




# Confirmatory Factor Analysis of Cognitive Load Ratings Supports a Two-Factor Model

Dayu Jiang<sup>a</sup> and Slava Kalyuga<sup>a</sup> 

<sup>a</sup>University of New South Wales

**Abstract** ■ The practically usable measures of cognitive load are important for evaluating learning and performance conditions, and obtaining empirical evidence in support of theoretical hypotheses. Subjective rating scales remain popular tools for measuring cognitive load, especially in realistic environments, despite the development of more technically sophisticated objective measures that are suitable mostly for laboratory settings only. In accordance with the traditional view of cognitive load theory, a number of studies suggested subjective rating scales aimed to differentiate three types of cognitive load – intrinsic, extraneous, and germane. However, according to a recently proposed modified theoretical model of cognitive load, germane load shares the identical theoretical ground with intrinsic load and therefore, is redundant. It has been hypothesized that rating scales based on the two-factor, intrinsic-extraneous model of cognitive load should be sufficient and valid tools for assessing levels of cognitive load. The reported study tested this hypothesis by conducting a confirmatory factor analysis of multidimensional rating data using the lavaan package in the programming language R. The results of the analysis supported the two-factor model of cognitive load.

**Keywords** ■ cognitive load theory, cognitive load measures, subjective rating scales, confirmatory factor analysis. **Tools** ■ R.

 [s.kalyuga@unsw.edu.au](mailto:s.kalyuga@unsw.edu.au)

 [10.20982/tqmp.16.3.p216](https://doi.org/10.20982/tqmp.16.3.p216)

**Acting Editors** ■

**Julie Albertosa** (Institut de Recherche Biomédicale des Armées)

■ **Lucile Chanquoy** (Université Côte d'Azur, CNRS, BCL, France)

**Reviewers**

■ Two anonymous reviewers

## Introduction

The practically usable measures of cognitive load experienced during performance or learning processes are important for evaluating realistic task conditions and obtaining empirical evidence in support of theoretical hypotheses and models. During the recent decades, subjective rating scales have been frequently used in multiple research studies within a cognitive load framework for measuring learner cognitive load, especially in experimental studies conducted in realistic classroom settings (for recent overviews of cognitive load measurement approaches and techniques, see DeLeeuw & Mayer, 2008; Leppink, Paas, van der Vleuten, van Gog, & van Merriënboer, 2013; Sweller, Ayres, & Kalyuga, 2011; Korbach, Brünken, & Park, 2018; F. Paas, Tuovinen, Tabbers, & van Gerven, 2003; Schmeck, Opfermann, van Gog, Paas, & Leutner, 2015). Even though more sophisticated objective techniques have been developed during this time (see Antonenko, Paas,

Grabner, & van Gog, 2010; Szulewski, Roth, & Howes, 2015), since they are mostly suitable only for laboratory-based experiments, subjective ratings have remained the most popular and widely used instruments.

### *Cognitive Load Theory and the Concept of Germane Load*

To briefly describe the nature of cognitive load as the phenomenon to be measured, the two most important components of human cognitive architecture need to be considered: working memory and long-term memory (Sweller et al., 2011; van Merriënboer & Sweller, 2005). In this model of human cognitive architecture, working memory is the component associated with limitations in processing capacity and cognitive resources. Working memory is responsible for conscious processing of novel sensory information and previously learned information from long-term memory. It is very limited in capacity and duration when dealing with novel information: only a few ele-



ments can be simultaneously held and processed in working memory, and only for a few seconds unless intentionally rehearsed (Baddeley, 1986; Cowan, 2001; Peterson & Peterson, 1959). Long-term memory is a permanent repository of organized knowledge structures that govern all our cognitive processes (Anderson, 1983; Kirschner, Sweller, & Clark, 2006). If a person has relevant knowledge in long-term memory, working memory limitations could be lifted when dealing with familiar task areas, because the available knowledge structures would allow chunking many elements of information into a single unit that is treated as a single element in working memory (Cowan, 2000; Cowan & Chen, 2009; Eysenck & Keane, 2015; Gilchrist & Cowan, 2012).

