AFRICAN JOURNAL OF THEORY AND PRACTICE OF EDUCATIONAL ASSESSMENT (AJTPEA)

Hard Print ISSN: 2545-5834
e-Print ISSN: 2992-362X

VOL. 12, JUNE, 2023

Development and Construct Validation of Attitude Scale to Computer-Based Testing using Undergraduates in Katsina State

Abdullahi Adamu Dan'inna PhD.

Department of Science and Vocational Education, Umaru Musa Yar'adua University, Katsina, Katsina State, Nigeria

Abstract

This study aimed at developing and validating Attitude Scale to Computer-Based Testing. Instrumentation design was used. Three objectives were established, and three research questions were generated and addressed. The population of the study comprised 34,050 undergraduates in the three universities in Katsina state. The sample comprised 1,000 subjects, derived using multistage procedure. The developed scale was used to generate data which were analysed using Factor Analysis. The findings for Exploratory Factor Analysis (EFA) revealed that, 67 items were developed initially, but only 50 were retained in the modified version. Five factors were restricted, which resulted in only 39 items. Also, Kaiser-Meyer-Olkin Measure of Sampling Adequacy *(KMO/MSA)* = .851 (χ^2 =17346.330, *df* = 1225, *p* =.000), indicating excellent sample size for factor analysis. Finally, Confirmatory Factor Analysis (CFA) was run and the data fit the model poorly, but after some necessary modifications, the final model, with only 24 items, had acceptable fit indices. For convergent validity, $\lambda \geq 0.64$, Average Variance Extracted *(AVE)* $\geq$ 0.508, and Composite Reliability *(CR)* ≥ 0.807, while for discriminant validity, $\sqrt{AVE} \geq 0.713$, and - $0.424 \leq R_{ij} \leq 0.068$. The researcher concluded that the final scale is valid. It was recommended among others that the final scale should be utilised by researchers in assessing students' attitudes to Computer-Based Testing.

Keywords: Development and validation attitude scale; Computer-based testing; Factor analysis of scales

Introduction

Scale development is the process of developing a valid and reliable measure of a construct in order to evaluate an attribute of interest. According to Tay and Jebb (2017), measuring psychological constructs presents distinct issues because they are often unobservable and cannot be tested directly, necessitating indirect assessment methods such as self-report. Similarly, these constructs are frequently quite abstract, making it challenging to evaluate which items accurately represent them and which do not so consistently. Furthermore, the constructions are frequently complex and may consist of multiple distinct components. As a result of these complexities, developing a measurement instrument can be a difficult task, and validation is especially critical during the scale development phase.

There are two basic techniques to scale construction: deductive and inductive. A deductive approach focuses on generating objects inside the construct's scope by applying theory and already-created conceptualisations. This method is useful when the construct's definition is clear and significant enough to generate an initial set of items. In contrast, an inductive technique is used when the construct's definition or dimensionality is unknown. In this situation, organisational incumbents are asked to submit descriptions of the notion, and a conceptualisation is then developed, which serves as the foundation for item generation (Tay & Jebb, 2017).

Validated instruments should provide sufficient proof of psychometric qualities. These include proof of validity and reliability. Validity refers to how well an instrument measures what it is designed to measure, and there are three fundamental techniques to validity of tests and measures. These include content validity, criterion-referenced validity, and construct validity. Similarly, reliability relates to how consistent assessment results are in assessing any variable of interest.

On the other hand, students' attitudes generally play an important role in the teaching/learning process to the extent that students' academic achievement depends largely on their attitudes towards that programme. Also, the way and manner by which students' academic achievement is assessed affect their performance as observed by other researchers (e.g. Butler, 2003; Karadeniz, 2009; Pomplun & Custer, 2005; Pomplun, Ritchie & Custer, 2006; Ricketts & Wilks, 2002; Wingenbach, 2000; and Yurdabakan & Uzunkavak, 2012).

According to Hosseini, Zainol Abidin, and Baghdarnia (2014), with the introduction of new technologies, computer-based testing (CBT) has begun to become more widespread and implemented in large-scale testing, despite the fact that limited computer accessibility and high cost have limited CBT's implementation in testing institutions.

In Nigeria, universities started organising screening tests using the traditional pen-and-paper testing (PPT) format since 2005, with the University of Ilorin being one of the first. During the 2008/2009 academic session, CBT was introduced in selecting prospective candidates for admission into the university (Alabi, Issa & Oyekunle, 2012). Although successfully implemented, this innovation presented its own hitches, particularly in terms of successful coordination of the exercise, which took place at several centre sites across the country with a high degree of irregularity, and thus paved the way for more universities to join later.

Universities in Katsina state were not left behind in this new innovation because in Umaru Musa Yar'adua University Katsina, CBT was introduced as a method of test administration during 2014 / 2015 academic session for assessing general courses, but later it applies to all courses in 100 and 200 levels, if the number of students exceeds 100. With the change of administration, the conduct of CBT was reverted to only general courses as before. Almost similar practice was witnessed in Federal University Dutsin-ma (FUDMA) and Al-Qalam University Katsina (AUK). Some students who take such computer-based assessments argue that their test scores do not accurately reflect their achievement due to their unfamiliarity with the testing mode. However, as institutions began to incorporate CBT into their examination systems alongside traditional paper-based testing systems, concerns arose about the comparability of test results from the two modalities of administration (Wang, 2004).

A careful search of literature by the researcher revealed that, although research were conducted on the relationship between students' attitudes to CBT and their academic achievement in various fields, the data collection instruments were developed by the researchers (e.g. Bulent, Yalman, & Selahattin, 2016; Christensen, & Knezek, 2009; Chua, 2012; Dammas, 2016; Gül, Çokluk, & Gül, 2015; Tella & Bashorun, 2012; Yurdabakan, & Uzunkavak, 2012; Jamil, 2012; Khoshsima, Hashemi Toroujeni, Thompson, & Ebrahimi, 2019), and a lot of them were reported to have only face validity aspect of content validity, thus lacking the basic psychometric analyses of standardised instruments. This problem, coupled with other issues, has called for the development and validation of an attitude scale to computer-based testing. Hence, the main aim of this research was to develop and validate an attitude scale to computer-based testing using undergraduates in Katsina state as the validation sample.

