	I was able to expand my prior knowledge with the learning content
GCL7 (Item 17)	Das Wissen, das ich durch das Lernmaterial erworben habe, kann ich schnell und sicher anwenden	I can apply the knowledge that I acquired through the learning material quickly and accurately

by positive correlations between the two constructs. Assuming the learning material is perceived as poorly presented (high ECL), it could also be perceived as complex (high ICL). The problem of significant correlations between ICL and ECL has been also shown in the validation study by Leppink et al. (2013). Although the present questionnaire uses different designations for the ICL ("learning content") and ECL ("learning material"), learners still have difficulties differentiating between the two loads. The negative correlations between ICL and GCL could be explained by motivational factors (Feldon et al., 2019). When the learning content is perceived as complex, a motivational deficit could cause learners to expend less mental effort to understand the information. Otherwise, when learners invest a high amount of mental effort (higher GCL), the learning content could be perceived as less complex. Negative correlations between ICL and learning performance as well as between ECL and learning performance were found indicating that a rather complex and poorly designed learning material harms learning (Sweller et al., 2019). These theory-consistent results can be mapped by the questionnaire. GCL was correlated positively with learning performance. Assuming that high GCL is associated with engaged learners investing cognitive resources to learning-relevant activities (intrinsic load; Paas and van Gog, 2006), the GCL items of our questionnaire are able to measure this construct.

Strengths and Limitations of Our Questionnaire

Overall, the results of this work confirm that the different types of cognitive load can be measured in a reliable and valid manner with the proposed questionnaire. Like any psychological scale, however, this questionnaire has strengths and weaknesses. In general, the development and validation analyses were conducted following recommendations by various methodologists in a transparent way. To ensure the best possible practicability, the theory-based items were formulated so that a variety of learning interventions across many learning media can be experimentally studied for their learning improvement. The multidimensional measurement of cognitive load helps instructional designers and practitioners to better understand why learning interventions may be more or less effective for learning considering different learning mechanisms. Differentiating between different sources of cognitive load can help to understand the sources of learning deficits. Either the learning material is too complex, learners have too little prior knowledge (ICL), the learning material is too poorly designed (ECL), or learners were unable to devote their cognitive resources to learning (GCL). Another strength of the questionnaire is its verification by means of three experimental studies. In these studies, the questionnaire proved itself for practical use in controlled, randomized experimental settings which are the central method in instructional design research (Sweller, 2021). In line with theoretical assumptions as well as empirical findings, the cognitive load questionnaire was able to reflect intentional changes in instructional design providing evidence for predictive validity (e.g., Kuncel et al., 2001). Accordingly, the scores of the cognitive load items can

predict learning performance as a criterion measure. Another positive aspect is that the experimental studies were performed under more realistic learning conditions (i.e., learning at home). The proposed cognitive load questionnaire was and is, therefore, also able to find effects in learning situations that are less artificial.

One weakness of the questionnaire is that just the German version has undergone the described construction and validation process. This is important because small changes in the item formulation stemming from translations may lead to different understandings (e.g., Arafat et al., 2016). Future studies should test the proposed instrument in its English version. Another critical point refers to the GCL measurement. In particular, *Study 2* revealed lower factor loadings of GCL compared to ICL and ECL. The authors decided to keep the GCL item battery as it was to keep the items close to the theoretical assumptions of CLT. Partially lower factor loadings of the GCL items nevertheless suggest that the relationship between the variables and the factor is weaker. Thus, further explanatory variables, therefore, seem to play a role here. For example, the individual's motivation to learn appear to be highly relevant in actively engaging in learning-related activities. Similarly, the *cognitive-affective theory of learning with media* (Moreno and Mayer, 2010) assumes that generative processing (similar to *germane processing*) is caused by motivational factors. Accordingly, the influence of cognitive processing on learning is mediated by motivational factors.

In addition, items used to measure GCL consistently showed lower reliability scores. However, these scores were still within the acceptable range. Especially in *Study 5*, where the active component of GCL was intentionally manipulated and tested, the questionnaire showed significant differences in GCL demonstrating the validity of the GCL items without changing perceptions of ICL and ECL. In this context, the questionnaire was able to detect medium effect differences. Finding even medium effect sizes is a strength of this questionnaire.

