ELECTRONIC ASSESSMENT ANXIETY SCALE: DEVELOPMENT, VALIDITY AND RELIABILITY

Dr. Osman TAT

ORCID: 0000-0003-2950-9647
Faculty of Education
Van Yuzuncu Yil University
Van, TURKIYE

Dr. Abdullah Faruk KILIC

ORCID: 0000-0003-3129-1763
Faculty of Education
Trakya University
Edirne, TURKIYE

Received: 24/10/2023 **Accepted:** 27/11/2023

ABSTRACT

The widespread availability of internet access in daily life has resulted in a greater acceptance of online assessment methods. E-assessment platforms offer various features such as randomizing questions and answers, utilizing extensive question banks, setting time limits, and managing access during online exams. Electronic assessment enables real-time monitoring, customization, and scalability of feedback, benefiting students, academic staff, and administrative personnel. However, students encounter specific challenges in the electronic assessment environments. These challenges include limited control over test settings and the isolated nature of taking exams without peers. Furthermore, the technological proficiency of both instructors and students, along with resource constraints (computers, mobile devices, internet), can impede the effective utilization of these assessment tools. Technical issues like slow internet connection or disconnections can have significant consequences, especially in online exams, posing difficulties for corrections. The main goal of this study is to develop a Likert-type scale capable of measuring anxiety related to technical issues, social isolation, and the test interface experienced in e-assessment contexts. The study consists of two separate groups: the first group comprising 359 participants and the second group consisting of 356 participants. Both groups include undergraduate and pedagogical formation certificate program students from a university in the Eastern Anatolia Region of Turkiye. Construct validity was assessed through exploratory and confirmatory factor analyses. Item parameters were examined using item analysis based on classical test theory. As a result of the study, a two-factor scale structure comprising 21 items measuring social and technical anxiety was developed. These two dimensions account for 59.89% of the total variance. The Cronbach's alpha coefficient for the entire scale was 0.93, the McDonald's omega coefficient was 0.93, and the construct reliability was 0.99.

Keywords: Anxiety, e-assessment, social isolation, technical problems.

INTRODUCTION

In the realm of education, assessment serves not only for grading but also as a means to identify the strengths and weaknesses of education, observe progress, and provide valuable feedback for the purpose of designing an appropriate learning plan for both teachers and students. Compared to other components, assessment has a greater impact on the nature and efficacy of the learning environment in this capacity (Ashton & Wood, 2006; Dermo, 2009; Hricko & Howell, 2006; Moore & Anderson, 2003; Robles & Braathen, 2002). The use of different technology instruments in educational settings has increased significantly in recent years.

These tools involve virtual learning environments, simulation software, virtual experiments, visualization of complex models, as well as communication tools including email, electronic forums, and instant messaging systems (Admiraal et al., 2014; Fill & Ottewill, 2006). The utilization of these tools has facilitated a shift in the learner profile in distance education, resulting in individuals who take more responsibility, engage in greater reflection, and show more independence (Donnelly & McSweeney, 2009; Heafner et al., 2015; Kidd & Morris, 2017; Palloff & Pratt, 2003). The changing nature of technology and the characteristics of learners have necessitated the development of new alternatives to conventional assessment methods. These new methods, despite being implemented in various contexts, have fundamentally brought techniques often referred to as e-assessment, online assessment, or remote assessment into our lives. "The term "e-assessment" is defined as the use of computers that have existed since the 1970s or modern mobile devices for the purpose of managing and evaluating exams. E-assessment consists many forms that can be conducted using both online and offline modes. The growing accessibility of internet usage in everyday life has contributed to an increased degree of willingness towards online assessment. The notion of conducting assessments online is interesting because it has the ability to solve difficulties such as test time, location, and cost (Graff, 2003; Hricko & Howell, 2006; Wandeler & Emmenegger, 2010). Moreover, online e-assessments are able to contribute to the examination and enhancement of online learning processes and outcomes (Daly et al., 2010). However, for online assessment to work well, decisions must be made during the planning phase about feedback systems, standards and rules, measurement tools, types of tests, and how exams are given that are in line with the behaviors that are desired. E-assessments created in this manner can be as successful as assessments performed in a classroom environment when conducted under supervision (Buchanan, 2004; Sanchez-Cabrero et al., 2021). Because e-assessment requires a higher degree of skill than traditional paper-and-pencil approaches, it may not be appropriate in the early stages of schooling. However, because today's pupils use smart phones, tablets, and laptop computers in their learning activities from a young age, it may be stated that such devices can also be used for evaluation reasons from a young age. E-assessment, which is not limited by age, can also be used not only for certain subjects or courses, but for almost all disciplines (Sainsbury & Benton, 2011; Sozen & Guven, 2019).

When compared to assessment given in a classroom context, evaluating learning in an online environment where students and teachers do not share the same physical space has a different meaning (Vonderwell & Boboc, 2013). Hence, while examinations and tests can be useful instruments for evaluating specific areas of learning, they should not be used as the main tools for assessing online programs. To get an accurate evaluation of the performance of online learners, it is important to use a range of assessment procedures together with traditional methods (Heafner et al., 2015; Palloff & Pratt, 2009). Discussions, group activities, and self-assessment assignments, in which students assume responsibility for their own learning, are widely used in the online environment to measure progress and development. The e-assessment mechanism, which includes approaches such as discussions and self-assessment, is classified into four major groups in the literature: peer, teacher, self-assessment and automated (Buchanan, 2004; Rovai, 2000). Each of these four main categories contains a wide range of assessment tools that may be used for formative and summative evaluation. Table 1 lists the measurement tools that may be utilized in formative and summative evaluation.

