

UNIJERPS
Unizik Journal of Educational Research and Policy Studies VOL.6;
https://unijerps.org
2021 Impact Factor: 5.641; 2020 Impact Factor: 5.355. January-June, 2021
Indexing and Impact Factor URL: http://sjifactor.com/passport.php?id=21363

ACADEMIC JOURNALS
expand your knowledge
http://www.academicjournals.org

USING FACTOR ANALYSIS IN THE DEVELOPMENT AND VALIDATION OF THE SENIOR SECONDARY SCHOOL ONE (SS1) MULTIPLE CHOICE MATHEMATICS TEST ITEMS IN OYO STATE, NIGERIA.

Osiesi, Prince Mensah.
Department of Educational Management and Business Studies,
Faculty of Education, Federal University of Oye-Ekiti, Oye, Ekiti State, Nigeria.
mensah.osiesi@fuoye.edu.ng; +2348151375430.

ABSTRACT

This study is instrumentation research that employed factor analysis in the development and Validation of the senior secondary school one (SS1) multiple choice mathematics test items. The purposive and convenient sampling techniques were adopted in selecting 100 students in Ibadan Metropolis of Oyo state, Nigeria. The Kudar-Richardson (KR-20) reliability coefficient of the MAT was 0.83 showing that it is internally consistent. Four research questions were raised in the study and the data collected was analysed at 5% significance level. The results revealed that the MAT test items were suitable for factor analysis as the Kaiser-Meyer Olkins' Measure of sampling adequacy (KMO) was above 0.67 and there is a strong relationship among the MAT test items as the Bartlett's Test of Sphericity was significant ($p = 0.00$); Seven components were extracted as they had Eigenvalues that is more than one (3.82, 1.99, 1.65, 1.52, 1.34, 1.17 and 1.08). These seven components are number bases, sets, statistics, algebra, geometry, measurement and probability; explaining a total of 62.83% of the variance. Lastly, the MAT test items have face, construct and content validity as the rotated factor loading matrixes ranged from 0.52 and 0.80. Hence, the study recommended the use of factor analysis in test item development and validation over other conventional techniques; priority be appropriated more to mathematics researches and teachers/examiners be regularly trained on the various statistical techniques (such as the factor analysis) in developing and validating test items for testing/examination purposes.


UNIJERPS
Unizik Journal of Educational Research and Policy Studies VOL.6;
https://unijerps.org
2021 Impact Factor: 5.641; 2020 Impact Factor: 5.355. January-June, 2021
Indexing and Impact Factor URL: http://sjifactor.com/passport.php?id=21363
ACADEMIC JOURNALS
expand your knowledge
http://www.academicjournals.org

Key Words: Development; Factor analysis; Mathematics; Multiple Choice Test items; Secondary Schools; Validation.

Introduction

Overtime, there comes the need for students' learning outcomes (their achievement or performances) to be measured and evaluated in order to ascertain if they are truly learning what they are supposed to learn. The procedures employed by the teacher and schools in attaining this, may include the use of tests, individual projects, assignments/homework, peer group projects and examinations. In many occasions, these assessment procedures do take the forms of oral testing, written tests (objective and theory) and a hands-on-practical and projects. Osiesi (2019) highlighted the essence of achievement tests to include: determining students' change in behavior, diagnosing the strength and weaknesses of students, promoting healthy rivalry among students, determining teachers' teaching efficiencies, assessing the level of students' competences, comparing students' performance in subject areas, assigning grades to students, For promotion of students, for students' guidance and counseling, for students' certification, for students' placement, for assessing the curriculum content and feedback Purposes.

Multiple-choice tests are becoming a usual means for assessing students learning; and there are many techniques inherent in educational and psychological literature. These are the factor analysis, classical test theory, cluster analysis, item response theory and model analysis (Ding & Beichner, 2009). According to Ajeigbe and Afolabi (2013), a typical multiple-choice objective test consists of questions sampled from the syllabus or subject outline within a well-defined ranged of cognitive levels. In multiple-choice questions, students are to choose the correct answer from a given number of possible answers, called foils. The incorrect options are termed distracters, while the correct option is called the key.

Kerlinger (1979) defined factor analysis as an analytic method of determining the number and nature of the variables that underlie larger numbers of variables or measures and it tells the researcher, in effect, what measures belong together and which one virtually measure the same thing. It is a method of investigating whether a number of variables of interest are linearly related to a smaller number of unobservable factors (Tryfos, 1997).