Research Objectives

Based on the above background, the objectives of the study were to:

1. develop items of the preoperational version of the Attitude Scale to Computer-Based Testing;
2. determine factor structure of the items in the preoperational version of the Attitude Scale to Computer-Based Testing; and
3. investigate construct validity indices of the modified preoperational versions of Attitude Scale to Computer-Based Testing.

Research Questions

Based on the above stated objectives, the following research questions were formulated for the study:

RQ1. How many items were developed for the preoperational version of the Attitude Scale to Computer-Based Testing?

RQ2. What is the factor structure of the items in the preoperational version of Attitude Scale to Computer-Based Testing?

RQ3. What are construct validity indices of the final version of Attitude Scale to Computer-Based Testing?

Research Methodology

This research adopted an instrumentation design (also known as new scale/construction design). It is a kind of design that involves developing a new instrument entirely. The primary purpose is to develop a valid and reliable measure of an existing construct by adhering to precise rules for construction, administration, scoring, and interpretation.

The target population of this research covered all the undergraduates in all the three universities in Katsina state. These universities were Al-Qalam University Katsina (AUK); a private university, Federal University Dutsin-ma (FUDMA); a federal university, and Umaru Musa Yar'adua University Katsina (UMYU); a state university. Altogether, there were 34,050 undergraduates in all these universities as of 2018/2019 session as shown in Table 1 below:

Table 1: Distribution of the Population based on Faculties and Departments in each of the Universities in Katsina State

S/N	University	No. of Faculties	No. of Depts.	No. of students
1	Al-Qalam University, Katsina	4	22	8,650
2	Federal University, Dutsin-ma	6	32	12,500
3	Umaru Musa Yar'adua University, Katsina	5	25	12,900

Total		15	79	34,050

Sources: Registry / Academic Unit of each University (2019/2020)

In scale development and validation, sample sizes of a few hundred or more are desirable as suggested by Adams and Wieman (2010). According to Costello and Osborne (2005) and DeVellis (2017), scale development in general, and factor analysis in particular, are large sample size methods, and Kyriazos (2018) emphasises the importance of this requirement in studies where Confirmatory Factor Analysis is used as a validation method. As a result, a variety of sample size recommendations for scale development in general, and factor analysis in particular, have been made. These requirements are often expressed in terms of either the minimum required sample size (N) or the minimal ratio of N to the number of variables under consideration (q).

Although it was recommended that N be at least 100, 200, 250, and so on, Comrey and Lee (1992) propose a rough rating scale for adequate sample sizes in factor analysis as follows: N = 100 (poor sample size), N = 200 (fair sample size), N = 300 (good sample size), N = 500 (very good sample size), and N = 1000 or more (excellent sample size). They recommend that researchers use a sample size of at least 500 when conducting factor analytic studies. As a result, this proposal was taken into account while selecting the sample size for this study. Some researchers have followed this recommendation in their studies (e.g., Bulent, Yalman, & Selahattin, 2016; Christensen & Knezek, 2009; Yurdabakan & Uzunkavak, 2012; Tella & Bashorun, 2012; Jamil, 2012). Table 2 gives a sample of 1000 undergraduates selected from the three universities as the validation sample.

Table 2: Distribution of the Sample Size Selected from each of the Universities

S/N	University	Population	Sample Size
1	Al-Qalam University Katsina	8,650	254
2	Federal University Dutsin-ma	12,500	367
3	Umaru Musa Yar’adua University Katsina	12,900	379
	Total	**34,050**	**1,000**

The researcher purposefully chose students in 100 and 200 levels because they are the ones taking CBT exams in almost all the universities in the study area. In addition, the researcher selected the samples in a systematic manner utilising the Multistage Random Sampling procedure. Two sampling methods were utilised. They were cluster sampling and simple random sampling. Initially, cluster sampling technique was utilised to choose the faculties that participated in the research, and within these faculties, departments were also chosen using same technique. The requisite number of students for the study were then selected from the faculties and departments using a simple random sampling procedure.

The scale developed and validated was used in generating the data regarding the attitude of undergraduates towards computer-based testing. The development aspect was based on the laid down procedures and theories guiding test construction, and the scale is a 5-point Likert scale, with five responses: Strongly Agree (SA), Agree (A), Undecided (U), Disagree (D), and Strongly Disagree (SD).

Procedure for the Development of Attitude Scale to Computer-Based Testing

Several suggestions were made by test experts regarding the definition, as well as the conceptualisation of the construct of the study (i.e. attitude to computer-based testing). Others made suggestions regarding the phases and procedures to be followed in attitude scale development. This study was guided by Tella and Bashorun (2012)'s definition and conceptualisation, which described students' attitudes to computer-based testing as their methods of thinking and sentiments about taking computer-based testing. In this context, students' attitudes towards computer-based testing are defined as an evaluative inclination or a student's preference for taking tests and examinations, utilising computers and other related gadgets, rather than the traditional oral or paper and pen approach.

The study was also guided by the steps/phases suggested by DeVellis (2003) and Dunn-Rankin (2004), and also emphasized by Bulent, Yalman, and Selahattin (2016). They include forming the item pool; asking for experts' views; conducting the trial application; applying the draft scale to the study group and performing factor analyses, and calculating the reliability of the scale. First of all, the guiding instrument in developing the item pool of the preoperational version of Attitude Scale to Computer-Based Testing was the Modified Fennema-Sherman Mathematics Attitude Scale, adapted and validated by Dan'inna (2016). This instrument has four subscales, representing four different dimensions of attitude to Mathematics. Since it measures students' attitude to Mathematics, the researcher felt that the nature and the direction of the items are useful when developing items for the preoperational version of Attitude Scale to Computer-Based Testing.