Table 1. Measurement Tools Utilized in e-Assessment

Uses	Measurement Tool
Submission of items for assessment	Essays: discursive, descriptive, analytical
	Reports: CBL, PBL
	Reviews: critical, analytical
	Media: image, audio, video, presentation
Automated Assessment	Multiple choice
	Short answer
Quizzes	Calculation
	Matching
Multimedia	Fill blanks
	True/false
	Matching
	Drag and drop
	Simulations
Online Discussions	Forums: case analysis, project development
	Debates
	Allocated roles: lead, summarize, provoke
	Role plays
Web Publishing	E-portfolios
	Webpages: blogs, wikis
	Shared documents: Google Documents

Source: (Benson, 2010)

E-assessment has various advantages for assessing learning. Most e-assessment platforms, for example, include capabilities such as randomizing questions and answers, applying huge question banks, setting up time limitations, and managing access in online examinations. Furthermore, e-assessment enables real-time monitoring, customization, and scalability of feedback to students, academic staff, and administrative staff (Dennick et al., 2010; Hricko & Howell, 2006; Jiao, 2015). The success of e-assessment, on the other hand, is dependent on a variety of criteria, including the sort of assessment used, the degree of feedback provided and student participation. A constructivist approach that stresses cooperation, inquiry, and mentoring is advocated in this regard for maximizing the benefits of e-assessment (Benson, 2010; Ramsaran-Fowdar et al., 2011; Rovai, 2000). When combined with real-time feedback, tasks allowing projects, portfolios, self-evaluation, and peer assessment in line with constructivist principles are especially successful instruments in e-assessment (Gaytan & McEwen, 2007). Nevertheless, e-assessment has a number of disadvantages, including software, hardware, and personnel costs, potential risks deriving from external threats and internal security policies, and technical issues that may pose a failure risk. These challenges must be considered for a successful implementation (Dennick et al., 2010; Hricko & Howell, 2006; Jiao, 2015).

The performance of individuals in tests has a significant role in shaping their educational possibilities and life circumstances since the beginning of the 20th century. Thus, the investigation of emotional reactions displayed by students during academic assessments has been a topic of academic interest (Pekrun et al., 2004). One of these emotional reactions is a psychological construct known in the literature as test or exam anxiety. The literature defines test anxiety as an extreme state of anxiety and stress related to assessment. Test anxiety, an important phenomenon that has been extensively studied from multiple viewpoints (Sarason, 1984), is known to have a negative influence on some students' test performance (Powers, 2001; Wine, 1971).

In some studies, which are investigate the level of test anxiety and its relationship with performance in e-assessment environments, it has been indicated that test anxiety in online assessments is either not significantly different from that in traditional classroom exams or slightly lower (Cassady & Gridley, 2005; Powers, 2001; Sanchez-

Cabrero et al., 2021; Stowell & Bennett, 2010). Cassady and Gridley (2005), on the other hand, stated in their study that students with less online experience are less likely to experience similar levels of comfort, and that anxiety levels are expected to rise in students who have not been systematically exposed to computer-based instructional processes. According to Stowell and Bennett (2010), e-assessment lowers anxiety in persons with severe claustrophobia but increases test anxiety in students with low classroom anxiety. The entrance procedure into the e-assessment system, trust in the system's correct functioning, the presence of proctors, or a lack of knowledge with online assessments are all mentioned as possible causes (Stowell & Bennett, 2010). Similarly, it has been found that students with high test anxiety perform worse when questions administered in different order compared to when the order of question administration is fixed. This suggesting that order of items may lead to additional challenges on students (Ortner & Caspers, 2011).

PURPOSE OF THE STUDY

Students encounter specific difficulties in the context of e-assessment, as opposed to typical tests delivered in a physical classroom setting. These difficulties include having little control over the test settings and taking the exam alone, without the presence of other pupils (Rovai, 2000). Furthermore, instructors' and students' technological abilities, as well as the restrictions of the suitable resources (computers, mobile devices, internet), might have a negative impact on the successful usage of evaluation tools. Connection speed, disconnection, or other technical problems can have serious effects, especially in online exams, that can be difficult to correct (Brink & Lautenbach, 2011; Senel & Senel, 2021). It is known that automated proctoring with camera or integrated in the e-assessment system, as well as records kept by the system regarding student movements, can influence students' emotional states (Divjak et al., 2022).

While some studies in the literature suggest that e-assessment conditions do not impose an additional challenge on students beyond test anxiety, numerous other studies have found that students in e-assessment environments are affected by various factors. These factors include insufficient technical infrastructure, the possibility of internet and power outages, inexperience in using computers or mobile devices for assessment, variations in the order of items, online proctoring, the absence of someone to consult during the exam, and social isolation. Test anxiety, computer anxiety, and computer competency anxiety may not be adequate to fully describe the difficulties students have in online assessment settings. However, a measurement tool which can assess the underlying structure with mentioned indicators has not yet been developed. The goal of this study is to develop a Likert type scale that might measure anxiety regarding technical problems, social isolation, and the test interface experienced in e-assessment contexts. The development of a valid and reliable assessment tool could open the way for further studies about the features of this structure or its relationship with different factors. This study aims to answer four research questions in this context:

1) How is the structure of the e-assessment anxiety scale?
2) How is the level of validity of the e-assessment anxiety scale?
3) How is the level of reliability of the e-assessment anxiety scale?
4) How are the parameters of items in scale according to classical test theory?

METHOD

We used the survey model for scale development. The survey model aims to describe the group's status in terms of the measured feature as it is (Karasar, 2014). The survey model was preferred since this study aimed to develop a measurement tool to measure electronic assessment anxiety.

Participants

There are two different study groups in the study. The first one consists of 359 participants. We used this sample to reveal the structure of the scale. We used exploratory factor analysis for this purpose. The participants in the first group consisted of undergraduate and pedagogical formation certificate program students studying at a university in the Eastern Anatolia Region of Turkiye. Among the first group participants, 32.03% (n=115) were male and 67.97% (n=244) were female. When analyzed according to their departments, it was observed that there were students from 14 different departments, such as social studies teaching, coaching, literature, and economics, and the highest number of students was in the English language teaching department with 76 students.

The second study group consists of 356 participants. We used this data set for confirmatory factor analysis (CFA) to examine model-data fit. In other words, CFA with the data of the second group to investigate the extent to which the structure of the scale fits in a similar group. The second group consisted of undergraduate and pedagogical formation certificate program students studying at a university in the Eastern Anatolia Region of Turkiye. Among the participants in the second group, 30.33% (n = 108) were male and 69.67% (n = 248) were female. When analyzed in terms of departments, it was observed that most participants were in classroom education, 16.85% (n = 60), and geography, 14.61% (n = 52), and there were participants from 16 different departments.

