International Journal of Progressive Education, Volume 20 Number 6, 2024

Turkish Adaptation Study of the Cognitive Load Scale: Reliability and Validity of the Cognitive Load Scale in Turkish Culture

Özden Demir [i]
Trabzon University

Zeynep Ayvaz Tuncel [ii]
Pamukkale University

Abstract

The purpose of this study is to perform Turkish adaptation of the "Cognitive Load Scale (CLS)" developed by Hwang, Yang, and Wang (2013) and to test its validity and reliability in Turkish culture. The Cognitive Load Scale was developed to determine the cognitive load experienced by learners during any learning and instructional activities. The scale consists of mental load and mental effort sub-factors that aim to determine the cognitive load experienced by learners during educational activities. The original scale consists of eight items and two sub-factors, with five items in the mental load sub-factor and three items in the mental effort sub-factor. Turkish adaptation of the scale was conducted on a sample of 376 pre-service teachers enrolled in two different education faculties. The Turkish scale, the validity and reliability of which was conducted with 376 pre-service teachers, was found to have a four-item and two-factor structure; the factor structures were valid; and internal consistency coefficients were found 0.80 for the total scale, .89 for the first sub-factor, and .78 for the second sub-factor. Besides, the confirmatory factor analysis revealed the following fit index values: ($\chi2$=2,2, sd=1, p>.01), RMSEA=.056 and $\chi2$/df=2,2 and RMR= .014, GFI= 0.997, AGFI= 0.971, NFI= 0. 99, CFI=0.99, IFI= 0.99, indicating the recommended criteria. It can be said that the adapted scale can be used in academic studies related to cognitive load.

Keywords: Cognitive Load; Intrinsic Cognitive Load; Extraneous Cognitive Load; Effective Cognitive Load; Mental Load; Mental Effort

DOI: 10.29329/ijpe.2024.1078.3

Submitted: 22/07/2023 **Accepted:** 15/10/2024 **Published:** 01/12/2024

----

[i] **Özden Demir,** Trabzon University, Faculty of Education, Department of Educational Sciences, Department of Curriculum and Instruction, Trabzon/Türkiye oooozden@gmail.com ORCID: 0000-0001-6293-5045

Correspondence: oooozden@gmail.com

[ii] **Zeynep Ayvaz Tuncel,** Pamukkale University, Faculty of Education, epartment of Educational Sciences, Department of Curriculum and Instruction, Denizli/ Turkey zatuncel@pau.edu.tr ORCID: 0000-0003-1817-9603

INTRODUCTION

The individual is an active participant who takes responsibility for his/her own learning in the process of learning to learn. The ability to regulate and control cognition is directly related to the learner's realization of meaningful and functional learning in the individual and active structure of learning. Such a process requires interactive teaching environments related to effective instructional design models that increase the effective cognitive load and decrease the extraneous cognitive load. Reorganizing different knowledge, attitudes, and skills and creating new knowledge, skills, and attitudes are highly important for learners when they face complex problems. The learners' organization network between the working memory and long-term memory is of crucial importance for the use of this cognitive architecture for the solution to novel problems. Hence, recent developments in the fields of instructional design, learning psychology, and cognitive neuroscience have clarified the socio-emotional and effortful features of human cognition and its complex structure that finds meaning in the individual. For this reason, the importance of cognition (features such as the interaction and organization of networks between working memory and long-term memory) is valuable in using instructional design structures for more effective and meaningful learning. In such a process, we encounter the cognitive load theory as an effective theory in the instructional design process (Sweller, Van Merriënboer & Paas, 2019; Paas & Van Merriënboer, 2020). Over the last three decades, the cognitive load theory (CLT) has evolved to understand the effects of instruction and instructional materials on students' achievement (Feldon, Callan, Juth & Jeong, 2019).

Limited capacity and duration of working memory are the focus of CLT (Sweller, Merriënboer & Paas, 2019). In this process, the sum of intrinsic cognitive load, extraneous cognitive load, and effective cognitive load should not exceed the working memory capacity so that learners can maintain their meaningful learning (Sweller & Sweller, 2006). For this reason, the determination of the unique cognitive load of each learner is important. Since it is not possible to measure cognitive load directly in a practical way, its measurement has recently become an issue addressed in the literature. In this respect, the purpose of this study is to perform Turkish adaptation of the "Cognitive Load Scale (CLS)", which was developed by Hwang, Yang, and Wang (2013) and which addresses reliable and valid indicators to address the types of cognitive load in the current literature, and to test its validity and reliability in Turkish culture.

Theoretical Framework

The Cognitive Load Theory

The cognitive load theory (CLT) is an instructional theory based on human cognitive structure and architecture and developmental psychology (Sweller, Ayres & Kalyuga, 2011; Sweller, Merriënboer & Paas, 2019). Five basic principles can be used to explain this theory (Sweller & Sweller, 2006). A general evaluation of these principles indicates the following structure: learners receive information from others, which is usually teachers, in the problem-solving process and analyze it using the *reorganization* principle. If they cannot receive information from others, they randomly produce new information and test its effectiveness by using randomness as the *formation principle*. Before it is stored indefinitely in our large, long-term memory using the principle of *information storage*, this information is processed in working memory, which has limited capacity and duration, using the principle of *narrow limits of change*. To manage appropriate action using the *environmental regulation and connection* principle, uncertain amounts of stored information can then be transferred back to working memory based on signals from the external environment. These five principles serve as the cognitive basis for instructional design (Sweller et al., 2011).

Three types of cognitive loads are included in the cognitive load theory: intrinsic load, extraneous load/ineffective load, and germane load/effective load (Sweller, Van Merrienboer & Paas, 1998). Intrinsic cognitive load depends on the nature (difficulty) of the content to be learned as well as the learner's experience level (Van Merrienboer and Ayres, 2005). Extraneous cognitive load is the loading of working memory as a result of poor instructional materials and designs. Extraneous cognitive load is high when the instructional design includes inappropriate content or different instructional materials negatively affecting the information processing process. With this notion, extraneous cognitive load is seen as an error in the whole instructional process (Van Merrienboer and Ayres, 2005), yet it is possible to regulate and control the extraneous cognitive load with the role of instructional designers. Germane or effective cognitive load emerges in the processes enabling the organization and regulation of mental structures. Forming effective cognitive load is of vital importance in forming functional meaning-making process in learning (Chen, Paas & Sweller, 2021). In such a process, the sum of intrinsic cognitive load, extraneous cognitive load and effective cognitive load should not exceed the capacity of working memory (Sweller & Sweller, 2006). In this process, CLT suggests that the human cognitive architecture develops to effortlessly assimilate primary knowledge (i.e., survival knowledge such as recognizing faces) rather than biologically secondary knowledge (i.e., cultural knowledge such as reading texts) (Sweller et al., 2011).

