Journal of Computer and Education Research
March 2025 Volume 13 Issue 25
http://dergipark.org.tr/jcer
(ISSN:2148-2896)

Research Article

Reliability and Validity of Turkish Version of the Multidimensional Cognitive Load Scale for Virtual Environments

Talha YILDIZ [1,*] Güldem Alev ÖZKÖK [2]

[1] Hacettepe University, Ankara, Türkiye, talhayildiz@hacettepe.edu.tr,
[2] Hacettepe University, Ankara, Türkiye, ozkok@hacettepe.edu.tr
[*] Corresponding Author: talhayildiz@hacettepe.edu.tr,

Article Info

Received: 16 December 2024
Accepted: 04 March 2025

Keywords: Cognitive load, multidimensional, virtual learning environments

10.18009/jcer.1601749

Publication Language: English

Abstract

This study aimed to adapt the "Multidimensional Cognitive Load Scale for Virtual Environments Scale (MCLSVE)" into Turkish, while evaluating its validity and reliability. A survey model was used for the adaptation process, utilizing the scale developed by Andersen and Makransky (2021), which is now referred to as the "MCLSVE-TR" The scale comprises five subscales: Intrinsic Load, Extraneous Load Instruction, Extraneous Load Interactions, Extraneous Load Environment, and Germane Load. The sample group for the study was 203 volunteer university students selected using a convenience sampling technique. Exploratory Factor Analysis (EFA) was conducted to determine the factor structure of the scale. As a result of EFA, it was revealed that the scale consisted of 18 items and 5 sub-dimensions, and these dimensions explained 82,34% of the total variance. In addition, Confirmatory Factor Analysis (CFA) confirmed the five-factor structure. Pearson correlation analysis to determine the relationship between scale factors, and Cronbach Alpha coefficient to determine the reliability level of scale factors were used. The findings confirmed that the Turkish adaptation of the MCLSVE is both valid and reliable.

To cite this article: Yıldız, T. & Özkök, G.A. (2025). Reliability and validity of Turkish version of the multidimensional cognitive load scale for virtual environments. *Journal of Computer and Education Research, 13* (25), 431-453. https://doi.org/10.18009/jcer.1601749

Introduction

One of the basic principles of Cognitive Load Theory (CLT), which has had a significant impact on the field of educational psychology, is that extraneous cognitive load should be reduced in order to free up sufficient cognitive resources for the actual learning to take place (Skulmowski & Xu, 2022). The amount of cognitive load (CL) is seen as an important individual characteristic for learning processes as it expresses the amount of information that the individual will acquire for a single time and for a short time (Barut Tuğtekin, 2020). Since its introduction in the 1980s, this theory has leveraged insights into human cognitive architecture to generate both experimental and instructional outcomes (Sweller, 2011).

There are different perspectives on how human cognitive structure acquires knowledge. The most common explanations involve the elements of the cognitive system, known as short-term memory (STM) or working memory (WM) and long-term memory (LTM). Secondary knowledge differs in terms of the CL they impose on WM.

The foundational premise of CLT asserts that WM has limitations in both capacity and duration. Consequently, individuals are only able to process a finite amount of information in WM and can retain it for a brief period (Baddeley, 1990). In contrast, LTM possesses a much larger capacity, as it organizes information into structures known as schemas (Chi et al., 1982). In essence, the human mind processes new information within WM, integrating it with existing knowledge, which is subsequently encoded into LTM.

Recently, CL is generally defined as Intrinsic Load (IL), Extrinsic Load (EL) and Germane Load (GL) (Sweller et al., 2011). IL is directly related to the internal complexity of the knowledge rather than the way in which the knowledge is acquired. Since IL is attributed to the complexity originating from the nature of the knowledge, it can change when the structure of the knowledge changes or when the learners' degree of expertise regarding that knowledge increases. For example, in the case of learning the syntax of a language, the learner has to analyze how each word in a sentence is related to the others. This results in a high IL on working memory because of the high degree of item interaction. On the contrary, if long lists of words have to be learned, a large number of items have to be assimilated, because the items do not have to be held in working memory at the same time. In this case, item interaction will be low. Element interaction cannot be determined only by analyzing the tasks or the learning material, as there may be many interactive elements for one learner and only one element for another, more expert learner (Schnotz & Kürschner, 2007).

CLT has defined EL in different ways that are not exactly equivalent. According to one definition, EL results from an unnecessarily high degree of element interaction in working memory, depending on the instructional format. For example, if a diagram is presented with integrated explanatory text, it is very difficult for the learner to ignore the text, even if they do not need it to understand it. The learner is forced to assimilate multiple elements of information at the same time, placing a heavy EL on working memory (Schnotz & Kürschner, 2007). In another definition, the EL is caused by unrelated cognitive activities.

Activities are seen as irrelevant if they are not directed towards schema acquisition and schema automation (Sweller 2005). The two definitions are not equivalent because cognitive activities that are irrelevant for learning do not necessarily involve high element interaction. As such, the second definition is broader than the first. Regardless of the specific way of defining EL, the theory assumes that EL interferes with learning and should therefore be reduced as much as possible by eliminating irrelevant cognitive activities (Leung et al., 1997). In this context, interactional, instructional and learning environment-based elements that may cause EL to be tried to be reduced.

Finally, GL is related to the mental resources allocated for the purpose of learning the knowledge (Ayres, 2018). GL is defined as the cognitive effort that the student must show to make sense of the content to be learned (Plass et al., 2010). GL is the CL resulting from cognitive activities in WM that are aimed at deliberate learning and go beyond simple task performance. If these activities were not located in WM, they would not cause CL. If they were not aimed at learning, they would not be germane. If they do not go beyond task performance, they are only part of the IL (Schnotz & Kürschner, 2007). The GL also depends on general learning orientations, affective and motivational aspects of learning. For example, students who adopt a deep learning approach will adopt a higher semantic load than students who adopt a surface learning approach (Marton & Saljö, 1984). Therefore, it is not enough to provide learning environments that enable learners to have the cognitive resources for GL (Schnotz & Kürschner, 2007).