Data Collection and Analysis

The data obtained from the first study group were analyzed in terms of EFA assumptions, and the data obtained from the second study group were analyzed in terms of CFA assumptions (Tabachnick & Fidell, 2019). There was no missing data in the first study group data set. Then, we examined the data set in terms of multicollinearity. We used tolerance (TV), condition index (CI), and variance inflation factor (VIF) statistics. The variance inflation factor should be less than 10, the tolerance value should be greater than 0.01, and the state index should be less than 30 (Kline, 2016; Tabachnick & Fidell, 2019) for the absence of multicollinearity. As a result of the analysis of the first data set, the VIF varied between 1.99-4.81, the TV varied between 0.21-0.5, and the CI varied between 1.00-65.5. Accordingly, the CI indicates there may be a multicollinearity problem. Therefore, we examined the inter-item correlation matrix. So, the lowest correlation between the variables was 0.24, and the highest one was 0.82. Accordingly, there is no multicollinearity between the variables. It is generally stated that the correlations between variables should be greater than 0.90 for multicollinearity problems (Tabachnick & Fidell, 2019).

For this reason, since the TV and VIF are within the appropriate range and the largest correlations between the variables are less than 0.90, it is concluded that there is no multicollinearity between the variables. Mahalonobis distance was calculated for multivariate outliers, and the significant individuals at $\alpha=0.001$ level were removed from the data set. Accordingly, 31 individuals were identified as outliers and were removed from the data set. Thus, a data set of 328 individuals was formed. Mardia's (1970) multivariate skewness and kurtosis coefficient can be used for another assumption of multivariate normal distribution. Since it is stated that Mardia's multivariate skewness coefficient gives better results than other methods (Uysal & Kilic, 2022), this method was preferred. Mardia's multivariate skewness coefficient was 34698.96, statistically significantly different from 0 ($p<0.05$). Therefore, unweighted least squares (ULS), which is strong against the violation of the multivariate normality assumption (Zygmont & Smith, 2014), was used as a factor extraction method in the EFA.

For the second study group, data were collected from 356 individuals; this dataset has no missing data. When the data set was analyzed in terms of multicollinearity, it was observed that the TV was in the range of 0.34-0.63, the CI was in the range of 1.00-33.72, and the VIF was in the range of 1.59-2.94. As stated above, CI indicated multicollinearity. Therefore, we examined the inter-item correlation matrix. So, the lowest correlation between the variables was 0.19, and the highest one was 0.71. Accordingly, there is no multicollinearity between the variables. Mahalonobis distance was calculated for multivariate outliers, and the significant individuals at $\alpha=0.001$ level were removed from the data set. Accordingly, 14 individuals were identified as outliers and were excluded from the data set. Thus, a data set of 342 individuals was formed. Mardia's (1970) multivariate skewness coefficient was found to be 34698.96 and statistically significantly different from 0 ($p<0.05$). Therefore, we used robust maximum likelihood (MLR) as an estimation method that is strong enough to violate the normality assumption (Brown, 2015) in CFA.

Factor software (Lorenzo-Seva & Ferrando, 2023) was used for EFA analysis. We used Mplus for CFA (Muthen & Muthen, 2012). We used the second group data set for the reliability coefficients and item analysis based on classical test theory (CTT). Cronbach's Alpha was calculated by psych (Revelle, 2022), McDonald's omega and structure reliability was calculated by semTools (Jorgensen et al., 2022) and stratified Alpha was calculated by sirt (Robitzsch, 2021) package in R language (R Core Team, 2022).

Scale Development Process

We adopted the deductive method in the scale development process for item writing. In this method, items are written using previously developed scales and a large literature review (Hinkin, 2005; Morgado et al., 2018). According to this method, a 69-item candidate scale form was first prepared and sent to three measurement and evaluation experts with doctorate degrees in their field. Similarly, opinions were received from one expert in computer and instructional technology education and one Turkish language and literature expert with doctorate degrees in the candidate scale form. The candidate scale form was made into 48 items in line with the opinions received.

FINDINGS

In this section, we examine the e-assessment scale's factor structure, validity, reliability, and CTT item analysis.

The Structure of the E-assessment Scale

We conducted EFA with Pearson correlation matrix on the first study group data set. Direct Oblimin, one of the oblique rotation methods, was used as the rotation method. In EFA, it was observed that the KMO value was 0.95. Accordingly, the data set is suitable for factorization (Kaiser, 1970). Bartlett's test of sphericity revealed that the correlation matrix obtained from the data set statistically significantly differed from the unit matrix ($\chi 2(171) = 3682.2$, $p<0.01$).

Firstly, we examined the dimensionality of the 48-item scale form by parallel analysis (PA), Minimum Average Partial (MAP) analysis, HULL method, explained variance ratio and eigenvalues to understand the number of dimensions. PA, MAP, and HULL methods suggested a two-dimensional structure in this first analysis. When the explained variance ratios were analyzed, it was observed that the eigenvalue of the first factor was 26.06. The explained variance ratio was 54.2%, while the eigenvalue of the second factor was 3.42, and the explained variance ratio was 7.12%. The number of factors with eigenvalues greater than one was seven. In this case, the proposed number of dimensions, 2, was thought to be a more appropriate solution for the data. Because the variance explained by the third factor is 3.3%, these factors can usually be ignored. In order to create simpler and reproducible structures, factors with an explained variance ratio of less than 5% can be considered insignificant (Kilic, 2022).

We observed that some items had factor loadings greater than 1, while others had cross-loadings for two-dimensional solutions. In this case, we tried to exploratory bi-factor models. However, some items did not load on any specific factor. In this case, we tried bi-factor (S-1) (Burns et al., 2020) models, but similarly, some items were included with items measuring different traits. In this case, since the variance explained by the first dimension was as high as 54%, it was thought that there might be a secondary factor in the data set. In order to have a secondary level factor, there should be at least 3 sub-factors (Brown, 2015). In this direction, as stated in the literature section, a three-factor structure was created as anxiety arising from social relations, anxiety arising from technical reasons, and anxiety arising from the exam. As a result of the EFA conducted in this way, it was seen that a 3-factor structure could be obtained. However, as a result of second-order CFA, it was observed that the error variance of the exam sub-factor was negative. While investigating the reason for this situation, it was observed that the items were highly correlated since they included both physiological and psychological reactions. Namely, while one of the items was "The thought that the electricity cut during the assessment makes me nervous.", another item was "The thought that the electricity cut during the assessment causes me to experience physiological changes (e.g., rapid heartbeat and breathing, sweating, trembling, dry mouth)." It was understood that the applied individuals gave similar answers to both items because they could not distinguish between these physiological and psychological reactions. For this reason, the items with physiological reactions were removed from the scale. As a result, a 19-item scale form was obtained.