Cognitive load is generally defined as a structure representing the load of performing a particular task on the cognitive system (Sweller, Van Merriënboer & Paas, 1998). Both are conceptualized as a task-based dimension (i.e., mental load) and a practice-based dimension (i.e., mental effort) affecting the learner's performance. We encounter the concepts of mental load and mental effort in this process (Sweller, Van Merriënboer & Paas, 1998; Hwang, Yang & Wang, 2013). Based on these arguments, teaching competence is best predicted by a combination of the mental effort employed and the level of tasks performed by learners (Van Merriënboer & Paas, 1998). Therefore, learners' mental effort while performing learning tasks is also an important cognitive component in cognitive load, especially in effective cognitive load. Mental effort refers to the amount of cognitive capacity or resources employed by the learner for performing the task (Sweller, Van Merriënboer & Paas, 1998). In addition, three main categories can be classified for the mental effort measurement techniques (Wierwille and Eggemeier, 1993), which includes subjective, physiological, and task- and performance-based indices, with a range of individual assessment techniques in each category. This cognitive architecture also consists of two separate memories that have different capacities and qualities. While Long Term Memory (LTM) is considered permanent and hypothetically unlimited, Working Memory (WM) is a temporary information processor with limited capacity (Cowen, 2001). Learning can be hindered by cognitive overload due to the limited capacity of WM (Sweller et al., 2019). Therefore, CLT aims to optimize the information processing load while learning biologically secondary information (Sweller et al., 2019). In this process, minimizing the intrinsic and extraneous (ineffective load) cognitive load and making arrangements related to instructional design to make germane load/effective load functional are considered to increase the level of meaningful learning in the instruction process. While intrinsic and extraneous cognitive load varies mostly depending on the instructional environment, effective cognitive load is related to processes directly related to learning, such as the creation and automation of schemas (Van Merrienboer and Ayres, 2005). While the intrinsic cognitive load is affected by the topic being learned, the extraneous cognitive load is affected by the content organization design of the topic being learned. The amount of effective cognitive load is considered to increase as appropriate instructional designs reduce extraneous cognitive load and contribute to learners creating more interactive schemas. For this reason, teachers to provide learners with guided learning as qualified instructional designers could contribute to the desired level of cognitive load experienced by students. Therefore, it is important to organize the Teaching Principles and Methods course, which is an important course in teacher education, in a way to contribute to increasing the effective cognitive load of instructional design and to provide pre-service teachers with the necessary competencies to minimize the intrinsic and extraneous cognitive load in instructional design. In this process, the determination of the level of cognitive load among pre-service teacher learners becomes an important topic.

Measurement of Cognitive Load

Although the multidimensional structure of cognitive load and the measurement of the complex relationship between performance, mental load, and mental effort remains unclear for researchers (Sweller, Van Merrienboer & Paas, 1998), studies have generally used two different methods, namely experimental and analytical methods. Experimental methods include the estimation of mental effort and performance, the collection of subjective data using rating scales, and the collection of performance-related data using primary and secondary task techniques. The majority of studies on cognitive load theory seem to have used rating scales, psychological and secondary task techniques, while analytical methods were not preferred much (Paas, Renkll & Sweller, 2004). An analysis of the literature (Paas, 1992 in Yang, Jen, Chang & Yeh, 2018; Cierniak, Scheiter & Gerjets, 2009; Leppink, Paas, Vleuten, Gog & Merriënboer, 2013; Anmarkrud, Andresen & Bråten, 2019; Hwang, Yang & Wang, 2013) shows that many direct and indirect measurement tools have been developed to determine learners' cognitive load levels. Observations done using the Mental Effort Rating Scale developed by Paas (1992) is a pioneer in the field (Paas, Van Merriënboer, & Adam, 1994; as cited in Yang, Jen, Chang, & Yeh, 2018). The scale requires participants to rate their cognitive effort on the learning task on a nine-point scale (1-very little to 9-very much). As they have shown a positive correlation with participants' performance, this scale as well as single-item cognitive load scales have been used in different studies. Besides, either general cognitive load or one of the cognitive load types is measured. In most cases, a single Likert item is used and the number of categories in the item typically varies. The number of items can be five (e.g., Salden, Paas, Broers, and Van Merriënboer, 2004), six (e.g., Cierniak et al., 2009) or nine (e.g., Eysink et al., 2009; Paas, 1992).

The scale developed by Cierniak et al. (2009) addresses three dimensions: the perceived difficulty of the learning content (ICL), perceived difficulty of the material presented (ECL), and concentration during the learning process (GCL). Similarly, a ten-item Likert-type scale was developed by Leppink et al. (2013 to address ICL, ECL, and GCL. Ayres (2006) also developed a rating scale to measure ICL, ECL, and GCL separately (in Eysink, De Jong, Berthold, Kollöffel, Opfermann & Wouters, 2009). On the other hand, although it is not clear to what extent workload and cognitive load refer to the same concept in different settings, the NASA-TLX is used to assess workload on a five to seven-point scale (Hilbert & Renkl, 2009). In addition, a ten-item scale was developed by Leppink et al. to address ICL, ECL, and GCL. This scale includes three items in ICL, three items in ECL, and four items in GCL. Hwang et al. (2013) developed the Cognitive Load Scale based on the measurements by Paas (1992) and Sweller, Van Merriënboer, and Paas (1998). The scale is a six-point Likert scale consisting of two sub-scales as mental load (five items) and mental effort (three items). Dönmez, Akbulut, Telli, Kaptan, Özdemir & Erdem (2022) developed a 13-item and three-factor scale to address intrinsic, extraneous, and effective cognitive load in computer-based learning environments. Cronbach's alpha coefficient was found.88 for the total scale, indicating that it is a functionally comprehensive scale addressing all three dimensions of Turkish culture. However, despite broader acknowledgments indicating the requirements of the integration of both cognitive and non-cognitive processes of learning functions (Plass & Kaplan, 2015), most cognitive load theory studies focus on the relationship between memory demands during learning, schema formation, and subsequent performance, without paying attention to the interactions between emotions or motivation and cognitive load (Feldon, Callan, Juth & Jeong, 2019)." The performance of working memory affects the cognitive load. Therefore, the necessary importance given to mental effort is believed to contribute to improving the level of germane and effective cognitive load. For this reason, the Cognitive Load Scale developed by Hwang et al. (2013) is functional for the evaluation of both mental load and mental effort as a whole. As stated by Sweller et al. (1998), the determination of learners' cognitive load levels is important in making effective cognitive load functional. In addition, another factor that makes the study valuable includes addressing the related scale in the context of Turkish culture and looking at its functionality in this direction.

This study performed the adaptation of the scale based on the theories by Sweller et al. (1998) and Pass (1992), which are accepted in the field of cognitive load, to Turkish culture and examined its validity and reliability using confirmatory factor analysis. Hence, the purpose of this study is to test the reliability and validity of the Turkish version of the Cognitive Load Scale (CLS). This scale was adapted so that it could be used in Turkey to determine the cognitive load tendencies (mental load and effort sub-scales) in the activities performed by students in the learning process as well as different types of research to be conducted in this field. Hence, the study tests and discusses the reliability and validity of the Cognitive Load Scale in Turkish culture.