As mentioned above, CLT has identified a number of factors that facilitate and hinder learning by reducing EL and managing GL over time. A similar approach is put forward by the Cognitive Theory of Multimedia Learning (CTML; Mayer, 2009). CTML is a theory that attempts to explain how people learn academic materials from words and graphics, referring to the human cognitive system (Mayer, 2021). According to CTML, there are information processing channels in the human cognitive system that process verbal and visual information. This structure progresses with human cognitive capacity, which is known to be limited, and active processing, that is, by organizing the material to be learned in consistent verbal and visual themes. As a result, it continues by integrating it with information previously stored in LTM (Mayer, 2024). Mayer and Moreno (2003) stated that they encountered the problem of CL in their studies on learning; meaningful learning requires the student to perform important cognitive operations, but the student's cognitive capacity is quite limited. In this context, instructional designers have stated that multimedia learning environments that can manage CL are needed (Clark, 1999; Sweller, 1999; van Merriënboer, 1997). Virtual learning environments, whether two-dimensional or three-dimensional, are environments where multimedia learning takes place. In other words, they are learning environments where words and images are presented together. However, recent research suggests that various design factors involved in virtual learning can lead to an increase in CL (Skulmowski & Xu, 2022). For example, feeling fully immersed in a virtual world during a learning task can create a whole new learning experience compared to learning with traditional media, but this immersion can also lead to students' cognitive resources being depleted in the experience itself rather than contributing to learning (Frederiksen et al., 2020).

The use of CLT and CTML in education has faced a number of challenges. First of all, CL is affected by individual differences (Sweller, 1988), so flexible teaching methods should be developed according to the needs of learners. It is stated that teaching complex mathematical and scientific concepts with multimedia tools may lead to misunderstandings (Mayer & Moreno, 2003). This situation necessitates good design of the content. Since the effective use of multimedia teaching materials is related to the effective use of technological tools, regional differences can lead to inequality of opportunity in education (Mayer, 2001). It is stated that instructors' lack of knowledge about CLT and CTML principles may make it difficult to apply these theories effectively and may have negative effects on the learning process (Sweller, 1994). Finally, current assessment approaches may not adequately reflect the complexity of cognitive processes and therefore may not accurately measure learning outcomes (Mayer & Moreno, 2003). In summary, CTML and CLT highlight three key instructional design processes: managing basic processing associated with the inherent characteristics of the educational content minimizing extraneous processing that does not contribute to understanding the core content and supporting students' engagement during learning by promoting effective processing (Mayer & Moreno, 2003).

Considering the impact of CLT on instructional design and learning understanding, measuring CL has become an important research topic (Andersen & Makransky, 2021). Efforts have been made to measure CL in this process. Because measuring CL is important for understanding how learners process information and how this process affects learning outcomes. In this context, different methods and tools have been used to measure CL.

CL was first tried to be measured with single-item subjective measurement tools (Cierniak et al., 2009; Klepsch et al., 2017). Although single-item measurement tools were quick and easy to use, the measurements made were generally limited when it came to understanding and distinguishing the type of CL (Andersen & Makransky, 2021).

Another method is performance-based measures. For example, if the learner has the necessary knowledge about a subject but performs poorly in performing a task required by this knowledge, this may indicate the presence of high CL. However, from this perspective, subtle analyses are required to distinguish between CL and other factors affecting performance (Sweller et al., 2011). The so-called dual-task methodology involves the learner performing one task while simultaneously engaging in a second task that requires cognitive resources. Here, while the level of performance on the first task provides information about CL, more CL on the first task leads to lower performance on the second task (Baddeley, 1990). Finally, learners are asked to verbalize their thought processes as they engage with the learning materials. By analyzing the information obtained verbally, information about the strategies followed by the learners in the learning process can be obtained (Ericsson & Simon, 1993).

The first psychometrically validated measurement tool to measure CL was developed by Leppink, et al (2013). The scale consists of 10 items measuring three dimensions of CL. The scale called the Cognitive Load Scale (CLS) includes three items measuring the IL related to the complexity of the subject, concept, and definitions arising from their own structure in the context of the field of statistics. In addition, the scale includes three items to measure the EL created by learning activities. Finally, the scale includes four items to determine the GL, i.e. how much the learning activities improve the student's understanding in the context of the learning subject. In the subsequent process, Andersen and Makransky (2021) first verified the validity of the CLS in their studies. Then, they addressed the EL that may arise from the teaching interventions used in the learning process, the learning environment used, and the interaction provided, as separate dimensions. As a result, they revealed that the MCLSVE is a valid and reliable scale.

It is clear that subjective ratings, physiological and performance-based methods generally aim to measure the total CL experienced by students; such a distinction cannot be made on the basis of physiological and performance-based methods (Schnotz & Kürschner, 2007). Regarding the use of subjective ratings, it is conceivable to develop questionnaire items that would allow a distinction to be made between different types of CL. However, when the literature is examined, there is a lack of a valid and reliable scale to measure CL in virtual learning environments in the context of Turkish education. In this context, this study aims to introduce the MCLSVE, which will allow differentiation between different types of CL and has psychometric validity, into Turkish culture. Within the scope of the study, the following research questions were determined and answers to these questions were sought:

1. Does MCLSVE adapted into Turkish have linguistic equivalence in Turkish?
2. Is the validity of MCLSVE adapted into Turkish sufficient to measure CL multidimensionally?
3. Is the reliability of MCLSVE adapted to Turkish adequate to measure CL multidimensionally?

Present Study

This study is scale adaptation research aimed at evaluating the CL perceived by university students in three-dimensional virtual environments in a multidimensional manner. The MCLSVE, which has been adapted into Turkish with validity and reliability studies conducted in this research, allows for the measurement of IL, interactional, environment-based, and instructional EL, as well as GL perceived by university students in three-dimensional virtual environments. This is achieved through five subscales of the MCLSVE, consisting of 18 items in total.

The first subscale, IL, pertains to the intricacy of the instructional material as perceived by the student and consists of 3 items. The second subscale, Extraneous Load from Instructions (EL ins), relates to the CL produced by processing unnecessary information due to instructional design decisions, and it also contains 3 items. The third subscale is Extraneous Load from Interactions (EL int), which deals with the processing of unnecessary information and includes 4 items. The fourth subscale addresses the CL arising from extraneous information (EL env), resulting from the learning environment itself, comprising 4 items. The fifth subscale focuses on the germane resources allocated to learning-related information, GL consisting of 4 items (Andersen & Makransky, 2021).

Using the MCLSVE, it is possible to measure the IL arising from the complexity of the subject being learned, and to manage and control the EL resulting from instructional decisions, the interactions provided, and the virtual learning environment itself. This enables effective management of GL. With this aim, the objective of this study was to adapt the

MCLSVE into Turkish in a scientifically accurate manner and to conduct validity and reliability assessments. According to Hambleton and Patsula (1999), when an adapted test is developed for cross-cultural or international assessments, creating an equivalent test in a second language is considered the most effective approach. Therefore, this study will provide valuable guidance for national scale development.

Method

The studies carried out during the adaptation process of the MCLSVE are explained under this title.