After the physiological expressions were removed from the scale, EFA was applied again to the 19-item scale form. In this case, it was observed that PA, MAP, and HULL methods suggested a 2-dimensional structure. When the eigenvalues and explained variance ratios were analyzed, it was observed that the eigenvalues of the first and second dimensions were 10.09 and 1.88, respectively. The explained variance ratios are 52.73% and 7.16%, respectively. The two dimensions explain 59.89% of the total variance. It can be said that this explained variance ratio is sufficient for social sciences (Buyukozturk, 2020). It was also observed that the number of factors with eigenvalues greater than 1 was two. Therefore, it was decided that the number of factors was 2, and the analysis continued. The EFA with direct oblimin rotation results are presented in Table 2.

Table 2. The EFA Results of the Electronic Assessment Anxiety Scale

Variable Name	Unrotated Solution		Rotated Solution		
	First Factor	Second Factor	First Factor	Second Factor	Communalities
i1	0.67	-0.22	0.64	0.08	0.49
i3	0.60	-0.09	0.43	0.22	0.37
i4	0.53	0.33	-0.18	0.75	0.39
i5	0.66	-0.15	0.54	0.17	0.46
i17	0.80	-0.37	0.92	-0.06	0.78
i18	0.78	-0.35	0.89	-0.05	0.73
i21	0.79	-0.19	0.66	0.18	0.65
i24	0.63	0.30	-0.08	0.76	0.49
i25	0.58	0.18	0.05	0.57	0.37
i26	0.72	0.03	0.33	0.44	0.52
i27	0.63	0.16	0.11	0.57	0.42
i31	0.68	0.20	0.07	0.65	0.50
i32	0.75	0.10	0.24	0.56	0.57
i34	0.67	0.30	-0.07	0.79	0.54
i36	0.77	0.12	0.23	0.59	0.61
i39	0.83	-0.06	0.51	0.38	0.68
i41	0.74	0.20	0.10	0.69	0.60
i42	0.80	0.09	0.29	0.57	0.64
i44	0.80	-0.32	0.85	0.01	0.74
Explained Variance Ratio	%52.73	%7.16			

The rotated results showed that eight items belong to the first factor and 11 items belong to the second one. The factor loadings of the items in the first-factor range between 0.33-0.92, and those in the second factor range between 0.44 and 0.79. Regarding cross-loading, the difference between the two-factor loadings is greater than 0.10 in all items that load on both factors. In addition, the inter-factor correlation was 0.73. Accordingly, it indicates that inter-factor correlation is high.

The first factor was named "Technical Anxiety," and the second factor was named "Social Anxiety" due to factor labeling performed by examining the items that loaded the first and second factors at the highest level. The items according to the dimensions are presented in Table 3.

Tablo 3. Scale Items*

Item	Technical Anxiety
i1	The thought that the electricity cut during the assessment makes me nervous.
i3	I feel nervous when I think my time will not be enough because I cannot type fast on the keyboard in open-ended questions.
i5	The possibility that the computer or mobile device I will use in the application will run out of charge makes me uneasy.
i17	The thought that the internet will be cut off if the application is online makes me uneasy.
i18	The thought of being unable to re-enter the system if the internet is disconnected and reconnected makes me uneasy.
i21	I feel uneasy because I cannot be sure my answers are saved.
i39	Knowing that my right to change my answers is limited makes me uneasy.
i44	The thought that my exam will be canceled due to the accidental closing of the application page increases my tension.
Item	**Social Anxiety**
i4	I worry that there is little opportunity to exchange opinions about the e-assessment task with my friends.
i24	In environments such as forums where I have to write my opinions, the thought that what I write will be seen by my friends makes me anxious.
i25	If my exam is scored automatically, I worry that it will be scored incorrectly.
i26	A countdown timer or stopwatch on my application screen makes me nervous.
i27	Taking a test alone makes me feel more nervous than in a classroom setting.
i31	A proctor monitoring my movements on my screen puts pressure on me.
i32	The possibility of forgetting my account information just before the application starts makes me nervous.
i34	I worry about being unable to find someone to consult about questions I think are wrong.
i36	If everyone is asked different questions in the practice, I feel uneasy thinking I will be asked difficult questions.
i41	I worry that my attention will be distracted in exams outside the classroom.
i42	I feel nervous because it is more difficult to understand what I read on the screen of my computer or mobile device.

*Note, *The items are only translated from Turkish to English.*

The technical anxiety dimension has items related to power failure, internet disconnection, or recording of answers (see Table 2). The social anxiety dimension relates to communication with friends or monitoring the screen by a supervisor. In this dimension, some items may not be thought to be related to social anxiety. However, they found a place in the social anxiety dimension based on the data obtained from the participants. For example, item 25 is ("If my exam is scored automatically, I worry that it will be scored incorrectly"). In this case, this item may have been included in the social dimension because he/she probably thought that their friends would see his/her scores. On the other hand, " The presence of a countdown timer or stopwatch on my application screen makes me nervous", item 26, which contains the statement " belongs to social anxiety. The fact that he/she was worried that his/her friends might see the countdown timer may have caused these items to be in the social anxiety dimension.

Confirmatory Factor Analysis of E-assessment Scale

According to the model-data fit indices obtained from CFA using the Pearson covariance matrix, the model-data fit is acceptable [$\chi 2$ (151) = 324.76, $p < 0.01$, $\chi 2/df$=2.15, RMSEA = 0.058 [90% C.I. 0.049-0.067], CFI = 0.931, TLI = 0.922, SRMR = 0.056]. The result $\chi 2/df$=2.15 suggests that the model-data fit is good (Anderson & Gerbing, 1984). The RMSEA value 0.058 indicates a good fit (Steiger, 2007). The recommended CFI value for indicating a good fit is at least 0.95. However, an acceptable fit can be indicated by a range of 0.90-0.95, as noted by Hu & Bentler (1998, 1999). Similarly, a TLI value in the range of 0.90-0.95 indicates a good fit. Based on the TLI and CFI values, it can be concluded that the model aligns with the data at an acceptable level. Since an SRMR value less than 0.08 is deemed acceptable (Hu & Bentler,

1999) and a value less than 0.05 indicates a good fit (Byrne, 1998), the model-data fit is acceptable. Figure 1 displays the path diagram resulting from the CFA. The inter-factor correlation is 0.73. The factor loadings of items in the technical anxiety dimension vary from 0.644 (i16) to 0.772 (i7). The factor loadings of items in the social anxiety dimension vary from 0.582 (i9) to 0.723 (i13). Thus, it can be inferred that items' commonly suggested factor loading to be greater than 0.40 (Howard, 2016) is also met.