The questionnaire also revealed a problem that has been discussed in CLT research for years. Thus, it seems that learners have difficulties differentiating between different cognitive load facets, indicating overlaps between involved processes (e.g., Ayres, 2018). The construct CLT, which in theory can differentiate between the types of cognitive load, cannot be transferred to reality with a perfect fit. Learners, who usually do not have such a comprehensive knowledge of CLT, can only partially differentiate between the complexity and the presentation format of the learning material. Another critical point relates to the theory-based construction of the questionnaire. All items were formulated very close to the theoretical assumptions of CLT, taking up both the foundational and the more recent literature (Sweller, 1988, 1994, 2021; Sweller et al., 2019). Concerning the ICL and ECL items, the formulation of the items could be well adapted to the theoretical description of the load types, as both constructs are clearly defined in the literature. In contrast, the definition of GCL has been frequently revised. This work defines GCL as a process variable ("*germane processing*") that involves the investment of cognitive resources to learning-relevant activities (i.e., dealing with the intrinsic load; Paas and van Merriënboer, 2020; Sweller, 2010). However, there are also definitions assuming that germane load refers to the redistribution of working memory resources from

extraneous to intrinsic activities (Sweller et al., 2019). These assumptions are relatively close to our underlying definition; however, ECL is not included in our items intended to measure GCL. The point is that the construction of a theory-based questionnaire highly depends on the most accurate definition possible of the construct to be measured. Since GCL is not always uniformly defined and the construction of a theory-based questionnaire highly depends on the most accurate definition of a construct, the formulation of GCL items can be debated and is dependent on the definition of questionnaire designers.

As also outlined by Klepsch et al. (2017), the most salient problem is that questionnaires require proper self-assessments. Self-rating scales are usually used in empirical settings with the assumption that learners can accurately estimate their experience retrospectively (Ayres, 2006). To accurately assess cognitive load experienced with a time delay to previous learning experiences can be considered as a metacognitive ability that is not equally developed in all individuals (Kelemen et al., 2000). In this context, repeating assessments during learning could be useful.

Conclusion

The presented questionnaire shows that cognitive load types can be measured separately in a reliable and valid manner. As a result, instructional designers and empirical researchers are better able to evaluate cognitive processes during learning. Besides the encouraging psychometric results of the questionnaire, theoretical ambiguities of CLT became apparent (e.g., de Jong, 2010; Moreno, 2010). It was shown again that the clear theoretical separation of the cognitive load types does not always show up in reality (see also Ayres, 2018). In particular, learners seem to have problems differentiating between the complexity (described as ICL) and the presentation format (described as ECL). One solution to overcome this ambiguity is to make learners aware of what is actually meant by the respective CLT types (Klepsch et al., 2017). However, this is not always possible in experimental settings and other measurement methods (e.g., electroencephalography) cannot accurately determine whether a cognitive load was triggered by the difficulty or by the presentation format. Furthermore, the role of germane load within CLT needs to be examined more closely. Consistent with Paas and van Merriënboer (2020), this work proposes to define GCL, not as a load per se, but rather as germane processing, whereby working memory resources are devoted to dealing with intrinsic load to construct and automate schemata in long-term memory. In particular, the factor analysis has shown that an accurate measurement, as is the case with the ICL and ECL, is all the more difficult with GCL. Similarly, Leppink et al. (2014) have expressed problems regarding the measurability of GCL and recommend measuring ICL and ECL to enhance the transparency and parsimony of CLT. Nevertheless, the proposed cognitive load questionnaire can be used in experimental settings to measure the cognitive load facets. However, one should be a little more careful with the GCL items. Understanding the nature of GCL seems to remain the “holy grail” in CLT research.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability Data is available on request from the corresponding author.

Declarations

Ethics Approval and Informed Consent Since the experiments constitute non-medical low-risk research, no special permission from an ethics committee was required. At the beginning of each study, participants were informed that the data will be used for research purposes only and that all data is collected anonymously. Thus, no identifying information was collected. Participants who prematurely stopped the survey were not included in the analyses and all of their data were deleted from the dataset. In general, the study was conducted in compliance with the guidelines of the German Research Foundation (DFG) and the German Psychological Society (DGPs).

Conflict of Interest The authors declare no competing interests.

References

Alisat, S., & Riemer, M. (2015). The environmental action scale: Development and psychometric evaluation. *Journal of Environmental Psychology, 43*, 13–23.

Antonenko, P., Paas, F., Grabner, R., & Van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review, 22*, 425–438.

Arafat, S. Y., Chowdhury, H. R., Qusar, M. M. A. S., & Hafez, M. A. (2016). Cross cultural adaptation & psychometric validation of research instruments: A methodological review. *Journal of Behavioral Health, 5*, 129–136.