Secondly, in the item development stage, the researcher identified and consulted some researchers' developed instruments. Some items were pooled and modified to suit the intended purpose. A lot of consultation was made with other researchers, and the developed scale was given to experts so as to have their contributions in order to ensure face validity of the scale. After effecting the experts' suggestion, the preoperational version of scale was made up to have five (5) subscales, with each subscale measuring specific aspect of students' attitudes towards computer-based testing. These subscales were:

General Computer Skills and Experience	(Item 1 – 13)
Confidence in Writing Computer-Based Tests	(Item 14 – 26)
Acceptance and Preference of CBT over PPT	(Item 27 – 39)
Anxiety / Stress in Writing Computer-Based Tests	(Item 40 – 52)
Perceived Impact of CBT on Academic Performance	(Item 53 – 67)

In each of these presumed subscales, there are items in both positive and negative direction. Only 23 out of 67 items were in negative direction. They included items 2, 5, 7, 9, 16, 17, 18, 23, 28, 30, 32, 39, 41, 42, 44, 47, 48, 51, 56, 63, 65, 66, and 67. The choice of these subscales was also guided by the recommendations of Akdemir, and Oguz (2008), Chua, Chen and Wong (1999), Mahar, Henderson and Deane (1997), and Olsen, Cox, Price, Strozeski and Vela (1990). According to these researchers, factors that influence students' attitude to computer-based testing may include computer experience, computer anxiety, and computer testing hardware and software. Others are confidence in using computer, confidence in using computer to write computer-based tests, preference of computer-based tests over paper-and-pencil tests, and stress associated with computer-based tests, etc.

Also, the preoperational version of the scale has a section for demographic data, such as name of the university, gender of the respondent, level of study, name of faculty and department, programme of study, identification number, and level of computer experience, if any. The scoring was done in such a way that each positive item earned a score, as follows: SA = 5, A = 4, U = 3, D = 2, and SD = 1. The scoring for each negative item was reversed, resulting in SA = 1, A = 2, U = 3, D = 4, and SD = 5.

Validation Procedure of the Preoperational Version

A good first step in the preliminary evaluation of the preoperational version of the scale was to have a group of students (say 10) respond to the items first as if they have a favourable attitude to CBT, and then as if they have an unfavourable attitude to CBT. As suggested by Domino and Domino (2006), the essence was to identify those items that show a distinct shift in the response of the students, and these are most likely useful items. On the other hand, items that cannot distinguish respondents with favourable attitude from those with unfavourable attitudes would be modified. This was done and based on the outcome, some items were already modified before inclusion.

The preoperational version of the Attitude Scale to Computer-Based Testing was then given to experts (Professors and Doctors) in the fields of Tests and Measurement/Psychometrics and Psychology, as well as those with experience in scale development, to look at the items and ensure face validity. The essence of this was for them to offer constructive criticisms and propose modifications on how the items were developed and designed. This was also done, and their suggestions were affected as appropriate.

The researcher then conducted the initial trial testing of the scale. A large sample size of 300, as suggested by some authorities, was selected randomly to include both male and female undergraduates. They were given the scale to respond after which the responses were collected back. The data generated was analysed with the help of IBM® SPSS® Statistics v.23 and IBM® SPSS® Amos v.23 software. Both Exploratory and Confirmatory Factor Analysis (EFA & CFA) procedures were followed in investigating the factor structure of the preoperational version of the Attitude Scale to Computer-Based Testing.

Validation Results

RQ1. How many items were developed for the preoperational version of the Attitude Scale to Computer-Based Testing?

Based on the results collected, it was obvious that initially, the preoperational version of the Attitude Scale to Computer-Based Testing contained up to 67 items, but as a result of the EFA done during the trial testing, only 50 items were maintained in the modified version, as provided in Table 3. The two versions are as found in the Appendix section.

Table 3: Number of Items Developed for the Preoperational and Modified Versions of Attitude Scale to Computer-Based Testing

Name of the Scale	Code	Number of Items
Preoperational version	ASCBT 1.0	67
Modified Preoperational version	ASCBT 1.1	50

Therefore, the answer to this research question is that the number of items developed for the preoperational version of Attitude Scale to Computer-Based Testing was initially 67, but only 50 were retained in the modified version after trial testing.

RQ2. What is the factor structure of the items in the preoperational version of Attitude Scale to Computer-Based Testing?

In this regard, EFA was conducted on the data generated using IBM® SPSS® Statistics v.23. The sampling adequacy was tested using the Kaiser Meyer-Olkin Measure of Sampling Adequacy

(KMO/MSA), while the strength of the correlation among majority of the items was tested using the Bartlett's Test of Sphericity statistic. The results were summarised in Table 4 below:

Table 4: KMO Measure of Sampling Adequacy and Bartlett's Test Results for the Preoperational Version of the Attitude Scale to Computer-Based Testing (ASCBT 1.0)

Type of Test	Statistics	Value
Kaiser-Meyer-Olkin Measure of Sampling Adequacy	KMO/MSA	.813
Bartlett's Test of Sphericity	Approx. Chi-Square	16788.548
	df	2211
	Sig.	.000

From the results in Table 4 above, it is clear that the sampling process for the study was relevant and adequate for factor analysis *(KMO/MSA = .813)* (Field, 2000; and Pallant, 2013). Again, the correlation among majority of the items was so strong ($\chi^2 = 16788.548$, $df = 2211$, $P = .000$). Therefore, these values showed that the data set for the trial testing had good fit for factor analysis.

From the EFA results, 15 factors were initially discovered, that explained *72.236%* of the total variance. When carefully observed, some items have lower factor loadings while some items have loaded to more than one component, and some components had only one item loading. So, when factor loadings below .50 were supressed, it was found that 50 items should be retained, and only six factors were identified with at least five items loading to them. Oblique rotation was then chosen, and the analysis was repeated with the 50 items, after removing all 17 items with factor loadings less than .50. The revised *KMO/MSA* = .852, χ^2 = 15329.098, *df* = 1225, and *p* = .000, and the six components explained approximately 64.246% of the entire variation, as shown in Table 5. The goal of the rotation was to maximise high item loadings while minimising low item loadings, to make results easier to read, and to offer a more parsimonious solution. So, based on the nature of the items, these six factors were named as:

Factor 1. Perceived Impact of Computer-Based Tests on Academic Performance
Factor 2. Confidence in Writing Computer-Based Tests
Factor 3. General Computer Appreciation and Experience
Factor 4. Acceptance/Preference of Computer-Based Tests over Paper-Based Tests
Factor 5. Basic Computer Application Skills; and
Factor 6. Anxiety / Stress in Writing Computer-Based Tests.

