KURAM VE UYGULAMADA EĞİTİM BİLİMLERİ EDUCATIONAL SCIENCES: THEORY & PRACTICE

Received: July 9, 2014
Revision received: May 1, 2016
Accepted: June 10, 2016
OnlineFirst: August 15, 2016

www.estp.com.tr
DOI 10.12738/estp.2016.5.0076 • October 2016 • 16(5) • 1531–1561

Research Article

The Turkish Version of Web-Based Learning Platform Evaluation Scale: Reliability and Validity Study*

Funda Dağ[1]
Kocaeli University

Abstract

The purpose of this study is to determine the language equivalence and the validity and reliability of the Turkish version of the *Web-Based Learning Platform Evaluation Scale* (*Web Tabanlı Öğrenme Ortamı Değerlendirme Ölçeği* [WTÖODÖ]) used in the selection and evaluation of web-based learning environments. Within this scope, the validity of the factor structure of this scale is examined on the basis of data collected from 482 students at 11 universities in Turkey. The results of the exploratory factor analysis (EFA) show that the 40-item, four-factor structure of the scale parallels the original scale. To determine whether this structure agrees with the sample data, first-level confirmatory factor analysis (CFA) was conducted. The findings indicate that the compliance level of the sample is good. In addition, Cronbach's alpha reliability coefficient for the entire scale was calculated as .95, whereas McDonald's omega coefficient, which is regarded more relevant for congeneric measurements, was calculated as .96. These results indicate that the Turkish version of this scale, with its 40-item, four-factor structure, is a valid and reliable instrument.

Keywords

Web-based learning • Evaluation scale • EFA • CFA • Validity and reliability

* This work was supported by a grant (No. 2012-033) from the Scientific Research Project Unit of Kocaeli University.
1 **Correspondence to:** Funda Dağ (PhD), Department of Computer Education and Instructional Technologies, Kocaeli University, Umuttepe, Kocaeli 41380 Turkey. Email: fundadag@kocaeli.edu.tr

Citation: Dağ, F. (2016). The Turkish version of Web-Based Learning Platform Evaluation Scale: Reliability and validity study. *Educational Sciences: Theory & Practice, 16*, 1531–1561.

The Internet, with its wide array of functions, has become a highly valuable source of information (Yiğit, Yıldırım, & Özden, 2000). In addition, it can be used as an effective teaching tool, owing to its ability to disseminate educational information in accordance with the scope and objectives of certain curriculums (Smith & Rogan, 1993). Studies have shown that the Internet is quickly becoming an important tool for the purpose of learning, teaching and obtaining information (Merrill & Goodman, 1972; Reiser & Gagne, 1983). In recent years, the number of web-based education applications has been increasing at a remarkable rate (Jackson, 2000), and many institutions and individuals prefer to provide education/training via the Internet, in accordance with the latest requirements. In this regard, personal/informal as well as institutional/formal web-based learning platforms have become widely used.

However, because of the wide variety of web-based learning applications and the lack of specific evaluation tools to effectively evaluate these web-based learning environments, it has been difficult for users (i.e. teachers/students/domain experts) to select the most suitable web-based learning applications among the many samples. Since studies continue to indicate that the educational and technological qualities of the web-based learning platforms can effectively enhance the quality of education and promote success (Dağ & Buluş Kırıkkaya, 2012; Jackson, 2000), the issue of how a web-based application should be selected/evaluated has become increasingly important.

Web-Based Learning Applications and Selection

Web-based learning applications, which consist of different components, are used to present teaching contents through interactive exercises and multimedia materials. In addition, based on these components, the teaching strategies determine how these contents are presented, whereas certain tools can improve access to such contents (Pahl, 2003). Previous studies have shown that web-based learning is an original approach that utilises technological features and interface components as instructional tools (Allen & Allen, 2002; Horton, 2000; Pahl, 2003).

In the development and evaluation of effective web-based learning applications, all of the technological, pedagogical and content-related components should be considered (Clark & Mayer, 2008; Dietinger, 2003). In the literature related to the development and usability of websites, especially in regard to their design criteria (Bevan, 2005; U.S. Dept. of Health and Human Services, 2006), some studies have focused on the development scale(s) for the design and evaluation of educational web-based applications (Çakıroğlu, Akkan, & Çebi, 2008; Hajerrouit, 2010; Henke, 2001). These studies generally examined the interface components of web-based learning applications and focused on design principles in terms of their usability factor (Hasan, 2014; Silius & Tervakari, 2003; Öztekin, Kong, & Uysal, 2010).

In general, previous studies on the design of web-based learning have focused on the technological dimensions of such applications. However, limited studies have evaluated all of the components of web-based learning platforms and the development of standard-based evaluation criteria and measurement tools (Ateş, 2013; Hsu, Yeh, & Yen, 2009). Moreover, some research related to the evaluation of web-based learning platforms only focused on a certain field of education (e.g. foreign language) (Liu, Liu, & Hwang, 2011; Yang & Chan, 2008). Thus, for effective and successful web-based learning, studies need to identify the standards for evaluating web-based learning platforms, especially in terms of educational, technological and pedagogical aspects (Dabbagh, 2005).

The measures set forth in the aforementioned studies related to the design, selection and evaluation of web-based learning applications appear to be similar (Ateş, 2013; Çakıroğlu et al., 2008; Hasan, 2014). In general, these measures can be categorised as those related to content (e.g. target-specific, accuracy, objectivity, comprehensiveness, timeliness, etc.) and those related to design (e.g. visual components for design, aesthetics, visual appeal, accessibility, guidance, navigation, ease of use, privacy and security, etc.). Moreover, Hajerrouit (2010) proposed two main evaluation criteria for the assessment of web-based learning applications: technological usability (e.g. page design, content design, site design, etc.) and pedagogical usability (e.g. intelligibility, utility, multimedia, collaboration, etc.) Accordingly, it can be stated that there is a lack of studies related to the criteria and scales necessary to evaluate web-based learning applications with all of their components.

Hsu et al. (2009) conducted a study on the development of a standard scale for evaluating the educational and technological features of web-based learning applications, including content, teaching strategies, teaching materials and interface. Within this scope, the researchers examined the validity and reliability of the scale items on the basis of a sample that included individuals who studied instructional design education and those who did not. As specified by the authors of the aforementioned study, the goal was to investigate (in different languages) whether the developed scale is a standard and objective one that is reliable and valid for evaluating web-based learning platforms.

The present study, which focuses on the adaptation of the original form developed by Hsu et al. (2009) to a different culture, is (to date) the first one in the literature. This study contributes to the literature in two ways. First, it develops a standard scale of which its reliability and validity is ensured in a different language. Second, it can be used in the selection and evaluation of widely used, web-based learning applications. The Turkish version of this scale was also used in a project (supported by the Department of Scientific Research and Projects at a university in Turkey) to evaluate and compare web-based teaching materials.

Research Objective

The objective of this study is to determine the language equivalence and the validity and reliability of the Turkish version of the *Web-Based Learning Platform Evaluation Scale* (*Web Tabanlı Öğrenme Ortamı Değerlendirme Ölçeği* (WTÖODÖ)), developed in English by Hsu et al. (2009), in order to evaluate web-based applications with all of their components.

Method

A quantitative approach was employed in this study to collect several types of information from the participants, including their attitudes, beliefs and opinions (McMillan & Schumacher, 2006). This section describes the pertinent details of the study group, the data collection tool, the procedures used to examine the validity and reliability of the WTÖODÖ (intended to be adapted to the Turkish language), the language equivalence procedure and the exploratory and confirmatory factor analysis procedures.

Participants

The participants consisted of 482 students enrolled in different programmes at various universities in Turkey. These students agreed to participate in the study on a voluntary basis. In this study, the criteria to determine the study group of the original scale were taken into consideration. Thus, the present study group also includes participants who have different knowledge and experience levels related to web-based learning platforms, especially in terms of education and technology. Within this scope, the participants consisted of third- or fourth-year students in the following disciplines: computer education; computer education and instructional technology; and computer engineering.

During the fall semester of the 2012–2013 academic year, data was collected through use of electronic forms that were submitted under the supervision of the university faculty members who agreed to participate in the study. Table 1 shows the distribution of the participants according to their respective university, department, gender and class year.

Table 1
Demographic Information of the Participants

University	Department	Gender		Class		Total
		Male	Female	3th	4th	
Abant İzzet Baysal University	Computer Education	1	2	2	1	3
	Computer Education and Instructional Technology	19	22	19	22	41
	Total	20	24	21	23	44 (9.1%)
Atatürk University	Computer Education	1	4	4	1	5
	Computer Education and Instructional Technology	14	18	25	7	32
	Total	15	22	29	8	37 (7.7%)
Başkent University	Computer Engineering	10	4	12	2	14
	Computer Education and Instructional Technology	4	4	7	1	8
	Total	14	8	19	3	22 (4.6%)
Dokuz Eylül University	Computer Education and Instructional Technology	18	8	14	12	26
	Total	18	8	14	12	26 (5.4%)
Fırat University	Computer Education	3	1	-	4	4
	Computer Education and Instructional Technology	52	45	-	97	97
	Total	55	46	0	101	101 (21%)
Gazi University	Computer Engineering	9	12	20	1	21
	Total	10	12	20	1	22 (4.4%)
Karadeniz Teknik University	Computer Education	2	1	3	-	3
	Computer Education and Instructional Technology	34	24	35	23	58
	Total	36	25	38	23	61 (12.7%)
Kocaeli University	Computer Engineering	21	27	2	46	48
	Computer Education	7	4	-	11	11
	Total	28	31	2	57	59 (12.2%)
Marmara University	Computer Education	24	10	25	9	34
	Total	24	10	25	9	34 (7.1%)
Uludağ University	Computer Education	0	3	3	-	3
	Computer Education and Instructional Technology	10	10	11	9	20
	Total	10	13	14	9	23 (4.8%)
Yüzüncü Yıl University	Computer Education and Instructional Technology	43	11	4	50	54
	Total	44	11	4	50	54 (11.2%)
Total	Gender	272 (56.4%)	210 (43.6%)	-	-	482 (100%)
	Class	-	-	186 (38.6%)	296 (61.4%)	482 (100%)
	Computer Education and Instructional Technology			336 (69.7%)		
	Computer Education			63 (13.1%)		
	Computer Engineering			83 (17.2%)		
	Final Total			482 (100%)		

The sample in the study consisted of 482 participants in their third (186 students, 38.6 %) and fourth (296 students, 61.4%) years from the departments of computer education and instructional technology (336 students, 69.7%), computer education (63 students, 13.1%) and computer engineering (83 students, 17.2%) from 11 different universities in Turkey. Among the students, 56.4% were male (272) and 43.6% were female (210). According to this data, it can be stated that there was a balanced distribution of participants according to their gender, class year and department.

