2019

The Evolution of Cognitive Load Theory and the Measurement of Its Intrinsic, Extraneous and Germane Loads: A Review

Giuliana Orru

Luca Longo

The Evolution of Cognitive Load Theory and the Measurement of Its Intrinsic, Extraneous and Germane Loads: A Review

Conference Paper · February 2019

DOI: 10.1007/978-3-030-14273-5_3

2 authors:

Giuliano Orru
Technological University Dublin - City Campus

Luca Longo
Technological University Dublin - City Campus

The evolution of Cognitive Load Theory and the measurement of its intrinsic, extraneous and germane loads: a review

Giuliano Orru, Luca Longo*
School of Computing, College of Health and Sciences,
Dublin Institute of Technology, Dublin, Ireland
Luca.Longo@dit.ie

Abstract. Cognitive Load Theory has been conceived for supporting instructional design through the use of the construct of cognitive load. This is believed to be built upon three types of load: intrinsic, extraneous and germane. Although Cognitive Load Theory and its assumptions are clear and well-known, its three types of load have been going through a continuous investigation and re-definition. Additionally, it is still not clear whether these are independent and can be added to each other towards an overall measure of load. The purpose of this research is to inform the reader about the theoretical evolution of Cognitive Load Theory as well as the measurement techniques and measures emerged for its cognitive load types. It also synthesises the main critiques of scholars and the scientific value of the theory from a rationalist and structuralist perspective.

Keywords: Cognitive Load Theory, Cognitive Load types; Intrinsic Load, Extraneous Load, Germane Load; Measures; Instructional Design; Efficiency.

1 Introduction

The construct of Cognitive Load (CL) is strictly related to the construct of Mental Workload (MWL). The former has evolved within Educational Psychology [1], while the latter within Ergonomics and Human Factors [2]. Despite their independent evolution within different disciplines, both are based upon the same core assumption: the limitations of the human mental architecture and the cognitive capacities of the human brain and its working memory [3] [4]. In a nutshell, as professor Wickens suggested [5], mental workload is equivalent to the amount of mental resources simultaneously elicited by a human during the execution of a task. In order to achieve an optimal performance, the working memory limits should not be reached [3] [4]. If this occurs, the mental resources are no longer adequate to optimally execute the underlying task. Within Ergonomics, the construct of Mental Workload has evolved both theoretically and practically. A plethora of ad-hoc definitions exist as well as several domain-dependent measurement techniques, measures and applications [1]. While abundance of research exists, the science of Mental Workload is still in its infancy because any of the proposed measures can generalise the construct itself. Similarly, within Educational Psychology, despite Cognitive Load Theory (CLT) is one of the most invoked learning theory for supporting instructional design [6], research on how to develop highly generalisable measures of Cognitive Load is limited. Also, it is unclear how its three types of load – intrinsic, extraneous and germane – can be measured and how they interact with each other. The aim of this paper is to provide readers with the theoretical elements underpinning the construct of Cognitive Load. This is done from an evolutionary perspective of the measurement techniques and measures emerged from the three types of load, accompanied with a critical discussion of their scientific value.

The remainder of the paper is structured as follows. Section 2 presents the key theoretical elements and assumptions of Cognitive Load Theory, as appeared in the literature. Section 3 focuses on a review of the measurement techniques and measures emerged for its intrinsic, extraneous and germane loads. Section 4 builds on this review by emphasising the open debate on the scientific value of CLT. Section 5 highlights new perspectives and research on CLT and the reconceptualization of its cognitive load types, as recently emerged in the literature. Section 6 summarises this study with final remarks suggesting novel research directions.

2 Cognitive Load Theory

Cognitive Load Theory (CLT) is a cognitivist learning theory aimed at supporting instructors in the development of novel instructional designs aligned with the limitations of the human cognitive architecture. In a nutshell, this architecture is the human cognitive system aimed at storing information, retrieving and processing it for reasoning and decision making [7]. CLT is based upon the assumption of *active processing* that views the learner as actively engaged in the construction of knowledge [8]. In other words, learners are actively engaged in a process of attention to relevant material and its organisation into coherent structures that are integrated with prior knowledge [9]. Another premise of CLT is the *dual-channel* assumption by which processing of information occurs in two distinct channels: an auditory and a verbal channel. The former processes auditory sensory input and verbal information while the latter processes visual sensory inputs and pictorial representations [10]. An essential component of this architecture is its memory that can store information for short and long term. According to another premise, *the limited capacity* assumption of CLT, the former memory, also referred to working memory, is conscious and limited, while the latter is unconscious and unlimited [2]. Baddeley [3] and Paivio [4], following Miller proposal [7] , support the view that when working memory has to deal with new information, it can hold just seven chunks at a time. However, if these chunks are related and if they have to be processed, human beings are capable to handle just two or three at the same time [11]. Expanding the capacity of working memory coincides with learning [2]. Learning take places by transferring pieces of information from working memory to long term memory [3] [4]. According to Schema Theory, this transfer of information allows the construction of knowledge, in long term memory, in the form of schema [12]. To construct a schema means to relate different chunks of information from a lower level to a higher level of complexity and to hold them as a single unit that can be understood as a single chunk of information [12]. In turn, schema can be retrieved to solve a problem, a task, or more generally to answer a question in educational contexts. Schema construction is believed to reduce the load in working memory [2]. The expansion of long term memory can be achieved by a reduction of the load of working memory. Leaving sufficient cognitive resources in working memory to process new information is one of the core objectives of educational instructional design. In fact, if the amount of information that has to be held in working memory lies within its limits, the learning phase is facilitated. Contrarily, if the amount of information overcomes these limits, an overload situation occurs and the learning phase is hampered.

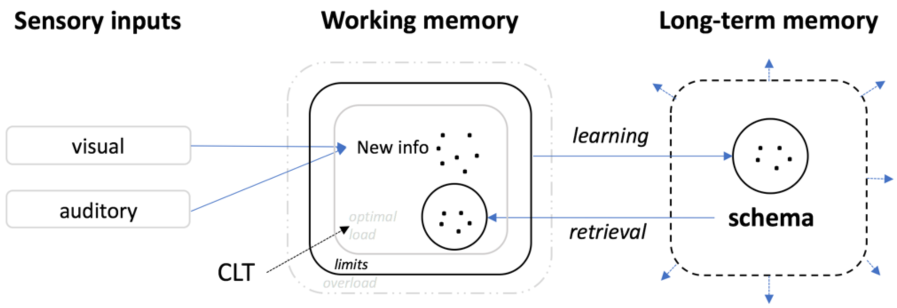

Figure 1: A representation of the mental architecture and the role of Cognitive Load Theory (CLT) in connection to working memory and schema construction

A core construct within CLT is Cognitive Load (CL), believed to be multidimensional. Intuitively it can be defined as the mental cost imposed by an underlying cognitive task on the human cognitive system [13]. It is possible to distinguish two types of factors that can interact with cognitive load: causal and assessment factors (figure 2). The formers affect cognitive load while the latter are affected by cognitive load. The causal factors include:

- features of the task (T) such as structure, novelty and pressure;
- the features of the environment (E) such as noise and temperature where a task (T) is executed and their interaction (ExT);
- the characteristic of a learner (L) such as capabilities, cognitive style and prior knowledge;
- the interaction between environment and learner characteristics (ExL);
- the interaction between task, environment, learner's characteristics Ex(TxL).

The assessment factors can be conceptualised with three dimensions: mental load, mental effort and mental performance. Mental load is imposed by the task (T) and/or by demands from the environment. It is a task-centred dimension, independent of the subject, and it is considered constant. Mental effort is a human-centred dimension that reflects the amount of controlled processing (capacity or resources allocated for task demands) in which the individual is engaged with [13]. It is affected by the task-environment interaction (ExT), the subject characteristics interaction with the environment (ExL) and the interaction of the learner with the task in the environment (Ex(TxL)). Similarly, the level of mental performance is affected by the factors that affect mental effort [4]. Other factors might affect cognitive load [14] [15] and research in the field has not produced a comprehensive list yet [16].