Factor analysis shows information on how test items are interrelated. It is a statistical technique that constructs a small set of artificial variables (usually referred to as components) through linear combinations of highly correlated observed variables (Ding & Beichner, 2009). Factor analysis is extensively used for developing and evaluating tests and scales (Pallant, 2007), wherein the test or scale developer reduces and refines a large number of test items/scales to smaller number of coherent test items or subscales. Usually, psychologists use it to describe theoretically a specific domain, gain insight into a domain and limit the number of variables (Ellis, 2017).

The main rationale in performing factor analysis is to reduce the number of variables (items) that truly explains and interprets the data. This can be accomplished in two steps: by factor extraction and factor rotation. The process of selecting the type of model and the number of factors to extract is called factor extraction. It precedes factor rotation which is aimed at obtaining a more simplified structure for ease of interpretability. Factor extraction consists of two approaches; the principal component analysis and the common factor analysis (Hatcher, 1994). The principal component analysis (PCA) and common factor analysis (CFA) although similar in features, serve different purposes (Gorsuch, 1990). The PCA is for components development devoid of the assumptions of causal relationship between the components and variables observed while the CFA examines the basic "factors" having assumed casual effects on the variables observed.

Vahedi and Farrokhi (2011) carried out a confirmatory factor analysis of the structure of abbreviated math anxiety scale. 298 students from Tabriz University in Iran were used in determining the validity and reliability of the scale. A two-factor solution provided a better fit to the data than a single factor. Similarly, Adediwura (2011) carried out a research on the development and confirmatory factor analysis of a scale for the measurement of gifted students' attitude towards mathematics. The results obtained showed that the mathematically gifted group performed significantly better than both the average group and the underachievers' group. Ugulu (2011; 2013) did a similar study on confirmatory factor analysis for testing validity and reliability of traditional knowledge scale to measure university students' attitudes. He used varimax

rotation and principal components factor analysis methods to obtain the scale factor structure. The value of his KMO value was found to be 0.681; the Bartlett's Test of Sphericity observed significance level was $p < 0.001$.

More so, Rossoni, Engelbert and Bellegard (2016) investigated how different methods of extraction, factor definition, and rotation of exploratory factor analysis affect the fit of measurement scales by meta-analysis of 23 studies. Their results indicate that the Principal Components method provides greater explained variance, provides greater reliability while the Quartimax provides lower correlation between factors. In the same vein, Pishghadam and Khosropanah (2011) in their study examined the construct validity of English Language teachers competency test using exploratory factor analysis (EFA) and found that the results revealed that the underlying factors of the test are what the test-designer claims to be. Cook, Steinberg, Sawaki, and Eignor (2008) examined a large-scale mathematics test for different groups of learners. They obtained a three-factor structure for all the groups. These factors were: Algebra, statistics, data analysis and probability; Geometry and measurement; and number bases.

In line with the noble results from these studies, there seems to be a lacuna in literature of researches that addressed the use of factor analysis as a statistical technique in the development and validation of Mathematics multiple choice test items, especially for the senior secondary school students in South West Nigeria; which is the crux of this present study. Mathematics is the scientific study of numbers, shapes, structures and change. It is a core subject at all levels of education; especially at the primary and secondary levels and the bedrock for major scientific, economic and industrial development of many nations (Osiesi, 2019).

A mathematics test is one designed to assess students' mathematical skills or capability. It is a tool used in ascertaining whether students have typically learned or acquired the expected mathematical experiences, knowledge and skills as detailed by the curriculum. Thus, students' achievement and performance in the subject can best be measured using mathematics test items. Mathematics teachers may have been facing challenges in developing, adopting or adapting multiple choice test items for students. Most times, the test items they use for assessing these

students may not truly be valid or reliable; necessitating the need for already developed and validated items. The study therefore developed and validated Mathematics Achievement Multiple Choice Test items using factor analysis.

Research Questions

1. Do the MAT test items meet the criteria for factor analysis?
2. What is the strength of the relationship between the MAT test Items?
3. How many factors emerged from the MAT test items?
4. Do the MAT test items have face, construct and content validity?