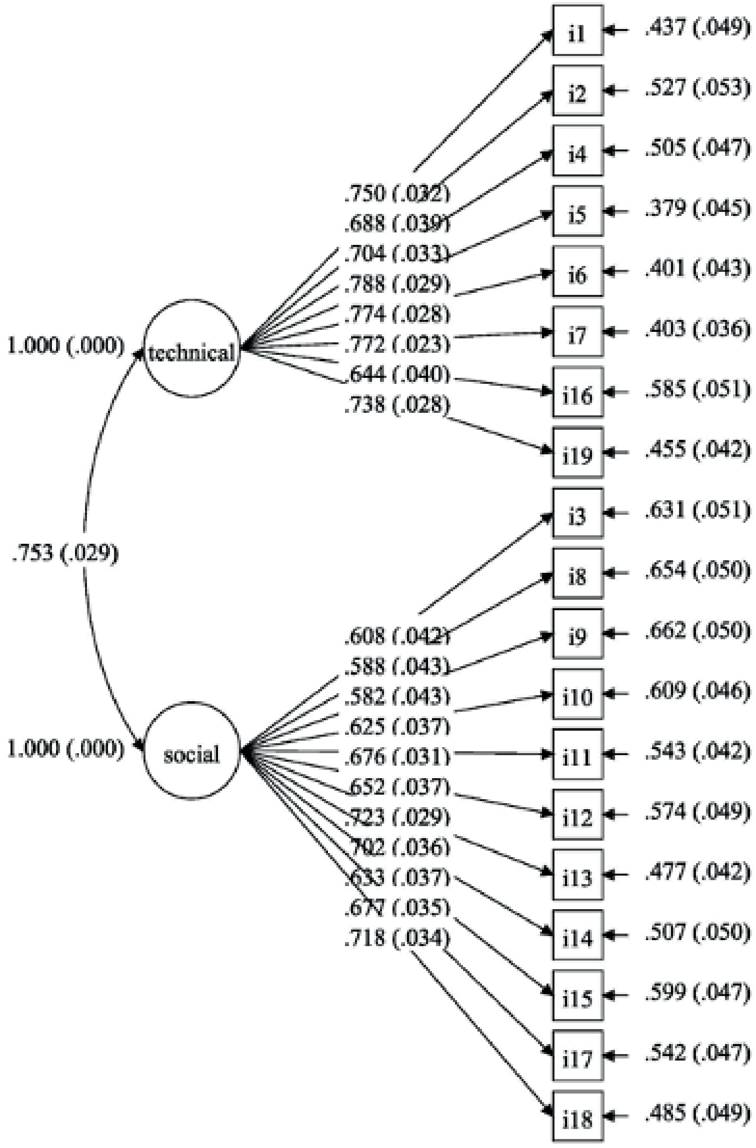


Figure 1. CFA path diagram

The Reliability of the E-assessment Scale

Since the model-data fit is acceptable in CFA, the E-assessment scale has a multi-dimensional structure. The reliability analysis included the calculation of Cronbach's Alpha, stratified Alpha, construct reliability and McDonald's omega coefficients. The Cronbach's alpha coefficient for the entire scale was 0.93, the McDonald's omega coefficient was 0.93, and construct reliability (Hair et al., 2019) was 0.99. Cronbach's alpha coefficient was 0.89 for technical anxiety and 0.89 for social anxiety. McDonald's omega coefficient was 0.89 and 0.90 for technical and social anxiety, respectively. The construct reliability was 0.99 for technical anxiety and 0.99 for social anxiety. The stratified alpha coefficient, calculated due to the scale's multidimensionality, was 0.93. Since Hair et al. (2019) recommended reliability coefficients for internal consistency to be greater than 0.70, it can be concluded that the internal consistency in this study is sufficient.

Item Analysis Based on Classical Test Theory for E-assessment Scale

In this section, we examined item analyses using classical test theory. Thus, Table 4 presents the mean, item endorsement rates, corrected item-total correlations, and lower and upper group analyses of the items according to classical test theory.

Table 4. The Results of Item Analysis for Scale†

Item	Mean	Standard deviation	Item endorsement rates	Skewness	Kurtosis	The average of lower and upper groups		t-value	Corrected item-total correlation
I1	3.76	1.02	0.75	-0.38	-0.72	2.88	4.58	-16.17**	0.61
I2	3.58	1.12	0.72	-0.56	-0.41	2.61	4.46	-15.18**	0.61
I3	2.86	1.09	0.57	0.08	-0.49	2.07	3.57	-10.74**	0.53
I4	3.59	1.18	0.72	-0.48	-0.75	2.60	4.54	-14.78**	0.60
I5	3.84	1.03	0.77	-0.60	-0.32	2.94	4.63	-15.22**	0.62
I6	3.94	1.02	0.79	-0.84	0.13	3.07	4.71	-13.58**	0.59
I7	3.60	1.10	0.72	-0.43	-0.70	2.56	4.48	-16.47**	0.66
I8	2.58	1.15	0.52	0.25	-0.76	1.84	3.40	-11.00**	0.51
I9	2.67	1.27	0.53	0.29	-0.88	1.86	3.65	-11.00**	0.52
I10	3.34	1.24	0.67	-0.34	-0.86	2.40	4.29	-13.47**	0.59
I11	2.79	1.30	0.56	0.18	-1.07	1.82	3.89	-14.15**	0.57
I12	3.10	1.29	0.62	-0.10	-1.03	2.14	4.16	-13.79**	0.56
I13	2.96	1.27	0.59	0.07	-1.03	1.99	4.25	-17.93**	0.67
I14	2.96	1.17	0.59	0.00	-0.74	2.04	3.95	-13.94**	0.62
I15	3.31	1.25	0.66	-0.30	-0.84	2.26	4.31	-14.10**	0.61
I16	3.43	1.13	0.69	-0.28	-0.75	2.44	4.32	-14.57**	0.63
I17	3.01	1.21	0.60	-0.04	-0.87	2.22	4.03	-12.45**	0.59
I18	3.13	1.21	0.63	-0.09	-0.88	2.17	4.23	-15.46**	0.65
I19	3.82	1.09	0.76	-0.64	-0.47	2.85	4.69	-14.46**	0.62
Sum Score	62.24	14.39	0.66	0.09	-0.38	44.77	80.13	-35.62**	-