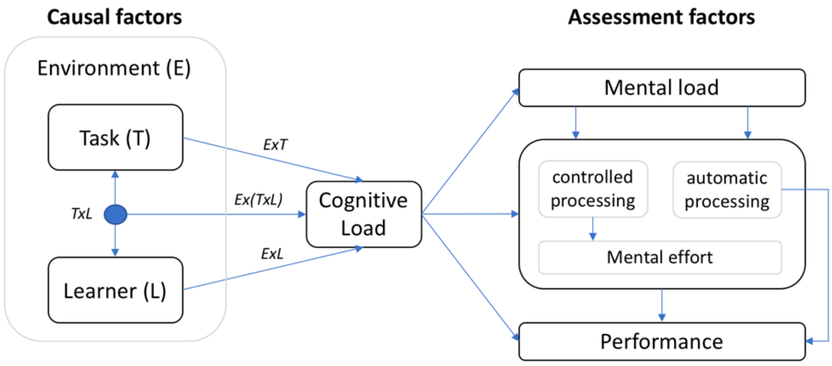

Figure 2: Causal factors and assessment factors according to [41].

Starting from the research of Halford et al. [17] on the difficulty in processing information with multiple elements at the same time during problem solving, Sweller defined the degree of complexity of these elements as `element interactivity' [18]. Starting from this definition, two types of cognitive load has emerged: the *intrinsic* and the *extraneous* loads. Intrinsic load refers to the numbers of elements that must be processed simultaneously in working memory (element interactivity) for schema construction. 'This type of load cannot be modified by instructional interventions because it is intrinsic to the material being dealt with. Instead, extraneous cognitive load is the unnecessary cognitive load and can be altered by instructional interventions' [2]. Sweller stated that the basic goal of Cognitive Load Theory is the reduction of extraneous load: this is a type of ineffective load that depends on the instructional techniques provided by the instructional format to complete a task [2]. This view is supported by Paas and colleagues that refers to extraneous load as the cognitive effect of instructional designs that hamper the construction of schema in working memory [19]. Beside, intrinsic and extraneous, Sweller defined another type of load: the *germane* load [2]. This is the extra effort required for learning (schema construction). It is possible to use this effort when intrinsic and extraneous loads leave sufficient working memory resources. This extra effort increases cognitive load, but it is connected to learning, thus, it facilitates schema construction. Germane load is the effective cognitive load and it is the result of those beneficial cognitive processes such as abstractions and elaboration that are promoted by 'good' instructional designs [20]. Reducing extraneous load and improving germane load by developing schema construction and automation should be the main goal of the discipline of instructional design. The three types of load emerged within Cognitive Load Theory, and their role, can be summarised in figure 3.

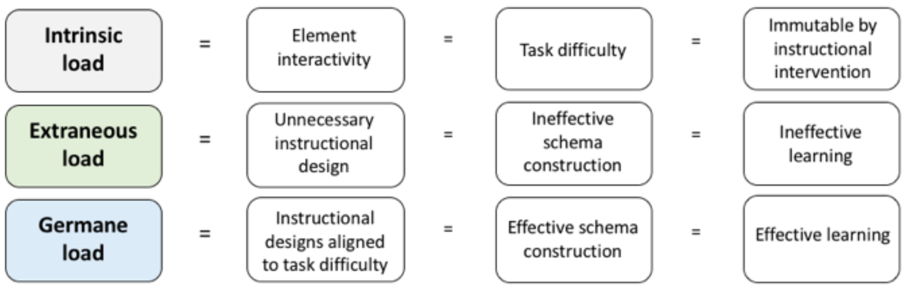

Figure 3: Definitions and role of the cognitive load types of Cognitive Load Theory.

Sweller and colleagues, with their attempt to define cognitive load within the discipline of Educational Psychology and for instructional design, believed that the three types of load are additive. This meant that the total cognitive load experienced by a learner in working memory while executing a task, is the sum of the three types of load, these being independent sources of load [2] (figure 4).

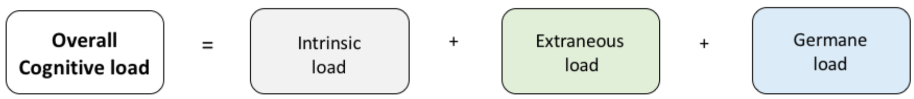

Figure 4: Additive definition of overall cognitive load.

Figure 5 depicts the relationship between the three types of cognitive load, as proposed in [21]. In condition A (overload), cognitive load exceeds the limits of the working memory of the learner due to an increment in the extraneous load. In turn errors are more frequent, longer task execution times occur, sometimes even leading to the inability to perform an underlying task. In condition B there is spare working memory capacity and the learners can perform optimally on an underlying task. With spare capacity, CLT proposes to increase the germane load in order to activate learning tasks, as in condition C.

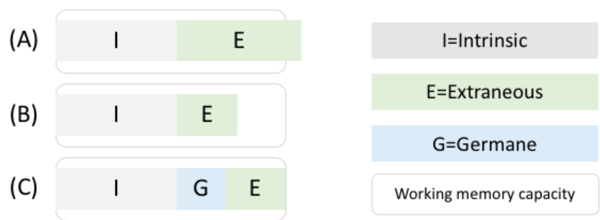

Figure 5: Relationship between the three types of cognitive load

3 Cognitive load types, measurement techniques and measures

The definition of the types of load within Cognitive Load Theory (CLT) are supported by empirical studies that have used three different classes of measures:

- *task performance measures* such as error rates, learning times and secondary task measures;
- *subjective measures* such as self-reporting questionnaires and rating scales;
- *physiological measures* such as eye movements and physical body responses.

Within Educational Psychology, the focus has always been on the first two classes: task performance and subjective measures. The reason is intuitive since physiological measurement techniques require special equipment to gather data, trainer operators to use this equipment and they are intrusive, most of the time not suitable for empirical experiments in typical classrooms. Additionally, evidence suggests they did not prove sufficient sensitivity to differentiate the three cognitive load types [13] envisioned in CLT. As a consequence, the next sections mainly focus on research studies that employed task performance and subjective measures.

Miwa et al. [21] developed a task-performance based method for cognitive load measurement built upon the *mental chronometry paradigm* [22], in line with the triarchic view of load [2]. The mental chronometry assumes that reaction time can reveal the quantity of intrinsic, extraneous and germane loads coming from the corresponding cognitive processes. The hypothesis behind their experiment is that if, in the 8x8 Reversi game (figure 4), the three types of cognitive load are manipulated by changing the presentation of the information to the players, it is possible to measure them by observing their reaction time of players between movements of discs on the board. To manipulate the intrinsic load, on one hand, an advisor computer agent provides some hints to participants on the possible subsequent move (low intrinsic load condition, figure 6 left). On the other hand, in another condition, no hints are provided (high intrinsic load condition, figure 6 right). In addition to this, in order to manipulate extraneous load, the white and black discs are changed with two different letters from the Japanese alphabet. Since these 2 letters are perceptually similar, they are expected to lead to higher perception and understanding exerted by participants. Eventually, the germane load is manipulated by altering the instructions presented to participants. In order to exert more germane load, each participant is requested to report, after the game, the heuristics learnt to play it. According to this paradigm, if the reaction time, in high intrinsic load conditions, is longer than the reaction time, in low intrinsic load conditions, then it can be considered as a valid indicator of intrinsic load. Here, learning corresponds to the development of effective strategies for moving discs that can lead a participant to win the game. These strategies imply that players control and regulate their cognitive processing by meta cognitive perspectives, thus increasing their germane load [21].

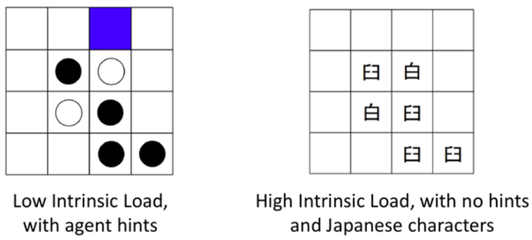

Figure 6: Low and high extraneous load conditions in the 8x8 Reversi game [35].