Data Collection Tool

The WTÖODÖ, developed in English by Hsu et al. (2009), was used to evaluate the web-based learning environment, under the required permissions from the authors. According to Hsu et al. (2009), the purpose of their study was to develop an evaluation scale and determine the design criteria for web-based learning platforms that cover various aspects of such learning, including instructional design and learning theories, learning tools and interface design.

The original scale consisted of 40 items and four factors, the latter of which included learning interface, teaching material, learning tool and instructional strategy. The components of the factor structure in the original scale are shown in Table 2.

Table 2
Components of the Factor Structure in the Original Scale

Factor (4 factors)	Category (22 indicators)	Items	Factor (4 factors)	Category (22 indicators)	Items
Instructional strategy (6 indicatosr) (10 items)	Instructional goal	1, 2, 8	Learning tool (5 indicators) (10 items)	System tool	21, 28
	Evaluation	3		Facilitation teaching	22, 23, 24, 29
	Assistance	4		Linking function	25, 26
	Teaching	9		Usability	27
	Communication	10		Navigation design	30
	Other strategies	5, 6, 7	Learning interface (5 indicators) (10 items)	Text	31, 32
Teaching material (6 indicators) (10 items)	Accuracy	11, 12		Image	33, 34
	Paragraph division	13		Animation	35, 36
	Appropriateness	14, 16, 17 20		Video	37, 38
	Range scheme	15		Overall interface design	39, 40
	Clear topic	18			
	Systematicness	19			

According to the researchers, they decreased the number of items (collected from the five different scales) from 275 to 56 by applying the Delphi technique that they completed during the first step of the study. Then, they implemented heuristic evaluation in three rounds (over the 56 items) and obtained the final scale form that included 40 items. In order to determine the validity and reliability of the scale, the researchers stated that they conducted exploratory factor analysis (EFA) on the final sample of 150 participants, which included individuals who studied instructional

design education and those who did not. They also stated that, during the EFA implementation, they used principal component analysis as a factorisation method and the rotation method as the varimax rotation method. At the end of their analysis, the researchers indicated that the total variance of the 40-item, four-factor scale was 67.497%, and the total variance rates of each of the scale's sub-factors were 15.959% for the instructional strategy factor, 18.817% for the instructional material factor, 13.394% for the learning tool factor and 19.327% for the learning interface factor. The factor loads of the items, according to the scale's sub-factors, fluctuated between .371 and .762 in the instructional technology factor, .624 and .736 in the teaching material factor, .532 and .684 in the learning tool factor and .594 and .812 in the learning interface factor. Furthermore, it was reported that the correlation between the four sub-factors and items of the scale changed between .700 and .784 ($p < .01$), which was significant. Finally, they reported that the Cronbach's alpha reliability coefficient of the scale was calculated as .924, and the reliability coefficients of the scale's sub-factors were calculated as .977 for the instructional strategy factor, .945 for the teaching material factor, .946 for the learning tool factor and .926 for the learning interface factor. The original scale was based on a five-point Likert scale ranging from 1 = not very important to 5 = very important.

Procedures

The necessary permission was obtained from the researchers who developed the original scale before working on the adaptation of the WTÖODÖ. During the implementation stage, the Turkish version of the scale was obtained by translating the original scale into Turkish, after which the language equivalence was examined. The validity study was initiated after determining the equivalence of both scales.

During the validity study of the WTÖODÖ's original form, EFA was conducted in order to formulate the scale's factor structure. Thus, although the validity and reliability analyses of the original scale were performed with fewer samples than those in the present study, EFA was still applied to examine the factor structure of the Turkish version of the scale. Moreover, in order to support the structure formed by EFA, confirmatory factor analysis (CFA) was conducted on the same sample (Kline, 1994; Van Prooijen & Van der Kloot, 2001).

In the reliability study, the internal consistency of the Turkish scale was examined. In this context, Cronbach's alpha coefficients for the overall factors and sub-factors of the Turkish scale were calculated and item analysis was conducted. Moreover, in the calculation of the internal consistency of the Turkish scale with Cronbach's alpha coefficient, McDonald's omega coefficient (McDonald, 1985; Yurdagül, 2006; Zinbarg, Revelle, Yovel, & Li, 2005), which is known to be suitable for congeneric measurements, was employed.

At the end of this application, the Turkish scale was used in the evaluation of three web-based teaching materials with different samples. The internal consistency coefficients for each scale were calculated, while the consistency of the Turkish scale in different evaluations was examined by calculating the total mean score based on the teaching materials' scale factors and the correlations between them.

Language equivalence procedure. This procedure was performed to identify whether there were any errors in the translation of the scale's items and to determine how much each item reflected its meaning. First, the forms used during the translation process were prepared for the experts, after which the Turkish translation of the original scale was performed independently by five experts who could speak both languages. Second, the translations were examined by two individuals in the translation group, after which a joint text was created. During the examination of the individual translations, the similarities between them facilitated the creation of a common text. The joint text and the original scale were then presented to the experts for their opinions. Third, one expert in the field of English language teaching performed a back translation based on other experts' recommendations. This was followed by another back translation by one expert in the field of education technology and another expert in the field of English language teaching. Fourth, the version was compared with the English back translation and compatibility between the original scale and the English translation was examined. As a result of this process, the scales were found to be compatible with one another. Fifth, the Turkish translation of the scale items was assessed by an expert in terms of its compliance with grammatical rules of the language.

In the final stage of the language equivalence process, the original scale and the language equivalence of its Turkish translation were piloted with 50 third-year students and 60 fourth-year students from the English Language Education Department at a state university in Turkey. The study group was organised according to the Solomon four-group design. Accordingly, the distribution and implementation process of the four groups related to the language equivalence study are shown in Table 3.

The students in Table 3 were grouped on the basis of their previous success in translation courses. The Turkish and English versions of the scale were presented to Groups 1 and 2 every two weeks, while the Turkish version was given to Group 3 and the English version was given to Group 4. Primarily, in the analysis of the collected data, a one-way variance analysis was performed to test whether there was a statistically significant result between the arithmetic means, especially in terms of the scale's sub-dimensions resulting from the second application of the four groups. Moreover, in order to test whether there was a significant difference between the first and second applications of Groups 1 and 2, a sample t-test and correlation analysis of the scale factors and items were conducted.

Table 3
Distribution and Implementation Process of the Groups related to the Language Equivalence Study

Group	1st implementation			2nd implementation	
	Form type	n	Time	Form type	n
Group 1	English	25	2 week	Turkish	25
Group 2	Turkish	25	2 week	English	25
Group 3			-	Turkish	30
Group 4			-	English	30

Validity procedure. The validity of the scale was tested with EFA, based on the data collected from the Turkish students. Then, CFA was applied to the scale as well as its factor structure to determine whether it confirmed the Turkish sample.

Reliability procedure. In order to analyse the consistency of the participants' responses regarding the individual scale items, the internal consistency reliability of the scale was determined, based on the reliability data obtained from a single sample (Büyüköztürk, 2010, p. 169; Şencan, 2005, p. 170). In determining the internal consistency, Cronbach's alpha reliability coefficients and McDonald's omega coefficients were calculated, after which the item analyses were performed. In addition, the correlations between the total scores of the scale factors and the correlations between them were examined.

Application of the Turkish Scale. The Turkish version of the scale was used to evaluate a group of web-based teaching materials with different samples at the end of the validity and reliability study. In this application, a new group of participants was involved in the evaluation of three different web-based teaching materials. At the end of the evaluation, Cronbach's alpha internal consistency coefficients and the total mean scores of the materials were calculated, based on the overall scale, its sub-factors and their correlations. Thus, the Turkish version was applied and the consistency of the scale (under different measurements) was investigated.

In this study, comparisons between the groups regarding the language equivalence analysis, the EFA applied for the validity and reliability of the scale and the analysis of the data obtained from the application of the Turkish scale were all performed with the SPSS 15 programme. For all of the analyses, the findings were evaluated at a 95% confidence interval with a $p = .05$ level of significance. Moreover, the CFA was performed with the LISREL 8.7 programme.