METHODS

Study Design and The Study Group

The Cognitive Load scale was developed by Hwang, Yang, and Wang (2013) to measure learners' cognitive load levels in any learning and instructional activities. Two main studies were conducted during the Turkish adaptation process of the scale: (1) translation study and (2) validity and reliability study. Expert linguists were consulted during the translation phase. Exploratory and confirmatory factor analysis methods were used during the validity study of the scale.

The study group for the Turkish adaptation of the Cognitive Load Scale consisted of 400 students who were enrolled in the teaching program at the Faculties of Education at Pamukkale University (n:214) and Trabzon University (n:186) in the 2021-2022 academic year. After data were organized and outliers and missing values were removed, the study data were from 376 participants including 170 students from TRU and 206 students from PU.

Measurement Tool

The measurement tool was developed by Hwang, Yang, and Wang (2013) based on the measurements by Paas (1992) and Sweller, Van Merriënboer and Paas (1998). It consists of eight items on a six-point Likert scale, which includes 5 items for "mental load" (items 1, 2, 3, 4, 5) and 3 items for "mental effort" (items 6, 7, and 8). Cronbach's alpha values of the two factors were found to be 0.86 and 0.85, respectively.

Data Analysis

Data were analyzed using a statistical program, and the significance level was accepted as .05 for the interpretation of the results. Descriptive analyses and group comparisons were performed using SPSS 21.0. Confirmatory factor analysis was conducted using AMOS. For validity, the construct validity of the scale adapted to Turkish was examined by exploratory factor analysis to evaluate its structure in Turkish culture. In addition, item-total correlations were examined to determine the power of the scale in distinguishing between people with high and low levels of cognitive load. On the other hand, item analysis was conducted based on the upper-lower group averages. For reliability of the scale adapted into Turkish, Cronbach's α correlation coefficient was calculated to test the consistency of the scale items with each other

FINDINGS

Scale adaptation studies aim to adapt scales developed in other cultures to different languages and cultures. Several national and international scale adaptation studies have been conducted and information on the process of cross-cultural scale adaptation has been provided in the literature. The studies conducted have many common points. In this regard, stages indicated by Cohen, Cohen, West, Aiken (2003), Field, (2005), Çokluk, Şekercioğlu, and Büyüköztürk (2012) were carried out in the Turkish adaptation process of the Cognitive Load Scale (CLS), which included the translation of the items from the original language into the target language, the determination of the equivalence of the items in the original form and the draft form, and the determination of the validity and reliability of the Turkish form obtained.

Translation Form

The translation stage is reported to constitute the most important step of the adaptation process, so as stated by Erkuş (2010) and Çokluk, Şekercioğlu, and Büyüköztürk (2012), the translator was determined considering the criteria which included possessing the knowledge of the target and source languages well, knowledge of the topic related to the scale, and experience in both cultures. In this regard, a lecturer specialized in the field of curriculum and instruction in the Faculty of Education at Pamukkale University translated the scale items from the original language to the target language, and the views of two more experts (in the foreign language and curriculum and instruction) were received for language validity.

Testing Item Equivalence

After the translation process was completed, judgmental and statistical methods were used to test the equivalence of the words and expressions in the draft form with the original scale. The "Single-Translation Method" among the judgmental methods was employed. Item equivalence is examined and evaluated according to the target language in the one-way translation method, which is the most important reason for choosing this method. In this way, in line with the common opinions of language experts who have a good command of the subject, it is possible to arrange expressions suitable for the target language into which the scale will be translated and adapt the structure of the expression in the source language to the target language (Hambleton & Bollwark, 1991; Cited in Cotiga, 2012; Erkuş, 2010). The single translation method was preferred in this respect. The Turkish translation received from the translator was assessed in terms of the vocabulary, concepts, and expressions used, and a draft form suitable for the target culture was prepared. Following this phase, two expert groups were formed consisting of lecturers from the Department of Curriculum and Instruction in the Faculty of Education at Pamukkale and Trabzon University; the translation was evaluated in terms of the concepts and expressions; and the scale was compared with the source language and evaluated. In addition, a Turkish language expert was consulted to evaluate the suitability of the Turkish translation of the draft form in terms of language. The scale items were evaluated one by one and necessary changes were made in line with the suggestions of the experts. Then the scale was administered to two students enrolled in the teaching program in the Faculty of Education at Trabzon University to evaluate the comprehensibility of the scale items as well as their suitability for the students who would be administered the scale. To collect data about item equivalence, the students were asked what each item meant. Considering how the students interpreted the items, revisions were made to the statements on the scale.

Validity Study

Construct validity

Exploratory factor analysis was performed to examine the structure of the Turkish adaptation of CLS in Turkish culture. Exploratory factor analysis aims to find new conceptually meaningful variables (factors) based on the relationships between variables by bringing together p number of related variables (Büyüköztürk, 2002; Field, 2005). In this way, it is possible to explain how the structure in question is based on the results obtained from the measurement tool created to measure an unknown structure (Erkuş, 2010).

As stated by Deniz (2007), exploratory factor analysis is a technique related to validity, and it is used for the adaptation of a scale to reveal its factors in the adapted culture. In this regard, this study aims to determine under which factors the items that make up the Turkish form of the CLS were grouped. In addition, the factor loadings of the CLS items were analyzed through factor analysis to determine the structure of the scale specific to Turkish culture.

Principle Component Analysis, which is frequently used in social sciences, was used as a factorization technique in the exploratory factor analysis. In addition, varimax orthogonal rotation was used to ensure that the correlation between factors was zeroed and thus the significance was enhanced for the interpretation of the factors; the lower limit of item eigenvalues was 1.00 for determining the number of factors (Tabachnick & Fidell, 2001; Field, 2005; Büyüköztürk, 2002).

Initially, the sample size was analyzed for the Exploratory Factor Analysis. The CLS was administered to 400 students enrolled in the Faculty of Education at Pamukkale University (PU) (n: 214) and Trabzon University (TRU) (n: 186). A total of 376 data forms TRU (n:170) and PU (n:206) were used for exploratory factor analysis after the collected data were organized and outliers and missing values were removed. In the first round of exploratory analysis, the eight-item scale had a one-factor structure, yet the break in the Scree Plot table was found to be in the second factor. The exploratory factor analysis was conducted again with two sub-factors using the varimax process, based on the theoretical basis of cognitive load and the two sub-factor structure of the original form of the scale. At the beginning of the first round of factor analysis of CLS with two sub-factors, the Kaiser-Meyer-Olkin (KMO) coefficient, and Bartlett's test of sphericity results were performed to determine whether the data were suitable for factor analysis; these values were found to be statistically significant (KMO =.84; Bartlett's test of sphericity $\chi 2$=1540,190 df =28 p<.001).