Participants

Due to accessibility concerns in selecting the study group, a convenience sampling method was employed. Consequently, a total of 203 associate and undergraduate students from various departments at Afyon Kocatepe University, who volunteered to participate, were included in the research sample. To ensure adequate factor recovery, Gorsuch (1974) categorized sample sizes over 200 as large and those under 50 as small. Similarly, Guilford (1954) also recommended a minimum of 200 participants. Therefore, the final sample for this study consisted of 203 university students. Since the virtual laboratory application is only used within the scope of General Chemistry I course, the number of students taking the related course is limited. In this context, the sample number, which was thought to represent the research population, was also limited in this sense. In other words, the main reason for determining the number of participants as 203 can be explained in this way.

Although their ages ranged between 19 and 21, most of the participants (>82%) were 18 years old. A total of 119 students (58.6%) reported using the Virtual Laboratory Application between 2-4 hours in total, while 82 students (41.4%) reported using the Virtual Laboratory Application for less than 2 hours.

Table 1 presents descriptive statistics that outline the personal characteristics of the students in the study group. As indicated in Table 1, the sample includes 68 males (33.5%) and 135 females (66.5%). All participants are second-year university students.

Table 1. Distribution of participants by gender

Gender	n	%
Male	68	33,50
Female	135	66,50
Total	203	100,00

Many of the participants were studying in the Department of Science Teaching (n= 81), followed by the Department of Chemistry (n= 29), Department of Molecular Biology and Genetics (n= 33), Department of Food Technology (42) and Department of Chemical Technology (18) (see Table 2). All of the participants stated that they knew and used the Virtual Laboratory Application within the scope of General Chemistry I course.

Table 2. Distribution of participants according to the departments they study

Department	n	%
Science Teaching	81	39,90
Chemistry	29	14,28
Molecular Biology and Genetics	33	16,26
Food Technologies	42	20,69
Chemical Technology	18	8,87
Total	203	100,00

Instruments

In this research, a personal information form and the MCLSVE developed by Andersen and Makransky (2021) were employed. Detailed information about the data collection tools is provided below.

Demographic Data Survey

In the study, a "Demographic Data Survey" created by the researchers was employed to collect demographic information about the participants. This survey included questions regarding the participants' age, gender, the department in which they were enrolled, and their experience with the virtual laboratory application.

Multidimensional Cognitive Load Scale for Virtual Environments (MCLSVE)

The MCLSVE is a scale created by Andersen and Makransky (2021), designed for use with university students. The scale allows for a multidimensional assessment of the CL perceived by university students in virtual learning environments. The instrument is composed of five sub-dimensions. These are (1) IL Subscale, (2) EL ins Subscale, (3) EL int Subscale, (4) EL env Subscale and (5) GL Subscale. The development study of the scale, whose original form consists of 18 Likert-type items, was carried out with the participation of 140 first-year biology students in Denmark. The analyses indicated that the reliability coefficients for the subscales of the original form of the scale were satisfactory, ranging from 0.81 to 0.85. Additionally, the scale demonstrated a five-factor structure that accounted for 4.1% of the variance in retention. The scale score is obtained by summing the scores of each item. There are no reverse scored items in the scale. A higher score on the scale is considered as higher CL. The scale takes an average of 4-5 minutes to complete and no cut-off score was set in the original development article.

Procedure and Data Analysis

In this study, the adaptation process was conducted in accordance with the principles outlined by Hambleton and Patsula (1999) that should be followed during the adaptation of measurement instruments. Accordingly, firstly, forward and backward translation technique was applied and language equivalence was ensured. In the forward translation stage, the scale was translated into Turkish by two different foreign language experts. Then, in the back translation stage, which was the second stage, the Turkish versions of the scale were crossed between two experts and translated back into English. After the forward and back translation stages, the original form of the scale and the translated forms of the scale based on Turkish culture and language structure were reviewed by two Turkish language experts who are proficient in foreign languages and the most appropriate expressions were selected and the scale was finalized. The concepts of virtual environments and CL discussed in this study are universal. It can be argued that the only situation that differs between cultures is practices that vary according to cultural structures or socio-economic status of countries. Although different applications are used or produced, at this point, it cannot be ignored that individuals are experiencing learning processes in virtual environments, especially after the Covid-19 pandemic. Therefore, it can be interpreted that these features are equivalent in different cultures. Then, field experts were consulted for content evaluation. Then, a pilot study was conducted with a small group. After all arrangements were made, the application was carried out in the sample group.

Within the scope of the validity and reliability studies of the Turkish version of the scale, the participants answered the scale after experiencing a virtual laboratory application. This application is a three-dimensional virtual learning environment used in chemistry and physics laboratory courses at the university level. The Virtual Laboratory Application is a three-dimensional web application developed in collaboration with the Scientific and Technological Research Council of Turkey (TÜBİTAK) and the Higher Education Council (YÖK). It facilitates the conducting of experiments in the general chemistry and general physics laboratory courses offered in various programs within the faculties of science and engineering as well as vocational schools of higher education. (CHE-YÖK, 2020) (see Figure 1).

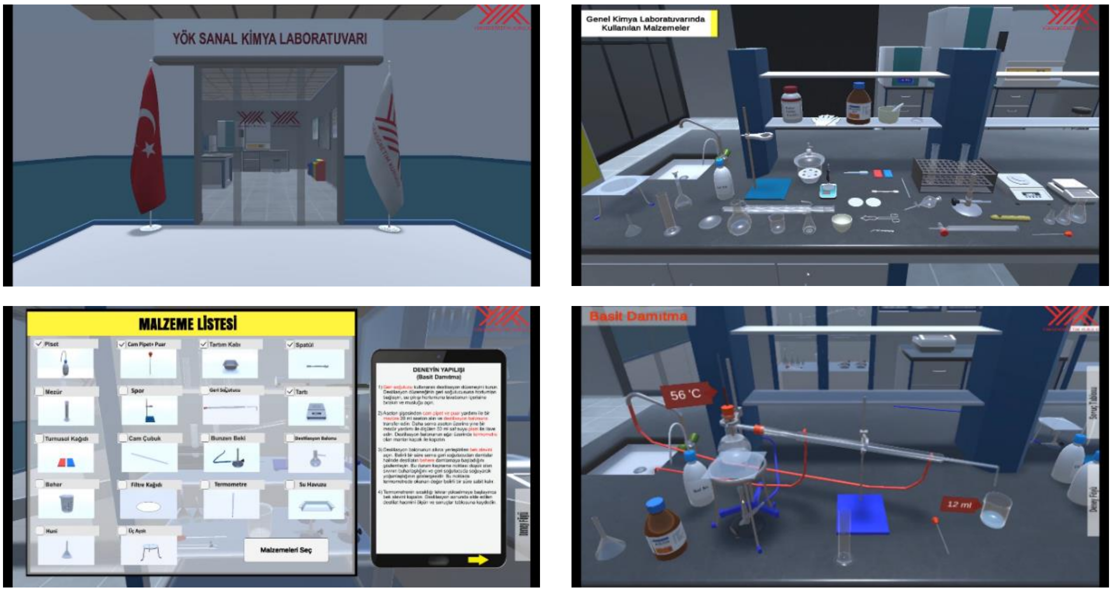

Figure 1. 3D virtual laboratory application screenshots

The participants of the study used the virtual laboratory application for 7 weeks within the scope of General Chemistry I course. During this period, the course instructors asked the students to experience 9 different experimental simulations on their own computers or tablets and report their results. Thus, the students performed the simulated experiments at their own learning pace and at times convenient for them. At the end of this process, the Turkish version of MCLSVE was administered to the participating students.