Table 5: Oblimin-Rotated Structure Matrix for the Modified Preoperational Version of the Attitude Scale to Computer-Based Testing (ASCBT 1.1)

	Components					
Items	**1**	**2**	**3**	**4**	**5**	**6**
Item 59	.939					
Item 55	.932					
Item 61	.929					
Item 57	.929					
Item 58	.925					
Item 60	.921					
Item 56	.895					
Item 54	.840					
Item 63	.665					
Item 64	.654					
Item 62	.651					
Item 53	.614					

Items	Components					
	1	2	3	4	5	6
Item 24		.811				
Item 16		.811				
Item 18		.802				
Item 22		.791				
Item 20		.769				
Item 26		.757				
Item 21		.729				
Item 23		.728				
Item 25		.725				
Item 17		.658				
Item 19		.639				
Item 14		.634				
Item 15		.583				
Item 6			.836			
Item 1			.833			
Item 2			.811			
Item 7			.795			
Item 4			.738			
Item 5			.737			
Item 3			.732			
Item 27				.862		
Item 34				.846		
Item 39				.808		
Item 28				.801		
Item 33				.735		
Item 32				.716		
Item 35				.712		
Item 8					.949	
Item 10					.928	
Item 12					.915	
Item 11					.912	
Item 9					.859	
Item 42						.839
Item 49						.787
Item 45						.777
Item 40						.773
Item 41						.759
Item 43						.671
Eigen Values	**8.678**	**7.705**	**4.945**	**4.444**	**3.824**	**3.158**
Total Variance Explained (%)	**17.355**	**14.150**	**9.889**	**8.888**	**7.648**	**6.316**
Cumulative Variance Explained (%)	**17.355**	**31.505**	**41.394**	**50.282**	**57.930**	**64.246**

Extraction Method: *Principal Component Analysis.*
Rotation Method: *Oblimin with Kaiser Normalization.*

Therefore, the answer to this research question is that the developed preoperational version of Attitude Scale to Computer-Based Testing was initially made to have 5 factors, but the modified version was found to have a structure of 6 factors, and they explained *64.246%* of the total variance.

RQ3. What are construct validity indices of the final version of Attitude Scale to Computer-Based Testing?

In this case, both EFA and CFA were run on the responses obtained from the final version (ASCBT 2.0). As mentioned earlier, the data generated for the validation/final study comprised

1000 cases, which were split into two subsets of 500 cases each, intended for EFA and CFA analyses separately, based on the recommendation of Hair, Black, Babin, and Anderson (2010) and Kline (2005). The first data set was to ***develop*** the measure (EFA), and the second data set was to ***validate*** the measure (CFA).

Exploratory Factor Analysis for the Modified Preoperational Version (ASCBT 2.0)

The results of the EFA were summarised in Tables 6 and 7. Initially, ten components with a minimum Eigen value of 1.00 were recovered, accounting for 68.321% of the overall variance. Items with factor loadings of less than 0.50 and cross-loaded on two or more factors of 0.50 or higher were excluded. The analysis was repeated, yielding a modified scale of 39 items organised into five different components, each with at least three item loadings. At this point, the Varimax (orthogonal) rotation approach was chosen because the data showed substantial correlations among the extracted factors, and the goal was to construct factor structures that were uncorrelated. The findings indicate that the five extracted components accounted for 49.495% of the total variance. The significant loading of all the items on the single factor indicates unidimensionality of the items, and the fact that no item had multiple (significant) cross loading was found to support the preliminary discriminant validity of the scale.

Table 6: KMO Measure of Sampling Adequacy and Bartlett's Test Results for the Modified Preoperational Version of the Attitude Scale to Computer-Based Testing (ASCBT 2.0)

Type of Test	Statistics	Value
Kaiser-Meyer-Olkin Measure of Sampling Adequacy	KMO/MSA	.851
Bartlett's Test of Sphericity	Approx. Chi-Square	17346.330
	Df	1225
	Sig.	.000

Table 7: Varimax-Rotated Component Matrix with Kaiser Normalization for the Modified Preoperational Version of the Attitude Scale to Computer-Based Testing (ASCBT 2.0)

	Components				
Items	**1**	**2**	**3**	**4**	**5**
Imp10	.941				
Imp9	.925				
Imp8	.924				
Imp7	.904				
Imp11	.898				
Imp6	.894				
Imp5	.867				
Cnf4		.754			
Cnf2		.741			
Cnf8		.731			
Cnf7		.727			
Cnf1		.716			
Cnf11		.693			
Cnf10		.693			
Cnf5		.687			
Cnf9		.648			

Items	Components				
	1	2	3	4	5
Cnf3		.589			
Bcs3			.730		
Bcs2			.728		
Bcs4			.713		
Bcs1			.692		
Bcs6			.690		
Bcs7			.688		
Bcs5			.660		
Bcs9			.570		
Bcs8			.569		
Prf7				.794	
Prf5				.788	
Prf6				.776	
Prf4				.772	
Prf3				-.515	
Prf1				-.512	
Prf2				-.510	
Anx1					.746
Anx4					.742
Anx6					.683
Anx2					.671
Anx3					.638
Anx5					.619
Eigen Values	7.305	6.144	4.539	3.433	3.327
Total Variance Explained (%)	14.610	12.288	9.077	6.866	6.654
Cumulative Variance Explained (%)	14.610	26.898	35.975	42.841	49.495

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Rotation converged in 7 iterations.

Based on the EFA results, it was discovered that the factor structure of ASCBT 2.0 was made up to have five distinct factors. These factors have 9 items (Basic Computer Skills and Experience), 10 items (Confidence in writing CBT), 7 items (Preference of CBT over PPT), 6 items (Anxiety during CBT), and 7 items (Perceived Impact of CBT on Academic Performance), making a total of 39 items in all.

To confirm the existence or otherwise of these factors in the scale, and to further validate the modified scale (ASCBT 2.0), there was the need to go for Confirmatory Factor Analysis (CFA), and this was conducted and presented in the next subsection.

Confirmatory Factor Analysis for the Modified Preoperational Version (ASCBT 2.0)

The second data set comprising another 500 cases was used to conduct the CFA. Based on the recommendation of Byrne (2001), the researcher developed an initial measurement model which included the identified five components as first-order factors, as shown in Figure 1. The model was tested using the Structural Equation Modelling (SEM) technique, employing the maximum likelihood method. The analysis was run using IBM® SPSS® Amos version 23. As explained by Anderson and Gerbing (1988), this method identifies the unidimensionality of each factor, indicating the presence of a single trait or construct underlying a group of measures. The five factors were correlated, with each item having a non-zero loading on its designated factors and virtually insignificant loadings on others, but the measurement error terms associated with each item were uncorrelated. This is because the model was tested with data from the best 39 items on the scale, the hypothesised five-factor model was shown to be non-fit, as illustrated in Figure 2.