After the administration, Kaiser Meyer-Oklin's (KMO) sampling adequacy measurement was performed to test the validity of the sample size statistically. The KMO value, which can take values between 0 and 1, is considered normal between 0.5 and 0.7, good between 0.7 and 0.8, very good between 0.8 and 0.9, and excellent above 0.9 (Field, 2005). Besides, a significant level of Bartlett's test of sphericity indicates that the sample size is good for factor analysis and the correlation matrix is suitable (Tabachnick & Fidell, 2001; Field, 2005; Büyüköztürk, 2002). However, an analysis of the items showed that item 5 was included in two sub-factors and the value between the two sub-factors was less than .20, so item 5 was removed from the scale and the analysis was performed again. In the second round of the two-subfactor analysis at the beginning of the factor analysis performed on the CLS, the Kaiser-Meyer-Olkin (KMO) coefficient and Bartlett's test of sphericity results were performed to determine whether the data were suitable for factor analysis; these values were found to be statistically significant (KMO =.81; Bartlett's test of sphericity $\chi 2$=1346,343 df =21 p<.001). However, an analysis of the items showed that item 8 was included in two sub-factors and the value between the two sub-factors was .20, so item 8 was removed from the scale and the analysis was performed again. In the third round of the two sub-factor analysis at the beginning of the factor analysis performed on the CLS, Kaiser-Meyer-Olkin (KMO) coefficient and Bartlett's test of sphericity results were examined to determine whether the data were suitable for factor analysis; the values were found to be statistically significant (KMO =.79; Bartlett's test of sphericity $\chi 2$=1086,653 df = 15 p<.001). However, an analysis of the items showed that item 1 had a value of .63 in the mental effort factor, not in the mental load factor. Item 1 was included in the mental load factor in the original scale, so the related item was removed from the scale and the analysis was performed again. In the fourth round of the two sub-factor analysis at the beginning of the factor analysis performed on the CLS, the Kaiser-Meyer-Olkin (KMO) coefficient and Bartlett's test of sphericity results were examined to determine whether the data were suitable for factor analysis; the values were found to be statistically significant (KMO =.73; Bartlett's test of sphericity $\chi 2$=908,627, df = 10 p<.001). However, an analysis of the items showed that item 2 received a value of .70 in the mental effort factor, not in the mental load factor. Item 2 was included in the mental load factor in the original scale, so the related item was removed from the scale, and analysis was performed again. In the fifth round of the two-subfactor analysis at the beginning of the factor analysis performed on CLS, Kaiser-Meyer-Olkin (KMO) coefficient and Bartlett's test of sphericity results were analyzed to determine whether the data were suitable for factor analysis; the values were found to be statistically significant (KMO =.642; Bartlett's test of sphericity $\chi 2$=675.02 df =6 p<.001). The total Cronbach's alpha reliability value of the two-factor and four-item forms was found to be .80 as a result of the administration of the CLS to 376 students. Cronbach’s alpha reliability values were .89 in the first sub-factor (2 items) and .78 in the second sub-factor (2 items). The first results of the factor analysis showed that the scale had two components with eigenvalues above 1.00. An analysis of the scree plot of the eigenvalues showed that the most significant break was in the second factor. Eigenvalue, the percentage contribution to the total variance, and the scree plot are reported to be the most commonly used criteria for the determination of the total number of factors (DeVellis, 2003). The determination of the appropriate number of factors, according to Cattel (as cited in DeVellis, 2003), is based on the criterion indicating the number of factors up to the point where the line graph takes a horizontal shape. In addition to these values, the original CLS includes two sub-factors. The CLS was developed with two main factors F1: Mental load and F2: Mental effort, so the factor analysis was limited to two components and performed again.

Table 1 demonstrates factors, factor loadings, factor eigenvalues, variance percentages explained by the factors, and Cronbach's alpha values; corrected item-total score correlations (r), common variances, and t-values of the items obtained from the factor analysis and reliability analyses, which resulted in a two-factor structure reached in five iterations.

Table 1. Factors, factor loadings, variance percentages explained by factors and item-total score correlation values in cognitive load scale (r)

Item No	F1	F2	$\bar{X}$	T	SD	r
I 3	,92		2,47	-17,883	1,409	,65*
I 4	,91		2,34	-21,075	1,513	,66*
I 6		,90	2,84	-19,768	1,374	,55*
I 7		,86	2,89	-22,071	1,504	,60*
Range	.91-.92	.86-.90	2,34 - 2,89	-17,883 and - 22,071	1,374- 1,504	,55*- ,66*
						Total
Variance %	44,976	41,267				86,243
Cronbach's alpha	.89	.78				.80

r: Item-total score correlations *sign indicates significance at .05 level. Note: for readability, factor r loadings less than .30 are not shown in the table. F1: Mental load and F2: Mental effort

Besides exploratory factor analysis (EFA), the factor structure of the CLS was also analyzed by confirmatory factor analysis (CFA). Exploratory factor analysis aims to discover the factor structure based on the relationships between variables. Confirmatory factor analysis, which examines model-data fit, tests the hypotheses about the relationship between variables (Tabachnick and Fidell, 2001). "Mental load" was the first component obtained from the analyses, and it included two items as I3 and I4. The items in this scale include "*Doing the activities in this course was troublesome for me*" and "*Doing the activities in this course was very frustrating for me*". The factor loadings of the two items in this sub-scale are between .91 and .92; the item-total score correlations are between .65 and .66, and the Cronbach's alpha internal consistency coefficient is .89. Exploratory factor analysis results showed that the second component of the CLS consisted of two items (Item no: 6 and 7). The factor loadings of the items in this factor called "mental effort" were between .86 and .90; the item-total score correlations were between .55 and .60; and Cronbach's alpha internal consistency coefficient was .78. The items in this scale included "*The way the instruction or content was presented during the course caused me to exert a lot of mental effort" and "I had to exert a lot of effort in completing the activity tasks or achieving the learning outcomes in this course".*

The two sub-scales explain 86.24% of the total variance. Cronbach's alpha internal consistency coefficient is .80 for the whole scale. Guttman Split Half values calculated as a test split technique to get an idea about the stability or consistency between the two halves of the scale are .88 for the "first" sub-scale, .78 for the "second" sub-scale, and .63 for the whole scale.

An analysis of Table 1 in terms of factor loading shows that factor loadings vary between .86 and .92. An analysis of the items loading on more than one factor shows that the items are generally loaded on the relevant sub-scales with significant differences (generally .30 and above). The arithmetic means of the four items range between 2.34 and 2.89 and standard deviations between 1.374 and 1.504.

In addition, the total scores obtained from the four items by the participants were ranked from the lowest to the highest for each item separately, and the lowest 27% and highest 27% groups were formed. Analyses aimed to examine whether the items distinguished these two groups from each other. This analysis showed that all the items were able to distinguish the groups significantly (p<.001).

Besides the reliability level of the two sub-scales, whether the scale was addible was also checked. The results are summarized in Table 2.