The correlation matrix was examined to determine the suitability of the data for factor analysis. To statistically examine the relationships between the variables in the data matrix, Bartlett's test of sphericity was utilized (Bartlett, 1950). In addition, the Kaiser-Meyer-Olkin (KMO) measure, derived from the correlation and partial correlation coefficients, was also applied to assess data appropriateness for factor analysis. The study employed the principal components method to extract the factors. To determine the optimal number of factors, eigenvalues greater than one were used as the selection criterion. Additionally, factor rotation was conducted to enhance clarity regarding which variables contributed to the formation of common factors, using the Varimax method for this process. Descriptive statistics were presented in terms of number (n), percentage (%), mean (X), standard deviation (SD), median (M), minimum (min), and maximum (max) values. The normality of the data associated with numerical variables was assessed using skewness and kurtosis, while relationships between numerical variables were analyzed using the Pearson correlation coefficient. A significance level of $p<0.05$ was deemed statistically significant. All these analyses were carried out using SPSS 23 and Lisrel 8.72 package programs.

Content Validity

To ensure the content validity of the MCLSVE, feedback was gathered from three doctoral-level academics specializing in computer and instructional technologies, as well as three teachers with master's degrees in the same field. All experts affirmed that the items included in the scale were necessary and appropriate. As a result, all items in their original form were utilized during the data collection process.

Small Group Practice

This is the last stage recommended before examining the psychometric properties of the scale translated into Turkish and is also called cognitive interviewing/information gathering. At this stage, the scale is usually administered to a sample (n=25-40 people) selected from the target population and the participants' opinions and feedback are obtained regarding the comprehensibility and acceptability of the scale items (Karaçam, 2019). In addition, this stage helps researchers to ensure that competent language and culturally neutral expressions are used in translation (Beaton et al., 2000).

Before the actual implementation, the MCLSVE was first applied to 25 students through face-to-face interviews. Opinions were obtained from the students regarding the comprehensibility of the items in the scale. Students were asked to indicate the items that they did not understand and did not find appropriate while filling out the scale. In this context, two items were corrected due to an expression error. After the pilot application, the actual application was started.

Ethical Principles

All processes related to this research were carried out in accordance with established scientific ethical principles. Initially, the developers of the scale were contacted, and the necessary permissions were secured. Informed consent forms were provided to participants in the study, and their involvement was entirely voluntary.

Finding

The mean scores and descriptive statistics of the items in the MCLSVE are presented in Table 3 and Table 4, respectively. Table 3 shows the mean scores of the 18 items in the scale. When the mean score values of the items are analyzed, it is seen that the 18th item has the highest mean and the 11th item has the lowest mean.

Table 3. Mean scores of the items in the MCLSVE

Items	$\overline{X}$	Ss	M
Item#1	4,89	± 1,93	5
Item#2	4,73	± 2,07	5
Item#3	5,36	± 2,23	5
Item#4	4,85	± 2,17	5
Item#5	4,69	± 2,26	5
Item#6	4,92	± 2,37	5
Item#7	4,90	± 2,20	5
Item#8	5,21	± 2,38	5
Item#9	5,25	± 2,31	5
Item#10	5,37	± 2,32	5
Item#11	4,46	± 2,07	5
Item#12	5,03	± 2,39	5
Item#13	4,91	± 2,30	5
Item#14	5,25	± 2,28	5
Item#15	6,18	± 1,99	6
Item-16	6,37	± 1,79	7
Item#17	6,42	± 1,88	7
Item#18	6,48	± 1,96	7

When Table 4 is examined, it is seen that the mean of the MCLSVE is 5.25 ± 1.43 points, the mean of IL Subscale is 5.00 ± 1.88 points, the mean of EL ins Subscale is 4.82 ± 2.03 points, the mean of E L int Subscale is 5.18 ± 2.03 points, the mean of EL env Subscale is 4.91 ± 2.00 points and the mean of GL Subscale is 6.36 ± 1.73 points. In addition, according to Table 4, when the relationship between the overall and sub-dimensions of the measurement tool is examined; IL ($r=0.848$, $p<.01$), EL ins ($r=0.899$, $p<. 01$), EL int ($r=0.904$, $p<.01$), EL env ($r=0.850$, $p<.01$) and GL ($r=0.155$, $p<.01$) subscales. In addition, the correlation coefficients of each factor with the other vary between .023 and .873 ($p<.01$). According to the results, correlation values ranging from weak to high correlation between the subscales were obtained ($p<0.05$).

Table 4. Descriptive statistics of the MCLSVE (N=203)

	$\overline{X}$	Ss	M	IL	EL ins	EL int	EL env	GL
IL	5,00	± 1,88	4,99	1				
EL ins	4,82	± 2,03	4,82	0.753**	1			
EL int	5,18	± 2,03	5,18	0.681**	0.821**	1		
EL env	4,91	± 2,00	4,91	0.607**	0.762**	0.873**	1	
GL	6,36	± 1,73	6,36	0.023	-0.131	-0.160*	-0.233**	1
Total	5,25	± 1,43	5,25	0.848**	0.899**	0.904**	0.850**	0.155*

(*p < .05. **p < .01)

Findings for Exploratory Factor Analysis (EFA)

As part of the study, EFA was performed to assess the construct validity of the MCLSVE (see Table 5). According to Table 5, KMO Barlett test was conducted to determine the adequacy of the distribution for factor analysis. The Barlett test result was obtained as 3251.960 (p<.05). This situation shows that the distribution is sufficient for factor analysis. It can be said that the amount of variance obtained as 82% in the study is sufficient. According to the EFA results presented in Table 5, item factor loadings vary between 0.554 and 0.970. This result shows that the factor loadings of all items are quite sufficient (>0.400). In this case, it was seen that there were no overlapping items in the scale. The factor loadings of the Intrinsic Load Subscale vary between 0.640 and 0.970, the factor loadings of the Extraneous Load instructions Subscale vary between 0.604 and 0.858, the factor loadings of the Extraneous Load interactions Subscale vary between 0.675 and 0.918, the factor loadings of the Extraneous Load environment Subscale vary between 0.554 and 0.656, and the factor loadings of the Germane Load Subscale vary between 0.869 and 0.936. It can be said that the 5 dimensions of the MCLSVE measure the sub-features. As a result of the exploratory factor analysis, it is seen that the MCLSVE is a valid measurement tool.