Cognitive load is defined as working memory load experienced when performing a specific task (Kalyuga, 2011; Sweller et al., 2011; van Merriënboer & Sweller, 2005). If cognitive load exceeds the available working memory capacity, learning and performance might be inhibited. Learning materials and tasks could be ineffective if they ignore the limited processing capacity of working memory. In cognitive load theory, two major types (dimensions) of load were initially introduced - intrinsic load and extraneous load (Sweller, 1994). Intrinsic load is the relevant, necessary load required for achieving a specific learning goal. The theory suggests that this load should be managed and kept within the available capacity of working memory (e.g., by segmenting or properly sequencing learning tasks). On the other hand, extraneous cognitive load is not relevant to achieving a learning goal, and it is caused by the cognitive activities the learner needs to perform due to a particular design of the learning task. For example, unnecessarily switching (splitting) attention between several interdependent sources of information that are placed at different locations (such as a diagram and separated explanatory text) may consume additional cognitive resources for maintaining some elements of information when searching for their co-referents in the other related sources (Kalyuga, Chandler, & Sweller, 1999; Mayer & Moreno, 1998). A similar unwarranted increase in extraneous cognitive load could be caused by processing redundant information that is not important for learning, such as re-describing verbally a self-explanatory diagram. Cognitive load theory recommends eliminating or reducing such redundant information (Kalyuga et al., 1999; Sweller et al., 2011; van Merriënboer & Sweller, 2005).

The concept of germane load was introduced to explain positive learning effects of additional learner activities that apparently increased demands on working memory, for example, when prompting learner self-explanations or increasing task variability during learning from worked examples (Paas & van Merriënboer, 1994b, 1994a; Sweller,

van Merriënboer, & Paas, 1998). Thus, in the traditional framework of cognitive load theory, germane cognitive load was treated as the additional (to intrinsic) relevant to learning load that further improved learning outcomes (e.g., enhanced learner transfer capabilities in a domain). However, Kalyuga (2011) argued that germane cognitive load should not be considered as an independent type of load, because it is theoretically inseparable from intrinsic load. The additional learner activities that are usually associated with germane load (such as generating self-explanations, imagining procedures, or studying a highly variable set of tasks instead of a set of similar tasks) could be readily incorporated into the definition of intrinsic load but with different learning goals to be achieved (e.g., achieving a higher-level goal of transferring the ability of applying knowledge in different situations, rather than lower-level goals of recalling or reproducing the learned procedures in similar situations). Accordingly, any activities that are relevant to learning are associated with corresponding specific learning goals and contribute to intrinsic cognitive load.

With this re-conceptualization, it has been proposed that a dual intrinsic-extraneous load typology of cognitive load (with intrinsic load defined relative to the corresponding learning goals) represents a straightforward and non-redundant approach (Kalyuga, 2011). With this approach, the aims of cognitive load theory become more clear-cut: to develop effective techniques for reducing extraneous cognitive load and managing (i.e. decreasing or increasing depending on the available cognitive resources) intrinsic cognitive load.

### *Measuring Cognitive Load*

In research studies within a cognitive load framework, cognitive load has been measured using various techniques. The existing methods for measuring cognitive load could be roughly classified along two dimensions: objective vs. subjective measures and direct vs. indirect measures. The first dimension concerns “whether the methods use subjective, self-reported data or objective observations of behavior, physiological conditions, or performance” while the second dimension involves “methods based on the type of relation of the phenomenon observed by the measure and the actual attribute of interest” (Brünken, Plass, & Leutner, 2003, p. 55). Accordingly, direct objective measures refer to the methods that can objectively measure learner characteristics that are directly related to cognitive (working memory) load when it is happening, for example, dual-task technique, eye tracking, or measures of brain activity such as fMRI (Jarodzka, Scheiter, Gerjets, & Van Gog, 2010; Korbach et al., 2018; Park & Brünken, 2015). Indirect objective measures (for example, physiological mea-



asures such as EEG, behavioral measures such as linguistic indices, or learning outcome measures) are objective measures that are regarded as indirect as they depend on tracing the processes that are assumed to be affected by cognitive load (Brünken et al., 2003). Direct subjective measures, including learners' rating of the experienced difficulty of the materials or exerted mental effort (Kalyuga et al., 1999), assume that cognitive load as a construct is directly connected with learning difficulty or invested mental effort and consequently, it could be indicated by learners' self-reports. Finally, indirect subjective measures rely on learners' ratings of subjective experiences that are presumably affected by or correlated with cognitive load, such as self-reported stress levels (Brünken et al., 2003).

Even though subjective measures of mental workload have long history of applications in research and practice (e.g., NASA-TLX index Hart, 2006; Hart & Staveland, 1988), subjective ratings of cognitive load associated with specific learning tasks and instructional materials have been used in research studies within a cognitive load perspective since 1990s, as they are easy to implement even in realistic classroom settings and do not interfere with main learning task activities (Paas, 1992; F. G. W. C. Paas, van Merriënboer, & Adam, 1994; Paas & van Merriënboer, 1994b, 1994a). Paas (1992) was first to suggest using a single-scale subjective measure of mental effort in educational psychology research. This simple scale and its variation - the task's difficulty scale (Kalyuga, Chandler, & Sweller, 1998) - have been used in multiple studies that generated various cognitive load effects (e.g., Diao, Chandler, & Sweller, 2007; Kalyuga et al., 1999, 2004; van Gog, Paas, & van Merriënboer, 2006).