Table 2. Cronbach's Alpha and Tukey's test of additivity results for the sub-factors of the cognitive load scale

Sub-scale	Cronbach's Alpha	Source of variance	Sum of squares	Mean of squares	F	Sd	p
Mental load	.89		3,251	3,251	7,533	1	,006
		(Nonadditivity)	2,268	2,268	5,317	1	,022
Mental effort	. 78		,357	,357	,482	1	,488
		(Nonadditivity)	3,843	3,843	5,242	1	,023

An analysis of Table 2 shows that the reliability values were 0.89 for the first sub-factor and 0.78 for the second sub-factor. Reliability coefficient values above 0.70 are accepted as high reliability for the scales (Tabachnick & Fidell, 2001). Accordingly, both sub-scales have a high level of reliability, and in terms of scoring, the nonadditivity feature was P=0.22 for the mental load. With a value of P=0.23 for the mental effort sub-scale, it was concluded that the scale was a Likert-type addible scale (Tukey Nonadditivity p>.05). An analysis of the variance analysis table for the mental load factor of the Cognitive Load Scale showed that the difference between measures was statistically significant with a value of P=.006 (P<.01), and the nonadditivity feature was not statistically significant with a value of P=.022 (P>.005). With a value of P=0.022 for the mental load factor, the scale was found to be a Likert-type addible scale (Tukey Nonadditivity p>.005). In other words, the two-item mental load sub-factor is addible, but the measurements include differences. An analysis of the variance table for the mental effort factor of the Cognitive Load Scale showed that the difference between measures was not statistically significant with a value of P=.488 (P<.01) and with a value of P=.023 (P>.005), the non-additivity feature was not statistically significant. For the mental effort factor, with a value of P=0.023, the scale was concluded to be a Likert-type addible scale (Tukey Nonadditivity p>.005). In other words, the two-item mental effort sub-factor is addible, but there are no differences between the measures. In addition, Table 3 presents the correlation results for the Cognitive Load Scale total and sub-factors (mental load and mental effort).

Table 3. Correlation analysis results of scale factors

	Cognitive load scale total	Cognitive load scale total	Mental effort
Cognitive load scale total	1	,864**	,845**
Cognitive load scale total	,864**	1	,460**
Mental effort	,845**	,460**	1

P**<.01

An analysis of Table 3 shows that the Cognitive Load Scale includes two sub-scales as the mental load sub-factor and the mental effort sub-factor

To test the total sub-factors of the Cognitive Load Scale adapted by Demir and Ayvaz-Tuncel, each scale was applied to 376 individuals, and the relationship between them was analyzed by Pearson Product-Moment Correlation Analysis. As seen in Table 3, a significant and positive correlation was found between the total score of the Cognitive Load Scale adapted by Demir and Ayvaz-Tuncel and mental load (r=.864, p<.01) and mental effort (r=.845, p<.01). There was also a significant and positive relationship between mental load and mental effort (r=.460, p<.01).

Confirmatory Factor Analysis

Several fit indices in CFA are used to assess the validity of the model. The most commonly used ones are the Chi-Square Fit Test, Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI), Root Mean Square Error (RMR or RMS), and Root Mean Square Error of Approximation (RMSEA). A ($\chi 2$/sd) ratio of less than 5 calculated by CFA is considered to be an indicator of a good fit of the model with the actual data (Sümer, 2000). GFI and AGFI values of higher than .90 and RMS or standardized RMS and RMSEA values of less than .05 are expected for model-data fit. On the other hand, GFI values higher than 0.85, AGFI values higher than 0.80 and RMS values lower than 0.10 are also accepted as criteria for model fit with actual data (Anderson and Gerbing, 1984; Marsh, Balla and McDonald, 1988).

Confirmatory factor analysis was performed to test the fit of the two-factor structure. Although model-data fit includes several statistics, χ2, χ2/df, RMSEA, NNFI, CFI, and GFI values are generally considered indicators (Sümer, 2000; Çokluk, Büyüköztürk, & Şekercioğlu, 2012; Hoe, 2008). The chi-square value calculated for model-data fit with the confirmatory factor analysis conducted to examine the extent to which the two-factor model of the CLS was compatible with the collected data was found to be significant (χ2=2,156, sd=1, p>.01).

An analysis of the literature indicates no consensus on which fit indices should be used for the evaluation of the model, yet more than one fit index is recommended to be used together. The recommended indices include the Chi-Square fit test (χ^2), chi-square and degrees of freedom ratio (χ^2 /sd), Root Mean Square Errors of Approximation (RMSEA), Square Root Mean Square Errors Standardized (S-RMS), Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI), Normed Fit Index (NFI), Non-Normed Fit Index (NNFI) and Comparative Fit Index (CFI). In this study, the fit indices and acceptable limit values used to determine the fit of the theoretical model to the data include a nonsignificant p-value, χ^2 / sd ratio of less than 3, indicating "good fit", and values up to 5 indicates "adequate fit". While S-RMS and RMSEA should be less than 0.10, GFI, CFI, AGFI, NFI, NNFI, and CFI should be greater than 0.90 (Byrne 1994: 147-149, Schermelleh-Engel, Moosbrugger and Müller 2003: 52, Sümer 2000: 49-74). It is recommended that the fit indices should be greater than 0.90 and the error indices should be less than 0.05. In other words, GFI, AGFI, and CFI values should be above .90 and SRMR and RMSA values should be below .05 (Schumacker & Lomax, 2004).

Table 4 demonstrates the fit indices, normal values and acceptable values used in CFA.

Table 4. The goodness of fit indices, normal values, and acceptable values used in CFA

Fit indices	Normal values	Acceptable value
χ2 "p" Value	p>0.05	-
χ2 /sd	<2	<5
GFI	>0.95	>0.90
AGFI	>0.95	>0.90
CFI	>0.95	>0.90
RMSEA	<0.05	<0.08
RMR	<0.05	<0.08
SRMR	<0.05	<0.08

(Schreiber, Nora, Stage, Barlow and King, 2006; Hooper, Coughlan and Mullen 2008).

Table 5 demonstrates the comparison of standard goodness-of-fit indices and research results.

Table 5. Comparison of standard goodness of fit indices and research results

Fit indices	Good fit	Acceptable fit	Fit Values obtained in the study
c^2	$0 \le c^2 \le 2df$	$2df \le c^2 \le 3df$	2.156
P value	0.05≤p≤1	0.01≤p≤0.05	.142
c^2/df	$0 \le c^2/df \le 2$	$2 \le c^2/df \le 3$	2.156
RMSEA	0≤RMSEA≤0.05	0.05≤RMSEA≤0.08	0.056
RMR			0.014
NFI	0.95≤NFI≤1.00	0.90≤NFI≤0.95	0.99
CFI	0.97≤CFI≤1.00	0.95≤CFI≤0.97	0.99
GFI	0.95≤GFI ≤1.00	0.90≤GFI≤0.95	0.99
AGFI	0.90≤AGFI≤1.00	0.85≤AGFI≤0.90	0.97
RFI	0.90<RFI<1.00	0.85< RFI <0.90	0.98

(Schermelleh-Engel-Moosbrugger, 2003)

A two-factor structure was analyzed for the Cognitive Load Scale. The structure examined includes two items in the cognitive load factor and two items in the mental effort factor. The model fit criteria were examined using confirmatory factor analysis and CMIN=2.156, DF=1, CMIN/DF=2.156, p>0.001, RMSEA=.056 and χ^2/df=2.156, RMR= .014, GFI= 0.997, AGFI= 0.971, NFI= 0.99, CFI=0.99, IFI= 0.99 were obtained. Fit indices of confirmatory factor analysis showed that the Turkish version of the CLS consisted of four items and two factors, and this model showed an excellent fit in terms of theory and statistics. An analysis of the β_0 values in Table 4 showed that I4 explained factor 1 the most, while with .744, I6 explained factor 2 the least. The standardized coefficients showing the relationship between the items and their factors ranged between .74 and .90, which were all significant at the .01 level. A general analysis of model fit indices showed that with RMSEA=0.56 and $\chi 2$/df=2.156 (Tabachnick & Fidell, 2001, Dorman & Knightley, 2006), the model showed a good fit.