The model fit indices as obtained from the results were as summarised in Table 8. These include the minimum Chi-square (χ^2), its degree of freedom (df), Bollen–Stine p (P), Chi-square normalised by degree of freedom (χ^2/df), Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI), Comparative Fit Index (CFI), Root Mean Square Error of Approximation RMSEA), Non-Normed Fit Index (NNFI) [also known as Tucker-Lewis Index (TLI)], and standardised Root Mean Square Residual (SRMR).

In order to improve the validity indices of the model, some modifications were necessary. In this regard, some items with low factor loadings in each subscale were dropped, and based on the suggestion outlined in the modification indices, a covariance between two items under Impact subscale (Imp7 and Imp8) was imposed. This resulted into a final measurement model with only 24 items as shown in Figure 3. The final model was then tested and all the fit indices obtained were acceptable as summarised in Table 9.

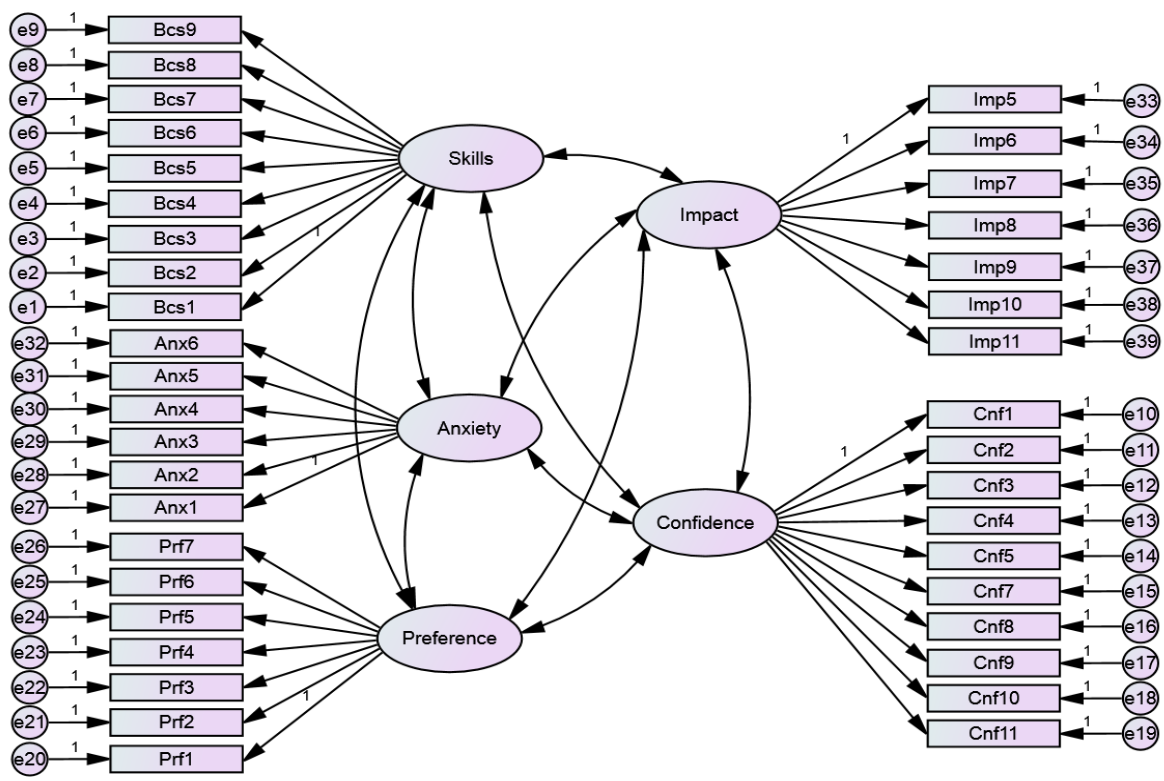

Figure 1: Initial measurement model (ASCBT 2.0)

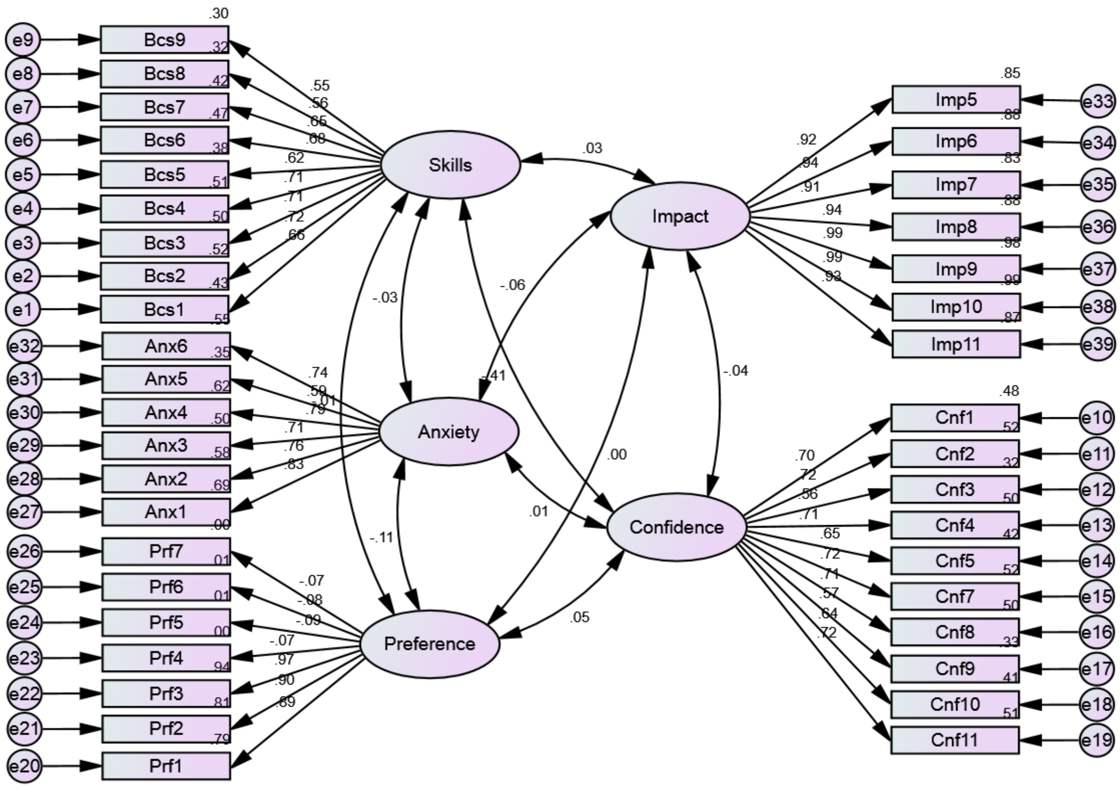

Figure 2: Standardised estimates for the initial measurement model (ASCBT 2.0)