Subjective ratings have been usually collected with Likert-type scales: participants are asked to estimate how easy or difficult were previous instructions to understand or learning task to perform (perceived task difficulty) or/and how much mental effort they invested into learning or performance (perceived mental effort) by choosing a number on the scale, ranging from 1 (extremely easy; least effort) to 7 or 9 (extremely difficult; most effort). Seven- or nine-point scales have been usually used. Such scales are typically presented immediately after a learning or problem-solving task, and require no more than a minute to run. For example, a simple technique that Schwamborn, Thillmann, Opfermann, and Leutner (2011) adopted to assess the amount of experienced cognitive load was to ask participants to rate both their invested mental effort and perceived task difficulty with two corresponding items using seven-point scales. Two similar subjective rating items were also used more recently by Schmeck et al. (2015).

Historically, subjective ratings of mental load were demonstrated to be able to provide sufficiently valid and

reliable estimates of perceived mental load in a non-intrusive way, and to correlate highly with objective measures (Braarud, 2001; Eggemeier, 1988; Gopher & Braune, 1984; Hill et al., 1992; Moray, 1982; Nygren, 1991; O'Donnell & Eggemeier, 1986; Xie & Salvendy, 2000). More recently, research in cognitive load theory has also shown that subjective rating measures are sensitive to variations in cognitive load between different designs of instructional materials (see F. Paas et al., 2003, for an overview). Of course, the validity of subjective ratings is based on the assumption that participants can adequately reflect and report on their cognitive processes during learning or performance. This assumption may not hold in some situations, for example, when dealing with younger children or when a significant share of automatic (subconscious) processes is involved. Also, participants' subjective understandings of the rating scales could be different. Another limitation of subjective scales is that usually they evaluate cognitive load retrospectively based on reporting participants' load during the previous learning session, rather than concurrently. Some studies argue in favour of concurrent measures as more accurately representing the real nature of cognitive load phenomena (Kalyuga & Plass, 2017).

Developing instruments that are capable of measuring not only overall cognitive load but also different constituent types of cognitive load has been a challenging direction of research (e.g., Cierniak, Scheiter, & Gerjets, 2009; DeLeeuw & Mayer, 2008). For example, Gerjets, Scheiter, Opfermann, Hesse, and Eysink (2009) suggested using variations of wording in rating scale items for this purpose. For evaluating intrinsic load, participants were required to rate the difficulty of the task domain. The level of extraneous load was evaluated using ratings of the difficulty of differentiating between important and unimportant information, and the difficulty of dealing with the environment. Finally, for ratings of germane load, participants were asked how much effort they made in understanding the content.

Gerjets, Scheiter, and Catrambone (2006) modified three selected items from the NASA-TLX scale that corresponded to task demands (the required mental and physical activity), navigational demands (the effort required for navigating in the learning environment), and the participant's effort (the required hard work) in an attempt to evaluate intrinsic, extraneous, and germane types of load respectively. However, they did not obtain conclusive evidence that these three items measured the corresponding types of cognitive load. Overall, there has been some scepticism in the field about the capability of the separate subjective rating scales to distinguish all three different types of cognitive load (e.g., Sweller et al., 2011).

A common criticism of the above versions of subjective rating scales is that using only one- or two-item scales



(for evaluating either overall cognitive load or each type of load) does not guarantee a sufficient level of reliability. As a reaction to this criticism, Leppink et al. (2013) proposed a multidimensional scale (“ten-item questionnaire for the measurement of IL, EL, and GL”) that included three items for intrinsic load, three items for extraneous load, and four items for germane load. The scale was validated in four experiments with participants learning statistics. In two experiments, the results of principal component analysis (a form of exploratory factor analysis, Experiment 1) and exploratory factor analysis (Experiment 3) supported the hypothesized component loadings between ten items and three factors. However, in the second and fourth experiments, the three-factor solution was only partially supported. In addition, the correlation between the germane load ratings and task performance scores was not statistically significant. These potential issues led Leppink, Paas, van Gog, van der Vleuten, and van Merriënboer (2014) to conduct two more experiments to examine whether the suggested three-factor cognitive load rating instrument could actually differentiate the three types of load. Even though they found consistent results in regard to the explicit differentiation of intrinsic and extraneous cognitive load, they also acknowledged that “support for the assumption that the third factor in the psychometric instrument represents or closely relates to germane cognitive load is limited” (Leppink et al., 2014, p. 40).