The research attempt by Miwa and colleagues [21] is indeed useful to investigate the discrimination between the three types of load and to provide guidelines on how to design experiments that contribute to the definition of cognitive load. Their preliminary findings suggest the three types of load are separable. However, it was executed in a highly controlled environment and not in more natural settings such as in a typical classroom, thus limiting the generalisability of their findings.

3.1 Subjective measures of cognitive load

Gerjets et al. [23] proposed two experiments on the use of hypermedia environments for learning on the topic of probability theory. In the first experiment, the validity of multimedia principles [10] in hypermedia environments has been tested. In the second experiment, an analysis of the ability of learners to impact their performance according to their prior experience was performed. A subjective 9-point Likert scale is employed to measure the three types of load during learning (table 1).

Table 1: Subjective rating scale to measure cognitive load types from [23]. Each item has to be rated on a 9 point Likert scale (1=extremely easy, 9=extremely difficult)

Type of load	**Scale**
Intrinsic Load	How easy or difficult do you consider probability theory at this moment?
Extraneous Load	• How easy or difficult is it for you to work with the learning environments? • How easy or difficult is it for you to distinguish important and unimportant information in the learning environments? • How easy or difficult is it for you to collect all the information that you need in the learning environment?
Germane Load	Indicate the amount of effort you exerted to follow the last example

Depending on the prior knowledge of learners, high intrinsic and extraneous loads should lead to poor learning outcomes while high germane load should lead to good learning outcomes. Unfortunately, in this study no evidence of this connection has been found. Consequently, authors claim that subjective rating scales are valid if learners are able to distinguish the types of cognitive load. In other words, in order to be sensitive to the differences in loads, learners should be aware of the cognitive process connected to the experienced load. To achieve this, training learners on Cognitive Load Theory can facilitate their understanding of the three types of load. However, this is not an easy condition to achieve. In fact, for instance, the level of difficulty (intrinsic load) could be due to the poor instructional design that increase the extraneous load, or due to the natural complexity of an underlying learning task. A novice learner could find this distinction really hard to understand and could not be able to comprehend if the own difficulty in learning can be attributed to the instructional design (extraneous load) or the complexity of the task (intrinsic load).

Corbalan et al. [24] hypothesises that, in order to prevent cognitive overload, it is possible to adapt task difficulty and the support of each newly selected learning task to the previous knowledge and experience of a learner and his/her perceived task load [24]. This can be done by employing some external agent, the learner him/herself, or both. The hypothesis was tested by employing two subjective rating scales, one for task load and one for germane load **(**as per table 2). This hypothesis was tested by performing an empirical 2x2 factorial design experiment with health sciences students. The design variables where the factors adaptation (absence or presence of the agent) and the control over task selection (program control or shared control program/learner).

Table 2: Measurements of Task and Germane load on a 7-point scale from [24] (1=extremely low, 7=extremely high)

Type of load	**Scale**
Task load	Rate your effort to perform the task
Germane Load	Evaluate the effort invested in gaining understanding of the relationship dealt with in the simulator task

Findings suggest that, on one hand, the presence of adaptation delivered more efficient learning and task involvement. On the other hand, shared control produced a higher task involvement when compared to program task selection. Learning here refers to good learning outcomes and lower effort exerted in the underlying learning tasks. Task involvement refers to good learning outcomes and higher effort exerted in the learning task. Both the cases prevented cognitive overload [24].

Ayres hypothesised that, by maintaining the extraneous and the germane loads constant, students can identify changes in the `element interactivity' within problems by means of subjective measures, and thus successfully quantify the intrinsic cognitive load [25]. In his study, extraneous and germane loads are maintained constant by not providing any instructional hint. Learners had to solve a set of four brackets-expansion problems without any explicit instructions (source of extraneous load) and without any didactic feedback (source of germane load). The bracket-expansion problems required a series of four operations in which the level of difficulty increased. Under this instructional condition, any change in the overall cognitive load is due to change in the element interactivity (source of intrinsic load). Intrinsic load is measured by a subjective measure, as depicted in table 3. After each operation, learners had to rate its difficulty. The hypothesis is that higher intrinsic load should correspond to more errors.

Table 3: Subjective rating scale of intrinsic load on a 7-point scale from [1]
(1=extremely easy, 7=extremely difficult)

Type of load	Scale
Intrinsic Load	How easy or difficult you found each calculation?

The authors tested their hypothesis with two experiments. In the first, students had low prior mathematical knowledge while in the second, participants had a wider range of mathematical skills. In the first experiment, students could recognise task difficulty since subjective intrinsic load was highly correlated with the errors committed by themselves. In the second experiment, although students did not commit many mistakes, they still could detect differences in task difficulty. These findings support the high sensitivity and reliability of the employed self-reporting measure. Additionally, the takeaway of this study is that, by keeping constant two sources of load out of three, it is possible to get a measure of the remaining dependent load. Transitively, it turns out that, by keeping constant the extraneous and the intrinsic loads, any change in cognitive load, irrespective of the measurement technique employed, corresponds to variations in the dependent variable, the germane load.

Gerjets et al. [26] investigated how to enhance learning through a comparison of two instructional designs on the same topic: how to calculate the probability of a complex event. The first design condition included worked out examples, while the latter included modular worked examples. To measure the experienced cognitive load of learners in each condition, a modified version of the NASA-Task Load Index was (table 4) [27]. Readers are referred to [27] for the original version.

Table 4: Modified version of the Nasa-TLX from [26] where each scale ranged from 0 to 100
(low level to high level)

Type of load	Scale
Task demands (Intrinsic Load)	How much mental and physical activity did you require to accomplish the learning task? (Thinking, deciding, calculating, remembering, looking, searching).
Effort (Extraneous Load)	How hard did you have to work to understand the contents of the learning environment?
Navigational demand	How much effort did you invest to navigate the learning environment?
Understanding	How successful did you feel in understanding the contents?
Stress	How much stress did you experience during learning?

Within Cognitive Load Theory, Sweller [2] stated that task demand is caused by the degree of element interactivity of the task (intrinsic load), while the effort is exerted to achieve an effective understanding of the instructional material. The navigational demands are related to those activities not strictly directed to learning. In this line, Gerjets and colleagues stated that the scale for task demands is aimed at quantifying the intrinsic load of the instructional material, the effort scale at quantifying the germane load and the scale for navigational demands is aimed at quantifying the extraneous load [26]. The hypothesis of the experiment is that the modular presentation of worked examples can increase the germane load more than their molar – as a whole – presentation. Unfortunately, findings did not provide evidence about any increment of the germane load. As a possible interpretation, the authors suggested that the instructional explanations provided during the task to increase the germane load, and the self-explanation derived using worked out examples, created a redundant information and the illusion of understanding hampered the learning instead of improving it. However, in a prior experiment, Gerjets et al. [26] have successful demonstrated that, in an example based learning, the modular presentation of worked examples can actually reduce the intrinsic load and improves the germane load more than the molar presentation of the same problem. The modular presentation provides a part-whole sequencing of the solution procedures whereas the molar presentation provides the solutions of the procedures as a whole. The segmentation of the presentation of the worked example led to a decrease in the degree of interactivity as well as in the number of simultaneous items. In turn, this led to a decrease in the intrinsic load. According to the authors, these findings are more relevant to novice learners, whereas the same instructional design could be redundant for more expert learners because the degree of their expertise increases. Consequently, in the case of expert learners, the molar presentation of solution procedures is a more appropriate instructional design. The modified version of the NASA-TLX, employed in this study, has been applied also in [28]. Here, authors focused on the effects of different kinds of computer-based graphic representations in connection to the acquisition of problem-solving skills in the context of probability theory. Despite different experiments, [29] [26] and [28] did not provide evidence on the reliability and validity of the subjective rating scale employed. Therefore, it can be only hypothesised that this scale is sensitive to the three types of load conceived within CLT.