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

Ayres, P. (2018). Subjective measures of cognitive load: What can they reliability measure? In R. Z. Zheng (Ed.), *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 9–28). Routledge.

Ayres, P., & Sweller, J. (2021). The split-attention principle in multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (pp. 199–211). Cambridge University Press.

Baddeley, A. (1992). Working memory. *Science, 255*, 556–559.

Barnette, J. J. (2000). Effects of stem and likert response option reversals on survey internal consistency: If You feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement, 60*, 361–370.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Statistical Psychology, 3*, 77–85.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Beege, M., Nebel, S., Schneider, S., & Rey, G. D. (2019). Social entities in educational videos: Combining the effects of addressing and professionalism. *Computers in Human Behavior, 93*, 40–52.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.
Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*, 78–117.
Biderman, M. D., Nguyen, N. T., Cunningham, C. J., & Ghorbani, N. (2011). The ubiquity of common method variance: The case of the Big Five. *Journal of Research in Personality, 45*, 417–429.
Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*, 230–258.
Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology, 49*, 109–119.
Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research, 27*, 509–540.
Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods, 15*, 17–32.
Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures, 12*, 25–44.
Chung, S., & Cheon, J. (2020). Emotional design of multimedia learning using background images with motivational cues. *Journal of Computer Assisted Learning, 36*, 922–932.
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.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
IBM Corp. (2021). *IBM SPSS Statistics for Windows* (Version 28.0) [Computer software]. IBM Corp.
Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10*, 7.
Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*, 51–57.
Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.
de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*, 105–134.
de Koning, B. B., Rop, G., & Paas, F. (2020). Effects of spatial distance on the effectiveness of mental and physical integration strategies in learning from split-attention examples. *Computers in Human Behavior, 110*, 106379.
DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). Sage Publications, Inc.
Dönmez, O., Akbulut, Y., Telli, E., Kaptan, M., Özdemir, İH., & Erdem, M. (2022). In search of a measure to address different sources of cognitive load in computer-based learning environments. *Education and Information Technologies, 27*, 10013–10034.
Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin, 81*, 358–361.
Eitel, A., Scheiter, K., & Schueler, A. (2013). How inspecting a picture affects processing of text in multimedia learning. *Applied Cognitive Psychology, 27*, 451–461.
Exline, J. J., Pargament, K. I., Grubbs, J. B., & Yali, A. M. (2014). The Religious and Spiritual Struggles Scale: Development and initial validation. *Psychology of Religion and Spirituality, 6*, 208–222.
Eyal, P., David, R., Andrew, G., Zak, E., & Ekaterina, D. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods, 54*, 1643–1662.
Feldon, D. F., Callan, G., Juth, S., & Jeong, S. (2019). Cognitive load as motivational cost. *Educational Psychology Review, 31*, 319–337.
Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*, 717–741.
Geary, D. C. (2005). *The origin of mind: Evolution of brain, cognition, and general intelligence*. American Psychological Association.
Geary, D. C. (2008). An evolutionarily informed education science. *Educational Psychologist, 43*, 179–195.

Gim Chung, R. H., Kim, B. S. K., & Abreu, J. M. (2004). Asian American multidimensional acculturation scale: Development, factor analysis, reliability, and validity. *Cultural Diversity and Ethnic Minority Psychology, 10*, 66–80.

Glass, G. V. (1966). Testing homogeneity of variances. *American Educational Research Journal, 3*, 187–190.

Greco, L. A., Baer, R. A., & Smith, G. T. (2011). Assessing mindfulness in children and adolescents: Development and validation of the Child and Adolescent Mindfulness Measure (CAMM). *Psychological Assessment, 23*, 606–614.

de Groot, A. (1965). *Thought and choice in chess* (2nd ed.). Mouton Publishers.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North-Holland.

Hegarty, M. (2005). Multimedia learning about physical systems. In R. E. Mayer (Ed.), *The cambridge handbook of multimedia learning* (pp. 447–466). Cambridge University Press.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research. *Educational and Psychological Measurement, 66*, 393–416.

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424–453.

Jackson, D. L., Gillaspy, J. A., Jr., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods, 14*, 6–23.

Jiang, D., & Kalyuga, S. (2020). Confirmatory factor analysis of cognitive load ratings supports a two-factor model. *The Quantitative Methods for Psychology, 16*, 216–225.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31–36.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*, 1–19.

Kalyuga, S., & Singh, A. M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28*, 831–852.