Table 5. EFA results of the MCLSVE

Factor	Item No	Factor Loading 1	2	3	4	5	Total Correlation	Explained Variance %	Cronbach Alpha
IL	1	0,970					0,876	31,17	0,899
	2	0,640					0,830		
	3	0,799					0,849		
EL ins	4		0,764				0,857	18,83	0,916
	5		0,604				0,756		
	6		0,858				0,843		
EL int	7			0,733			0,821	16,34	0,919
	8			0,918			0,834		
	9			0,847			0,814		
	10			0,675			0,790		
EL env	11				0,656		0,821	8,38	0,892
	12				0,627		0,778		
	13				0,577		0,792		
	14				0,554		0,799		
GL	15					0,869	0,804	7,56	0,879
	16					0,875	0,819		
	17					0,936	0,869		
	18					0,923	0,858		
Scale								82,34	0,86
KMO=0.936 DF=153 $\chi2$=3251,960210 p<0.001									

Findings for Confirmatory Factor Analysis (CFA)

CFA was performed with data obtained from each subscale for the structural validity of the MCLSVE. The cut-off values in CFA analysis (Hu & Bentler, 1999; Kline 2015) were evaluated according to Table 6.

Table 6. Boundary values in CFA analysis

Indexes	Limit Values
	Excellent ≤ 3≤ Good ≤ 5
RMSEA	Excellent ≤ 0.05 ≤ Good ≤ 0.08
SRMR	Excellent ≤ 0.05 ≤ Good ≤ 0.08
CFI	Excellent ≥ 0.95 ≥ Good ≥ 0.90
IFI	Excellent ≥ 0.95 ≥ Good ≥ 0.90
GFI	Excellent ≥ 0.95 ≥ Good ≥ 0.90
TLI	Excellent ≥ 0.95 ≥ Good ≥ 0.90

In Table 7, χ^2/df, RMSEA, SRMR, IFI, TLI, CFI and GFI were used to evaluate the factor validity of the models within the scope of CFA. The model obtained for the MCLSVE (χ^2/df=1.794) consists of five dimensions. The fit indices for this model show that the model is acceptable. CFA was applied to the MCLSVE, which consists of 18 items and five dimensions. The model is visually presented in Figure 2. Each of the path coefficients of the dimensions on the 18 questions was found to be statistically significant ($p<0.05$). Accordingly, IL Subscale consists of items 1-3, EL ins Subscale consists of items 4-6, EL int Subscale consists of items 7-10, EL env Subscale consists of items 11-14 and GL Subscale consists of items 15-18.

CFA was applied to determine whether the five-factor structure obtained as a result of the exploratory factor analysis of the MCLSVE would be confirmed in the Turkish sample. The goodness of fit values calculated for the validity of the MCLSVE are given in Table 7. In the literature, it is accepted as a criterion for the model's fit with real data that NFI, IFI, GFI, CFI and TLI values are around 0.90 and above and RMSEA and SRMR values are below 0.08 (see Table 7).

Table 7. Fit values of the model of the MCLSVE

Scale	(χ2/sd)	RMSEA	SRMR	IFI	CFI	GFI	TLI
Model	1.794	0.063	0.485	0.963	0.924	0.862	0,912

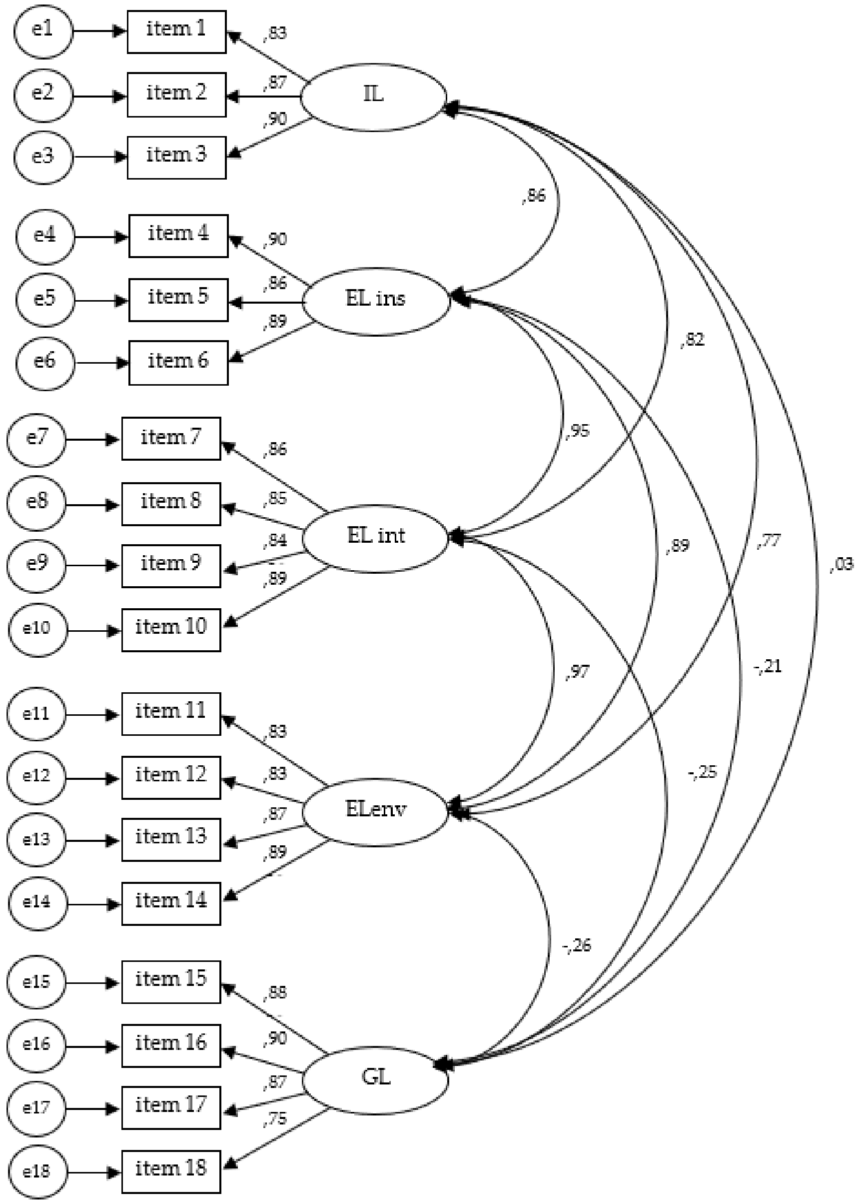

Figure 2. CFA model of MCLSVE

When the measurement model presented in Figure 2 is examined, the theoretical structure of the five-factor scale consisting of 18 items proposed as a result of EFA was confirmed as a result of CFA. In other words, as a result of CFA, it was evaluated that the scale met the criteria for construct validity. The factor loadings, variance ratios and t values of the items in the scale are reported in Table 8.