Galy et al. [30] tested the additivity between the intrinsic, the extraneous and the germane loads by manipulating three factors believed to have an effect on each of them. In detail, this study assumed that task difficulty is an indicator of intrinsic load, time pressure of extraneous load and the level of alertness of germane load. The effect on the experienced overall cognitive load is connected to the manipulation of the extraneous and intrinsic loads which are respectively estimated by the self-reporting of notions of tension (time pressure) and mental effort (task difficulty). The level of alertness is measured by the French paper-and-pencil version of the Thayers's Activation-Deactivation Checklist [31]. Questions are listed in table 5. For each word in the deactivation list, each student had to tick one from the "not at all", "don't know", "little" and "much" labels. These labels are respectively mapped to weights (1, 2, 3 and 4). The responses were counted up to have a measure of four factors: general activation (GA), deactivation sleep (DS), high activation (HA), and general deactivation (GD). The GA/DS ratio yielded an alertness index.

Table 5: Self report scales of cognitive loads types from [30]. Intrinsic and extraneous load are in the scale 0 to 10 (low time pressure/mental effort to high effort/considerable effort)

Type of load	Scale
Intrinsic load	Rate the mental effort (task difficulty) you experienced during the task
Extraneous load	Rate the tension (time pressure) you experienced during the task
Germane load	Select one of the following responses ("not at all", "don't know", "little" and "much") for each of 20 listed adjectives: active, energetic, vigorous, full of, lively, still, quiet, placid, calm, at rest, tense, intense, clutched up, fearful, jittery, wide-awake, wakeful, sleepy, drowsy, tired.

The experimental task consisted of a memory recalling activity with 2 digit numbers (low difficulty) or 3 digit numbers (high difficulty) in four conditions: low difficulty and low time pressure, low difficulty and high time pressure, high difficulty and low time pressure, high difficulty and high time pressure. The difference in cognitive load due to variations in task difficulty and time pressure with respect to the different levels of alertness can be taken as an indicator of differences in the contribution of germane load. In low difficulty and low time pressure conditions, germane load is believed to be substantially inexistent, but in high difficulty and high time pressure conditions, it is assumed that the learner has to employ specific strategies to execute the memory task and thus generating germane load. Authors believed that germane load, as a function of alertness, corresponds to the subject's capability to select strategies to be employed while performing the learning task. However, the implementation of these strategies is determined by the amount of free cognitive resources determined by task difficulty and time pressure [30]. Consequently, the authors claimed that alertness is a germane load factor depending on the quantity of working memory resources left by the intrinsic and extraneous load experienced.

Leppink et al. [32] developed a new instrument for the measurement of intrinsic, extraneous and germane loads. The authors consider the critique of Kalyunga et al. [33] about the expertise reversal effect and its consequences on the learning and on the different types of load. According to this, the same instructional feature may be associated with germane load for a learner and with extraneous load for another learner, depending on the level of expertise and on the level of prior knowledge. To develop a more sensitive instrument to detect changes in cognitive load types, they proposed a multi-item questionnaire (table 6). Authors conducted experiments in four lectures of statistics, asking to rate difficult or complex formulas, concepts and definitions using the scales in table 6. In a number of studies, Leppink and colleagues verified 7 hypotheses regarding the reliability of the new instrument compared with other instruments, used in the past, to measure intrinsic load [25], extraneous load [34], germane load [35] and for overall cognitive load [36]. They also tested five hypotheses connected to the expected relationship between prior knowledge and intrinsic load, and between prior knowledge and learning outcomes. Through an exploratory analysis, it has emerged that the reliability of the rating scale was positive, the extraneous load and the germane load elements were negatively correlated and the elements that were supposed to measure intrinsic load were not correlated to germane load.

Table 6: Multi-subjective rating scales of cognitive load types from [32] in the scale 0 to 10 (0=not at all, 10=completely).

Type of load	Scale
Intrinsic Load	• The topic/topics covered in the activity was/were very complex • The activity covered formulas that I perceive as very complex • The activity covered concepts and definitions that I perceived as very complex
Extraneous Load	• The instruction and/or explanation during the activity were very unclear. • The instruction and/or explanation were, in terms of learning, very ineffective. • The instruction and/or explanations were full of unclear language.
Germane Load	• The activity really enhanced my understanding of the topic(s) covered • The activity really enhanced my knowledge and understanding of statistics • The activity really enhanced my understanding of the formulas covered • The activity really enhanced my understanding of concepts, definitions

Eventually, the elements that were expected to measure intrinsic load had moderate correlation with extraneous load. The validity of the scales was verified by comparing the subjective ratings with the learning outcomes assessed by a performance test. As hypothesised, a high prior knowledge corresponded to a low intrinsic load. Extraneous cognitive load was higher when a problem was solved by an unfamiliar format and germane load was higher when a problem was solved by a familiar format. There is partial evidence that higher germane load, as measured by multiple subjective scales, lead to higher results on post-task test performance.

Leppink and colleagues [37] criticised their own previous study [32] mentioning the uncertainty of their multiple subjective rating scales to represent the three different types of cognitive load. The main reasons of their critique are three: 1) the correlation between germane load and the learning outcomes, in the task performance, was lower than expected and not statistically relevant 2) the previous experiments were all focused on a single topic, namely statistic 3) the manipulations applied in [32] did not lead to the expected differences in the measurement of the three different cognitive load types [37]. In summary, their psychometric instrument might have measured only the level of expectation instead of the actual invested effort devoted in the complexity of the activity (intrinsic load), its ineffective explanations (extraneous) and its understanding (germane). To evaluate a more direct relation between the three types of load and the learning outcomes, a randomized experiment was performed, with bachelor students who received a description of the Bayes theorem. To measure the three different types of load, the authors changed the order of the rating scales and added three items to it (as per table 7). These items were supposed to contribute to the evaluation of the internal consistency of the theoretical assumption that the three types of load are separated, additive and independent. Findings suggest that two items improved the internal consistency of the mental effort for intrinsic and extraneous loads but not for germane load, suggesting its re-definition [6].

Table 7: Informed subjective rating questionnaire proposed in [39] available to learners while rating different learning scenarios on a 7-point Likert scale (1=very low, 7=Very high)

Type of load	Scale
Intrinsic Load	I invested a very high mental effort in the complexity of this activity.
Extraneous Load	I invested a very high mental effort in unclear and ineffective explanations and instructions in this activity.
Germane Load	I invested a very high mental effort during this activity in enhancing my knowledge and understanding.

Zukic and colleagues [38] focused on the assessment of the validity of the instrument developed in [37] as well as its internal consistency and its capability to correlate with learning outcomes. In their study, the correlations between intrinsic and extraneous load and between extraneous and germane load were statistically significant. A low degree of experienced intrinsic load and a high degree of reported germane load could explain the improvement of the learning outcomes. Additionally, a regression analysis verified that the items associated to the germane load could actually explain the perceived learning. Eventually, a confirmatory factor analysis supported the development of a three-dimensional model that includes the three types of load. The main take away of this study is that germane load can be measured as an independent source of load.

Klepsch et al. proposed an alternative way to measure the three load types reliably and validly [39]. The novelty of their approach is the use of two forms of ratings: informed and naïve. According to this, they conducted an experiment with two different group of learners. The first, the informed rating group, was trained on how to differentiate the three types of load through a theoretical explanation of CLT and its assumptions. The second, the naïve rating group, did not receive the training on CLT. Learners were asked to rate 24 learning scenarios grouped in 5 different domains (language learning, biology, mathematics, technology and didactics). To detect changes in the cognitive load experienced by the two groups of learners, only one type of cognitive load was manipulated at a time. The learners in group one received the questionnaire in table 8, while those in group two received the questionnaire in table 9 Both the groups received also an additional question on perceived overall cognitive load adapted from [36]. The participants in the informed ratings group correctly discriminated intrinsic, extraneous and germane loads in line with the expectations. However, participants, in the Naïve ratings group, correctly discriminated only the intrinsic and the extraneous loads but they were not able to differentiate germane load.