Table 6 shows the standard and non-standard β coefficient values for the Cognitive Load Scale.

Table 6. Standard and nonstandard path (β0: standard path coefficients β1: nonstandard path coefficient) coefficient values for the cognitive load scale

Items	Path	Factor	β_0	β_1	S.E.	C.R.	P
m3	<---	Mental load (F1)	,887	1			<0,001
m4	<---	Mental load (F1)	,902	1,092	0,081	13,501	<0,001
m6	<---	Mental effort (F2)	,744	1			<0,001
m7	<---	Mental effort (F2)	,867	1,275	0,136	9,389	<0,001

(β_0: Standard path coefficients β_1: Nonstandard path coefficients F1: Mental load, F 2: Mental effort)

An analysis of the two-factor model tested with CFA showed that fit indices values obtained were RMSEA=.056 and $\chi 2$/df=2.2 and RMR= .014, GFI= 0.997, AGFI= 0.971, NFI= 0. 99, CFI=0.99, IFI= 0.99, which were found within the specified limits, indicating excellent model fit. An analysis of Table 6 shows that the path coefficients of all items under Mental load (F1) and Mental effort (F 2) were found statistically significant in the confirmatory factor analysis (p<0.001). The path coefficient value obtained for item 3 was β_1= 1, β_1= 1,092 for I4, β_1= ,744 for I6, and β_1= ,867 for I7. An analysis of path coefficients showed that I4 is the item with the highest effect on F1 (β0=,90). An analysis of the standardized path coefficients shows that I7 is the item with the highest effect on F 2 (β_0=,86).

Figure 1 contains standardized path coefficients

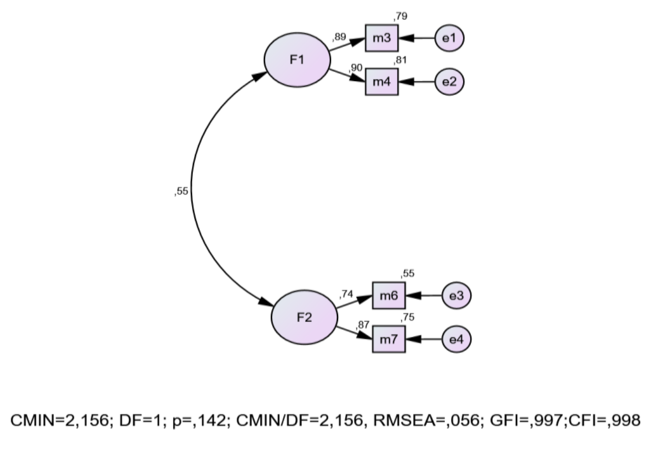

Figure 1. Standardized path coefficients

Figure 2 contains non-standardized path coefficients

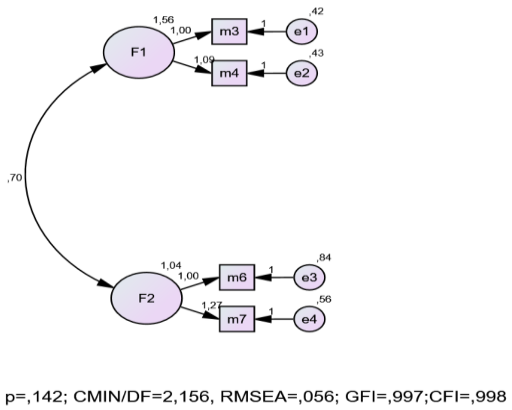

Figure 2. non-standardized path coefficients

Criterion validity

At this phase, the study included a sample group (n=259) enrolled in departments of two different Education Faculties in the second semester of the 2022-2023 academic year. In addition to the Cognitive Load Scale adapted by Demir and Ayvaz-Tuncel (2023), the Cognitive Load scale developed by Dönmez, Akbulut, Telli, Kaptan, Özdemir, and Erdem (2022) was administered as the data collection tool. The Cognitive Load scale is a 13-item and three-factor scale developed by Dönmez et al. (2022) to address intrinsic, extraneous, and effective cognitive load in computer-based learning environments. It was developed after confirmation with 193 undergraduate students immediately after they attended online webinars and 99 undergraduate students after they attended face-to-face classes. The 13-item scale, the validity and reliability of which was performed on 292 participants, was found to have a three-factor structure; the factor structures were found to be valid; and the internal consistency coefficient was found 0.88, which explained 57.51% of the total variance. Cronbach's alpha internal consistency coefficients were .87 for intrinsic cognitive load, .81 for extraneous cognitive load, and .82 for effective cognitive load. In addition, confirmatory factor analysis results showed that the fit index values ($\chi2$= 119.18, RMSEA=0.069, $\chi2$/df= 1.92, SRMR= 0.0783, GFI= 0.914, AGFI= 0.971, NFI= 0.926, NNFI= 0.953, CFI= 0.962) fit the recommended criteria. Criterion validity is cross-checked by looking at the correlation of the scores obtained by the participants from the scale to be developed with both another test measuring the same behavior and a test measuring another related behavior (Büyüköztürk, 2008: 169). For this purpose, the Cognitive Load Scale (Dönmez, Akbulut, Telli, Kaptan, Özdemir and Erdem, 2022) was used to test criterion validity.

To test the Cognitive Load Scale adapted by Demir and Ayvaz-Tuncel (2023), both scales were applied to 259 individuals, and the relationship between them was analyzed by Pearson Product-Moment Correlation Analysis. As seen in Table 7, a significant and positive correlation was found between the total Cognitive Load Scale adapted by Demir and Ayvaz-Tuncel (2023), and the intrinsic cognitive load (r=.631, p<.01), extraneous cognitive load (r=.713, p<.01) and effective cognitive load (r=.546, p<.01) of the Cognitive Load Scale developed by Dönmez et al. (2022). Similarly, a significant and positive relationship was found between the mental load factor of the Cognitive Load Scale adapted by Demir and Ayvaz-Tuncel and the extraneous cognitive load factor of the Cognitive Load Scale developed by Dönmez et al. (2022) (r=.336, p<.01), and between the mental effort factor and the extraneous cognitive load factor (r=.217, p<.01).