*Note, *p<0,01 †Scale items are in Appendix 1.*

Table 4 presents the mean, standard deviation, skewness, kurtosis coefficients, and upper and lower group averages for each item and total test score. It also highlights the t-values obtained from the t-test. The corrected item-total correlations are presented in the last column of Table 4.

i1 had the highest mean (3.76), while i8 had the lowest (2.58). These items are presented as follows. i1 is " The thought that the electricity cut during the assessment makes me nervous." while i24 is " In environments such as forums where I have to write my opinions, the thought that what I write will be seen by my friends makes me anxious.".

The skewness coefficient ranges from -0.84 (i6) to 0.29 (i9). Similarly, the skewness of the total score was 0.09, so the scale items followed a normal distribution (Tabachnick & Fidell, 2019). Additionally, the kurtosis coefficient ranged from -1.07 (i11) to 0.13 (i6), and the kurtosis coefficient of the total score was -0.38. As such, the variables do not significantly deviate from normal distribution (Chou & Bentler, 1995). Therefore, we utilized a t-test to analyze the mean differences between the upper 27% and lower 27% groups. The mean values of both items and the total score were found to be statistically significantly different between the groups. Thus, it can be concluded that the individuals in the lower and upper groups differ statistically significantly regarding both items and the total score. When examining the corrected item-total correlations, it became evident that they generally exceeded the recommended cut-off point of 0.30 (Buyukozturk, 2020). It can be concluded that the items possess a satisfactory level of discrimination based on both the statistically significant difference between the means in the lower and upper groups and the sufficiency of the factor loadings.

DISCUSSIONS AND CONCLUSION

As a result of a study on e-assessment anxiety, a two-factor scale structure consisting of 21 items measuring social and technical anxiety was developed. The EFA (sample 1) and CFA (sample 2) data indicate that the e-assessment anxiety scale's scores are adequately valid. As a result of both exploratory and confirmatory factor analyses, the scores from the scale can be used to measure the entire scale (by summing all items) and two dimensions, technical and social anxiety, individually. Researchers studying e-assessment anxiety can create an e-assessment anxiety score by summing the item scores from the data collected using the scale. On the other hand, researchers seeking to conduct an in-depth study on a specific type of anxiety, such as technical anxiety or social anxiety, can gather responses to items in the respective dimensions. Despite the high correlation between the dimensions, obtaining a total score is meaningful. After analyzing the data obtained from the scale, it was noted that both the sub-dimensions of the scale and the scale as a whole demonstrated sufficient reliability in terms of internal consistency. Therefore, it can be stated that researchers conducting similar studies utilizing the e-assessment anxiety will have similar results.

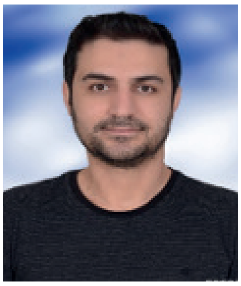
One limitation of this study is that the data was collected from the faculty of education and pedagogical formation students of a university in the Eastern Anatolia Region. Measuring the e-assessment concerns of individuals from different regions or age groups with this scale may lead to low validity results. Additionally, it is important to note that the results of this study are limited to this specific population and may not be generalizable to other populations. For this reason, researchers intending to use the scale should focus on using it at the university level. Conducting separate validity studies with data obtained from their samples can help establish the validity of the data.

The development of the e-assessment anxiety scale with a two-factor structure, focusing on separate measurements of social and technical anxiety, offers significant theoretical and practical implications. This scale provides researchers with the flexibility to measure overall e-assessment anxiety or focus on specific dimensions, aligning with the multifaceted nature of anxiety as a construct. Practically, the scale is useful in educational settings for identifying students with high levels of e-assessment anxiety, aiding in the development of targeted interventions to improve their academic performance. The scale's validation in university settings also indicates its potential applicability in similar educational environments.

However, the study's limitation, particularly its focus on a specific population from the Eastern Anatolia Region, highlights the need for further validation across diverse demographic groups. This would enhance the scale's generalizability and utility in various cultural and educational contexts. The scale's development also contributes to the broader understanding of anxiety assessment, emphasizing the complexity of anxiety experiences and the importance of considering different anxiety dimensions in research and clinical practice for a more comprehensive evaluation and targeted interventions.

Authors' Note: This study was supported by the Scientific Research Projects Coordination Unit of Van Yuzuncu Yil University under the project number SYD-2021-9537.

BIODATA and CONTACT ADDRESSES of AUTHORS



Osman TAT is an Assist. Prof. Dr. of Educational Measurement and Evaluation at the Faculty of Education, Van Yuzuncu Yil University. Dr. Tat gained his Ph.D. in Department of Educational Measurement and Assessment at Hacettepe University, in January 2020. His academic interest areas are item response theory, hierarchical linear models, exploratory and confirmatory factor analyses, structural equation modeling, data mining and text mining. He has 8 journal articles published in international indexes, 1 international book, 2 international book chapters and other national and international articles, papers submitted to international conferences.

Osman TAT
Department of Educational Science, Education Faculty
Address: Van Yuzuncu Yil University, 65080, Van, Turkiye
Phone: +90 0432 444 50 65-21781
E-mail: osmantat@yyu.edu.tr

Abdullah Faruk KILIC is an Assoc. Prof. Dr. of Educational Measurement and Evaluation at the Education Faculty, Trakya University. Dr. Kilic gained his Ph.D. in Educational Measurement and Evaluation at October 2019. His academic interest areas are dimensionality, exploratory and confirmatory factor analysis, item weighting, and reliability. He has over ten journal articles published in international indexes, international book chapters, national and international articles, and papers submitted to international meetings.