As mentioned in the previous section, Kalyuga (2011) argued that the conceptualization of germane cognitive load was not supported by either theoretical considerations or empirical findings, and this concept seemed to be redundant given that two-factor categorization (intrinsic and extraneous types of load) was sufficient for interpreting the findings from the perspective of cognitive load theory (see also Sweller, 2010, for supporting ideas). One of the implications of the modified dual (intrinsic-extraneous) model for measurement of cognitive load is that there is no need for separate measures of all three traditional types of cognitive load, but only for its two types - intrinsic and extraneous. However, this suggestion has remained mostly a theoretical conjecture as there have been no empirical or psychometric evidence available to support it. The study reported below represents an attempt to provide such evidence.

## Method

### Participants

The participants were 60 Year 1 undergraduate students of English as a foreign language at a technological university in China. They were 21 years old on average and had more than 7 years’ experience of formal English learning when

the study was conducted. The participants were required to rate the level of cognitive load experienced during their learning by responding to a cognitive load rating questionnaire that was developed from the instrument designed by Leppink et al. (2013).

### Procedures

The learning task required participants to understand a short conversation between a passenger and a staff at an airport check-in counter. The instruction was delivered in three phases: vocabulary learning, sentences learning, and passage learning. The experiment was originally designed to compare alternative visual and auditory presentation formats in order to test specific predictions of cognitive load theory. The results were published in Jiang, Kalyuga, and Sweller (2018), however the confirmatory factor analysis of the collected subjective ratings of cognitive load was not conducted comprehensively in that study. Parts of the original data related to the subjective ratings were used to run the confirmatory factor analysis.

### Instrument

The questionnaire was distributed to the participants after they had completed the three learning phases. It should be noted that the original Leppink et al.’s (2013) questionnaire had three items on intrinsic cognitive load, three items on extraneous cognitive load, and four items on germane cognitive load. However, considering that there were three learning phases in this study, the researchers modified the questionnaire (see Appendix) to include twelve items, with four items for each type of cognitive load (Items CL1-CL4 on intrinsic load; Items CL5-CL8 on extraneous load; and Items CL9-CL12 on germane load). Within each cognitive load category, one item was used for generally rating the whole learning process, and the remaining three items focused on the three specific learning phases respectively.

Although the items CL9-CL12 were originally designed to rate the germane cognitive load according to Leppink et al. (2013), the necessity and validity of including the germane cognitive load in measuring instruments is contentious according to the modified dual model of cognitive load (Kalyuga, 2011; Sweller, 2010). Therefore, it was assumed that the last four items could actually target intrinsic cognitive load from a different perspective. Since for each of these four items, in contrast to the rest of the questionnaire, the lower quality ratings were associated with lower rating numbers (such as 0 – “not at all”), the responses to items CL9-CL12 were re-coded into the reverse scale before analyses. In accordance with the modified dual model of cognitive load, it was hypothesized that the first four (1-4) and last four (9-12) items could be used to rate intrinsic cognitive load and the middle four items (5-



**Table 1** ■ Means, standard deviations, skewness, kurtosis, and correlations for the questionnaire items

Item	Mean	SD	Skewness	Kurtosis	$R^2$
CL1	2.83	1.15	0.39	-0.22	0.61
CL2	2.97	1.10	0.36	-0.27	0.53
CL3	3.03	1.06	0.36	-0.09	0.63
CL4	2.85	1.33	0.83	0.65	0.67
CL5	3.90	1.10	-0.03	-0.74	0.36
CL6	3.92	1.20	-0.19	-0.25	0.43
CL7	3.98	1.17	-0.28	-0.50	0.29
CL8	3.87	1.26	0.55	-0.27	0.37
CL9	3.20	1.02	0.26	-0.43	0.25
CL10	3.10	0.99	0.22	-0.02	0.27
CL11	3.20	1.02	-0.12	-0.51	0.25
CL12	3.50	1.14	70.44	-0.20	0.46

**Table 2** ■ Covariance estimates, standard errors, *t* values, and *p* values for the questionnaire items

Item	Covariance estimates	Standard errors	<i>t</i> Value	<i>p</i> Value
CL1	.49	.11	4.36	<.001
CL2	.53	.11	4.64	<.001
CL3	.45	.10	4.50	<.001
CL4	.41	.12	3.33	=.001
CL5	.79	.18	4.39	<.001
CL6	.86	.20	4.13	<.001
CL7	.90	.20	4.42	<.001
CL8	.75	.23	3.22	=.001
CL9	.93	.17	5.41	<.001
CL10	.83	.16	5.38	<.001
CL11	.95	.17	5.42	<.001
CL12	.81	.16	5.10	<.001

8)- to rate extraneous cognitive load.