Table 8: Fit Indices for the Initial Measurement Model of the Modified Preoperational Version (ASCBT 2.0)

Fit Index	Value Obtained	Recommended	Source
χ^2	5268.016		
Df	692		
P	.000		
χ^2/df	7.613	≤ 3.00	Byrne (2001)
GFI	0.694	≥ 0.90	Hair, Anderson, Tatham, and Black, (1998)
AGFI	0.655	≥ 0.80	Chau and Hu (2001)
CFI	0.738	≥ 0.95	Hu and Bentler (1999)
RMSEA	0.115	≤ 0.08	Brown and Cudeck (1993)
NNFI (TLI)	0.719	≥ 0.95	Hu and Bentler (1999)
SRMR	0.090	< 0.08	Hair, Anderson, Tatham, and Black, (1998)

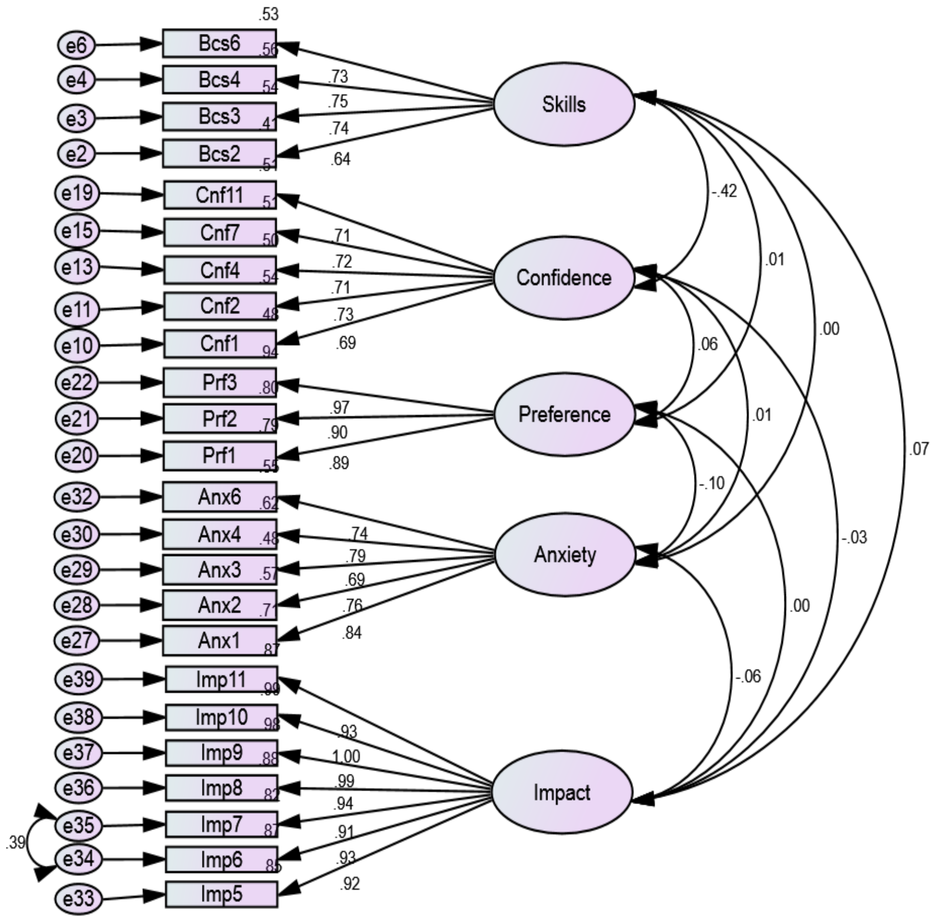

Figure 3: Standardised estimates for final measurement model (ASCBT 2.1)

Table 9: Fit Indices for the Final Measurement Model of the Final Version (ASCBT 2.1)

Fit Index	Value Obtained	Recommended	Source
χ^2	626.222		
Df	241		
P	.000		
χ^2/df	2.598	≤ 3.00	Byrne (2001)
GFI	0.911	≥ 0.90	Hair, Anderson, Tatham, and Black, (1998)
AGFI	0.889	≥ 0.80	Chau and Hu (2001)
CFI	0.966	≥ 0.95	Hu and Bentler (1999)
RMSEA	0.057	≤ 0.08	Brown and Cudeck (1993)
NNFI (TLI)	0.961	≥ 0.95	Hu and Bentler (1999)
SRMR	0.035	< 0.08	Hair, Anderson, Tatham, and Black, (1998)

Assessment of Convergent Validity Indices for the Final Version (ASCBT 2.1)

Convergent validity is the extent to which different items measuring the same construct agree with one another. Hair, Black, Babin, and Anderson (2010) recommend using Factor Loadings (λ), Average Variance Extracted (AVE), and Composite Reliability (CR) to measure an instrument's convergent validity. For factor loadings ($\lambda \geq 0.5$), average variance extracted (AVE ≥ 0.5), and composite reliability (CR ≥ 0.7), a general rule of thumb applies. The study found that $\lambda \geq 0.64$, AVE ≥ 0.508, and CR ≥ 0.807, confirming the convergent validity of the ASCBT 2.1 (refer to Table 10 below).