Methods

The study adopted the instrumentation research. According to Kpolovie (2010), instrumentation research is the science of test development. The study population consisted of all senior secondary school one students (SS1) in Oyo State, Western Nigeria. Purposive and convenient sampling was used in selecting only one hundred SS1 students in Ibadan Metropolis of the state. This was because the test items so developed by the researcher was mainly for the Class (since the items were curbed from the SS1 mathematics textbook) and the researcher was very familiar with the schools, teachers and terrains in the metropolis. These students that was purposefully used to respond to the test items was from the International school, university of Ibadan, Ibadan, Oyo state, Nigeria. A total of twenty test items (as final draft) were administered to these hundred students after pilot testing with an initial fifty test items, which were later on refined and reduced to twenty valid and reliable items, with the help of the KR-20 reliability approach (r = 0.83). The criteria adopted for choosing these twenty items for final inclusion in the final draft of the MAT were: The items with difficulty index range of -1.0 to +0.1 (SSI, 2003), discrimination index range of -3.0 to +3.0 (Ojerinde, Popoola, Ojo & Onyeneho, 2012), and the vulnerability to guessing, not exceeding 0.35 (Opasina, 2009).

The research instrument was the Mathematics Achievement Test (MAT) which has two sections (A & B). Section A captured the demographic information of the students such as school name, age, class and gender. Section B has items intended to determine the cognitive level of achievement of the learners' in Mathematics. The final draft of the

instrument has twenty multiple choice test items of four options lettered A to D. Students' right response to an item was scored 1 while a wrong response was scored 0. This was administered on six different occasions (due to time constraint, as the researcher avoided tampering with the regular school/classroom interaction and lesson schedules of the sampled school) with careful observation and monitoring of students as they respond to the test items. The item sheets were collected after each student has finished and a total of 100 item sheets were counted for completion/verification. Scoring of the items was done on SPSS, version 20. The Data was analysed at 5% significance level.

Presentation of Results

Research Question 1: Do the MAT items meet the criteria for factor analysis?

Table 1: KMO and Bartlett's Test

Kaiser-Meyer-Olkin	**Measure of Sampling Adequacy.**	.67
	Approx. Chi-Square	425.57
Bartlett's Test of Sphericity	Df	190
	Sig.	.00

Table 1 depicts the KMO. It shows the MAT items were suitable for factor analysis as the Kaiser-Meyer Olkins' Measure of sampling adequacy (KMO) was above 0.67; which is an acceptable value for principal components factor analysis.

Research Question 2: What is the strength of the relationship between the MAT test Items?

Table 1 shows the Bartlett's Test of Sphericity which reveals the strength of the relationships among the test items and was found to be significant ($p = 0.00$). This signifies a strong relationship among the MAT test items.

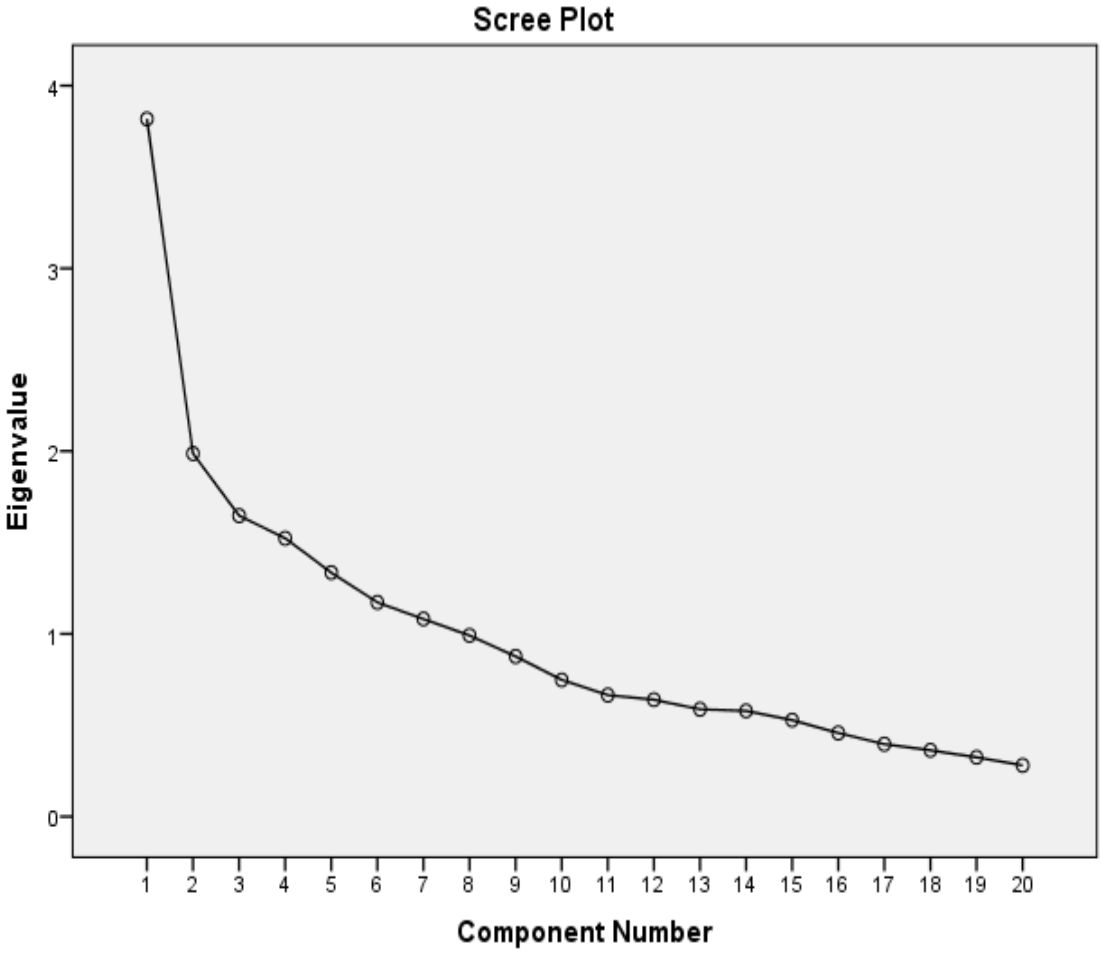
Research Question 3: How many factors emerged from the MAT test items?