Table 8: Informed subjective rating questionnaire proposed in [39] available to learners while rating different learning scenarios on a 7-point Likert scale (1=very low, 7=Very high)

Type of load	Scale
Intrinsic Load	During this task, Intrinsic Load was…
Extraneous Load	During this task, Extraneous Load was…
Germane Load	During this task, Germane Load was…

Table 9: First version of the Naïve rating scales questionnaire proposed in [39] 7-point Likert scale (1=completely wrong, 7=absolutely right)

Type of load	Scale
Intrinsic Load	• For this task, many things needed to be kept in mind simultaneously • This task was very complex
Germane Load	• For this task, I had to highly engage myself • For this task, I had to think intensively what things meant.
Extraneous Load	• During this task, it was exhausting to find the important information • The design of this task was very inconvenient for learning • During this task, it was difficult to recognize and link the crucial information

A reliability analysis of the scales was executed, by task, using the Cronbach alpha measure based on the formula presented in [40]. This allowed to compute the mean of several given alpha values based on sampling distribution. The validity of the measure was analysed by comparing the ratings of learners with the expectations for each type of load for each task. A very low reliability was detected for all the tasks, in the informed ratings group, this being an indicator of the capability of learners to differentiate the types of load separately. However, in the naïve ratings groups, reliability was high, suggesting how the three types of load were not clearly separable. In particular, germane load was the dimension that was not discriminable across the two groups. Starting from this unsatisfying finding, the authors developed a new scale for germane load (table 10).

Table 10: Second version of the Naïve rating questionnaire proposed in [39] with a new scale for the germane load on a 7-point Likert scale (1=completely wrong, 7=absolutely right)

Load type	Scale
Germane Load	• I made an effort, not only to understand several details, but to understand the overall context • My point while dealing with the task was to understand everything correctly • The learning task consisted of elements supporting my comprehension of the task

Subsequently, they evaluated the overall new questionnaire with a larger sample. A new experiment was conducted with a group of students who received 8 tasks, one at a time, designed to induce more or less germane load. Here, in contrast to the first study, and in line with the idea of doing experiments in more realistic learning environments, each learning task was designed to induce changes in the three types of load. For intrinsic load, the degree of interactivity of the tasks was manipulated. For extraneous load, different learning formats were considered, some employing text and pictures together, some individually and some with additional non-relevant information. Eventually, germane load was manipulated by creating tasks aimed at eliciting different degrees of deeper learning processes. A reliability and validity analysis, conducted as in the first experiment, confirmed that it is possible to measure the three types of load separately, in line with the triarchic theory of load [2].

3.2 Task performance and self-reported measures

Deleeuw and Mayer tested the separability of the three types of load in a multimedia lesson on the topic of electric motors [41]. Two experiments were executed: one with a pre-question on the content of the lesson aimed at motivating learners to focus on deeper cognitive processing, and one without. Authors manipulated extraneous load providing redundant instructional designs to learners. Similarly, they manipulated intrinsic load through changes on the complexity of the sentences that explained the lesson. Eventually, they examined the differences in the germane load by comparing students with high scores on a test of problem solving transfer, against students with lower scores. The authors evaluated the sensitivity of the response time to a secondary task during learning for measuring the extraneous load, the effort ratings during learning for measuring the intrinsic load, and the self-reported difficulty rating, after learning, for measuring the germane load (as per table 11). In details, the secondary task consisted of a visual monitoring task where learners had to recognise a periodic change of colour and to press the space bar each time this colour change took place.

Table 11: Subjective rating scales and secondary task reaction time proposed by [41] on a 9-point Likert Scale (1=extremely low, 9=extremely high for intrinsic load and 1=extremely easy, 9=extremely difficult for germane load)

Types of load	**Scale/measure**
Intrinsic Load	Your level of mental effort on this part of the lesson.
Germane Load	How difficult this lesson was.

+

Extraneous Load	Measured by the response time to a secondary task. At each of eight points in an animated narration, the background colour slowly changes (pink to black). Learner is required to press the spacebar of the computer as soon as the color changes.

The findings of the experiment supported the triarchic theory of cognitive load [2]. Students who received redundant information needed longer reaction time than students who did not receive redundant instructional design. The explanation about the electric motor has been provided by learners using different sentences with different levels of complexity. The scale for intrinsic load reflected higher effort for high complexity sentences and lower effort for low complexity sentences. Students who reported a lower and a higher transfer reflected their difficulty by the rating scale provided: low transfer reflected high difficulty, high transfer reflected low difficulty. Thus, the authors showed that these different measures of load (reaction time, effort and difficulty) are sensitive to different types of load (extraneous, intrinsic, germane) [41]. The three different variables analysed (redundancy and complexity of statement, high or slow capacity of transfer to solve a problem) are strongly correlated with the three different types of load, thus providing evidence for their good sensitivity. Eventually, authors recommended a replication of their research study in other contexts and with different students because the measurement of the three cognitive load types might be often intrusive, creating an artificial learning situation. In addition, the study did not account for the prior knowledge of learners, (most of them had a low prior knowledge) as an important variable that could influence the overall perception of load.

Cerniak et al. [34] hypothesised how the split attention effect, proposed by Sweller [2], could be mediated not only by a reduction of extraneous cognitive load but also by an increase of germane load (germane load explanation) [42]. An experiment, conducted in a learning context on physiological processes of a nephron, was aimed at testing the above the research hypothesis. Authors employed the reaction time on a secondary as a task performance measure, in order to detect variations in the overall cognitive load between learners who received an integrated format of instructional designs and learners who received a split source format. The former learners were expected to experience less overall load because the integrated format was believed to decrease their extraneous load. The latter learners were expected to experience more overall load due to the split attention effect believed to increase their extraneous load, as suggested in [43]. In the experiment, learners had to press the space bar of the keyboard of a computer every time a stimulus appeared on the screen (for example the change of a colour). The longer time required to react to this secondary task, the higher cognitive load exerted on the primary task. Eventually, subjective ratings were applied to measure the three types of load, as per table 12.

Table 12: Subjective rating scales for the cognitive load types proposed by [34] on a 6-point Likert Scale (1=not at all, 6=extremely)

Type of load	Scale
Intrinsic Load	How much difficult was the learning content for you?
Extraneous Load	How difficult was it for you to learn with the material?
Germane Load	How much did you concentrate during learning?

Findings showed that there is no difference in the overall cognitive load between learners who received the split source format and those who received the integrated source format. As a consequence, the former learners increased their extraneous load and decreased their germane load, whereas the latter learners decreased their extraneous load and increased their germane load. This confirms that the extraneous and germane loads partially mediate the split attention effect. However, authors brought forward a critique whereby there could be a possible confusion between the two different questions designed for intrinsic and extraneous loads. Learners could have the impression to answer the same question. In fact, in a new learning context, learners might not be able to identify the source of difficulty that means the content or the instructional material delivered. The authors spotted a high correlation between the extraneous and the germane loads through an analysis of the learning outcomes. However, they did not state that these measures of loads were aimed at tackling different working memory resources. As a consequence, the relation between learning processes and working memory capacity was not demonstrated.

4 Synthesis and observations on the scientific value of Cognitive Load Theory

According to the literature review conducted in the previous sections, it appears evident that the three types of load envisioned in Cognitive Load Theory – intrinsic, extraneous and germane – have been mainly measured by means of subjective rating scales. This has more a practical explanation because self-reporting scales are easier to use and they do not influence the primary task when compared to secondary task measures. They can be administered post-learning tasks and they are aimed at representing a perceptual subjective experience of a learner for an entire learning session. This is in contrast to secondary task measures which, even if more sensitive to variations of cognitive load, they are more intrusive since they alter the natural execution of a learning task. A number of researchers brought forward critiques on Cognitive Load Theory in relation to its theoretical clarity [44] [45] and its methodological approach [46]. According to these critiques, the assumptions of CLT appear circular because its three types of load are believed not to be empirically measurable. Empirical research is based on observed and recorded data and it derives knowledge from actual experience rather than from a theory, a belief or a logic coming from first principles 'a priori'. This is the case of subjective rating scales aimed at measuring the cognitive load types. Regardless of the way these scales relate to the evaluation of the different cognitive load types, all of them underlie the phenomenon they are pretending to measure in their premises or suppositions, namely the definitions of intrinsic, extraneous and germane loads (figure 7, left). In other words, the premises of CLT – its cognitive load types – are believed to be confirmed by the data coming from their measurements circularly, without empirical evidence.