Table 10: Convergent Validity Indices of the Final Measurement Model of the Final Version (ASCBT 2.1)

Factor (Construct)	**Items**	**Factor Loading (λ)**	**CR**	**AVE**
Basic Computer Skills and Experience **(Skills)**	Bcs2	0.64	0.807	0.511
	Bcs3	0.74		
	Bcs4	0.75		
	Bcs6	0.73		
Confidence in Writing CBT **(Confidence)**	Cnf1	0.69	0.838	0.508
	Cnf2	0.73		
	Cnf4	0.71		
	Cnf7	0.72		
	Cnf11	0.71		
Preference of CBT over PPT **(Preference)**	Prf1	0.89	0.942	0.844
	Prf2	0.90		
	Prf3	0.97		
Anxiety during CBT **(Anxiety)**	Anx1	0.84	0.876	0.587
	Anx2	0.76		
	Anx3	0.69		
	Anx4	0.79		
	Anx6	0.74		
Impact of CBT on Academic Performance **(Impact)**	Imp5	0.92	0.984	0.895
	Imp6	0.93		
	Imp7	0.91		
	Imp8	0.94		
	Imp9	0.99		
	Imp10	1.00		
	Imp11	0.93		

Assessment of Discriminant Validity Indices for the Final Version (ASCBT 2.1)

Discriminant validity is the extent to which items distinguish between constructs or measure distinct constructs. This can be evaluated by looking at the correlations between measures of possibly overlapping constructs. According to Compeau, Higgins, and Huff (1999) and

Ramayah, Lee, and Boey (2011), the rule of thumb is that items should load more heavily on their own constructs in the model, and the average variance shared between each construct and its measures should be greater than the variance shared between the construct and other constructs.

When displayed in a matrix, the diagonal components (in bold) should indicate the square root of the average variance extracted (AVE), and the off-diagonal elements should represent the correlations between the constructs. To show discriminant validity, diagonal elements should be larger than off-diagonal elements (Chiu & Wang, 2008). The results of this investigation verified the discriminant validity of the final version (ASCBT 2.1) (see Table 11 below).

Table 11: Discriminant Validity Indices of the Final Measurement Model of the Final Version (ASCBT 2.1)

	Skills	Confidence	Preference	Anxiety	Impact
Skills	**0.715**				
Confidence	-0.424	**0.713**			
Preference	0.007	0.055	**0.919**		
Anxiety	0.005	0.013	-0.104	**0.766**	
Impact	0.068	-0.033	-0.001	-0.059	**0.946**

Note: Diagonal elements (in bold) are the square root of the average variance extracted (AVE). Off-diagonal elements are the correlations among the constructs.

Therefore, the answer to this research question is that the construct validity indices of the final version of the Attitude Scale to Computer-Based Testing were as follows:

i. For convergent validity, $\lambda \geq 0.64$, $AVE \geq 0.508$, and $CR \geq 0.807$.

ii. For discriminant validity, $\sqrt{AVE} \geq 0.713$, and $-0.424 \leq R_{ij} \leq 0.068$

These two sets of indices account for the construct validity of the final version of the Attitude Scale to Computer-Based Testing.

Discussions

From the results, it was found that the final version of the Attitude Scale to Computer-Based Testing (ASCBT 2.1) had acceptable convergent validity and discriminant validity indices. These two parameters together account for the construct validity of the scale. Therefore, the final scale (ASCBT 2.1) has the ability to measure the construct it was designed to measure, that is, attitude towards computer-based testing among undergraduates in Nigerian universities. The implication of this finding is that, researchers willing to carry out researches involving assessment of students' attitudes towards computer-based testing can now make use of the final version of the scale in doing so, and by so doing, a valid measure of the construct can be obtained.

Summary and Conclusion

From the results so far presented, it can be established that an instrument titled "Attitude Scale to Computer-Based Testing" (ASCBT) was developed and validated. The preoperational version (ASCBT 1.0) had 67 items with 5 subscales, but only 50 items with 6 subscales were retained in the modified version (ASCBT 1.1) after the trial testing. This modified version, after some modifications, was named ASCBT 2.0 and was used to generate data for the main validation. The EFA conducted has revealed that only 39 items should be retained, while the CFA/SEM confirmed further that only 24 items of the scale had acceptable fit indices in the model developed by the researcher. Hence, the final version (ASCBT 2.1) was found to have only 24 items with 5 subscales that explained a reasonable percentage of variance in the measurement of the target construct, i.e., attitude towards computer-based testing among undergraduates in Katsina state.

In conclusion therefore, based on the outcome of this study, the researcher concluded that the final version of the Attitude Scale to Computer-Based Testing (ASCBT 2.1) was found to have acceptable construct validity indices. So, researchers interested in investigating students' attitude to computer-based testing and who are in need of standard instruments can make good use of this one.

References

Adams, W. K., & Wieman, C. E. (2010). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education, iFirst Article*, 1 – 24. ISSN: 0950-0693 (print) / ISSN: 1464-5289 (online). DOI: 10.1080/09500693.2010.512369

Akdemir, O., & Oguz, A. (2008). Computer-based testing: An alternative for the assessment of Turkish undergraduate students. *Computers & Education 51*, 1198 – 1204.

Alabi, A. T., Issa, A. O., & Oyekunle, R. A. (2012). The use of computer based testing method for the conduct of examinations at the University of Ilorin. *Ife Journal of Educational Leadership Administration and Planning, 1*(1), 226 – 236. ISSN: 1592 – 3113.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modelling in practice: A review and recommended two-step approach. *Psychological Bulletin 103*(3), 411 – 423.

Brown, M. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds). *Testing Structural Equation Models.* Sage, Newbury Park, CA.

Bulent, B., Yalman, M. & Selahattin, G. (2016). Attitude scale towards web-based examination system (MOODLE)-Validity and reliability study. *Educational Research and Reviews, 11*(17), 1641 – 1649. DOI: 10.5897/ERR2016.2850

Butler, D. L. (2003). The impact of Computer-based testing on students' attitudes and behaviour. *The Technology Source*. Retrieved November 2012, from http://ts.mivu.org/default.asp?show=article&id=1034

Byrne, B. M. (2001). *Structural equation modelling with AMOS: Basic concepts, applications, and programming.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Chau, P. Y. K., & Hu, P. J. H. (2001). Information technology acceptance by individual professionals: A model comparison approach. *Decision Sciences, 32*(4), 699 – 719.

Chiu, C. M. & Wang, E. T. G. (2008). Understanding web-based learning continuance intention: The role of subjective task value. *Inf. Manag. 45*(3), 194 – 201.