Results

Findings of the Language Equivalence Procedure

According to Table 3, the Shapiro–Wilk test was used to evaluate the suitability of the data for the parametric analysis and the normal distribution of the data. In

addition, Levene's test was carried out to test the homogeneous distribution of the data according to the groups and factors specified in the original scale. When the data was evaluated with the Shapiro–Wilk test, it was observed that the significance value of each group and each factor of the scale was $p \geq .08$. Regarding Levene's test, the finding that the significance value of each factor was $p \geq .45$ indicates that the data can be analysed parametrically.

Furthermore, as shown in Table 3, a one-way analysis of variance was conducted in terms of the sub-dimensions of the scale in order to examine whether there was a difference between the scores obtained from the applications in Groups 3 and 4, the scores obtained from the final application in Groups 1 and 2 and the total average scores. The mean test scores regarding the sub-dimension of the scale and the standard deviation values are presented in Table 4.

Table 4
Mean Test Scores regarding the Sub-dimensions of the Scale and the Standard Deviation Values

Sub-Factor	Group	n	$\bar{x}$	Std. Deviation	Std. Error
Instructional strategy	Group 1	25	40,88	3.44	.69
	Group 2	25	40.72	4.71	.94
	Group 3	30	40.77	5.46	.100
	Group 4	30	38.43	4.92	.90
	Total	110	40.15	4.79	.46
Teaching material	Group 1	25	41.64	4.80	.96
	Group 2	25	41.12	4.64	.93
	Group 3	30	41.27	5.05	.92
	Group 4	30	39.43	5.67	1.04
	Total	110	40.82	5.09	.49
Learning tool	Group 1	25	39.92	4.31	.86
	Group 2	25	41.16	3.91	.78
	Group 3	30	39.50	5.20	.95
	Group 4	30	38.00	4.14	.76
	Total	110	39.56	4.53	.43
Learning interface	Group 1	25	40.24	3.88	.78
	Group 2	25	42.64	4.50	.90
	Group 3	30	41.47	5.84	1.07
	Group 4	30	39.20	4.47	.82
	Total	110	40.84	4.88	.47

According to Table 4, there are small differences between the mean scores of the groups regarding the sub-dimensions of the scale. The variance analysis results, which were applied to test whether these differences are statistically significant, are presented in Table 5.

Table 5
The Variance Analysis Results regarding the Differences between the Mean Scores of the Scale's Sub-dimensions

		Sum of squares	Df	Mean square	*F*	*p*
Instructional strategy	Between groups	121.26	3	40.42	1.80	.152
	Within groups	2382.41	106	22.48		
	Total	2503.67	109			
Teaching material	Between groups	82.73	3	27.58	1.07	.367
	Within groups	2743.63	106	25.88		
	Total	2826.36	109			
Learning tool	Between groups	140.36	3	46.79	2.37	.075
	Within groups	2092.70	106	19.74		
	Total	2233.06	109			
Learning interface	Between groups	182.47	3	60.82	2.67	.051
	Within groups	2412.59	106	22.76		
	Total	2595.06	109			

As shown in Table 5, there was no statistically significant difference between the groups since the mean level of instructional strategy, teaching material, learning tool and learning interface was .05. Since there was no significant difference between the group averages, it was concluded that there was no difference between the English and Turkish versions of the scale in terms of language equivalence.

As a result of the application of the scale in Groups 1 and 2, a related group *t*-test was conducted on the data of Group 2 in order to test whether there was a significant difference between the Turkish and English versions of the scale. The findings show that the sub-dimensions and items are expected to meet the conditions of the significant correlation value as well as the non-significant *t*-value (or at least one of them) (Akbulut, 2010, p. 52). The results are presented in Table 6.

Table 6
Mean, Standard Deviation, Correlation and t-values of the Turkish and English Versions regarding the Sub-dimensions of the WTÖODÖ

Sub-factor (Turkish)	$\bar{x}$	Std. Deviation	Sub-factor (English)	$\bar{x}$	Std. Deviation	*r*	*t*
Instructional strategy	40.72	4.713	Instructional strategy	39.08	3.39	.49*	1.92
Teaching material	41.12	4.64	Teaching material	40.32	5.11	.73**	1.12
Learning tool	41.16	3.91	Learning tool	38.36	5.43	.61**	3.22**
Learning interface	42.64	4.50	Learning interface	39.84	5.90	.56**	2.78**

n = 25,*$p < .05$, **$p < .01$.

According to Table 6, there was no significant correlation in the sub-dimensions of the instructional strategy ($t = 1.92$) and teaching material ($t = 1.12$), based on the related samples t-test conducted on the scores of the Turkish and English versions in Group 2. However, there was a significant correlation in the sub-dimensions of learning tool ($t = 3.22$) and learning interface ($t = 2.78$) at the $p < .01$ level. Accordingly, it can be stated that the Turkish and English versions of the scale are

equivalent, owing to the presence of high-level correlation coefficients for the same sub-dimensions and the significant correlation between the total mean scores of the two dimensions of the scale.

Table 7
Mean, Standard Deviation, Correlation and t-values of Turkish and English Versions regarding the Items of the WTÖODÖ

Items (Turkish)	$\bar{x}$	Std. Deviation	Items (English)	$\bar{x}$	Std. Deviation	*r*	*t*
1	4.40	.50	1	4.44	.51	-.23	-.25
2	4.20	.58	2	4.04	.54	.25	1.16
3	4.28	.46	3	4.00	.58	-.16	1.77
4	3.88	.88	4	3,96	.94	.55**	-.46
5	3,96	.98	5	3,76	.52	-,02	.89
6	4,04	.89	6	3,88	.73	.53**	1
7	4,08	.64	7	3,92	.10	.27	.79
8	4,28	.54	8	3,88	.67	-.37	2
9	3,80	.96	9	3,96	.68	-.34	-.60
10	4,08	.86	10	4.08	.57	-.01	0
11	3.72	1.10	11	4.00	.87	.48*	-1.37
12	4.12	.83	12	4.16	1.11	.11	-.15
13	4.20	.65	13	4.12	.78	.53**	.57
14	4.32	.69	14	4.12	.88	.07	.93
15	4.20	.71	15	3.80	.96	.62**	2.62*
16	4.36	.57	16	4.12	4.12	.32	1.66
17	4.24	.66	17	3.96	4.24	-.24	1.27
18	4.04	.79	18	4.04	3.96	-.00	0
19	4.44	.51	19	4.24	.60	.46*	1.73
20	4.16	.62	20	4.08	.81	.30	.46
21	4.08	.76	21	3.96	.84	.27	.62
22	4.08	.70	22	3.72	.89	-.23	1.44
23	4.12	.78	23	3.92	.86	.32	1.04
24	4.04	.86	24	3.84	.85	.56**	1.23
25	4.28	.84	25	3.56	.58	-.59**	2.82**
26	3.88	.78	26	3.88	.93	.44*	0
27	4.08	.64	27	4.08	.53	.40*	1.55
28	4.32	.69	28	4.24	.93	.01	.35
29	4.16	.62	29	3.96	.89	.09	.96
30	4.20	.71	30	4.20	.65	.27	0
31	4.64	.49	31	4.40	.58	.24	1.81
32	4.16	.75	32	3.88	.67	.46*	1.90
33	4.32	.75	33	3.96	.74	.48*	2.38*
34	4.36	.57	34	4.12	.78	-.01	1.24
35	4.16	.85	35	4.04	.94	-.17	.44
36	4.36	.70	36	4.20	.87	.08	.75
37	4.20	.76	37	3.84	.85	-.46*	1.30
38	4.36	.64	38	3.76	1.01	.59**	3.67**
39	3.96	.98	39	3.80	.76	.44*	.85
40	4.36	.70	40	4.08	.86	-.12	1.19

n = 25, *$p < .05$, ** $p < .01$.

The mean, standard deviation, correlation and t-values obtained from the related sample *t*-test were applied to the scores of the same group (Group 2) to test whether there was a significant difference between the items in the Turkish and English versions of the scale. The results are presented in Table 7.

The correlation and t-value significance levels were examined in order to identify the equivalence of the scale items in terms of language. Accordingly, as shown in Table 7, while items 1–14, 16–24, 26–32, 34–37 and 39 met non-significant *t*-value conditions, items 15, 25, 33 and 38 did not. However, they met the condition of significant correlation. Based on this data, it can be stated that the items in the Turkish and English versions of the scale are equivalent.

Findings of the Explanatory Factor Analysis

EFA was applied to the data from the 482 Turkish students in order to form the factor patterns in the examination of the scale's validity. The Kaiser–Meyer–Olkin (KMO) measure and Bartlett's sphericity test were applied to test the compatibility of the data before performing the EFA. The KMO value was found to be .95, and since it was above .90 (Tavşancıl, 2005, p. 50), it was concluded that the data for the EFA was a perfect fit. In addition, the obtained chi-square value, as a result of Bartlett's sphericity test, was found to be significant (χ^2 (482) = 10823.39; $p < .01$). In this regard, it was accepted that the data came from multivariate normal distribution. As in the original scale, in the application of EFA, principal components analysis was used as the factoring method and varimax orthogonal rotation was used as the rotation method (Hsu et al., 2009).

In the first analysis conducted without rotation, it was found that there were seven components whose eigenvalues were greater than 1. The contribution of the first component to the total variance coefficient was 39.88%, the contribution of the other components to the total variance coefficients varied from 4.78% to 2.64% and the contribution of the seven components to the total variance coefficient rate was 60.25%. When the total variance table and screen-plot graph were examined, it was found that the eigenvalues of the components (starting from the fifth component) were close to 1 (or less) and their contribution to the variance was relatively small. Thus, an additional analysis of the four factors and the factor structure in the original scale was conducted (Hsu et al., 2009).