Table 2: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.82	19.09	19.09	3.82	19.09	19.09	2.14	10.68	10.68
2	1.99	9.93	29.03	1.99	9.93	29.03	2.06	10.28	20.99
3	1.65	8.24	37.26	1.65	8.24	37.26	1.98	9.90	30.85
4	1.52	7.62	44.88	1.52	7.62	44.88	1.89	9.45	40.30
5	1.35	6.68	51.56	1.34	6.68	51.56	1.61	8.04	48.35
6	1.17	5.86	57.42	1.17	5.86	57.42	1.46	7.30	55.64
7	1.08	5.40	62.83	1.08	5.40	62.83	1.44	7.18	62.83
8	.99	4.96	67.78						
9	.88	4.38	72.17						
10	.75	3.74	75.90						
11	.67	3.33	79.23						
12	.64	3.20	82.43						
13	.59	2.94	85.36						
14	.58	2.89	88.25						
15	.53	2.64	90.89						
16	.46	2.29	93.18						
17	.40	1.98	95.16						
18	.36	1.81	96.98						
19	.32	1.62	98.60						
20	.28	1.40	100.00						

Extraction Method: Principal Component Analysis.

Table 2 depicts the Total variance explained. It shows the number of components to be extracted. Seven (7) components were extracted as they had Eigenvalues that is more than one (3.82, 1.99, 1.65, 1.52, 1.34, 1.17 and 1.08). These seven components explain a total of 62.83% of the variance.



The scores plot portrays the Eigenvalues (quality scores) as displayed above. From the above graph, the foremost 7 components have Eigenvalues greater than 1. This 7 components are termed "strong factors". Immediately after the 7 components i.e from components 8 onwards, the Eigenvalues **gradually drops off**; suggesting 7 factors do underlie the items.

Research Question 4: Do the MAT Test items have face, construct and content validity?