**Statistical Analyses**

The reliability of the scale as measured by Cronbach’s alpha was .78. In the situations where the factors are theoretically defined and the loadings of indicator items onto specific factors are hypothesized before testing, confirmatory factor analyses are usually recommended (Wang & Wang, 2012). A confirmatory factor analysis of the collected data was carried out using the programming language R on the RStudio software (RStudio team, 2016) to test the factorial validity of the instrument.

**Results**

The confirmatory factor analysis was conducted by using the lavaan package (Rosseel, 2012). As the data was stored in a SPSS file, the R package foreign (R Core Team, 2017) was used to access the questionnaire data. Prior to carrying out the confirmatory factor analysis, the mean and standard deviation, skewness, kurtosis, and squared

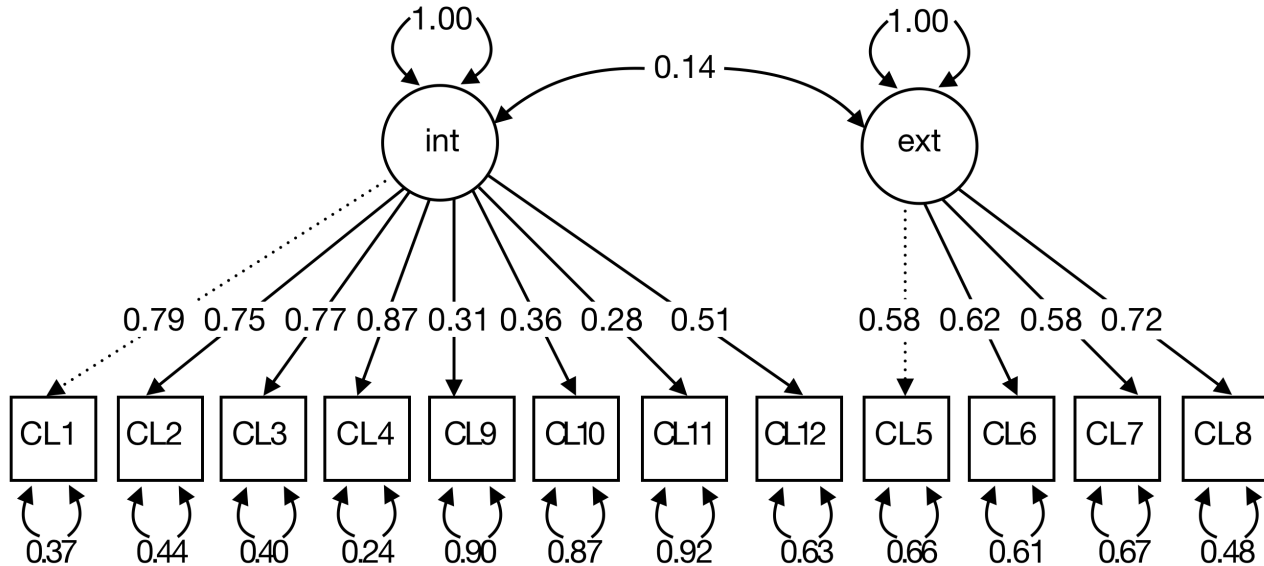
multiple correlations ( $R^2$ ) of each item in the subjective rating questionnaire were examined using the describe and smc functions from the psych package (R Core Team, 2018).

Table 1 shows the mean, standard deviation, skewness, kurtosis, and squared multiple correlations ( $R^2$ ) of each item. The data approximate a normal distribution as the absolute values of skewness and kurtosis are smaller than the cut-offs of 2 and 7 respectively (Curran, West, & Finch, 1996). In addition, the squared multiple correlations ( $R^2$ ) were above the accepted level (.25) of item reliability.

A two-factor model was specified according to the hypothesis by commanding the appropriate lavaan syntax, and a confirmatory factor analysis was conducted through the cfa function. The essential index values were summarized and reported. The  $\chi^2$  statistics for the two-factor model was  $\chi^2 = 54.069, df = 53 (p = .433)$ , indicating that the two-factor model was a good fit. Both  $CFI = 0.994$  and  $TLI = 0.993$  were greater than .95, also indicating an ideal fit (Westland, 2015). As both RMSEA (0.018) and



Figure 1 ■ Factor loading paths of indicators onto latent variables



SRMR (0.077) indices were less than 0.08, the two-factor model fit was acceptable (Westland, 2015). Factor loadings details including covariance estimates, standard errors, t values, and p values of the items, are summarized in Table 2. The factor loadings of each indicator onto the two latent variables are displayed in Figure 1. It can be seen that all items except one have factor loadings greater than 0.30 (CL11 has a loading of 0.28), thus generally satisfying the conventionally accepted cut-off value of factor loading (Kim & Mueller, 1978; Brown, 2006).