In the repeated analysis of the four factors, it was found that the contribution of the factors to the variance was 14.40% for the first factor, 13.47% for the second factor, 12.65% for the third factor, 11.29% for the fourth factor and 51.80 % for all four factors. On the basis of the factorial structure emerging from the analysis, it was found that the majority of the items were similarly grouped under the same factors

as those in the original scale (Hsu et al., 2009). However, 11 items (i11, i12, i14, i20, i26, i27, i29, i31, i32, i33 and i40) were found to be overlapping. In addition, among these items, 7 (i12, i13, i14, i28, i29, i30 and i31) were found to be under different factors from the factors in the original scale structure. The overlapping items and the ones specified under the factors that differed from the factors in the original scale are presented in italics in Table 8.

In the resulting structure from the EFA, the items were primarily examined in terms of the acceptance level of the factor load values and overlapping. The acceptance level of the factor load values was taken as .32 (Tabachnick & Fidell 2001, p. 194 as cited in Çokluk, Şekercioğlu, & Büyüköztürk, 2012). The high rate of common variance that the factors explain in one item (close to 1 or above .66) increases the total variance rate described in relation to the resulting structure. The lower limit in the common factor variance can be accepted as .20. Accordingly, the factor load value of the structure formed as a result of the EFA was kept as .32 and all of the items (40 items in total) remained in the scale. Before deciding on the scale items that were overlapping, the overlaps in the items, their contributions to the common factor variance and the impact of the items on the conceptual structure of the scale were evaluated. Although there were 11 overlapping items (i11, i12, i14, i20, i26, i27, i29, i31, i32, i33 and i40), considering that the common factor variance described with the factors in these items was above (> .37) and which is well above the accepted lower limit of .20, they were kept in the scale. In addition, since the factor load values of these items were above .32 (>.34), the elimination of these items would have changed the conceptual structure of the scale, according to experts' opinions (Şencan, 2005, p. 392). The factor load values for each item in the scale and the common factor variance are shown in Table 8.

Table 8
The Factor Load Values of the WTÖODÖ based on the EFA

No	Item	Teaching Material	Learning interface	Instructional strategy	Learning tool	Comm. Fac. Vari. (h^2)
i18	Öğretim materyalindeki ünite/bölüm/ konu başlıkları net ve açıktır.	.701				.61
i19	Öğretim materyalinin organizasyon yapısı açık, anlaşılır ve sistematiktir.	.615				.61
i17	Öğretim materyalinin niteliği (düzey/ doğruluk/güncellik vb.) uygundur ve öğrencinin kapasitesi ile örtüşür.	.606		.399		.59
i16	Öğretim materyalinin niceliği (sayısı/ miktarı/süresi) uygundur ve öğrencinin kapasitesi ile örtüşür.	.592		.426		.60
i28	*Site, hızlı hata yönergesi (düzeltme/ ayıklama için) sağlar.*	*.569*			*.388*	*.55*
i15	Öğretim materyallerinin sunum planı uygundur ve materyaller birbirleri ile ilişkilidir.	.561		.352		.57
i31	*Metinler açık bir biçimde okunabilir.*	*.519*	*.491*			*.59*
i30	*Gezinme açık ve kolay anlaşılırdır.*	*.515*	*.409*			*.52*
i11	Öğretim materyali hatasız ve eksiksizdir.	.488		.387		.49
i29	*Site, yöneticiden destek almak için uygun bir mekanizma sağlar.*	*.470*			*.465*	*.51*
i20	*Öğretim materyalinin sunduğu örnek olaylar ve durumlar, öğrencinin bilişsel yetenekleriyle örtüşür.*	*.418*	*.330*	*.326*	*.330*	*.50*
i37	Video kalitesi net ve iyidir.		.690			.62
i38	Video iletimi akıcıdır ve duraklama yapmamaktadır.		.686			.59
i35	Canlandırma tasarımı (animasyon) bilgiyi açık ve anlaşılır şekilde iletir.		.665	.322		.61
i36	Canlandırma tasarımı (animasyon) öğrenme isteğini arttırır.		.599			.57
i39	Arayüz tasarımı memnuniyet verici ve sanatsaldır.		.595		.433	.56
i34	Grafikler ve metin birbirini tamamlar ve öğrencinin anlayışını geliştirmeyi destekler.		.565			.53
i33	*Görseller bilgiyi açık ve anlaşılır şekilde iletir.*	*.445*	*.540*			*.59*
i13	*Öğretim materyalindeki paragraflar açık ve anlaşılırdır.*	*.408*	*.528*	*.344*		*.57*
i40	*Arayüz tasarımı yaratıcıdır.*		*.480*		*.467*	*.47*
i32	*Kelimeler çoğunlukla bilgi aktarıcı özelliktedir.*	*.352*	*.426*			*.36*
i12	*Öğretim materyali tarafsızdır.*	*.335*	*.386*	*.323*		*.37*
i03	Site, sınıf için değerlendirme amaçlı alıştırma uygulamaları sunar.			.617		.45
i06	Site, öğrenmeye yardımcı olacak farklı ortamlar kullanır.			.60		.44
i05	Site, öğrenenin daha iyi anlayabilmesi için örnek olay ve durumlar sunar.		.327	.596		.48
i01	Site, öğretim hedefini açıkça belirtir.			.579		.46
i02	Site, öğrenilmesi gereken bilgi ve teknikleri belirtir.			.577		.48

Table 8
The Factor Load Values of the WTÖODÖ based on the EFA

i08	Sitede sunulan içerik, öğretim hedeflerine uygundur.	.415		.566		.53
i07	Site, motivasyonu arttırmak için yeni ve ilgi çekici stratejiler uygular.			.560	.359	.51
i09	Site öğrenenlerin geçmiş öğrenme deneyim ve bilgilerini etkili şekilde tamamlar.			.501	.362	.46
i04	Sitede sıkça sorulan sorular bölümü mevcuttur.			.461		.28
i10	*Site öğrenen iletişimi ve etkileşimine imkan sağlar. (örn: çevrimiçi tartışma listeleri)*			*.445*	*.432*	*.41*
i14	*Öğretim materyali öğrenme motivasyonunu arttırır.*	*.330*	*.343*	*.383*	*.349*	*.50*
i24	Site, öğrenme kaydı tutar.				.717	.57
i25	Menü bağlantıları düzgün görünür.	.464			.624	.64
i21	Site pratik öğrenme araçları (Ör: çevrim içi not etme aracı) sağlar.				.585	.48
i26	*Menü kategorileri düzgün ve uygundur.*	*.541*			*.571*	*.65*
i27	*Site, öğrenme sürecinin yönetilmesine imkan sağlar.*	*.482*			*.540*	*.58*
i22	Site, arama fonksiyonları sağlar.				.495	.43
i23	Site, indirme işlemleri için ilgili yazılımları sağlar.				.488	.39
Expl. Vari.	%51.802	%14.397	%13.468	%12.647	%11.289	

Note. Factoring method: Principal components analysis; Rotation method: Varimax orthogonal rotation.

The findings show that the majority of the items described as theoretical were placed under the factors in the original scale (Hsu et al., 2009). However, 7 of the 40 items in the original scale (i12, i13, i14, i28, i29, i30 and i31) were placed under factors that differed from those in the original scale structure. In addition, the factor load values of these items were above the acceptance level (>.32) and they had factor load values that were close to one another (see Table 9). Whereas Item 28 (*Site, hızlı hata yönergesi (düzeltme/ayıklama için) sağlar'*) and Item 29 (*Site, yöneticiden destek almak için uygun bir mekanizma sağlar*) are located under the learning tool factor in the original scale, in the present study, Item 28 (with its load value of .569) and Item 29 (with its load value of .470) are located under the teaching material factor. Similarly, Item 30 (*Gezinme açık ve kolay anlaşılırdır*), with a load value of .515, and Item 31 (*Metinler açık bir biçimde okunabilir*), with a load value of .519, are located under the teaching material factor in this study, even though they were placed under the learning interface factor in the original scale. Moreover, Item 12 (*Öğretim materyali tarafsızdır*), with a load value of .386, and Item 13 (*Öğretim materyalindeki paragraflar açık ve anlaşılırdır*), with a load value of .528, are placed under the learning interface factor in this study, even though they are located under the teaching material factor in the original scale. Finally, although Item 14, with a load value of .383, was located under the teaching material factor in the original

scale, it was placed under the instructional strategy factor in this study as a result of the analysis. Further information about these factors is presented in Table 9.

Table 9
The Factors under which the Items are Placed in the Turkish Version of the Scale compared to Those in the Original Scale

Item	Item No.	f1- Teaching material	f2- Learning interface	f3- Instructional strategy	f4- Learning tool	Comm. Vari. Contr.	The factor in the original scale	The factor in the Turkish scale
Site, hızlı hata yönergesi (düzeltme/ayıklama için) sağlar.	i28	.569	.236	.142	.388	.55	Learning tool	Learning tool
Site, yöneticiden destek almak için uygun bir mekanizma sağlar.	i29	.47	.214	.169	.465	.51	Learning tool	Learning tool
Gezinme açık ve kolay anlaşılırdır.	i30	.515	.409	.193	.23	.52	Learning tool	Learning interface
Metinler açık bir biçimde okunabilir.	i31	.519	.491	.209	.189	.59	Learning interface	Learning interface
Öğretim materyali tarafsızdır.	i12	.335	.386	.323	.032	.37	Teaching material	Teaching material
Öğretim materyalindeki paragraflar açık ve anlaşılırdır.	i13	.408	.528	.344	.059	.57	Teaching material	Teaching material
Öğretim materyali öğrenme motivasyonunu arttırır.	i14	.33	.343	.383	.349	.50	Teaching material	Instructional strategy

According to Table 9, Items 28 and 29 were placed under the learning tool factor, like in the original scale, after considering their factor load values, their contributions to the common variance and the relations of their contents to their conceptual structure. Item 30, placed under the learning tool factor in the original, was located under the learning interface factor in the present study, based on the consensus of the experts, after considering its factor load value, its contribution to the common variance, its relation to the conceptual structure and its non-valid load value of .23. Item 30 is an item representing the navigation design indicator, one of the indicators of the learning tool factor, based on the conceptual structure of the scale (see Table 2). Following the experts' opinions, Item 30 was placed under the learning interface factor since navigation could be an important criterion of learning interface, it could be related to the learning tool factor and it has no relation with the indicators representing the teaching material factor. Moreover, Item 31 was placed under the learning interface factor, as in the original scale, after considering its factor load value, its contribution

to the common variance and the relation of its content with the conceptual structure.