Table 8. MCLSVE CFA result *t* and R^2 values for the items

Items	*t*	R^2
Item#1	14.22	0.689
Item#2	14.63	0.757
Item#3	15.42	0.810
Item#4	14.49	0.810
Item#5	14.09	0.740
Item#6	14.97	0.792

Item#7	15.45	0.740
Item#8	13.95	0.723
Item#9	15.20	0.706
Item#10	15.32	0.792
Item#11	15.45	0.689
Item#12	14.18	0.689
Item#13	14.30	0.757
Item#14	14.93	0.792
Item#15	13.97	0.774
Item#16	14.61	0.810
Item#17	16.15	0.757
Item#18	16.83	0.563

According to the item parameter values reported in Table 8, all t values of the items were significant at the 0.01 level (T>2.58). Item multiple correlation coefficient squares (R^2) ranged between 0.563 and 0.810. When the t and R^2 values of the items related to the measurement model were examined, it was seen that the highest contribution was provided by item#3 in the IL factor, item#4 in the EL ins factor, item#10 in the EL int factor, item#14 in the EL env factor and item#16 in the GL factor. The lowest contribution was provided by item#1 in the IL factor; item#5 in the EL ins factor; item#9 in the EL int factor; item#11, item#12 in the EL env factor; and item#18 in the GL factor.

Findings for Reliability and Item Analyses

Cronbach Alpha coefficients of the the MCLSVE was calculated as 0.798. It was calculated as 0.814 for the IL Subscale, 0.814 for the EL ins Subscale, 0.796 for the EL int, 0.814 for the EL env Subscale and 0.946 for the GL Subscale (see Table 8). In this framework, it is possible to conclude that the reliability of the obtained measurements in terms of stability is extremely high. This indicates that the results are consistent and dependable, affirming the robustness of the instrument used in the analysis.

Table 9. The means, Cronbach's alpha (α) coefficients, standard deviations, and bivariate correlations for the five dimensions of the MCLSVE and one affect scale

Scale	M	α	Sd	IL	EL ins	EL int	EL env	GL
IL	26,53	0,814	1,88	-				
EL ins	26,71	0,798	2,03	0,75	-			
EL int	26,35	0,796	2,03	0,68	0,82	-		
EL env	26,62	0,814	1,99	0,60	0,76	0,873	-	
GL	25,16	0,946	1,72	0,02	- 0,13	- 0,16	- 0,23	-
MCLSVE	26,27	0,798	1,42	0,85	0,90	0,90	0,85	0,1

In this scenario, it can be concluded that the reliability of the measurements obtained in terms of stability is both good and very good. The effects among the items in the MCLSVE and its subscales are detailed in Table 9. Upon analyzing Table 10, it is evident that the path coefficients for all subscales across the 18 items are statistically significant ($p<0.05$). This indicates that all subscales exert a highly significant effect on the items.

Table 10. Evaluation of the effects between the items in the MCLSVE and the subscales

			β	*se*	*zβ*	*t*	*p*
Item#1	<--	Intrinsic Load	1,000		0,830		
Item#2	<--	Intrinsic Load	1,145	0,083	0,870	13,786	<0.001
Item#3	<--	Intrinsic Load	1,265	0,086	0,897	14,700	<0.001
Item#4	<--	Extraneous Load instructions	1,000		0,895		
Item#5	<--	Extraneous Load instructions	0,902	0,064	0,861	14,061	<0.001
Item#6	<--	Extraneous Load instructions	0,894	0,058	0,887	15,503	<0.001
Item#7	<--	Extraneous Load interactions	1,000		0,859		
Item#8	<--	Extraneous Load interactions	0,971	0,064	0,850	15,286	<0.001
Item#9	<--	Extraneous Load interactions	0,949	0,067	0,839	14,206	<0.001
Item#10	<--	Extraneous Load interactions	0,991	0,067	0,888	14,865	<0.001
Item#11	<--	Extraneous Load environment	1,000		0,870		
Item#12	<--	Extraneous Load environment	0,875	0,056	0,891	15,607	<0.001
Item#13	<--	Extraneous Load environment	0,929	0,075	0,833	12,392	<0.001
Item#14	<--	Extraneous Load environment	0,921	0,068	0,833	13,469	<0.001
Item#15	<--	Germane Load	1,000		0,880		
Item#16	<--	Germane Load	0,962	0,061	0,903	15,815	<0.001
Item#17	<--	Germane Load	0,927	0,070	0,869	13,326	<0.001
Item#18	<--	Germane Load	0,854	0,075	0,752	11,377	<0.001

Discussion and Conclusion

In the literature, it is stated that e-learning has five main interrelated dimensions: “assessment and evaluation”, “learning environments”, “teaching models”, “teaching areas” and “teaching tools” (Gürcan & Özyurt, 2020). Especially in the recent period, the Covid-19 pandemic, which has deeply affected our lives in every field, has also affected the field of education and training, and the usage area of e-learning has expanded even more in the world (Serbest et al., 2023). It can be said that this situation makes it even more important to examine the existing relationships between the dimensions determined for e-learning. Virtual learning environments allow the use of multimedia in different ways. However, when the relevant literature is examined, the critical importance of CL especially in multimedia-rich virtual learning environments is emphasized (Akan & Keskin, 2023) and the necessity of measuring it accurately is pointed out (Akhter, 2017; Costley, 2020; Huang et al., 2020; Skulmowski & Xu, 2022). The main purpose of this study is to adapt the MCLSVE into

Turkish and Turkish cultural context and to assess its validity and reliability. It is expected that the adaptation of the MCLSVE will not only enhance the existing literature but also provide valuable support to instructional designers working with virtual learning environments in their efforts to effectively measure CL.