Table 3:
Component
Matrix

Test Items	Components						
	1	2	3	4	5	6	7
Item1	.38	-.53	-.17	.42	.18	.18	.10
Item2	.43	-.31	-.09	.32	-.54	-.11	.05
Item3	.37	-.21	-.52	.07	.52	-.11	-.07
Item4	.44	.19	-.21	-.15	-.44	-.42	.31
Item5	.32	.20	.02	.35	-.27	.49	-.22
Item6	.51	-.19	-.14	-.43	-.02	.11	-.14
Item7	.42	-.09	-.12	-.00	-.16	.44	.37
Item8	.68	-.14	-.01	.11	-.00	-.22	-.02
Item9	.47	-.00	.29	-.49	-.30	-.05	.11
Item10	.56	-164	-.41	.08	-.04	.08	-.26
Item11	.44	.19	-.03	-.23	.09	.37	.38
Item12	.21	-.36	.48	-.30	.38	.19	.25
Item13	.46	-.34	.47	.27	.07	-.32	-.08
Item14	.39	-.17	.57	.23	.07	-.06	.13
Item15	.32	.41	-.19	.21	.36	-.25	.31
Item16	.51	.34	.06	-.05	.20	-.20	-.04
Item17	.46	.62	.15	.16	.15	.22	-.11
Item18	.43	.13	.33	-.05	-.02	.09	-.51
Item19	.50	.15	-.17	-.45	.09	-.10	-.27
Item20	.14	.61	.13	.34	-.01	-.03	.13

Extraction Method: P
Component Analysis
a. 7 components extracted.

Table 4: Rotated Component Matrixa

	Component						
	1	2	3	4	5	6	7
Item1	-.18	.35	-.14	.67	.10	.23	.13
Item2	-.03	.25	-.16	.20	.71	.07	.25
Item3	.21	-.04	.16	.79	-.07	.02	-.18
Item4	.27	-.04	.28	-.09	.74	.15	-.19
Item5	-.02	.01	.09	.05	.10	.18	.75
Item6	.62	.05	-.16	.21	.09	.26	.03
Item7	.02	-.01	-.02	.20	.22	.68	.18
Item8	.32	.40	.19	.34	.35	.06	.06
Item9	.54	.28	-.04	-.32	.26	.32	-.04
Item10	.36	-.03	-.02	.56	.26	.08	.28
Item11	.22	.02	.25	.01	-.03	.68	.04
Item12	.16	.56	-.16	-.02	-.39	.40	-.24
Item13	.09	.80	.04	.17	.17	-.18	.03
Item14	-.03	.74	.11	-.04	.04	.20	.09
Item15	-.03	-.00	.73	.25	.07	.09	-.18
Item16	.38	.17	.52	.07	.05	.02	.03
Item17	.22	.08	.65	-.06	-.12	.14	.48
Item18	.48	.30	.09	-.06	-.10	-.14	.47
Item19	.74	-.07	.16	.14	.04	.04	-.00
Item20	-.17	.01	.64	-.19	.08	.00	.25

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 11 iterations.

Table 3 and 4 showed the seven components which were further subjected to rotated component matrix for the selection of the MAT items which shows the construct validity of the instrument. Nineteen items emerged under the seven structured rotated factor matrix using Varimax with Kaiser Normalization rotation method. To enhance the interpretability of the factors, only variables with factor loadings of $p \geq 0.5$ were considered. Since the rotated factor loading matrixes ranges from 0.52 and 0.80, it can be concluded that the MAT has face, construct and content validity.

Discussion

The finding of this study showed that the MAT items were suitable for factor analysis as the Kaiser-Meyer Olkins' Measure of sampling adequacy (KMO) was above 0.67 and the Bartlett's Test of Sphericity was significant (p= 0.00). This is in support of Ugulu (2011), who asserted that the suitability of the data for factor analysis can best be tested by Kaiser-Mayer-Olkin (KMO) coefficient and Barlett Sphericity Test, and also George and Mallery (2001) on the strength of the relationship among the MAT test items which should be $p < 0.001$.

Another finding of the study was that seven factors were inherent in the items with a total of 62.83% of the variance explained. This finding strongly agrees with the assertion of Tabachnick and Fidell (2001) who reported that a value of around 62.89% explained variability is considered as sufficient variance explanation in social sciences studies. These seven factors are includes number bases, sets, statistics, algebra, geometry, measurement and probability. This finding agrees partly with the findings of Cook et al (2008) that obtained a three factor structure (Number sense; statistics, probability and data analysis; and geometry and measurement) from a Mathematics assessment items used in comparing students with and without disabilities.

More so, the study did reveal that the MAT test items were valid in terms of face, construct and content as the rotated factor loading matrixes ranges from 0.52 and 0.80. In general, the finding is consistent with the findings of Vahedi and Farrokhi (2011); Adediwura

(2011);Pishghadam and Khosropanah (2011); Pallant, 2007 and Ugulu (2013) who reported similar results.