Meanwhile, Items 12 and 13 were placed under the teaching material factor, as in the original scale, considering their factor load values, their contributions to the common variance and their relation with the conceptual structure. Owing to its higher factor load value of .383, Item 14 was placed under the instructional strategy factor instead of the teaching material factor, after consulting with the experts and considering its factor load value, its contribution to the common variance and its relation with the conceptual structure. In addition, Item 14 was placed under this factor since it was related to Items 5, 6 and 7, which represented the other strategies' indicators (see Table 2) found in the instructional strategy factor.

Finally, after determining the factors for all of the items, the load values ranged from .335 to.701 for the teaching material factor, from .409 to .690 for the learning interface factor, from .383 to .617 for the instructional strategy factor and from .488 to .717 for the learning tool factor. Accordingly, the load values of the factors under which the items were placed were evaluated, ranging from good to excellent. In addition, whereas 10 items were placed under each of the four factors in the original scale (see Table 2), in this study, 9 items (i11, i12, i13, i15, i16, i17, i18, i19, and i20) were placed under the teaching material factor, 11 items (i30, i31, i32, i33, i34, i35, i36, i37, i38, i39, and i40) were placed under the learning interface factor, 11 items (i1, i2, i3, i4, i5, i6, i7, i8, i9, i10, and i14) were placed under the instructional strategy factor and 9 items (i21, i22, i23, i24, i25, i26, i27, i28, and i29) were placed under the learning tool factor.

Based on the findings of the EFA, the final form of the 40-item, four-factor Turkish scale included a total variance explanatory rate of 51.802%, a KMO value of .95 and a Bartlett's sphericity test chi-square value (χ^2) of 10823.39 ($p < .001$, df = 482). Moreover, it can be stated that the total variance of the Turkish version is at a good level, since the variance rate falls between 40% and 60% (Tavşancıl, 2005).

Findings of the Confirmatory Factor Analysis

CFA is a technique used to verify a structure, which has been previously defined and delimited, as a conceptual structure or model (Maruyama, 1998 as cited in Çokluk et al., 2012). In this regard, CFA can be used to test the existing correlations between the items and factors obtained by EFA as well as to identify whether the items are sufficiently represented by the associated factors (Özdamar, 2002 as cited in Çokluk et al., 2012). The extent to which the recommended model matches the available data is determined by the compatibility index values (calculated at the end of the analysis) and the fit index criteria referred to in the literature regarding compatible indexes (Çokluk et al., 2012, p. 272).

For the WTÖODÖ, first-level CFA was applied for the four-factor structure to confirm the accuracy of the factor structure, according to the data obtained from the Turkish sample. Considering the conceptual structure of the scale, the alignment errors were determined to improve the adaptation of the model, after which the model was found to best fit the criteria. Before each modification process, attention was given to the fact that the factor loads of the latent variables representing each factor within the conceptual structure of the scale should be high. In addition to the adaptability of the model, even its variance should be low and its factor correlations should not be too high (>.85) (Çokluk et al., 2012, pp. 268–274).

The compliance index values, obtained before and after the modification of the scale and verified at the end of the CFA, are presented in Table 10.

Table 10
Comparison of the Compliance Index Values of the Turkish WTÖODÖ

Fit index*	Good compliance condition	Acceptable compliance condition	Scale model, as a result of the CFA and modification
x^2	$0 \leq x^2 \leq 2df$	$.05 \leq p \leq 2df$ $05 \leq p \leq 2df$	1421.99
p	$0 \leq x^2 \leq 2df$	$.01 \leq p \leq .05$	.000
x^2 / *df*	$0 \leq x^2 / df \leq 2$	$2 \leq x^2 / df \leq 3$	1.98
RMSEA	$0 \leq$ RMSEA $\leq .05$	$.05 \leq$ RMSEA $\leq .08$	.045
RMR			.026
SRMR	$0 \leq$ SRMR $\leq .05$	$.05 \leq$ SRMR $\leq .10$	.040
NFI	$.95 \leq$ NFI ≤ 1	$.90 \leq$ NFI $\leq .95$	.97
NNFI	$.97 \leq$ NNFI ≤ 1	$.95 \leq$ NNFI $\leq .97$	.99
CFI	$.97 \leq$ CFI ≤ 1	$.95 \leq$ CFI $\leq .97$	.99
GFI	$.95 \leq$ GFI ≤ 1	$.90 \leq$ GFI $\leq .95$	.87
AGFI	$.90 \leq$ AGFI ≤ 1	$.85 \leq$ AGFI $\leq .90$	.85
RFI	$.90 \leq$ RFI ≤ 1	$.85 <$ RFI $< .90$ $85 <$ RFI $< .90$	.97
PGFI			.76

Note. * Resource: Schermelleh-Engel, Moosbrugger, and Müler (2003).

The structure validity of the 40-item, four-factor scale was tested by conducting CFA. Upon examining the modification proposals, the analysis was reiterated by adding the 15 proposals that do not disrupt the conceptual structure of the criteria (see Table 2), that will have significant contributions on the x^2 value and that will have meaningful proximity (see Figure 1). As shown in Table 10, according to the compliance index values obtained for the final model, the value of x^2 / *df* was less than 2 (1.98) and it had good compliance. In addition, the root mean square error of approximation (RMSEA) value was less than .05 (.045), which also indicated good compliance. Furthermore, the goodness-of-fit index (GFI) value was .87, the adjusted goodness-of-fit value was .85, the root mean square residual (RMR) was .026 and the standardised RMR was .040, the non-normed fit index (NNFI) was .99 and the

comparative fit index (CFI) was .99, all of which indicated good compliance. Finally, the parsimonious goodness-of-fit index (PGFI) value, which gives an idea about the simplicity of the proposed model, was increased from .72 to .76, especially when its proximity to 1 was examined (Çokluk et al., 2012, p. 270).

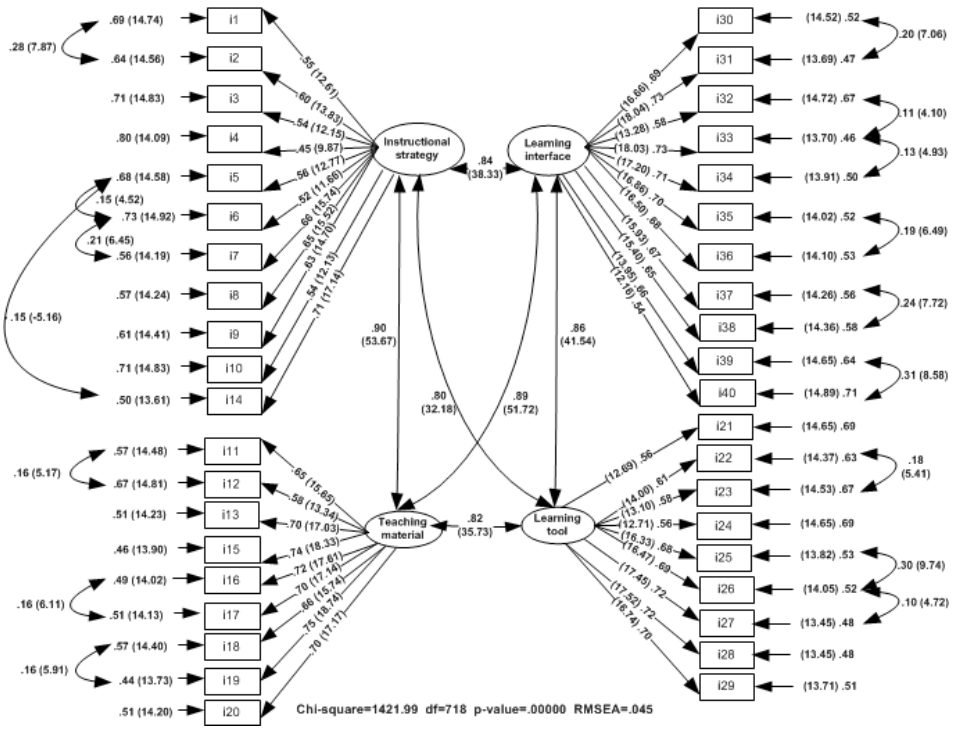

Figure 1. Confirmatory Factor Analysis (CFA) model.