Table 7. Correlation results between the cognitive load scale adapted by Demir and Ayvaz-Tuncel (2023) and the Cognitive load scale developed by Dönmez et al. (2022)

	Cognitive load scale adapted by Demir and Ayvaz-Tuncel (2023) total	Mental load	Mental effort	Cognitive load scale developed by Dönmez et al. (2022) total	Intrinsic cognitive load	Extraneous cognitive load	Effective cognitive load
Cognitive load scale total	1	,054	,116	,103	,631**	,713**	,546**
Mental load	,054	1	,376**	,823**	-,142*	,336**	-,296**
Mental effort	,116	,376**	1	,836**	-,003	,217**	-,088
Cognitive load scale developed by Dönmez et al. (2022) total	,103	,823**	,836**	1	-,086	,332**	-,229**
Intrinsic cognitive load	,631**	-,142*	-,003	-,086	1	,084	,409**
Extraneous cognitive load	,713**	,336**	,217**	,332**	084	1	-,082
Effective cognitive load	,546**	-,296**	-,088	-,229**	,409**	-,082	1

P**<.01, P*<.005

Test-retest reliability

The test-retest method was used to statistically test the stability of the Turkish form of the CLS in terms of the qualities it measures over time. Test-retest reliability aims to test the stability of a measurement tool in a certain time interval (Çokluk et al., 2012; Erkuş, 2010). The test-retest reliability coefficient of the scale was determined by administering it to 342 students who were enrolled at the Faculty of Education in two universities in Turkey within a six-month interval. Pearson Product-Moment correlation coefficient was performed to test the stability between the students' scores from both administrations. Accordingly, [r=.304, p<.001] was found between the two administrations of the CLS. In addition, confirmatory factor analysis was performed on the Cognitive Load Scale again, and the values related to the measurement model were found (CMIN: .067, df:1, CMIN/df: .067, NFI: 1,000, GFI: 1,000, RFI: .999, CIFI:1,000, RMSEA:. 000).

DISCUSSION, INTERPRETATION AND CONCLUSION

Knowledge of pre-service teachers' cognitive load within the scope of the activities carried out in the teaching principles and methods course is believed to contribute to the design of the teacher education process to be provided to them. For this reason, this study followed a way to adapt a reliable and valid scale to contribute to the measurement of pre-service teachers' cognitive load.

Many fit indices such as the Chi-Square fit test ($\chi 2$), Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI), Comparative Fit Index (CFI), Normed Fit Index (NFI), Non-Normed Fit Index (NNFI) and Root Mean Square Error of Approximation (RMSEA) are used in the process of the determination of cognitive load tendencies (Şimşek, 2007; Hoe, 2008).

The Cognitive Load Scale was developed by Hwang, Yang, and Wang (2013). Considering the original scale, a two-subfactor structure was conducted in the scale adaptation study. In this respect, as a result of the exploratory factor analysis, Kaiser-Meyer-Olkin (KMO) coefficient and Bartlett's test of sphericity results were analyzed and these values were found to be statistically significant (KMO =.642; Bartlett's test of sphericity $\chi 2$=675,02 df =6 p<.001).

According to Klein (1998) and Wiersma (2000), reliability is the consistency of a measurement tool in any measurement. Calculation of Cronbach's alpha is the most commonly used reliability method in studies (Dorman & Knightley, 2006; Smolleck, Zembal-Saul & Yoder, 2006; Johnson, Stevens & Zvoch, 2007). Cronbach's alpha internal consistency coefficient was found .80 for the total scale. In addition, Cronbach's alpha reliability coefficients were calculated as .89 in the first sub-factor and .78 in the second factor, thus a reliable measurement tool with four items and two factors was obtained.

Confirmatory factor analysis was employed to test the fit of the two-factor structure. Although many statistics are available for model-data fit, $\chi 2$, $\chi 2/df$, RMSEA, NNFI, CFI, and GFI values are generally considered indicators (Sümer, 2000; Çokluk, Büyüköztürk, & Şekercioğlu, 2012; Hoe, 2008). The chi-square value calculated for model-data fit with the confirmatory factor analysis conducted to examine the extent to which the two-factor model of the CLS was compatible with the collected data was found ($\chi 2$=2,156, sd=1, p>.01). When the two-factor model tested as a result of the CFA was analyzed, it was found that the fit index values obtained with RMSEA=.056 and $\chi 2/df$=2.156 and RMR= .014, GFI= 0.997, AGFI= 0.971, NFI= 0. 99, CFI=0.99, IFI= 0.99 complied with the recommended criteria.

The standardized coefficients showing the relationship between the items and their factors ranged between .42 and .84, and all of them were significant at the .01 level. In general, the model showed a good fit with RMSEA=0.56 and $\chi 2/df$=2.156 (Tabachnick & Fidell, 2001; Dorman & Knightley, 2006). The results of the confirmatory factor analysis conducted to test the original factor structure of the CLS, which was also supported by expert opinions, confirmed that the four items in the scale showed a valid structure on the faculty of education students. These values indicate the adequacy of the data fit of the model (Kelloway, 1998; Heubeck & Neill, 2000; Corral & Calvete, 2000; Ingles, Hidalgo & Mendez, 2005; Şimşek, 2007; Hoe, 2008).

In line with the characteristics of the items in the factors, the first factor was determined as "Mental load" and the second factor was determined as "Mental effort". This scale, which is valid and reliable with the results obtained, is also a reliable measurement tool adapted by confirmatory factor analysis in our country to determine education faculty students’ cognitive load tendencies.

Considering the results of the validity and reliability studies conducted with education faculty students in this study, the scale has the features to measure the cognitive load tendencies of education faculty students with a two-factor structure:

- It has a valid and reliable structure,

-The results obtained from the actual administration of the scale can provide the necessary feedback on students' self-perception of their cognitive load tendencies,

-The adapted measurement tool can be examined at a meta-analytic level in future administrations and studies including different samples,

- "CLS" is considered to be used in studies involving different types of research to determine the self-perceptions of education faculty students about their cognitive load tendencies.

Conflicts of Interest: No potential conflict of interest was declared by the authors

Funding Details: There has been no financial support for this work. No financial support was received for the research, authorship, and/or publication of this article.

CRediT Author Statement: Prof. Dr. Özden Demir: Writing, original draft, visualization, data curation, review & editing, conceptualization, supervision. Associate Professor Zeynep Ayvaz Tuncel: Writing, original draft, visualization, data curation, review & editing, conceptualization, supervision.

Ethical Statement: Ethics Committee Permission: Trabzon University, Social and Human Sciences Scientific Research and Publication Ethics Histor: 16.05.2022 and Number: E-81614018-000-2200018936

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

REFERENCES

Anderson, J. C. & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness of fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*(2), 155-173.

Anmarkrud, Ø., Andresen, A. & Bråten, I. (2019). Cognitive load and working memory in multimedia learning: Conceptual and measurement issues. *Educational Psychologist*, *54*(2), 61-83. https://doi.org/10.1080/00461520.2018.1554484

Buyukozturk, S. (2002). Factor analysis: Basic concepts and its use in scale development. *Educational Administration in Theory and Practice, 32*(32), 470-483.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Sage Publications, California.

Chen, O., Paas, F. & Sweller, J. (2021). Spacing and interleaving effects require distinct theoretical bases: A systematic review testing the cognitive load and discriminative-contrast hypotheses. *Educational Psychology Review, 33,* 1499-1522. https://doi.org/10.1007/s10648-021-09613-w .

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. https://doi.org/10.1016/j.chb.2008.12.020

Cohen, J., Cohen, P., West, S., G. & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis in the behavioral sciences* (Third Edition). Mahwah, NJ.