Kelava, A., & Moosbrugger, H. (2020). Deskriptivstatistische Itemanalyse und Testwertbestimmung [Descriptive statistical item analysis and test score determination]. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 143–158). Springer.

Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition, 28*, 92–107.

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., Ayres, P., & Chandler, P. (2011). Contemporary cognitive load theory research: The good, the bad and the ugly. *Computers in Human Behavior, 27*, 99–105.

Klepsch, M., & Seufert, T. (2020). Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instructional Science, 48*, 45–77.

Klepsch, M., & Seufert, T. (2021). Making an effort versus experiencing load. *Frontiers in Education, 6*, 645284.

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology, 8*, 1997.

Korbach, A., Brünken, R., & Park, B. (2018). Differentiating different types of cognitive load: A comparison of different measures. *Educational Psychology Review, 30*, 503–529.

Krabbe, P. (2017). *The measurement of health and health status: Concepts, methods and applications from a multidisciplinary perspective*. Academic Press.

Krell, M. (2017). Evaluating an instrument to measure mental load and mental effort considering different sources of validity evidence. *Cogent Education, 4*, 1280256.

Krieglstein, F., Beege, M., Rey, G. D., Ginns, P., Krell, M., & Schneider, S. (2022a). A systematic meta-analysis of the reliability and validity of subjective cognitive load questionnaires in experimental multimedia learning research. *Educational Psychology Review, 34*, 2485–2541.

Krieglstein, F., Schneider, S., Beege, M., & Rey, G. D. (2022b). How the design and complexity of concept maps influence cognitive learning processes. *Educational Technology Research and Development, 70*, 99–118.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*, 162–181.

Lance, C. E., & Vandenberg, R. J. (2002). Confirmatory factor analysis. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 221–254). Jossey-Bass.

Lecerf, T., & Canivez, G. L. (2018). Complementary exploratory and confirmatory factor analyses of the French WISC–V: Analyses based on the standardization sample. *Psychological Assessment, 30*, 793–808.

Leopold, C. (2021). The imagination principle in multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The cambridge handbook of multimedia learning* (pp. 370–380). Cambridge University Press.

Leopold, C., & Mayer, R. E. (2015). An imagination effect in learning from scientific text. *Journal of Educational Psychology, 107*, 47–63.

Leopold, C., Mayer, R. E., & Dutke, S. (2019). The power of imagination and perspective in learning from science text. *Journal of Educational Psychology, 111*, 793–808.

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*, 1058–1072.

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

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*, 936–949.

Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology, 86*, 114–121.

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health, 23*, 151–169.

Mansikka, H., Virtanen, K., & Harris, D. (2019). Comparison of NASA-TLX scale, modified Cooper-Harper scale and mean inter-beat interval as measures of pilot mental workload during simulated flight tasks. *Ergonomics, 62*, 246–254.

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177–198). Sage Publications Inc.

Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning, 33*, 403–423.

Mayer, R. E., & Fiorella, L. (2021). Principles for managing essential processing in multimedia learning: Segmenting, Pre-training, and Modality Principles. In R. E. Mayer & L. Fiorella (Eds.), *The cambridge handbook of multimedia learning* (pp. 243–260). Cambridge University Press.

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

Mayer, R. E., Mathias, A., & Wetzell, K. (2002). Fostering understanding of multimedia messages through pre-training: Evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology: Applied, 8*, 147–154.

McDonald, R.P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*, 276–282.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*, 412–433.

Meichsner, F., Schinköthe, D., & Wilz, G. (2016). The caregiver grief scale: Development, exploratory and confirmatory factor analysis, and validation. *Clinical Gerontologist, 39*, 342–361.

Mertler, C. A., & Vannatta, R. A. (2001). *Advanced and multivariate statistical methods: Practical applications and interpretation*. Pyrczak Publishing.

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

Möller, H. J. (2014). Self-rating scales. In G. Alexopoulos, S. Kasper, H. J. Möller, & C. Moreno (Eds.), *Guide to assessment scales in major depressive disorder* (pp. 23–34). Adis.

Moreno, R. (2010). Cognitive load theory: More food for thought. *Instructional Science, 38*, 135–141.

Moreno, R., & Mayer, R. E. (2010). Techniques that increase generative processing in multimedia learning: Open questions for cognitive load research. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 153–178). Cambridge University Press.

Moreno, R. E., & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 9–28). Cambridge University Press.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*, 396–402.