Considering the scale of the CFA model presented in Figure 1, it can be stated that the t-values of the scale items vary between 9.87 and 18.33 and they are found to be significant at the $p < .01$ level (Çokluk et al., 2012). Accordingly, since all of the scale items are valid indicators of the related factors, it can be stated that they are significant and acceptable. According to Figure 1, the error variances of the items ranged from .46 to .75. Furthermore, the lambda-x values, standardised to the items of the scale, ranged from .45 to .75. On the basis of these findings, the error variance of the scale factor can be considered as low, whereas their factor loads can be considered as medium or high (Kline, 2005 as cited in Çokluk et al., 2012). In addition, when the correlations between the scale factors in Figure 1 were examined, it was found that the correlations of the instructional strategy–teaching material and teaching material–learning interface factors were slightly more than .85. As for the cause of these correlations being more than .85, the items representing these factors are thought to test the conceptual structures that are close to one another.

Finally, according to the compliance index values and the CFA model, it can be stated that the Turkish version of the WTÖODÖ, consisting of 40 items and four factors, is a good model with all of its compliance statistics and a valid scale with all of its factor structures.

Reliability Analysis

According to Cronbach's alpha internal consistency coefficient for the overall scale and its sub-factors as well as McDonald's omega reliability coefficient (McDonald, 1985; Yurdagül, 2006; Zinbarg et al., 2005), the 40-item, four-structure WTÖODÖ is accepted as a valid scale. In addition, the total item correlation of more than .30 is an indicator that the items distinguish the similar features well and that the internal consistency of the scale is generally high (Büyüköztürk, 2010).

Table 11
Cronbach's Alpha and McDonald's Omega Reliability Coefficients as well as the Total Item Correlations related to the WTÖODÖ

Factor	Item	Total item correlation*,**	Cronbach Alpha (α)	ω	Factor	Item	Total item correlation*,**	Cronbach Alpha (α)	ω
Instructional strategy	i1	.552	.86	.85	Learning tool	i21	.517	.87	.82
	i2	.584				i22	.573		
	i3	.49				i23	.524		
	i4	.411				i24	.464		
	i5	.504				i25	.616		
	i6	.491				i26	.635		
	i7	.605				i27	.64		
	i8	.613				i28	.645		
	i9	.59				i29	.630		
	i10	.508			Learning interface	i30	.661	.91	.90
	i14	.674				i31	.695		
Teaching material	i11	.618	.90	.89		i32	.537		
	i12	.527				i33	.70		
	i13	.663				i34	.662		
	i15	.691				i35	.66		
	i16	.679				i36	.65		
	i17	.663				i37	.645		
	i18	.632				i38	.622		
	i19	.712				i39	.577		
	i20	.68				i40	.538		
Reliability coefficients for the overall scale								.95	.96

*p < .01, ** N = 482.

Accordingly, when Table 11 is examined, Cronbach's alpha and McDonald's omega reliability coefficients for the scale and factors are found to be more than .80. Thus, it can be stated that the scale and factors are highly reliable (.80 < α < 1) (Özdamar, 1999 as cited in Çokluk et al., 2012). Moreover, as shown in Table 11, the total item correlation values of the scale range from .411 to .70. On the basis of these

findings, it can be stated that the scale items are reliable, their discrimination degrees are high and the WTÖODÖ is a reliable scale. The correlations between the scale factors are presented in Table 12.

Table 12
The Correlation Values between the Mean (Average) Scores, Standard Deviation and Factors of the WTÖODÖ

Factor	$\bar{x}$	Std. Deviation	$r_{teaching\ material}$	$r_{learning\ interface}$	$r_{instructional\ strategy}$	$r_{learning\ tool}$
Teaching material	38.454	5.26	--	--	--	--
Learning interface	46.434	6.34	.766**	--	--	--
Instructional strategy	45.652	5.76	.762**	.710**	--	--
Learning tool	36.788	5.16	.695**	.700**	.660**	--

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

According to Table 12, the mean scores obtained from the scale were respectively 45.652 (SD = 5.76) for the instructional strategy factor, 36.788 (SD = 5.16) for the learning tool factor, 38.454 (SD = 5.26) for the teaching material factor and 45.652 (SD = 5.76) for the learning interface factor. The significant correlations between all of the scale factors, and the positive and high relation between the factors, verify the structures obtained as a result of the explanatory and confirmatory factor analyses. This shows that the scale is, in fact, formed of independent factors.

Application of the Turkish Version of the WTÖODÖ

The Turkish version of the WTÖODO, in line with the findings, was used to evaluate three web-based teaching materials at different times. These teaching materials include similar content but different instructional materials developed for an 8th grade science and technology course. These materials were examined by 51 individuals, including three experts and 48 fourth-year students in the science education department, in order to ensure that the comparison and evaluation of these materials, especially regarding their instructional and technical features, were in accordance with the standards. Subsequently, the participants were asked to evaluate the three materials separately by using the WTÖODÖ. The total mean scores, standard deviation values and the correlations of the items, based on the scale's sub-factors, are presented in Table 13.

Cronbach's alpha coefficients calculated for the scale and the sub-factors for each material were found to be .80 or above. According to Table 13, the Turkish version of the WTÖODÖ can be regarded as a consistent scale in different evaluations, since the correlations between the scores are significant ($p < .001$), the correlations between the scores of different items received are from the same sub-factor ($p < .001$) and the correlations are positive and high in the evaluations of each of the three materials. In sum, given the results of the validity and reliability analyses, the Turkish version of the WTÖODÖ is a valid and reliable evaluation tool.

Table 13.
The Total Mean Scores, Standard Deviation and Correlation Values regarding the Sub-factors of the WTÖODÖ

	Factor	$\bar{x}$	Std. Deviation	Material 1				Material 2				Material 3			
				r_{IS}	r_{TM}	r_{LT}	r_{LI}	r_{IS}	r_{TM}	r_{LT}	r_{LI}	r_{IS}	r_{TM}	r_{LT}	r_{LI}
Material 1	Instructional strategy (IT)	39.76	5.96	--	--	--	--	--	--	--	--	--	--	--	--
	Teaching material (TM)	33.27	5.74	.787**	--	--	--	--	--	--	--	--	--	--	--
	Learning tool (LT)	29.86	5.17	.448*	.482**	--	--	--	--	--	--	--	--	--	--
	Learning interface (LI)	39.63	6.66	.634**	.653**	.378**	--	--	--	--	--	--	--	--	--
Material 2	Instructional strategy (IT)	40.43	6.09	.623**	--	--	--	--	--	--	--	--	--	--	--
	Teaching material (TM)	33.51	5.36	--	.476**	--	--	.671**	--	--	--	--	--	--	--
	Learning tool (LT)	29.53	7.40	--	--	.631**	--	.674**	.692**	--	--	--	--	--	--
	Learning interface (LI)	40.04	6.99	--	--	--	.359**	.647**	.768**	.621**	--	--	--	--	--
Material 3	Instructional strategy (IT)	41.51	6.86	.690**	--	--	--	.617**	--	--	--	--	--	--	--
	Teaching material (TM)	34.00	5.33	--	.465**	--	--	--	.344**	--	--	.787**	--	--	--
	Learning tool (LT)	31.18	6.42	--	--	.648**	--	--	--	.578**	--	.825**	.710**	--	--
	Learning interface (LI)	40.33	6.75	--	--	--	.399**	--	--	--	.460**	.790**	.724**	.769**	--

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

Conclusion

The purpose of this study was to adapt the *Web-Based Learning Platform Evaluation Scale*, developed by Hsu et al. (2009), to the Turkish language and determine if it can be utilised as a standard, objective scale for evaluating web-based learning platforms. In line with this purpose, the factor structure was created after conducting EFA of the data obtained from the Turkish participants and testing the structure by CFA.

Within the scope of this research, a translation, a back translation, a pilot study and a language equivalence study were performed in order to identify whether the original scale and the Turkish version of the scale were equivalent in terms of language. On the basis of the findings, it can be stated that the Turkish scale and the English scale are equivalent to one another.

The scale structure of the original scale, consisting of 40 items and four factors, was first tested by EFA, based on data collected from 482 students enrolled at various universities in Turkey. The results show that the Turkish version of the scale, including the four factors of instructional strategy, teaching material, learning tool and learning interface, explains more than half of the total variance (51.80%).

In the structure formed after conducting EFA, 11 items were identified as overlapping. Before deciding on the removal of these overlapping items from the scale, their factor load values and the impact of the items on common factor variance were examined. On the basis of the consensus of experts, these 11 items were kept in the scale since the factor load values of the 11 overlapping items were more than .32 (>.34), which is well above the accepted lower limit of .20. The 7 items among which there were also overlapping ones were placed under the factors that differed from those in the original scale. In addition, these items were placed in accordance with the factor load values of the items, their contribution rates to common variance and the experts' opinions regarding the correlation between the conceptual structure of the study and the items. As a result of all of these stages, the Turkish scale included a total of 40 items, including 11 items under the instructional strategy factor, 9 items under the teaching material factor, 9 items under the learning tool factor and 11 items under the learning interface factor.

It was concluded that the model, formed as a result of the first-level CFA, showed a good fit. The compliance indexes of the model were also examined and the chi-square value (χ^2 (718, N = 482) = 1421.99, $p < .000$) was found to be significant. In addition, the rate of the chi-square value to the degree of freedom was less than 2, while the root mean square error of approximation (RMSEA) value was .045, the standardised S-RMR was .040, the goodness-of-fit index (GFI) value was .87, the adjusted goodness-of-fit (AGFI) value was .85, the comparative fit index (CFI) was .99, the non-normed fit index (NNFI) was .99 and the normed fit index (NFI) was .97, all of which had an acceptable goodness of fit.