Abdullah Faruk KILIC
Department of Educational Science, Education Faculty
Address: Trakya University, 22030, Edirne, Turkiye
Phone: +90 284 212 08 08 - 1278
E-mail: afarukkilic@trakya.edu.tr

REFERENCES

Admiraal, W., Janssen, T., Huizenga, J., Kranenburg, F., Taconis, R., & Corda, A. (2014). E-Assessment of Student-Teachers' Competence as New Teachers. Turkish Online Journal of Educational Technology - TOJET, 13(4), 21–29. Retrieved from https://eric.ed.gov/?q=e-assessment&pr=on&ft=on&id=EJ1043184

Ashton, H., & Wood, C. (2006). Use of Online Assessment to Enhance Teaching and Learning: The PASS-IT Project. European Educational Research Journal, 5(2), 122–130. https://doi.org/10.2304/eerj.2006.5.2.122

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. Psychometrika, 49(2), 155–173. https://doi.org/10.1007/BF02294170

Bahar, M., & Asil, M. (2018). Attitude towards e-assessment: influence of gender, computer usage and level of education. Open Learning: The Journal of Open, Distance and E-Learning, 33(3), 221–237. https://doi.org/10.1080/02680513.2018.1503529

Benson, R. (2010). Online learning and assessment in higher education: A planning guide. Oxford: Chandos Publishing.

Brink, R., & Lautenbach, G. (2011). Electronic assessment in higher education. Educational Studies, 37(5), 503–512. https://doi.org/10.1080/03055698.2010.539733

Brown, T. A. (2015). Confirmatory factor analysis for applied research (2nd ed.). The Guilford.

Buchanan, E. A. (2004). Online Assessment in High Education, 163–176. https://doi.org/10.4018/978-1-59140-178-0.CH008

Burns, G. L., Geiser, C., Servera, M., Becker, S. P., & Beauchaine, T. P. (2020). Application of the Bifactor S – 1 Model to Multisource Ratings of ADHD/ODD symptoms: An appropriate bifactor model for symptom ratings. Journal of Abnormal Child Psychology, 48(7), 881–894. https://doi.org/10.1007/s10802-019-00608-4

Buyukozturk, S. (2020). Sosyal bilimler icin veri analizi el kitabi: Istatistik, arastirma deseni, SPSS uygulamalari ve yorum (28. Baski). Pegem Akademi.

Byrne, B. M. (1998). Structural equation modeling with lisrel, prelis, and simplis: Basic concepts, applications, and programming. Psychology Press. https://doi.org/10.4324/9780203774762

Cassady, J. C., & Gridley, B. E. (2005). The Effects of Online Formative and Summative Assessment on Test Anxiety and Performance. The Journal of Technology, Learning and Assessment, 4(1). Retrieved from https://ejournals.bc.edu/index.php/jtla/article/view/1648

Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 37–55). Sage.

Daly, C., Pachler, N., Mor, Y., & Mellar, H. (2010). Exploring formative e-assessment: using case stories and design patterns. Assessment & Evaluation in Higher Education, 35(5), 619–636. https://doi.org/10.1080/02602931003650052

Dennick, R., Wilkinson, S., & Purcell, N. (2010). Online eAssessment. AMEE guide. Assessment: Vol. 39. Dundee: AMEE.

Dermo, J. (2009). e-Assessment and the student learning experience: A survey of student perceptions of e-assessment: A survey of student perceptions of e-assessment. British Journal of Educational Technology, 40(2), 203–214. https://doi.org/10.1111/j.1467-8535.2008.00915.x

Divjak, B., Zugec, P., & Pazur Anicic, K. (2022). E-assessment in mathematics in higher education: a student perspective. International Journal of Mathematical Education in Science and Technology, 1–23. https://doi.org/10.1080/0020739X.2022.2117659

Donnelly, R., & McSweeney, F. (2009). Applied e-learning and e-teaching in higher education. Hershey PA: Information Science Reference.

Fill, K., & Ottewill, R. (2006). Sink or swim: taking advantage of developments in video streaming. Innovations in Education and Teaching International, 43(4), 397–408. https://doi.org/10.1080/14703290600974008

Gaytan, J., & McEwen, B. C. (2007). Effective Online Instructional and Assessment Strategies. American Journal of Distance Education, 21(3), 117–132. https://doi.org/10.1080/08923640701341653

Graff, M. (2003). Cognitive Style and Attitudes Towards Using Online Learning and Assessment Methods. Electronic Journal of E-Learning, 1(1).

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate data analysis (8th ed.). Cengage.

Hinkin, T. R. (2005). Scale development principles and practices. In R. A. Swanson & E. F. Holton III (Eds.), Research in organizations: Foundations and methods of inquiry (pp. 161–179). Berret-Koehler.

Heafner, T. L., Hartshorne, R., & Petty, T. M. (2015). Exploring the effectiveness of online education in K-12 environments. Advances in early childhood and K-12 education (AECKE) book series. Hershey PA: Information Science Reference.

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? International Journal of Human-Computer Interaction, 32(1), 51–62. https://doi.org/10.1080/10447318.2015.1087664

Hricko, M., & Howell, S. L. (2006). Online assessments and measurement: Foundations and challenges / Mary Hricko, Scott L. Howell [editors]. Hershey, Pa., London: Information Science Publishing.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychological Methods, 3(4), 424–453. https://doi.org/10.1037/1082-989X.3.4.424

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 37–41. https://doi.org/10.1080/10705519909540118

Jiao, H. (2015). Enhancing students' engagement in learning through a formative e-assessment tool that motivates students to take action on feedback. Australasian Journal of Engineering Education, 20(1), 9–18. https://doi.org/10.7158/D13-002.2015.20.1

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). semTools: Useful tools for structural equation modeling (0.5-6) [Computer software]. https://CRAN.R-project.org/package=semTools

Karasar, N. (2014). Bilimsel arastirma yontemi (26. Baski). Nobel Yayinlari.

Kaiser, H. F. (1970). A second generation little jiffy. Psychometrika, 35(4), 401-415. https://doi.org/10.1007/BF02291817

Kilic, A. F. (2022). Acimlayici faktor analizinde boyut sayisina karar verme: Yontemlere kisa bir bakis. Pamukkale University Journal of Social Sciences Institute, 51(Ozel sayi 1), O305-O318.

Kidd, T. T., & Morris, L. R. (Eds.) (2017). Advances in the educational technologies and instructional design (AETID) book series. Handbook of Research on Instructional Systems and Educational Technology. Hershey PA: Information Science Reference.

Kline, R. B. (2016). Principle and practice of structural equation modeling (4th ed.). The Guilford.