In addition, the fact that human cognitive processes, related to the same instructional design, can be regarded as germane load in one case and as extraneous load in another case, it means that CLT can account for nearly every situation [45]. This critique also refers to the `Expertise Reversal Effect' [47]. In fact, on one hand, some instructional design, such as written explanations followed by a graphic element to enhance its understanding, can be useful for a novice learner, by reducing the extraneous load and increasing the germane load. On the other hand, the same graphical aid can be useless for an expert learner because it can reduce germane load and increase extraneous load. In fact, for an expert, it can be redundant to read instructional designs just registered and automatized in own memory, hampering understanding and learning. Depending on the degree of expertise, the same instructional design can lead to germane or extraneous load, emphasising the circularity of CLT (figure 7, right). The theoretical differences regarding the types of load are based on the subjective experienced load of learners, implying that they are able to differentiate them by their own. This issue, as discussed above, depends on the way the questions are formulated, and on the familiarity of learners on the different cognitive load types, and their prior knowledge. All these variables are not easy to monitor and control, they can create confusion on the source of the supposed experienced load. Under a strict scientific view, the evaluation of this supposed load does not come from the experience of the learners, rather from the principles of the theory is based upon.

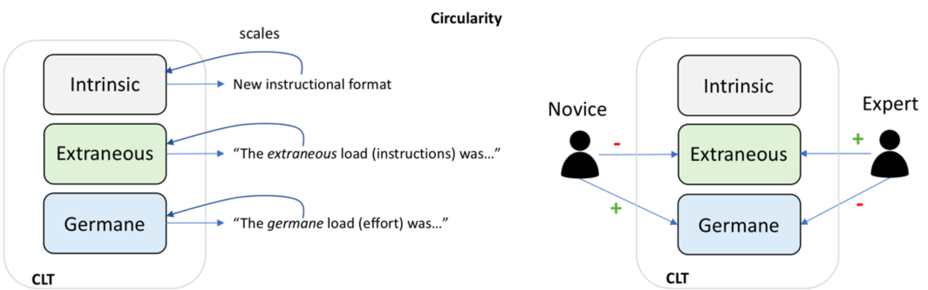

Figure 7: The circularity of the load types of Cognitive Load Theory (left) and the 'expertise reversal effect' by which different cognitive processes can be regarded differently (right)

To analyse the scientific value of CLT, two different methodological approaches have been followed: the rationalism of Karl Popper [48] [49] and the structuralist approach of the theories of Joseph Sneed [50]. Under the former approach, it is not possible to consider CLT scientific because its basic principles, namely the three different types of load, cannot be tested by means of any experimental method, consequently they are not falsifiable [46] (Figure 8, left). To be scientific, the measures should be sensitive to the different types of load. From a strict rationalist point of view, a measure is scientific if it does not presuppose the assumptions that it shall measure in its rationale [46]. However, as previously discussed, most of the subjective rating scales, conceived for the cognitive load types, contain the variables they pretend to measure. This implies that the logic of the questions influences the logic of the answers. In turn, the measures of the loads can be obtained 'a priori', by setting the questions to validate the theory they are pretending to verify, and not through any authentic experience of cognitive load. CLT should provide empirical evidence about the cognitive load types. Unfortunately, this has not convincingly emerged in the literature of Educational Psychology, Instructional Design and Cognitive Load Theory, justifying the scepticism regarding the possibility to measure the three different types of load.

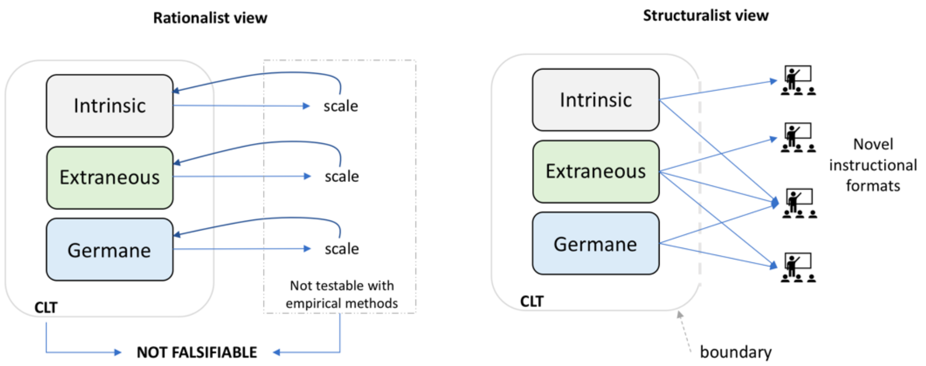

Figure 8: The rationalist view of Cognitive Load Theory (left) and the structuralist view (right)

The second methodological approach to analyse CLT is based upon structuralism [51] [50] [52] [53]. Under its logic, the scientific value of the theoretical principles of CLT does not depend on their empirical validity. Rather it depends upon their effectiveness to form the ground of the structure of a theory that consents to derive specific predictions on how detailed instructional manipulations can affect learning outcomes [46]. The structuralist analysis considers the fundamental assumptions of CLT as theoretical axioms. The empirical content of these axioms is valid in the context of the theory if they contribute to expand the theory itself (Figure 8, right). Regardless whether it is possible to validate some research predictions or not, these predictions can still expand the theory. In fact, CLT has been extensively adopted for the design of several new instructional formats, expanding its boundaries [46]. As discussed in the previous sections, several research experiments have been performed in different learning contexts. In each of this, the intrinsic, extraneous and germane loads have been manipulated, individually or in pair by employing the traditional experimental/control group design. In turn, the cognitive load of learners and their learning outcomes where analysed [6]. If this analysis showed that learning has been actually facilitated, and statistical power held, then it means that a new instructional design was conceived as it actually promoted one or more types of load. Similarly, starting from the study of the 'Goal free effects' compared to the traditional ways to solve a problem (means analysis), Sweller and his colleagues have produced various novel findings and approaches to inform instructional design. Yet, Plass et al. [54] provided a complete list of CLT effects such as the 'Worked completion effect' [55], the 'Split attention effect' [56] the 'Redundancy effect' [57], the 'Modality effect' [58], the 'Expertise reversal effect' [47] and the 'Collective working memory effect' [11]. As a consequence, according to a structuralist point of view, Sweller stated that CLT has been developed and evolved as a consequence of these contributions and experiments [6]. They defend the fact that the three types of load were not elaborated a priori, rather they have been developed according to experimental findings that are falsifiable in their nature. In fact, it is still possible to replicate the experiments and obtain opposite findings. However, what cannot be considered falsifiable is only the definition of the three types of load employed in different experiments because the measures adopted are not considered scientific. In short, Sweller and colleagues strongly support the view that CLT is actually built upon empiricism [59]. As educational psychologist, Sweller and Chandler [60] share the same ultimate goal in the context of cognition and instruction: the generation of new, helpful instructional techniques aimed at improving learning.

5 Reconceptualization of cognitive load types

As a consequence of the critiques related to the theoretical development of CLT and after several failed attempts to find a generally applicable measurement technique as well as the development of measures for the three different types of load, the theory has been re-conceptualised using the notion of element interactivity. This refers to the numbers of elements that must be processed simultaneously in working memory for schema construction and their interactions [18]. In this update of CLT, the element interactivity now defines the mechanisms not only of intrinsic load, but also of extraneous load [6]. In detail, the extraneous load is related to the degree of interactivity of the elements of the instructional material used for teaching activities, and instructional designs should be aligned to this. These designs should not focus on enhancing the number of items to be processed by learners, otherwise the resulting load could be considered extraneous. In other words, when instructional designs do not add instructions that increase the number of elements that must be processed within working memory, then the germane load of learners can be triggered. In this case, existing instructions can facilitate the use of working memory allocated for the intrinsic load. Additionally, germane load is no longer an independent source of load, it is a function of those working memory resources related to the intrinsic load of the task. In turn, intrinsic load depends on the characteristic of the task, extraneous load on the characteristic of the instructional material, on the characteristic of the instructional design and on the prior knowledge of learners. Eventually, germane load depends on the characteristics of a learner which equates to the resources of working memory allocated to deal with the intrinsic load [6] (figure 9).