Cronbach's alpha coefficient for the Turkish scale was calculated as .95 for the overall scale, .86 for the instructional strategy factor, .90 for the learning material factor, .87 for the learning tool factor and .91 for the learning interface factor. In addition, since the factor loads of the scale items were not equal to one another, McDonald's omega coefficient for the overall scale and its sub-factors was calculated as .96 for the overall scale, .85 for the instructional strategy factor, .89 for the learning material factor, .82 for the learning tool factor and .90 for the learning interface factor. Accordingly, the adapted scale and related factors were found to be highly reliable ($.80 < \alpha < 1$) (Özdamar, 1999 as cited in Çokluk et al., 2012). Finally, on the basis of the research findings, it can be concluded that the Turkish version of the *Web-Based Learning Platform Evaluation Scale*, with its 40 items and four-factor structure, can be used as a valid and reliable tool to evaluate web-based learning environments, including their technological, pedagogical and content-wise components.

Discussion

In order to use an evaluation tool with a language that differs from its original language, the adaptation of the original scale to the target language should be systematically achieved and the psychometric features of the scale should be evaluated (Şencan, 2005). Owing to the situations arising from the structural features of the languages and the differences related to the concepts used in the scale, it may be insufficient to translate the scale from its original language into that of the target group's language. In scale adaptation, concept and language equivalence can be achieved through the double-translation method (Şencan, 2005) and subsequent statistical analyses (Hambleton, 2005). In the present study, the language and concept equivalence of the Turkish version of the WTÖODÖ and the original scale was ensured by applying the double-translation method and performing statistical analyses.

In addition, another important stage in scale adaptation studies is the identification of psychometric features of the scale. The data collection and analysis methods used are critical factors that affect the validity of the adapted scale (Hambleton, 2005). In the present study, the validity and reliability analyses were conducted on the findings obtained from a sample consisting of those with experience in teaching design education and those without such experience, which was a similar approach to the one used in the original scale. In this way, one of the critical factors that would have affected the validity of the adaptation was eliminated.

In social science research, it is difficult to arrive at a definite conclusion since it is dependent on many acceptances. However, it should be noted that researchers should exceed the minimum number that the data analysis requires (Karasar, 2010). For example, Comrey and Lee (1992) stated that 50 is too weak, 100 is weak, 200 is moderate, 300 is good, 500 is very good and 1000 is excellent for an adequate sample size in factor analysis. Consequently, Çokluk et al. (2012) indicated that a sample size of at least 300 should be a general rule (Çokluk et al., 2012, p. 206). The sample size can also be identified according to dependent criteria such as item and factor numbers. Based on these approaches, consistent results with samples of more than 150 can be obtained in the case of factor loads with limits of .60 and the number of items being at least four for each factor. A sample size of around 150 is thought to be sufficient in the cases of factor loads with limits of .40 and the number of items being 10 for each factor (Çokluk et al., 2012, p. 207).

According to Kline (2005), in CFA applications, the number of items of a sample being 10 times more is said to be suitable and it should not be less than 200. In addition, a sample size of more than 200 is considered to be adequate for CFA application, provided that there should be three items for each factor (Çokluk et al., 2012, p. 266). When evaluated according to this information, the WTÖODÖ, with

its original scale consisting of four factors and 10 items under each factor, can be said to be quite sufficient, especially when it is evaluated according to dependent criteria such as sample size (N = 150) and factor number. However, the sample size (N = 482) in the validity and reliability study of the Turkish version of the scale can be said to be highly sufficient for conducting both EFA and CFA. CFA can also be applied in studies of scale translation from a different language in order to verify the predetermined factor structure (Çokluk et al., 2012, p. 283). Owing to the impossibility of a comparison of EFA and CFA, on the basis of the inappropriate EFA applications and cultural differences (Van Prooijen & Van der Kloot, 2001), the factor structure formed by EFA may not be verified with the data received from a different sample (e.g. Dunn et al., 2006; Hooper, Marotta, & Depuy, 2009; Öztürk, 2011). In order to examine such cases, Van Prooijen and Van der Kloot (2001) stated that conducting CFA of the factor structures on the same dataset can be more effective.

In the literature, there are studies on adaptation with the application of CFA on the same dataset (e.g. Meunier & Roskam, 2009) and scale development (e.g. Küçük, Yılmaz, Baydaş, & Göktaş, 2011; Lee & Lehto, 2013; Li, 2016) in order to support the factor structure created by EFA. In the present study, the validity of the Turkish version of the original scale was examined by taking the above-mentioned approaches into consideration, even though they were implemented with a fewer number of participants. Another reason was that CFA was not conducted to verify the factor structure of the original scale. As a result of the EFA, there was a need to evaluate whether the factorial structure formed by applying CFA on the same dataset can be verified since there were 11 overlapping items in the Turkish scale, seven of which were placed under the factors that differed from the ones in the original scale. Thus, the number of items under the factors in the Turkish scale differed from that in the original scale.

It can also be concluded that there are effects of cultural differences when the 11 overlapping items are placed under different factors according to the factor structure in the Turkish scale, even though the language and conceptual equivalence have been ensured. The cultural validity of a scale depends on many features such as its compliance with information and the experiences in the group where the evaluation is performed (Şencan, 2005, p. 804). For instance, Item 30 (*Gezinme açık ve kolay anlaşılırdır*) had to be placed under the learning interface factor and under the teaching material factor since it had a factor load value of .409 with the former and a factor load value of .515 with the latter. Unlike the validity results in the original scale, it had a very low factor load value of .23 under the learning tool factor. This indicates that the sample to which the original scale was applied and the sample in this Turkish version interpreted some of the concepts differently. The results might also be similar when the other overlapping items and the ones placed under the different factors are examined. Accordingly, it can be concluded that students in the Turkish sample interpreted the

content and the platforms (i.e. learning management systems, mobile media, blogs or websites) where the content was presented as a whole, whereas the sample of the original scale interpreted the content and the presentation individually. This situation may have arisen from the cultural differences and/or the differences related to the experiences/perceptions of the two different samples related to web-based learning.

In this study, in the scope of the reliability analysis, Cronbach's alpha and McDonald's omega internal consistency coefficients were calculated for the Turkish scale and its factors. Besides, three different web-based teaching materials were evaluated with the scale by a different sample and statistical analyses were conducted to determine the consistency of the scale with these measurements. Within this scope, the Turkish version of the scale was aimed to establish a solid basis for further research by presenting more than one piece of evidence related to its validity and reliability.

Web-based learning consists of a wide range of platforms ranging from web-based instructional sites to learning management systems (Baker, 2003). Thus, it is important to identify the important critical factors for successful web-based learning applications for different users (e.g. students, teachers, designers, developers, etc.) and to create a multidisciplinary evaluation framework for web-based learning platforms with effective pedagogical and usability features (Silius & Tervekari, 2003). In this regard, it is important to develop standard scales that are tested in terms of validity and reliability for the selection and evaluation of web-based learning applications.

When considered from this point of view, the systematic and standardised process (Hsu et al., 2009) is considered to be important in the development of the scale items in order to identify the objective design principles for the web-based learning platforms as well as to develop an effective evaluation scale. For this reason, the Turkish version of the scale in this study contributes to the literature since it examines the standard and objective structure of the original scale in a different culture. This study is the first to adapt an original scale to a language of a different culture. Accordingly, this scale development study has practical implications since it can be used to create standard, objective evaluation tools in other web-based learning environments.

Suggestions

On the basis of the validity analysis conducted with the data obtained from the Turkish sample, it can be recommended that a new extensive scale be developed by adding new items or arranging the existing items. Moreover, in this study, the validity and reliability analyses were conducted on a sample consisting of university students. However, although there was an extensive study to ensure language equivalence through double translation and a pilot study during the application of the scale, the scale items and concepts can have different meanings for the participants, especially when the overlapping items in the scale

and the placement of some of the items under different factors are taken into consideration. For this reason, the validity and reliability of the scale can be examined with a sample of selected education and web-technology experts in order to assess the conceptual structure of the scale in-depth. Furthermore, with a different perspective, the implementation of EFA and CFA on the same sample in this study can pose a problem for the generalisation of different samples. Thus, future studies should verify the factorial structure on the basis of data collected from a different sample. Finally, future research should evaluate different web-based learning platforms, with samples consisting of users from different educational and technological backgrounds in order to evaluate the competency of the Turkish scale.

References

Allen, M. W., & Allen, M. (2002). *Michael Allen's guide to e-learning*. Hoboken, NJ: John Wiley and Sons.

Akbulut, Y. (2010). *Sosyal bilimlerde SPSS uygulamaları (sık kullanılan istatistiksel analizler ve açıklamaları SPSS çözümleri)* [SPSS Applications in social science (SPSS statistical analysis and explanation of commonly used solutions)]. İstanbul, Tukey: İdeal Kültür Yayıncılık.

Ateş, A. (2013). Eğitsel web sitelerini değerlendirmeye yönelik bir ölçek önerisi [A scale proposal for evaluation of the educational web sites]. *Eğitim Teknolojileri Araştırmaları Dergisi*, *4*(1). Retrieved from http://www.et-ad.net/

Baker, R. K. (2003). A framework for design and evaluation of internet-based distance learning courses: Phase one–framework justification, design and evaluation. *Online Journal of Distance Learning Administration*, *6*(2). Retrieved from http://www.westga.edu/~distance/ojdla/

Bevan, N. (2005, July). *Guidelines and standards for web usability*. Paper presented at the HCI International, Las Vegas, Nevada, USA.

Büyüköztürk, Ş. (2010). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni SPSS uygulamaları ve yorum* [Handbook of data analysis for socil science: Statistics, research design, SPSS applications and comments]. Ankara, Turkey: Pegem Akademi.