The data collected from participants with a sample size of 203 were analyzed. A correlation matrix was examined to determine the suitability of the data for EFA. A correlation coefficient threshold of 0.30 was used to confirm the appropriateness of including items in their respective factor structures (Hair et al., 1998). Bartlett's test of sphericity was used to test the statistical significance of the correlations between the variables and the results showed a statistical significance of 3251.960 ($p<.05$). In addition, the KMO coefficient, which evaluates the suitability of the data set for factor analysis, showed excellent sampling adequacy by giving a value of 0.94 (Cerny & Kaiser, 1977).

EFA used the principal components method to derive factors. The criterion for factor selection was based on retaining factors with eigenvalues above one. Varimax rotation was applied to clarify the items contributing to each common factor. The higher the variance explained by the scale, the higher the construct validity (Tavşancıl, 2005). When the literature is examined, the variance level between 40% and 60% is accepted (Scherer, 1988). The total variance explained by all factors in the scale is 82.34%. According to this result, it is seen that the variance level explained by the scale is quite high. As a result of the rotation, it was concluded that the scale has a five-factor structure. The factor loadings of the items in the IL factor ranged between .640 and .970, the factor loadings of the items in the EL ins factor ranged between .604 and .858, the factor loadings of the items in the EL int factor ranged between .675 and .918, the factor loadings of the items in the EL env factor ranged between .554 and .656, and the factor loadings of the items in the GL factor ranged between .869 and .936. As a result of the analysis conducted on the final form of MCLSVE-TR, it was seen that there were no overlapping items in the scale and there were no items with factor loadings below .30. Even if EFA results yielded positive results regarding the theoretical basis of the scale, there is a need for CFA. For this reason, CFA was conducted.

In CFA, fit indices such as Chi-square fit test, RMSEA, CFI, TLI and GFI were used to assess the factor validity of the model. Acceptable thresholds were determined: RMSEA $\leq$ 0.08, IFI, TLI, CFI $\geq$ 0.90 and GFI $\geq$ 0.85 (Büyüköztürk, 2013). The resulting model for the MCLSVE, which covers five dimensions, showed a fit index of $\chi^2/df = 1.794$ and met the previously mentioned acceptance criteria for fit indices. This result indicates that the scale is based on a solid theoretical foundation (Brown, 2006). The internal consistency coefficient of the scale was (α = .81) in the IL factor (α = .80), (α = .80) in the EL ins factor, (α = .80) in the EL int factor, (α = .81) in the EL env factor, and (◎ = .94) in the GL factor. Since it is recommended that the α value should be at least 0.60 to 0.70 in the reliability analysis (Anderson, 1988), it was concluded that the psychometric quality of the scale was at a high level. When EFA and CFA results were analyzed together, it was evaluated that the developed scale was a valid, reliable and theoretically sound scale.

When the related literature is examined, it can be seen that there are scale studies to measure CL multidimensionally in different contexts similar to the purpose of the current study. Dönmez et al. (2022), through a multi-stage correlational study, proposed a 13-item three-factor scale to address intrinsic, extrinsic and germane CL in computer-based learning environments. EFA with maximum likelihood explained 58 percent of the variance with factor loadings above .49 and internal consistency coefficients above .81. Convergent and discriminant validity indices were found to be at acceptable levels. Moreover, achievement was found to be positively related to the germane loadings and negatively related to the intrinsic and extrinsic loadings. The factor structure was then validated with two different participant groups in virtual learning and face-to-face learning environments. In another study, Krieglstein et al. (2023) developed a new instrument to measure three types of CL and validated it based on five empirical studies. In Study 1, principal component analysis revealed a three-component model, which was then confirmed using confirmatory factor analysis (Study 2). Finally, in three experiments (Studies 3-5), the scale was shown to be sensitive to changes in CL, supporting its predictive validity. The quality of the CLS was emphasized by satisfactory internal consistencies across all studies. In summary, it was emphasized that the proposed scale can be used in experimental settings to validly and reliably measure different types of CL. Choi and Lee (2022) proposed a scale model to measure CL multidimensionally as EL, IL and GL in the context of online learning. This three-factor model was found to be consistent with previous research on CL in offline learning environments. It was stated that item fit statistics for all items were acceptable. It can be seen that the results of the related scale studies on multidimensional measurement of CL are similar to the findings of this study.

In educational psychology, CLT, which assumes that learning is related to a CL placed on the learner's WM (Sweller, 2020), is seen as one of the most effective frameworks. The increasing interest in CLT in educational sciences has brought along the need to measure the types of CL (Krieglstein et al., 2022). However, little empirical research has been conducted on the validity of a CL measurement tool suitable for virtual learning environments in higher education (Choi & Lee, 2022). In this context, in the light of the findings obtained, it can be said that the scale adapted in this study fills an important gap in the relevant literature and can be used in future studies with its psychometric properties. However, it should be kept in mind that the current study has some limitations in addition to the mentioned contributions. It is emphasized that it is not always correct to prefer scale adaptation over scale development and that it may be both easier and more appropriate to develop a new scale (Hambleton & Patsula, 1999). In this context, it should be taken into consideration that adaptation studies may yield different results across cultures (Serbest et al., 2023). In addition, a limitation of this study is that the sample was limited to second-year university students who experienced a specific virtual learning environment in a specific university. In future studies, it would be useful to test the factor structure of the scale in a wider age group from different universities. Furthermore, participants were examined for the presence of perceived multidimensional CL in the virtual learning environment. It would be useful to use the results of this study in studies where 2D or 3D virtual learning environments are experienced in different contexts. Moreover, in future research, the measurement model can be tested with large sample groups within the framework of item response theory. Researchers who will use the MCSVEL-TR can test the test-retest reliability of the scale in the light of the current findings and also compare the total scores obtained by students in large sample groups according to various variables.

Ethical Committee Permission Information

Name of the board that carries out ethical assessment: Hacettepe University Scientific Research Ethics Committee for Social and Human Sciences

The date and number of the ethical assessment decision: 22.10.2024- 00003838706

Author Contribution Statement

Talha YILDIZ: *Conceptualization, methodology, data collection, processing, analysis, interpretation, supervision, review-writing and editing.*

Güldem Alev ÖZKÖK: *Conceptualization, methodology, data analysis, interpretation, supervision, review-writing and editing.*

References

Akan, S., & Keskin, S. (2023). Etkileşimli öğretimsel videoların başarı, bilişsel yük ve video kapılma üzerine etkisi [The effects of interactive instructional videos on achievement, cognitive load and video engagement]. *Erzincan University Journal of Education Faculty, 25(*2), 198-208.

Akhter, F. (2017). Virtual learning environment: how well designed multimedia lowers the learners'cognitive load. *Journal of International Business Research, 16*(1), 1-6.