### Discussion

Because of their simplicity, practicality, and non-intrusiveness into learning processes, subjective rating scales have been used to measure cognitive load (including different types of load) in the majority of studies within a cognitive load framework, assuming that there is a direct relation between participants' self-reports and actual levels of cognitive load. Most of publications in cognitive load theory since 1998 have mentioned three additive types of cognitive load - intrinsic, extraneous, and germane. While Sweller (2010) and Kalyuga (2011) proposed theoretical reasons against this triarchic framework and in favour of the modified dual model, this paper provides psychometric empirical support for this model.

The reported study performed a confirmatory factor analysis of the subjective rating data collected with a multidimensional cognitive load questionnaire in the area of

learning English as a foreign language listening skills by using the programming language R. The confirmatory factor analysis was conducted through the lavaan package by executing relatively simple commands. A variety of fit indices (chi-square, CFI, TLI, RMSEA, and SRMR) satisfied the cut-off values of acceptable model fit (Kim & Mueller, 1978; Brown, 2006; Westland, 2015). The results of the confirmatory factor analysis supported the hypothesized two-factor (intrinsic and extraneous) model as an acceptable fit. Even though Mardia (1974) claimed that small samples did not necessarily distort the normality of data, it should be acknowledged that the relatively small sample size of this study (60) could be a potential limitation in interpreting the findings. Future research should adopt larger samples to conduct factor analysis on the categories of cognitive load in a broader range of subject domains.

As a dynamic characteristic of working memory, cognitive load changes on the short timescale of working memory operation (Kalyuga & Plass, 2017). Therefore, ideally, it is necessary to develop measures of cognitive load at local levels and in real time. If subjective rating scales are used during learning or performance tasks rather than retrospectively, they may potentially serve as such local-level tools, although they could be intrusive and external to the learning or performance processes, especially in their multi-item and multi-dimensional forms investigated in this paper. Even though eliminating one dimension could be helpful in reducing the interference that



the ratings placed on learning activities, technically more sophisticated methods such as psychophysiological techniques are better suitable for measuring cognitive load in real time. However, they can be used mostly in laboratory settings. As a simple alternative, Kalyuga and Plass (2017) suggested using concurrent verbal reports as a method for measuring cognitive load in real time at the local level. Potentially, this method could be integrated with the subjective ratings techniques by using selected rating questions (items) as prompts during the concurrent verbal reports and parts of pre-report instructions to participants.

### Conclusion

The dual (two-dimensional) model of cognitive load makes the cognitive load theory framework more straightforward and transparent by eliminating the redundant concept of germane load, and rendering unnecessary the efforts in developing separate psychometric measures for all three types of cognitive load. The reported study provided empirical psychometric evidence for the validity of the dual model, thus allowing to further clarify the basic concepts of cognitive load theory.

### References

- Anderson, J. (1983). Retrieval of information from long-term memory. *Science*, 220, 25–30. doi:10.1126/science.6828877
- Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review*, 22, 425–438. doi:10.1007/s10648-010-9130-y
- Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- Braarud, P. O. (2001). Subjective task complexity and subjective workload: Criterion validity for complex team tasks. *International Journal of Cognitive Ergonomics*, 5, 261–273. doi:10.1207/S15327566IJCE0503\_7
- Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Brünken, R., Plass, J., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38, 53–61. doi:10.1207/S15326985EP3801\_7
- 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. Retrieved from <http://doi.org/10.1016/j.chb.2008.12.020>
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–114. Retrieved from <http://doi.org/10.1017/S0140525X01003922>
- Cowan, N., & Chen, Z. (2009). How chunks form in long-term memory and affect short-term memory limits. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 86–101). Washington, DC: Psychology Press.
- Curran, P., West, S., & Finch, J. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16–29. doi:10.1037/1082-989X.1.1.16
- DeLeeuw, K., & Mayer, R. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, 100, 223–234. doi:10.1037/0022-0663.100.1.223
- Diao, Y., Chandler, P., & Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology*, 120, 237–261.
- Eggemeier, F. (1988). Properties of workload assessment techniques. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 41–62). New York: Elsevier.
- Eysenck, M., & Keane, M. (2015). *Cognitive psychology: A student's handbook*. Washington, DC: Psychology Press.
- Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction*, 16, 104–121. Retrieved from <http://doi.org/10.1016/j.learninstruc.2006.02.007>
- Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F., & Eysink, T. (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, 360–370. doi:10.1016/j.chb.2008.12.015
- Gilchrist, A. L., & Cowan, N. (2012). Chunking. In V. Ramachandran (Ed.), *The encyclopedia of human behavior, 2nd edition* (pp. 476–483). San Diego, CA: Academic Press.
- Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors*, 26, 519–532. doi:10.1177/001872088402600504
- Hart, S. (2006). Nasa-task load index (nasa-tlx); 20 years later. In H. Factors & E. Society (Eds.), *Proceedings of*