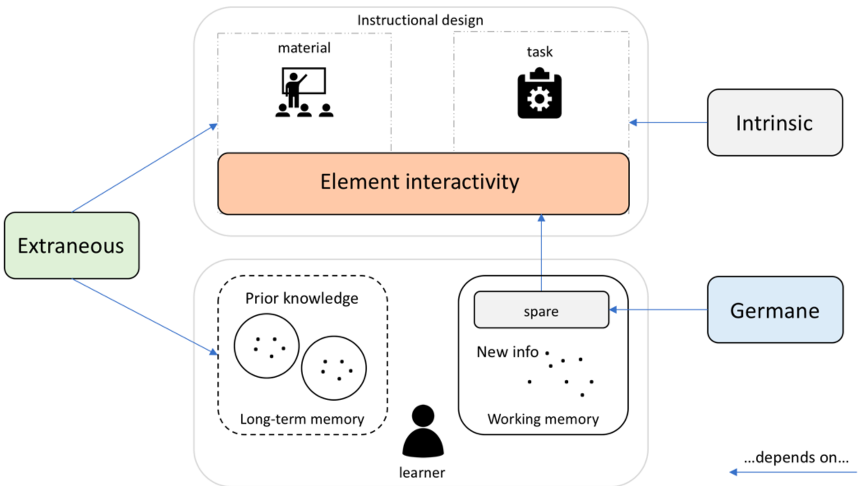

Figure 9: Redefinition of the cognitive load types and their roles.

The main theoretical contradiction before the reconceptualization of CLT was the additivity and the compensability of germane and extraneous loads. Here, the critical point is that, if extraneous load decreases, while keeping intrinsic load constant, then germane load should increase too. However, the measures for the three cognitive loads appeared in the last 30 years, confirm that this compensation does not have empirical evidence: the total cognitive load does not remain constant but changes [6]. After the reconceptualization (figure 9), germane load is related to that part of working memory that deals with the degree of element interactivity of the task. It can be promoted by creating instructional design aligned to it but it also depends on the intrinsic load, and as a consequence, it is not clearly measurable [6]. In fact, germane load now forms a balanced whole with extraneous load without creating logical and empirical contradictions. If intrinsic load remains constant but extraneous load changes, the overall cognitive load changes too because more or less working memory resources are devoted to deal with the degree of element interactivity. At a given level of knowledge and expertise, intrinsic load cannot be altered without changing the content of the material presented to learners altogether. Extraneous load, instead, can be altered by changing the instructional procedures. Yet, germane load coincides to those working memory resources allocated to deal with the degree of element interactivity inherent to an underlining learning task. Although germane load has now a fundamental role to deal with intrinsic load, the additivity of CLT still holds in the two remaining theoretical assumptions: the intrinsic and the extraneous load. According to this reconceptualisation, most of the critiques related to the circularity of CLT do not longer stand according to Sweller. Additionally, Sweller and colleagues, as in [59], consider the unidimensional subjective rating scale of mental effort proposed by Paas et al. [13] a valid measure of overall cognitive load. In fact, if intrinsic load is kept constant, it is feasible to measure the extraneous load by only altering the instructional designs between an experimental and a control group. It is also possible to measure one type of load keeping the other constant, and the overall load measured would be an indicator of the modified type of load: extraneous or intrinsic.

6 Final remarks

The measurement of the cognitive load types envisioned in the Cognitive Load Theory is a critical challenge for its theoretical development and its scientific value. After the literature review conducted in the previous sections, and after the presentation of the critiques that brought to the reconceptualization of the cognitive load types, the reader is left with two possible interpretations. On one hand, germane load is not clearly measurable by a common and standardised way, consequently its theoretical independence is denied, and after its reconceptualisation, it is a function of intrinsic load. On the other hand, there is evidence, triggered by the proposal of novel multiple subjective rating scales [38] [39], that the three types of load are measurable, even the most challenging, namely the germane load. Sweller and colleagues believe that germane load exists but it is not measurable [59]. He suggested that one of the most reliable way to measure the overall cognitive load is the unidimensional subjective scale of Mental Effort [13]. However, the fact that unidimensional scale has been widely employed within Educational Psychology, does not mean it is always the most appropriate. For example, within the discipline of Human Factors (Ergonomics), there exist a plethora of empirical studies that all point to the multi-dimensional nature of the construct of Mental Workload (cognitive load for educational psychologists) [16] [61] [62] [63] [64] [65]. There is also an emerging body of knowledge, within Computer Science, that is employing more formal non-linear approaches for modelling mental workload as a multi-dimensional construct [66] [67] [68]. Similarly, applications of mental workload as a multi-dimensional concept can be found in Human Computer Interaction [69] [70] [71] [72].

Learning is a complex process, it is hard to evaluate it mostly because it is perceived as a subjective one. Similarly, cognitive load it is a complex construct and it is assumed it can be modelled end evaluated through quantitative criteria to satisfy the empirical exigencies of scientific research. This is an existing methodological gap and it is the reason why, so far, there is little evidence of generally applicable subjective measurement techniques and measures for the three types of load and for overall cognitive load. According to Klepsch and colleagues, their informed rating scale is a novelty in CLT research and it seems to be a valid method for measuring the three types of load [39]. They believe it is the most logical approach because if learners are informed, then the evaluation of the experienced load can be done with a higher degree of awareness. However, in our view, this might bring back the issue of circularity, suggesting that we are leading learners to understand Cognitive Load Theory as well as its assumptions and influence them to rate their subjective experience to fit our expectations. Cognitive load is a complex construct and indeed CLT has had a significant impact for instructional design. Circularity is also an important issue that should be avoided in favour of empiricism and falsifiability of measures. We believe that, with advances in technologies and the availability of cheap sensors and non-invasive instruments for gathering responses of the human brain and bodies, physiological measures of mental workload might finally shed further light on the complex but fascinating problem of cognitive load modelling.

References

1. Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*(3), 251–296.
2. Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics*, *58*(1), 1–17. doi:10.1080/00140139.2014.956151
3. Baddeley, A. (1998). Recent developments in working memory. *Current Opinion in Neurobiology*, *8*(2), 234–238. doi:10.1016/S0959-4388(98)80145-1
4. Paivio, A. (1986). Mental Representation: A Dual-Coding Approach Oxford Univ. *Press, New York*.
5. Hart, S. G., Wickens, C. D. (2012). Workload assessment and prediction. *MANPRINT: an approach to systems integration*, 257-296. Springer, Dordrecht.
6. Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review*, *22*(2), 123–138.
7. Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.
8. Clark, R. C., Nguyen, F., Sweller, J., & Baddeley, M. (2006). Efficiency in learning: Evidence-based guidelines to manage cognitive load. *Performance Improvement*, *45*(9), 46–47.
9. Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, *38*(1), 43–52.