Andersen, M. S., & Makransky, G. (2021). The validation and further development of a multidimensional cognitive load scale for virtual environments. *Journal of Computer Assisted Learning, 37*(1), 183-196.

Anderson, L. W. (1988). Attitudes and their measurement. In J. P. Keeves (Ed.), *Educational research, methodology and measurement: An international handbook.* Pergamon.

Ayres, P. (2018). Subjective measures of cognitive load—What can they reliably measure? In *Cognitive load measurement and application—A theoretical framework for meaningful research and practice* (1st ed.). Routledge.

Baddeley, A. (1999). *Human memory.* Allyn & Bacon.

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

Barut Tuğtekin, E. (2020). Çoklu ortamla öğrenmede farklı etiketleme yaklaşımlarının başarı, bilişsel yük ve motivasyona etkisi [Effect of different tagging approaches on achievement, cognitive load and motivation in multimedia learning]. Anadolu Üniversitesi: Yayımlanmamış doktora tezi.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186-3191.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* Guilford.

Büyüköztürk, Ş. (2013*). Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum (18. bs.) [Data analysis handbook for social sciences: Statistics, research design, SPSS applications and interpretation (18th ed.)].* Pegem Akademi.

Cerny, B. A., & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate behavioral research, 12*(1), 43-47.

Chi, M.,Glaser, R.,& Rees, E. (1982). Expertise in problemsolving. InR. Sternberg (Ed.), *Advances in thepsychologyofhumanintelligence* (pp. 7–75). Erlbaum.

Choi, Y., & Lee, H. (2022). Psychometric properties for multidimensional cognitive load scale in an E-learning environment. *International Journal of Environmental Research and Public Health, 19*(10), 5822.

Cierniak, G., Gerjets, P., & Scheiter, K. (2009). Expertise reversal in multimedia learning: Subjective load ratings and viewing behavior as cognitive process indicators. *Proceedings of the Annual Meeting of the Cognitive Science Society, 31*(31), 1906–1911.

Clark, R. C. (1999). *Developing technical training* (2nd ed.). International Society for Performance Improvement.

Costley, J. (2020). Using cognitive strategies overcomes cognitive load in online learning environments. *Interactive Technology and Smart Education, 17*(2), 215-228.

Council of Higher Education [CHE-YÖK] (2020). Yök sanal laboratuvar projesi"nin tanıtımı yapıldı: 2024. Retrieved from https://www.yok.gov.tr/Sayfalar/Haberler/2020/yok-sanal-laboratuvar-projesi-tanitildi.aspx

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.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data (Rev. ed.).* Bradford Books/ MIT Press.

Frederiksen, J. G., Sørensen, S. M. D., Konge, L., Svendsen, M. B. S., Nobel-Jørgensen, M., Bjerrum, F., & Andersen, S. A. W. (2020). Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. *Surgical Endoscopy, 34*(3), 1244-1252.

Gorsuch, R. L. (1974). *Factor analysis.* Saunders

Guilford, J. P. (1954). *Psychometric methods (2nd ed.).* McGraw-Hill

Gürcan, F., & Özyurt, Ö. (2020). E-öğrenme araştırmalarındaki temel eğilimler ve bilgi alanları: 2008-2018 yılları arasında yayımlanan makalelerle konu modelleme analizi [Emerging trends and knowledge domains in e-learning researches: topic modeling analysis with the articles published between 2008-2018]. *Journal of Computer and Education Research, 8*(16), 738-756.

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., (1998). *Multivariate Data Analysis (5th ed.).* Prentice Hall

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology, 1,* 1-13.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Huang, C. L., Luo, Y. F., Yang, S. C., Lu, C. M., & Chen, A. S. (2020). Influence of students' learning style, sense of presence, and cognitive load on learning outcomes in an immersive virtual reality learning environment. *Journal of Educational Computing Research, 58*(3), 596-615.

Karaçam, Z. (2019). Ölçme araçlarının Türkçeye uyarlanması [Adaptation of scales to Turkish]. *Ebelik ve Sağlık Bilimleri Dergisi, 2*(1), 28-37.

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology, 8,* 1–8.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4. Edition). Guilford Publications, Inc.

Krieglstein, F., Beege, M., Rey, G. D., Sanchez-Stockhammer, C., & Schneider, S. (2023). Development and validation of a theory-based questionnaire to measure different types of cognitive load. *Educational Psychology Review, 35*(1), 9.

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*(4), 1058–1072.

Leung, M., Low, R., & Sweller, J. (1997). Learning from equations or words. *Instructional Science, 25*, 37–70.

Marton, F., & Saljö, R. (1984). Approaches to learning. In F. Marton, D. Hounsell, & D. Entwistle (Eds.), *The experience of learning* (pp. 39–58). Scottish Academic Press.

Mayer, R. E. (2001).*Multimedia learning.* Cambridge University Press.
Mayer, R. E. (2009). *Multimedia learning (2nd ed.).* Cambridge University Press.
Mayer, R. E. (2021). *Multimedia learning (3rd ed.).* Cambridge University Press.
Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review, 36*(1), 8.
Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38,* 43–52.
Plass, J. L., Moreno, R., & Brünken, R. (2010). *Cognitive load theory.* Cambridge University Press.
Scherer, R. F. (1988). Dimensionality of coping: Factor stability using the ways of coping questionnaire. *Psychological Report, 62,* 76-770.
Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19,* 469-508.
Serbest, Y., Aydın, M. K., & Kuş, M. (2023). Üniversite dijital eğitim ortamını değerlendirme ölçeği (ÜDEODÖ): Uyarlama, geçerlik ve güvenirlik çalışması [Assessing university digital educational environment (AUDEE scale): Adaptation, validity and reliability study]. *Journal of Computer and Education Research*, *11*(21), 356-375.
Skulmowski, A., & Xu, K. M. (2022). Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational Psychology Review, 34*(1), 171-196.
Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science, 12*(2), 257-285.
Sweller, J. (1994). *Instructional design in technical areas.* Australian Society for Educational Technology.
Sweller, J. (1999). *Instructional design in technical areas.* ACER Press.
Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19–30). University Press.
Sweller, J. (2011). *Cognitive load theory.* In psychology of learning and motivation (Vol. 55, pp. 37-76). Academic Press.
Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68,* 1–16.
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitiveload theory.* Springer.
Tavşancıl, E. (2005). *Tutumların ölçülmesi ve SPSS ile veri analizi [Measurement of attitudes and data analysis with SPSS].* Nobel Yayın.
van Merriënboer, J. J. G. (1997). *Training complex cognitive skills*. Educational Tech. Press.