Lorenzo-Seva, U., & Ferrando, P. J. (2023). Factor (Version 12.04.02) [Computer software]. Universitat Rovira i Virgili.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. Biometrika, 57(3), 519-530.

McDonald, A. S. (2001). The Prevalence and Effects of Test Anxiety in School Children. Educational Psychology, 21(1), 89–101. https://doi.org/10.1080/01443410020019867

Moore, M. G., & Anderson, W. G. (2003). Handbook of distance education. Mahwah, N.J., London: Lawrence Erlbaum Associates.

Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2018). Scale development: Ten main limitations and recommendations to improve future research practices. Psicologia: Reflexão e Critica, 30(1), 3. https://doi.org/10.1186/s41155-016-0057-1

Muthen, L. K., & Muthen, B. O. (2012). Mplus statistical modeling software: Release 7.0. Los Angeles, CA: Muthen & Muthen.

Ortner, T. M., & Caspers, J. (2011). Consequences of Test Anxiety on Adaptive Versus Fixed Item Testing. European Journal of Psychological Assessment, 27(3), 157–163. https://doi.org/10.1027/1015-5759/a000062

Palloff, R. M., & Pratt, K. (2003). The virtual student: A profile and guide to working with online learners / Rena M. Palloff, Keith Pratt. The Jossey-Bass higher and adult education series. San Francisco, Great Britain: Jossey-Bass. Retrieved from https://eric.ed.gov/?id=ED474182

Palloff, R. M., & Pratt, K. (2009). Assessing the online learner: Resources and strategies for faculty (1st ed.). Jossey-Bass guides to online teaching and learning. San Francisco: Jossey-Bass.

Pekrun, R., Goetz, T., Perry, R. P., Kramer, K., Hochstadt, M., & Molfenter, S. (2004). Beyond test anxiety: Development and validation of the test emotions questionnaire (TEQ). Anxiety, Stress & Coping, 17(3), 287–316. https://doi.org/10.1080/10615800412331303847

Powers, D. E. (2001). Test Anxiety and Test Performance: Comparing Paper-Based and Computer-Adaptive Versions of the Graduate Record Examinations (Gre©) General Test. Journal of Educational Computing Research, 24(3), 249–273. https://doi.org/10.2190/680W-66CR-QRP7-CL1F

Putwain, D. W. (2008). Deconstructing test anxiety. Emotional and Behavioral Difficulties, 13(2), 141–155. https://doi.org/10.1080/13632750802027713

Ramsaran-Fowdar, R. R., Baguant, P., & Fowdar, S. (2011). A Critical Analysis of e-Assessment with Particular Emphasis on the use of Different Types of Online Quizzes. The International Journal of Learning: Annual Review, 18(1), 191–202. https://doi.org/10.18848/1447-9494/CGP/v18i01/47378

R Core Team. (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.r-project.org/

Revelle, W. (2022). psych: Procedures for psychological, psychometric, and personality research (2.2.9) [Computer software]. Northwestern University. https://cran.r-project.org/package=psych

Robitzsch, A. (2021). sirt: Supplementary Item Response Theory Models. https://cran.r-project.org/package=sirt

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. Personality and Individual Differences, 42(5), 893–898. https://doi.org/10.1016/j.paid.2006.09.017

Robles, M., & Braathen, S. (2002). Online Assessment Techniques. Delta Pi Epsilon Journal, 44(1), 39–49. Retrieved from https://eric.ed.gov/?id=EJ646729

Rovai, A. P. (2000). Online and traditional assessments: what is the difference? The Internet and Higher Education, 3(3), 141–151. https://doi.org/10.1016/S1096-7516(01)00028-8

Sainsbury, M., & Benton, T. (2011). Designing a formative e-assessment: Latent class analysis of early reading skills. British Journal of Educational Technology, 42(3), 500–514. https://doi.org/10.1111/j.1467-8535.2009.01044.x

Sanchez-Cabrero, R., Casado-Perez, J., Arigita-Garcia, A., Zubiaurre-Ibanez, E., Gil-Pareja, D., & Sanchez-Rico, A. (2021). E-Assessment in E-Learning Degrees: Comparison vs. Face-to-Face Assessment through Perceived Stress and Academic Performance in a Longitudinal Study. Applied Sciences, 11(16), 7664. https://doi.org/10.3390/app11167664

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. Journal of Personality and Social Psychology, 46(4), 929–938. https://doi.org/10.1037//0022-3514.46.4.929

Senel, S., & Senel, H. C. (2021). Remote Assessment in Higher Education during COVID-19 Pandemic. International Journal of Assessment Tools in Education, 8(2), 181–199. Retrieved from https://eric.ed.gov/?id=EJ1303929

Sozen, E., & Guven, U. (2019). The Effect of Online Assessments on Students' Attitudes Towards Undergraduate-Level Geography Courses. International Education Studies, 12(10), 1. https://doi.org/10.5539/ies.v12n10p1

Stowell, J. R., & Bennett, D. (2010). Effects of Online Testing on Student Exam Performance and Test Anxiety. Journal of Educational Computing Research, 42(2), 161–171. https://doi.org/10.2190/EC.42.2.b

Vonderwell, S. K., & Boboc, M. (2013). Promoting Formative Assessment in Online Teaching and Learning. TechTrends, 57(4), 22–27. https://doi.org/10.1007/s11528-013-0673-x

Wandeler, B., & Emmenegger, B. (2010). Soziokulturelle Animation: Professionelles Handeln zur Forderung von Zivilgesellschaft, Partizipation und Kohasion. Luzern: Interact.

Wine, J. (1971). Test anxiety and direction of attention. Psychological Bulletin, 76(2), 92–104. https://doi.org/10.1037/h0031332

Tabachnick, B. G., & Fidell, L. S. (2019). Using multivariate statistics (7th ed.). Pearson.

Uysal, I., & Kilic, A. F. (2022). Cok degiskenli normallik: Testler ne kadar dogru ne kadar guclu? In F. Nayir & S. Poyrazli (Eds.), Egitim Bilimlerinde Guncel Arastirmalar. Ani.

Zygmont, C., & Smith, M. R. (2014). Robust factor analysis in the presence of normality violations, missing data, and outliers: Empirical questions and possible solutions. The Quantitative Methods for Psychology, 10(1), 40-55. https://doi.org/10.20982/tqmp.10.1.p040