10. Mayer, R. E. (2005). Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. *The Cambridge handbook of multimedia learning*, 169–182.
11. Kirschner, F., Paas, F., & Kirschner, P. A. (2011). Superiority of collaborative learning with complex tasks: A research note on an alternative affective explanation. *Computers in Human Behavior*, *27*(1), 53–57.
12. Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in Problem Solving. 142. In R. Sternberg (Ed) Advances in the psychology of human intelligence (pp 7-75). Hillsdale: Lawrence Erlbaum
13. Paas, F. G., Van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and motor skills*, *79*(1), 419–430.
14. Longo, L., & Barrett, S. (2010). A Computational Analysis of Cognitive Effort. In *Intelligent Information and Database Systems, Part II* (pp. 65–74).
15. Longo, L., & Barrett, S. (2010). Cognitive Effort for Multi-agent Systems. In *Int Conf. on Brain Informatics, BI* (pp. 55–66).
16. Longo, L., & Leva, M. C. (2017). *Human Mental Workload: Models and Applications: First International Symposium, H-WORKLOAD 2017, Dublin, Ireland, June 28-30, 2017, Revised Selected Papers* (Vol. 726). Springer.
17. Halford, G. S., Maybery, M. T., & Bain, J. D. (1986). Capacity Limitations in Children's Reasoning: A Dual-Task Approach. *Child Development*, *57*(3), 616–627. doi:10.2307/1130340
18. Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295–312.
19. Paas, F. G., & Van Merriënboer, J. J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of educational psychology*, *86*(1), 122.
20. Gerjets, P., & Scheiter, K. (2003). Goal configurations and processing strategies as moderators between instructional design and cognitive load: Evidence from hypertext-based instruction. *Educational psychologist*, *38*(1), 33–41.
21. Miwa, K., Kojima, K., Terai, H., & Mizuno, Y. (2016). Measuring cognitive loads based on the mental chronometry paradigm. In *COGNITIVE 2016 : The Eighth International Conference on Advanced Cognitive Technologies and Applications* (pp. 38–41).
22. Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological psychology*, *26*(1–3), 3–67.
23. Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., & Eysink, T. H. S. (2009). Learning with hypermedia: The influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior*, *25*(2), 360–370. doi:10.1016/j.chb.2008.12.015
24. Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2008). Selecting learning tasks: Effects of adaptation and shared control on learning efficiency and task involvement. *Contemporary Educational Psychology*, *33*(4), 733–756. doi:10.1016/j.cedpsych.2008.02.003
25. Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, *16*(5), 389–400.

26. Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction*, *16*(2), 104–121. doi:10.1016/j.learninstruc.2006.02.007
27. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (Vol. 52, pp. 139–183). Elsevier. doi:10.1016/S0166-4115(08)62386-9
28. Scheiter, K., Gerjets, P., & Catrambone, R. (2006). Making the abstract concrete: Visualizing mathematical solution procedures. *Computers in Human Behavior*, *22*(1), 9–25. doi:10.1016/j.chb.2005.01.009
29. Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing Instructional Examples to Reduce Intrinsic Cognitive Load: Molar versus Modular Presentation of Solution Procedures. *Instructional Science*, *32*(1/2), 33–58. doi:10.1023/B:TRUC.0000021809.10236.71
30. Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology*, *83*(3), 269–275. doi:10.1016/j.ijpsycho.2011.09.023
31. Thayer, R. E. (1978). Toward a psychological theory of multidimensional activation (arousal). *Motivation and Emotion*, *2*(1), 1–34.
32. Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. doi:10.3758/s13428-013-0334-1
33. Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of Expertise and Instructional Design. *Human Factors*, *40*(1), 1.
34. 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*(2), 315–324. doi:10.1016/j.chb.2008.12.020
35. Salomon, G. (1984). Television is" easy" and print is" tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of educational psychology*, *76*(4), 647.
36. Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology*, *84*(4), 429.
37. Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction*, *30*, 32–42. doi:10.1016/j.learninstruc.2013.12.001
38. Zukić, M., Đapo, N., & Husremović, D. (2016). Construct and Predictive Validity of an Instrument for Measuring Intrinsic, Extraneous and Germane Cognitive Load. *Universal Journal of Psychology*, *4*(5), 242–248. doi:10.13189/ujp.2016.040505
39. Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.01997

40. Rodriguez, M. C., & Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological methods*, *11*(3), 306.
41. DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, *100*(1), 223–234. doi:10.1037/0022-0663.100.1.223
42. Kester, L., Kirschner, P. A., & van Merriënboer, J. J. G. (2005). The management of cognitive load during complex cognitive skill acquisition by means of computer-simulated problem solving. *British Journal of Educational Psychology*, *75*(1), 71–85. doi:10.1348/000709904X19254
43. Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, *43*(1), 16–26.
44. Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*(4), 469–508.
45. De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional science*, *38*(2), 105–134.
46. Gerjets, P., Scheiter, K., & Cierniak, G. (2009). The Scientific Value of Cognitive Load Theory: A Research Agenda Based on the Structuralist View of Theories. *Educational Psychology Review*, *21*(1), 43–54. doi:10.1007/s10648-008-9096-1
47. Sweller, J., Ayres, P. L., Kalyuga, S., & Chandler, P. (2003). The expertise reversal effect. The expertise reversal effect. *Educational Psychologis*t, 38 (1), 23-31.
48. Popper, K. (2005). *The logic of scientific discovery*. Routledge.
49. Popper, K. (2014). *Conjectures and refutations: The growth of scientific knowledge*. routledge.
50. Sneed, J. D. (2012). *The logical structure of mathematical physics* (Vol. 35). Springer Science & Business Media.
51. Balzer, W., Moulines, C., & Sneed, J. (1987). *An Architectonic for Science–The Structuralist Program. Dordrecht: Reidel*. Springer). Traducción al castellano de P. Lorenzano (2012): Una arquitectónica para la ciencia–El programa estructuralista. Universidad Nacional de Quilmes: Bernal (Argentina).
52. Stegmüller, W. (2013). *The structuralist view of theories: a possible analogue of the Bourbaki programme in physical science*. Springer Science & Business Media.
53. Westermann, R. (1988). Structuralist reconstruction of psychological research: Cognitive dissonance. *German Journal of Psychology,* 12(3), 218-231.
54. Plass, J. L., Moreno, R., & Brünken, R. (2010). *Cognitive load theory*. Cambridge University Press.
55. Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*(1), 59–89.
56. Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of educational psychology*, *80*(4), 424.
57. Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, *8*(4), 293–332. doi:10.1207/s1532690xci0804_2
58. Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of educational psychology*, *87*(2), 319.

59. Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory, explorations in the learning sciences, instructional systems, and performance technologies 1*. New York, NY: Springer Science+ Business Media.
60. Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and instruction*, *8*(4), 351–362.
61. Longo, L. (2015). A defeasible reasoning framework for human mental workload representation and assessment. *Behaviour and Information Technology*, *34*(8), 758–786.
62. Longo, L. (2014). *Formalising Human Mental Workload as a Defeasible Computational Concept* (PhD Thesis). Trinity College Dublin.
63. Longo, L. (2016). Mental Workload in Medicine: Foundations, Applications, Open Problems, Challenges and Future Perspectives. In *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 106–111). doi:10.1109/CBMS.2016.36
64. Wickens, C. D. (2017). Mental workload: assessment, prediction and consequences. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 18–29). Springer.
65. Hancock, P. A. (2017). Whither workload? Mapping a path for its future development. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 3–17). Springer.
66. Rizzo, L., Dondio, P., Delany, S., & Longo, L. (2016). Modeling Mental Workload Via Rule-Based Expert System: A Comparison with NASA-TLX and Workload Profile (Vol. 475, pp. 215–229). doi:10.1007/978-3-319-44944-9_19
67. Moustafa, K., Luz, S., & Longo, L. (2017). Assessment of mental workload: a comparison of machine learning methods and subjective assessment techniques. In *International Symposium on Human Mental Workload: Models and Applications* (pp. 30–50). Springer.
68. Rizzo, L. M., & Longo, L. (2017). Representing and inferring mental workload via defeasible reasoning: a comparison with the NASA Task Load Index and the Workload Profile. Proceedings of the 1st Workshop on Advances In Argumentation In Artificial Intelligence co-located with {XVI} International Conference of the Italian Association for Artificial Intelligence (AI*IA 2017), (pp. 126-140).
69. Longo, L., & Dondio, P. (2015). On the Relationship between Perception of Usability and Subjective Mental Workload of Web Interfaces. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2015, Singapore, December 6-9, Volume I* (pp. 345–352).
70. Longo, L. (2012). Formalising Human Mental Workload as Non-monotonic Concept for Adaptive and Personalised Web-Design. In *UMAP* (pp. 369–373).
71. Longo, L. (2017). Subjective Usability, Mental Workload Assessments and Their Impact on Objective Human Performance. In *IFIP Conference on Human-Computer Interaction* (pp. 202–223). Springer.
72. Longo, L. (2018). Experienced mental workload, perception of usability, their interaction and impact on task performance. *PloS one*, *13*(8), e0199661.