- the human factors and ergonomics society 50th annual meeting* (pp. 904–908). Washington, DC: HFES.
- Hart, S., & Staveland, L. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload advances in psychology* (Vol. 52, pp. 139–183). doi:[10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hill, S., Laveccchia, H., Byers, J., Bittner, A., Zaklad, A., & Christ, R. (1992). Comparison of four subjective workload rating scales. *Human Factors*, *34*, 429–439. doi:[10.1177/001872089203400405](https://doi.org/10.1177/001872089203400405)
- Jarodzka, H., Scheiter, S., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*, 146–154.
- Jiang, D., Kalyuga, S., & Sweller, J. (2018). The curious case of improving foreign language listening skills by reading rather than listening: An expertise reversal effect. *Educational Psychology Review*, *30*, 1139–1165. doi:[10.1007/s10648-017-9427-1](https://doi.org/10.1007/s10648-017-9427-1)
- Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, *23*(1), 1–19. doi:[10.1007/s10648-010-9150-7](https://doi.org/10.1007/s10648-010-9150-7)
- Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors*, *40*, 1–17. doi:[10.1518/001872098779480587](https://doi.org/10.1518/001872098779480587)
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, *13*, 351–371. doi:[10.1002/\(SICI\)1099-0720\(199908\)13:4](https://doi.org/10.1002/(SICI)1099-0720(199908)13:4)
- Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors*, *46*, 567–581.
- Kalyuga, S., & Plass, J. (2017). Cognitive load as a local characteristic of cognitive processes: Implications for measurement approaches. In R. Zheng (Ed.), *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 59–74). Washington, DC: Routledge.
- Kim, J., & Mueller, C. (1978). *Factor analysis: Statistical methods and practical issues*. New York: Sage Publications.
- Kirschner, P., Sweller, J., & Clark, R. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*, 75–86. Retrieved from [http://doi.org/10.1207/s15326985ep4102\\_1](http://doi.org/10.1207/s15326985ep4102_1)
- Korbach, A., Brünken, R., & Park, B. (2018). Differentiating different types of cognitive load: A comparison of different measures. *Educational Psychology Review*, *30*, 503–529. doi:[10.1007/s10648-017-9404-8](https://doi.org/10.1007/s10648-017-9404-8)
- Leppink, J., Paas, F., van der Vleuten, C., van Gog, T., & van Merriënboer, J. (2013). Development of an instrument for measuring different types of cognitive load. *Behaviour Research*, *45*, 1058–1072. doi:[10.3758/s13428-013-0334-1](https://doi.org/10.3758/s13428-013-0334-1)
- Leppink, J., Paas, F., van Gog, T., van der Vleuten, C., & van Merriënboer, J. (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](https://doi.org/10.1016/j.learninstruc.2013.12.001)
- Mardia, K. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya: The Indian Journal of Statistics, Series B*, *36*, 115–128.
- Mayer, R., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Experimental Psychology*, *90*, 312–320. doi:[10.1037/0022-0663.90.2.312](https://doi.org/10.1037/0022-0663.90.2.312)
- Moray, N. (1982). Subjective mental workload. *Human Factors*, *24*, 25–40. doi:[10.1177/001872088202400104](https://doi.org/10.1177/001872088202400104)
- Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, *33*, 17–33. doi:[10.1177/001872089103300102](https://doi.org/10.1177/001872089103300102)
- O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (vol. (pp. 1–49). Chapter 42, . Wiley: 2.
- Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*, 429–434. doi:[10.1037/0022-0663.84.4.429](https://doi.org/10.1037/0022-0663.84.4.429)
- Paas, F. G. W. C., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*, 419–430. doi:[10.2466/pms.1994.79.1.419](https://doi.org/10.2466/pms.1994.79.1.419)
- Paas, F., Tuovinen, J., Tabbers, H., & van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*, 63–71. doi:[10.1207/S15326985EP3801\\_8](https://doi.org/10.1207/S15326985EP3801_8)
- Paas, F., & van Merriënboer, J. J. G. (1994a). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*, 351–371. doi:[10.1007/BF02213420](https://doi.org/10.1007/BF02213420)
- Paas, F., & van Merriënboer, J. J. G. (1994b). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach.