Clark, R. C., & Mayer, R. E. (2008). *E-learning and the science of Instruction: Proven guidelines for consumers and designers of multimedia learning*. Hoboken, NJ: John Wiley and Sons.

Çakıroğlu, Ü., Akkan, Y., & Çebi, A. (2008, January). *Eğitsel içerikli web sitelerinin standardizasyon kriterlerinin belirlenmesi ve uygulanması* [Determination of criteria for standardization of educational content website and implementation]. Paper presented at Akademik Bilişim, Çanakkale, Turkey.

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş., (2012). S*osyal bilimler için çok değişkenli istatistik* [Multivariate statistics for social science]. Ankara, Turkey: Pegem Akademi.

Dabbagh, N. (2005). Pedagogical models for e-learning: A theory-based design framework. *International Journal of Technology in Teaching and Learning, 1*(1), 25–44. Retrieved from http://livinglearning.co/ml/unit_three/module3.1/3.1/6.Pedagogical%20Models%20for%20E-Learning.pdf

Dağ, F., & Buluş Kırıkkaya, E. (2012). Sekizinci sınıf doğal süreçler ünitesi için web tabanlı öğretim materyali hazırlığı: Farklı alan uzmanlarının materyal hakkındaki görüşleri [Preparation of web-based teaching materials for natural processes in 8th grade: Different views on the material of professionals]. *e-Journal of New World Sciences Academy (NWSA)*, *7*(1), Article No: IC0487.

Dietinger, T., (2003). *Aspects of e-learning environments* (Doctoral dissertation, Graz University of Technology, Styria, Austria). Retrieved from http://www.iicm.edu/thesis/tdieting_diss.doc

Dunn, J. G. H., Dunn, J. C., Gotwals, J. K., Vallance, J. K. H., Craft, J. M., & Syrotuik, D. G. (2006). Establishing construct validity evidence for the sport multidimensional perfectionism scale. *Psychology of Sport and Exercise, 7*(1), 57–79. http://dx.doi.org/10.1016/j.psychsport.2005.04.003

Hajerrouit, S. (2010). Developing web-based learning resources in school education: A user-centred approach. *Issues in Information Science and Information Technology, 6,* 115–135. Retrieved from http://www.ijello.org/Volume6/IJELLOv6p115-135Hadjerrouit688.pdf

Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3–38). Mahwah, NJ: Lawrence Erlbaum Associates.

Hannafin, M. J., & Kim, M. C. (2003). In search of a future: A critical analysis of research on web-based teaching and learning. *Instructional Science, 31*(4–5), 347–351.

Hasan, L. (2014). Evaluating the usability of educational websites based on students' preferences of design characteristics. *International Arab Journal of e-Technology*, *3*(3). Retrieved January 2014 from http://www.iajet.org/

Henke, H. (2001). *Evaluating web-based instruction design*. Human-Computer interaction course project paper. Retrieved from http://www.chartula.com/evalwbi.pdf

Hooper, L. M., Marotta, S. A., & Depuy, V. (2009). A confirmatory factor analytic study of the Posttraumatic Growth Inventory among a sample of racially diverse college students. *Journal of Mental Health*, *18*(4), 335–343. http://dx.doi.org/10.1080/09638230802522502

Horton, W. (2000). *Designing web-based training: How to teach anyone anything anywhere anytime*. Hoboken, NJ: John Wiley and Sons.

Hsu, C. M., Yeh, Y. C., & Yen, J. (2009). Development of design criteria and evaluation scale for web-based learning platforms. *International Journal of Industrial Ergonomics, 39*(1), 90–95. http://dx.doi.org/10.1016/j.ergon.2008.08.006

Jackson, G. B. (2000). How to evaluate educational software and websites. *TechKnowLogia, 129,* 57–58. Retrieved from http://http://www.techknowlogia.org/

Karasar, N. (2010). *Bilimsel araştırma yöntemi* [Scientific research method]. İstanbul, Turkey: Nobel Yayıncılık.

Kline, P. (1994). *An easy guide to factor analysis*. London, UK: Routledge.

Kline, R. B. (2005). *Principles and practices of structural equation modeling*. New York, NY: Guilford Press.

Küçük, S., Yılmaz, R., Baydaş, Ö., & Göktaş, Y. (2015). Ortaokullarda artırılmış gerçeklik uygulamaları tutum ölçeği: Geçerlik ve güvenirlik çalışması [Augmented reality applications attitude scale in secondary schools: Validity and reliability study]. *Education & Science, 39*(176), 383–392. http://dx.doi.org/10.15390/EB.2014.3590

Lee, D. Y., & Lehto, M. R. (2013). User acceptance of YouTube for procedural learning: An extension of the Technology Acceptance Model. *Computers & Education*, *61*, 193–208. http://dx.doi.org/10.1016/j.compedu.2012.10.001

Li, B. (2016). Identifiable but Changeable: Capturing the features of teacher identity. *International Journal Social Science & Education, 6*(2), 225–234. Retrieved from http://ijsse.com/

Liu, G. Z., Liu, Z. H., & Hwang G. J. (2011). Developing multi-dimensional evaluation criteria for English learning websites with university students and professors. *Computers & Education*, *56,* 65–79. http://dx.doi.org/10.1016/j.compedu.2010.08.019

Maruyama, G. M. (1998). *Basics of structural equation modelling*. Thousand Oaks, CA: Sage.

McDonald, R. (1985). *Factor analysis and related methods.* Hillsdale, NJ: Erlbaum.

McMillan, J. H., & Schumacher, S. (2006). *Research in education: Evidence-based inquiry.* Boston, MA: Pearson.

Merrill, M. D., & Goodman, R. I. (1972). *Selecting instructional strategies and media.* Washington, DC: National Special Media Institutes.

Meunier, J-C., & Roskam, I. (2009). Validation of the preschool and primary school form of a questionnaire assessing parents' childrearing behavior. *Journal of Clinical Child & Adolescent Psychology*, *38*(1), 166–175. http://dx.doi.org/10.1080/15374410802575370

Özdamar, K. (2000). *Paket programlar ile istatistiksel veri analizi: Çok değişkenli analizler* [Statistical data analysis with packaged software: Multivariate analysis]. Eskişehir, Turkey: Kaan Kitapevi.

Öztekin, A., Kong, Z. J., & Uysal, O. (2010). UseLearn: A novel checklist and usability evaluation method for elearning systems by criticality metric analysis. *International Journal of Industrial Ergonomics*, *40*(4), 455–469. http://dx.doi.org/10.1016/j.ergon.2010.04.001

Pahl, C. (2003). Managing evolution and change in web-based teaching and learning environments. *Computers & Education, 40*(2), 99–114. http://dx.doi.org/0.1016/S0360-1315(02)00100-8

Öztürk, M. A. (2011). Confirmatory factor analysis of the educators' attitudes toward educational research scale. *Educational Sciences: Theory & Practice*, *11*, 737–747.

Reiser, R. A., & Gagne, R. M. (1983). *Selecting media for instruction*. Englewood Cliffs, NJ: Educational Technologies Publications.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Psychological Research*, *8*(2), 23–74. Retrieved from http://user.uni-rankfurt.de/~kscherm/schermelleh/mpr_Schermelleh.pdf

Schumacker, R. E., & Lomax, R. E. (1996). *A beginner's guide to structural equation modeling.* New Jersey, NJ: Lawrence Erlbaum Associates.

Silius, K., & Tervakari, A-M. (2003, May). A*n evaluation of the usefulness of web-based learning environments. The evaluation tool into the portal of Finnish virtual university*. Paper presented at mENU 2003 International Conference on University Networks and E-learning, Valencia, Spain. Retrieved from http://www.mit.jyu.fi/OPE/kurssit/TIES462/Materiaalit/Silius_Tervakari.pdf

Smith, P., & Rogan, T. J. (1993). *Instructional Design*. New York, NY: Macmillan.

Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik* [Reliability and validity in the social and behavioral measures]. Ankara, Turkey: Seçkin Yayınevi.

Tavşancıl, E. (2005). *Tutumların ölçülmesi ve SPSS ile veri analizi* [Measurement of attitudes and data analysis with SPSS]. Ankara, Turkey: Nobel Yayınları.

U.S. Dept. of Health and Human Services (2006). *The research-based web design & usability guidelines*, *enlarged/expanded edition*. Washington, DC: U.S. Government Printing Office.

Van Prooijen, J. W., & Van Der Kloot, W. A. (2001). Confirmatory analysis of explanatively obtained factor structures. *Educational and Psychological Measures*, *61*(5), 777–792. http://dx.doi.org/10.1177/00131640121971518

Yang, Y. T. C., & Chan, C. Y. (2008). Comprehensive evaluation criteria for English learning websites using expert validity surveys. *Computers & Education*, *51,* 403–422. http://dx.doi.org/10.1016/j.compedu.2007.05.011

Yiğit, Y., Yıldırım, S., & Özden, Y. (2000). Web tabanlı internet öğreticisi: Bir durum çalışması [Web based internet tutorial: A case study]. *Hacettepe University Journal of Education, 19,* 166–176.

Yurdugül, H. (2006). The comparison of reliability coefficients in parallel, tau-equivalent, and congeneric measurements. *Ankara University Journal of Faculty of Educational Sciences*, *39*(1), 15–37. http://dx.doi.org/10.1501/Egifak_0000000127

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's, β and McDonald's ω: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 1–11. http://dx.doi.org/10.1007/s11336-003-0974-7