Conclusion

In the item analysis phase of test development, statistical methods such as the factor analysis are used to identify potential item inadequacies. Items that are performing very poorly are to be expunged before the final draft of the items. In other cases, items may still be usable, after modest changes are made to improve their performance on future testing. Teachers should have enough valid and reliable items for the assessment and evaluation of students at one point or the other. Test developers are to make available these items from time to time**.**

Recommendations

This paper recommends the use of factor analysis in test item development and validation over other conventional techniques. Top priority should be appropriated to more researches in mathematics as it is fundamental subject. There should be regular training and development of teachers and examiners on the various statistical techniques (such as the factor analysis) used in developing and validating test items for testing/examination purposes.

Regardless of the contributions arising from this study, it has some limitations, which suggest future studies: The sample size used in this study is small, as such could be expanded for a more robust result and statistical manipulations and convergent and discriminant validity tests between the factors obtained was not done

References

Adediwura A.A (2011). The development and confirmatory factor analysis of a scale for the measurement of gifted students' attitude towards mathematics. World Journal of Education vol . 1, no. 1.

Ajeigbe, T. O. & Afolabi, E. R. I (2013): A study of Item Response and Anchor Bias in Economics Objective Tests among Senior Secondary School Students in Osun State. Academic Journal of Interdisciplinary Studies, Vol. 2 No 3 p. 173-179.

Cook, L. L., Steinberg, J., Sawaki, Y., & Eignor, D. R. (2008). Using factor analysis to compare the

internal structure of a state standards-based math assessment. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.

Ding L. and Beichner R. (2009). Approaches to data analysis of multiple-choice questions**.** Physics Education Research **5**, 020103. DOI: 10.1103/PhysRevSTPER.5.020103

Ellis J. L (2017). Factor analysis and item analysis. Nijmegen.

George D, Mallery P (2001). SPSS for windows: step by step. Allyn & Bacon, USA.

Gorsuch R. (1990). Common factor analysis versus component analysis: Some well and little known facts, Multivar. Behav. Res. **25**, 33.

Hatcher L. (1994). *A Step-By-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling* SAS Institute, Cary, NC.

IRT from SSI (2003). BILOG-MG, MULTILOG, PARSCALE, TESTFACT edited by Math; Ida du toit SSI. Scientific software international.

Kerlinger, F. N. (1979). *Behavioura lresearch: A conceptual approach*. Dallas Holt: Rinehart and Winston.

Kpolovie, P. J. (2010). Advanced research methods. Owerri: Spring Field Publisher Ltd.

Ojerinde D., Poppola K., Ojo F. & Onyeneho P. (2012). Introduction to item response theory. Abuja: Mike Press Ltd.

Opasina O. C. (2009). Development and validation of alternative to practical Physics using item response theory model. Unpublished Ph.D Thesis, University of Ibadan.

Osiesi M. P (2019). Construction, administration and the standardisation of Mathematics multiple choice test items for junior secondary School two (jss 2) in oyo state, Nigeria. British Journal of Education Vol.7, Issue 11, pp.29-40.

Pallant J. (2007). SPSS- Survival Manual, 3rd Edition. A Step by Step Guild to Data Analysis. McCraw Hill, Open University Press.

Pishghadam R. and Khosropanah F. (2011). Examining Construct Validation of the English Language Teachers Competency Test. International Education Studies 4(3). DOI:10.5539/ies.v4n3p194.

Rossoni L., Engelbert R. and Bellegard N.L. (2016). Normal science and its tools: Reviewing the effects of exploratory factor analysis in management. Rev. Adm. (São Paulo) vol.51 no.2 São Paulo. https://doi.org/10.5700/rausp1234.

Tabachnick BC, Fidell LS (2001). Using multivariate statistics. Boston: Allyn & Bacon.

Tryfos, P. (1997). *Methods of business analysis and forecasting: Text and cases*. New York: John Wiley & Sons.

Ugulu I, Aydin H (2011). Research on students' traditional knowledge about medicinal plants: Case study of high schools in Izmir, Turkey. J. Appl. Pharmaceut. Sci. 1(9):43-46.

Ugulu I. (2013). Confirmatory factor analysis for testing validity and reliability of traditional knowledge scale to measure university students' attitudes. Educational Research and Reviews. Vol. 8(16), pp. 1399-1408.DOI: 10.5897/ERR2013.1537.

Vahedi S. and Farrokhi F. (2011). A confirmatory factor analysis of the structure of abbreviated math anxiety scale. Iran J Psychiatry, 6(2): 47–53.