- Journal of Educational Psychology*, 86, 122–133. doi:[10.1037/0022-0663.86.1.122](https://doi.org/10.1037/0022-0663.86.1.122)
- Park, B., & Brünken, R. (2015). The rhythm method: A new method for measuring cognitive load — an experimental dual task study. *Applied Cognitive Psychology*, 29, 232–243.
- Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193–198. doi:[10.1037/h0049234](https://doi.org/10.1037/h0049234)
- R Core Team. (2017). Foreign: Read data stored by ‘minitab’ (Version 0.8-70). Retrieved from <https://CRAN.R-project.org/package=foreign>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Austria: R Foundation for Statistical Computing. Vienna. Retrieved from <https://www.r-project.org/>
- Rosseel, Y. (2012). Lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36.
- 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. doi:[10.1007/s11251-014-9328-3](https://doi.org/10.1007/s11251-014-9328-3)
- Schwaborn, A., Thillmann, H., Opfermann, M., & Leutner, D. (2011). Cognitive load and instructionally supported learning with provided and learner-generated visualizations. *Computers in Human Behaviour*, 27, 89–93. doi:[10.1016/j.chb.2010.05.028](https://doi.org/10.1016/j.chb.2010.05.028)
- 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. doi:[10.1007/s10648-010-9128-5](https://doi.org/10.1007/s10648-010-9128-5)
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. doi:[10.1007/978-1-4419-8126-4](https://doi.org/10.1007/978-1-4419-8126-4)
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296. doi:[10.1023/a:1022193728205](https://doi.org/10.1023/a:1022193728205)
- Szulewski, A., Roth, N., & Howes, D. (2015). The use of task-evoked pupillary response as an objective measure of cognitive load in novices and trained physicians: A new tool for the assessment of expertise. *Academic Medicine*, 90, 981–987. doi:[10.1097/ACM.0000000000000677](https://doi.org/10.1097/ACM.0000000000000677)
- van Gog, T., Paas, F., & van Merriënboer, J. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction*, 16, 154–164.
- van Merriënboer, J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent development and future directions. *Educational Psychology Review*, 17, 147–177. doi:[10.2307/23363899](https://doi.org/10.2307/23363899)
- Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. doi:[10.1002/9781118356258](https://doi.org/10.1002/9781118356258)
- Westland, J. (2015). *Structural equation models: From paths to networks*. doi:[10.1007/978-3-319-16507-3](https://doi.org/10.1007/978-3-319-16507-3)
- Xie, B., & Salvendy, G. (2000). Prediction of mental workload in single and multiple task environments. *International Journal of Cognitive Ergonomics*, 4, 213–242. doi:[10.1207/S15327566IJCE0403\\_3](https://doi.org/10.1207/S15327566IJCE0403_3)



### Appendix: Subjective Rating Scale

All of the following questions refer to the learning activities that have just finished. Please respond to each of the questions on the following scale (0 meaning *not at all the case* and 10 meaning *completely the case*).

No.	Items	Scales
CL1	The topic covered in the lesson was very complex.	0 1 2 3 4 5 6 7 8 9 10
CL2	The vocabulary learning activity was very complex.	0 1 2 3 4 5 6 7 8 9 10
CL3	The sentence learning activity was very complex.	0 1 2 3 4 5 6 7 8 9 10
CL4	The passage learning activity was very complex.	0 1 2 3 4 5 6 7 8 9 10
CL5	The instructions and explanations during the lesson were, in terms of listening learning, very ineffective.	0 1 2 3 4 5 6 7 8 9 10
CL6	The vocabulary learning activity was very unclear.	0 1 2 3 4 5 6 7 8 9 10
CL7	The sentence learning activity was very unclear.	0 1 2 3 4 5 6 7 8 9 10
CL8	The passage learning activity was very unclear.	0 1 2 3 4 5 6 7 8 9 10
CL9	The lesson really enhanced my understanding of the topic covered.	0 1 2 3 4 5 6 7 8 9 10
CL10	The lesson really enhanced my understanding of the vocabulary covered.	0 1 2 3 4 5 6 7 8 9 10
CL11	The lesson really enhanced my understanding of the sentenced covered.	0 1 2 3 4 5 6 7 8 9 10
CL12	The lesson really enhanced my understanding of the passage.	0 1 2 3 4 5 6 7 8 9 10

### Citation

Jiang, D., & Kalyuga, S. (2020). Confirmatory factor analysis of cognitive load ratings supports a two-factor model. *The Quantitative Methods for Psychology*, 16(3), 216–225. doi:[10.20982/tqmp.16.3.p216](https://doi.org/10.20982/tqmp.16.3.p216)

Copyright © 2020, *Jiang and Kalyuga*. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 17/06/2019 ~ Accepted: 05/05/2020