Corral, S. & Calvete, E. (2000). Machiavellianism: dimensionality of the Mach IV and its relation to self-monitoring in a Spanish sample. *The Spanish Journal of Psychology, 3*(1), 3-13. https://doi.org/10.1017/S1138741600005497

Cotigă, M. I. (2012). Development and validation of a Romanian version of the expanded version of Positive and Negative Affect Schedule (PANAS-X). *Procedia - Social and Behavioral Sciences, 33,* 248 – 252. https://doi.org/10.1016/j.sbspro.2012.01.121

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87-114. https://doi.org/10.1017/S0140525X01003922

Çokluk, Ö., Şekercioğlu, G. & Büyüköztürk, Ş. (2012). *Multivariate SPSS and LISREL applications for social sciences.* Pegem Academy Publishing, Ankara.

Deniz. Z. (2007). Psychometric measurement tool adaptation. *Journal of Ankara University Faculty of Educational Sciences, 40*(1), 1-16. https://doi.org/10.1501/Egifak_0000000180

DeVellis, R. F. (2003). *Scale development theory and applications*. SAGE Publications International Educational and Professional Publisher, Chapter 5 pp: 60-96.

Dorman, J. P. & Knightley, W. M. (2006). Development and validation of an instrument to assess secondary school students' perceptions of academic tasks. *Educational Studies, 32*(1), 47-58. https://doi.org/10.1080/03055690500415951

Dönmez, O., Akbulut, Y., Telli, E., Kaptan, M., Özdemir, İ. H. & Erdem, M. (2022). In search of a measure to address different sources of cognitive load in computer-based learning environments. *Education and Information Technologies*, *27*(7), 10013-10034. https://doi.org/10.1007/s10639-022-11035-2.

Erkuş, A. (2010). The conflict of operations with the meanings of the Turkish equivalents of psychometric terms. *Journal of Measurement and Evaluation in Education and Psychology 1*(2), 72-77.

Eysink, T. H. S., De Jong T., Berthold, K., Kollöffel, B., Opfermann, M. & Wouters, P. (2009). Learner performance in multimedia learning arrangements: An analysis across instructional approaches. *American Educational Research Journal, 46,* 1107–1149. https://doi.org/10.3102/0002831209340235

Feldon, D. F., Callan, G., Juth, S. & Jeong, S. (2019). Cognitive load as motivational cost. *Educational Psychology Review*, *31*, 319-337. https://doi.org/10.1007/s10648-019-09464-6

Field, A. (2005). *Discovering statistic using SPSS for Windows*. SAGE Publications, London.

Heubeck, B. G. & Neill, J. T. (2000). Confirmatory factor analysis and reliability of the mental health inventory for Australian adolescents. *Psychological Reports, 87*(2), 431-440. https://doi.org/10.2466/pr0.2000.87.2.431

Hilbert, T. S. & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Computers in Human Behavior, 25,* 267–274. https://doi.org/10.1016/j.chb.2008.12.006

Hoe, S. L. (2008). Issues and procedures in adopting structural equation modeling technique. *Journal of Applied Quantitative Methods, 3(*1), 76-83.

Hooper, D., Coughlan, J. & Mullen, M. R. (2008). Structural equation modelling: guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53-60.

Hwang, G. J., Yang, L. H. & Wang, S. Y. (2013). A concept map-embedded educational computer game for improving students' learning performance in natural science courses. *Computers & Education*, *69,* 121-130. https://doi.org/10.1016/j.compedu.2013.07.008

Inglés, C. J., Hidalgo, M. D.& Méndez, F. X. (2005). Interpersonal difficulties in adolescence: a new self-report measure. *European Journal of Psychological Assessment*, *21*(1), 11-22. https://doi.org/10.1027/1015-5759.21.1.11

Johnson, B., Stevens, J. J.& Zvoch, K. (2007). Teachers' perceptions of school climate: A validity study of scores from the revised school level environment questionnaire. *Educational and Psychological Measurement, 67*(5), 833-844. https://doi.org/10.1177/0013164406299102

Kelloway, E. K. (1998). *Using LISREL for structural equation modeling: A researcher's guide*. SAGE Publications, London.

Leppink, J., Paas, F., Van der Vleuten, C. P, Van Gog, T. &Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*, 1058-1072. https://doi.org/10.3758/s13428-013-0334-1

Marsh, H. W., Balla, J. R. & Mc Donald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391-410. https://doi.org/10.1037/0033-2909.103.3.391

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), 429-434. https://doi.org/10.1037/0022-0663.84.4.429

Paas F., Van Merriënboer J. & Adam, J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79,* 419–430. https://doi.org/10.2466/pms.1994.79.1.419

Paas, F. & Van Merriënboer, J. J. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science, 29*(4), 394-398. https://doi.org/10.1177/09637214209221

Paas, F. G. W. C., Renkll, A. & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*(1-2), 1-8. https://www.jstor.org/stable/41953634

Salden, R. J., Paas, F., Broers, N. J. & Van Merrienboer, J. J. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in air traffic control training. *Instructional Science, 32,* 153-172. https://doi.org/10.1023/B:TRUC.0000021814.03996.ff

Schermelleh-Engel, K. & Helfried, M. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness of fit measures. *Methods of Psychological Research Online, 8* (2), 23-74.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A.& King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: a review. *The Journal of Educational Research, 99*(6), 323-38. https://doi.org/10.3200/JOER.99.6.323-338

Schumacker, R. E. & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. Psychology Press.

Smolleck, L. D., Zembal-Saul, C. & Yoder, E. P. (2006). The development and validation of an instrument to measure preservice teachers' self-efficacy in regard to the teaching of science as inquiry. *Journal of Science Teacher Education, 17*(2), 137-163. https://doi.org/10.1007/s10972-006-9015-6

Sümer, N. (2000). Structural equation models: basic concepts and examples. *Turkish Psychology Writings, 3*(6), 49-74.

Sweller, J. & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology, 4*(1), 434–458. https://doi.org/10.1177/147470490600400135

Sweller, J., Ayres., P. & Kalyuga, S. (2011). *Cognitive load theory*. Springer, New York.

Sweller, J., Van Merriënboer, J. J. G. & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–297. https://doi.org/10.1023/A:1022193728205

Sweller, J., Van Merriënboer, J. J. & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Şimşek, Ö. F. (2007). *Introduction to structural equation modeling, basic principles and LISREL applications*. Equinox, Ankara

Tabachnick, B. G. & Fideli, L. S. (2001). *Using multivariate statistics (Fourth Edition).* Ally and Bacon, Boston.

Van Merrienboer, J. J. G. & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development*, *53*(3), 5-13. https://link.springer.com/content/pdf/10.1007/BF02504793.pdf

Yang, C., Jen, C. H., Chang, C. Y. & Yeh, T. K. (2018). Comparison of animation and static-picture based instruction: Effects on performance and cognitive load for learning genetics. *Journal of Educational Technology & Society 21*(4), 1-11. https://doi.org/10.30191/ETS.201810_21(4).0001 https://www.jstor.org/stable/26511533