Christensen, R. W. & Knezek, G. A. (2009). Construct validity for the teachers' attitudes toward computers questionnaire. *Journal of Computing in Teacher Education, 25*(4), 143 – 155.

Chua Y. P. (2012). Replacing paper-based testing with computer-based testing in assessment: Are we doing wrong? *Procedia – Social and Behavioural Sciences, 64,* 655 – 664.

Chua, S. L., Chen, D., & Wong, A. F. L. (1999). Computer anxiety and its correlates: A meta-analysis. *Computers in Human Behaviour*, *15*, 609 – 623.

Compeau, D., Higgins, C. A., & Huff, S. (1999). Social cognitive theory and individual reactions to computing technology: A longitudinal study. *MIS Q, 23*(2), 145 – 158.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis.* Hillsdale, NJ: Erlbaum.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*, 1 – 9.
Dammas, A. H. (2016). Investigate students' attitudes towards computer-based test (CBT) at Chemistry course. *Archives of Business Research, 4*(6), 58 – 71.
Dan'inna, A., A. (2016). Construct validation of the modified Fennema-Sherman mathematics attitude scale among senior secondary school students in Katsina state. *Journal of Educational Research and Policies, 11*(2), 93 – 102.
DeVellis, R. (2003). *Scale development: Theory and applications*. London: SAGE.
DeVellis, R. F. (2017). *Scale development: Theory and applications (4th edition).* Thousand Oaks, CA: Sage.
Domino, G. & Domino, M. L. (2006). *Psychological testing: An introduction*. New York: Cambridge.
Dunn-Rankin, P. (2004). *Scaling methods.* London: Routledge.
Ebrahimi, M. R., Hashemi Toroujeni, S. M., & Shahbazi, V. (2019). Score equivalence, gender difference, and testing mode preference in a comparative study between computer-based testing and paper-based testing. *iJET, 4*(7), 128 – 143. Available @ https://doi.org/10.3991/ijet.v14i07.10175
Field, A. (2000). *Discovering Statistics using SPSS for Windows*. London – Thousand Oaks – New Delhi: Sage publications.
Gül, E., Çokluk, Ö., & Gül, Ç. D. (2015). Development of an attitudes scale toward online assessment. *Procedia - Social and Behavioural Sciences, 174*, 529 – 536.
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis.* Upper Saddle River, NJ: Prentice-Hall.
Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, B. C. (1998). *Multivariate data analysis*. Eaglewood Cliff, NJ: The Prentice Hall International.
Hosseini, M., Zainol Abidin M. J., & Baghdarnia, M. (2014). Comparability of test results of computer based tests (CBT) and paper and pencil tests (PPT) among English language learners in Iran. *Procedia - Social and Behavioural Sciences, 98,* 659 – 667
Hu, L. T., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling, 6*, 1 – 55.
Jamil, M. (2012). Perceptions of university students regarding computer assisted assessment. *TOJET: The Turkish Online Journal of Educational Technology, 11*(3), 267 – 277.
Karadeniz, S. (2009). The impacts of paper, web and mobile based assessment on students' achievement and perceptions. *Scientific Research and Essay, 4*(10), 984 – 991. Retrieved August 2012, from http://www.academicjournals.org/sre
Khoshsima, H., Hashemi Toroujeni, S. M., Thompson, N., & Ebrahimi, M. R. (2019). Computer-based (CBT) vs. Paper-based (PBT) testing: Mode effect, relationship between computer familiarity, attitudes, aversion and mode preference with CBT test scores in an Asian private EFL context. *Teaching English with Technology*, *19*(1), 86 – 101. Retrieved from http://www.tewtjournal.org
Kline, R. B. (2005). *Principles and practice of structural equation modelling (2nd ed.).* New York: Guilford.
Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in General. *Psychology, 9*, 2207 – 2230. https://doi.org/10.4236/psych.2018.98126

Mahar, D., Henderson, R., & Deane, F. (1997). The effects of computer anxiety and computer experience on users' performance of computer based tasks. *Personal and Individual Differences, 22*(5), 683 – 692.

Olsen, J. B., Cox, A., Price, C., Strozeski, M., & Vela, I. (1990). Development implementation and validation of a computerized test for state wide assessment. *Educational Measurement: Issues and Practice, 9*(2), 7 – 10.

Pallant, J. (2013). S*PSS survival manual. A step by step guide to data analysis using SPSS, 4th edition*. Allen & Unwin. Available @ www.allenandunwin.com/spss.

Pomplun, M., & Custer, M. (2005). The score comparability of computerised and paper-and-pencil formats for K-3 reading tests. *Journal of Educational Computing Research*, *32*(2), 153 – 166.

Pomplun, M., Ritchie, T., & Custer, M. (2006). Factors in paper-and-pencil and computer reading score differences at the primary grades. *Educational Assessment*, *11*(2), 127 – 143.

Ramayah, T., Lee, J. W. C., & Boey, J. C. J. (2011). Network collaboration and performance in tourism sector. *Serv. Bus. 5*(4), 411 – 428.

Ricketts, C., & Wilks, S. J. (2002). Improving students' performance through computer-based assessment: Insights from recent research. *Assessment and Evaluation in Higher Education*, *27*(5), 475 – 479. DOI: 10.1080/ 02602930220000009348

Tay, L., & Jebb, A. (2017). Scale development. In S. Rogelberg (Ed), *The SAGE encyclopaedia of industrial and organizational psychology, 2nd edition*. Thousand Oaks, CA: Sage

Tella, A., & Bashorun, M. T. (2012). Attitude of undergraduate students towards computer-based test (CBT): A case study of the University of Ilorin, Nigeria. *International Journal of Information and Communication Technology Education, 8*(2), 33 – 45.

Wang, S. (2004). *Online or paper: Does delivery affect results? Administration mode comparability study for Stanford diagnostic reading and mathematics tests*. USA: Harcourt Assessment Inc.

Wingenbach, G. J. (2000). Agriculture students' computer skills and electronic exams. *Journal of Agricultural Education, 41*(1), 69 – 78. DOI: 10.5032/ jae.2000.01069

Yurdabakan, I., & Uzunkavak, C. (2012). Primary school students' attitudes towards computer-based testing and assessment in Turkey. *Turkish Online Journal of Distance Education, 13*(12), 177 – 188.