Ouwehand, K., van der Kroef, A., Wong, J., & Paas, F. (2021). Measuring cognitive load: Are there more valid alternatives to Likert rating scales? *Frontiers in Education, 6*, 702616.

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., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review, 24*, 27–45.

Paas, F., & Sweller, J. (2021). Implications of cognitive load theory for multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The cambridge handbook of multimedia learning* (pp. 73–81). Cambridge University Press.

Paas, F., & van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction, 16*, 87–91.

Paas, F., & van Merriënboer, J. J. G. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science, 29*, 394–398.

Paas, F., Renkl, A., & Sweller, J. (2003). 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.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193–198.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903.

Pouw, W., Rop, G., de Koning, B., & Paas, F. (2019). The cognitive basis for the split-attention effect. *Journal of Experimental Psychology: General, 148*, 2058–2075.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*, 287–297.

Ringnér, M. (2008). What is principal component analysis? *Nature Biotechnology, 26*, 303–304.

RStudio Team (2022). RStudio: Integrated Development for R (Version 2022.07.2) [Computer software].

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

Schneider, S., Krieglstein, F., Beege, M., & Rey, G. D. (2021). How organization highlighting through signaling, spatial contiguity and segmenting can influence learning with concept maps. *Computers and Education Open, 2*, 100040.

Schreiber, J. B. (2021). Issues and recommendations for exploratory factor analysis and principal component analysis. *Research in Social and Administrative Pharmacy, 17*, 1004–1011.

Schroeder, N. L., & Cenkci, A. T. (2018). Spatial contiguity and spatial split-attention effects in multimedia learning environments: A meta-analysis. *Educational Psychology Review, 30*, 679–701.

Schroeder, N. L., & Cenkci, A. T. (2020). Do measures of cognitive load explain the spatial split-attention principle in multimedia learning environments? A systematic review. *Journal of Educational Psychology, 112*, 254–270.

Shea, M., Wong, Y. J., Nguyen, K. K., & Gonzalez, P. D. (2019). College students' barriers to seeking mental health counseling: Scale development and psychometric evaluation. *Journal of Counseling Psychology, 66*(5), 626–639.

Sibley, C., Coyne, J., & Baldwin, C. (2011). Pupil dilation as an index of learning. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55*, 237–241.

Sonderen, E. V., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS ONE, 8*, e68967.

Suresh, K. P. (2011). An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *Journal of Human Reproductive Sciences, 4*, 8–11.

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed likert items. *Journal of Marketing Research, 45*, 116–131.
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.
Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312.
Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*, 123–138.
Sweller, J. (2016). Working memory, long-term memory, and instructional design. *Journal of Applied Research in Memory and Cognition, 5*, 360–367.
Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development, 68*, 1–16.
Sweller, J. (2021). *The role of evolutionary psychology in our understanding of human cognition: Consequences for cognitive load theory and instructional procedures*. Advance online publication.
Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185–233.
Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.
Sweller, J., Van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292.
Tubbs-Cooley, H. L., Mara, C. A., Carle, A. C., & Gurses, A. P. (2018). The NASA Task load index as a measure of overall workload among neonatal, paediatric and adult intensive care nurses. *Intensive and Critical Care Nursing, 46*, 64–69.
Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: The role of controlled search. *Memory & Cognition, 41*, 242–254.
Wang, B., Ginns, P., & Mockler, N. (2022). Sequencing tracing with imagination. *Educational Psychology Review, 34*, 421–449.
Wildt, A. R., & Mazis, M. B. (1978). Determinants of scale response: Label versus position. *Journal of Marketing Research, 15*, 261–267.
Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*, 806–838.
Zambrano, J., Kirschner, F., Sweller, J., & Kirschner, P. A. (2019). Effects of prior knowledge on collaborative and individual learning. *Learning and Instruction, 63*, 101214.

Authors and Affiliations

Felix Krieglstein[1] · Maik Beege[2] · Günter Daniel Rey[1] · Christina Sanchez-Stockhammer[3] · Sascha Schneider[4]

[1] Psychology of Learning with Digital Media, Institute for Media Research, Faculty of Humanities, Chemnitz University of Technology, Chemnitz, Germany

[2] Digital Media in Education, Department of Psychology, University of Education, Freiburg, Germany

[3] English and Digital Linguistics, Institute for English and American Studies, Faculty of Humanities, Chemnitz University of Technology, Chemnitz, Germany

[4] Educational Technology, Institute of Education, Faculty of Arts and Social Sciences, University of Zurich, Zurich, Switzerland