Available online at www.sciencedirect.com

ScienceDirect

Procedia Social and Behavioral Sciences 2 (2010) 5292–5298

Procedia
Social and Behavioral Sciences

WCES-2010

Level determination exam (SBS-2008) the determination of the validity and reliability of 7th grade mathematics sub- test

Duygu Anıl[a] *, Cem Oktay Güzeller[b], Ömay Çokluk[c], Güçlü Şekercioğlu[c]

[a] *Faculty of Education,Hacettepe University, Ankara, 06532, Turkey*
[b] *Faculty of Education,Akdeniz University, Antalya, 07058, Turkey*
[c] *Faculty of Educational Sciences,Ankara University, Ankara, 06590, Turkey*

Received November 11, 2009; revised December 1, 2009; accepted January 22, 2010

Abstract

The aim of this study is to study the validity and reliability of the SBS-2008 7th grade Mathematics sub-test. The psychometric characteristics of the items and tests have been studied with this aim. Exploratory and confirmatory factor analysis along with structural validity has been studied under the validity study whilst the internal consistency (KR-20) reliability and two halves reliability (test halving) has been evaluated. At the end of the exploratory factor, it has been monitored that the 12th item of the Mathematics sub-test consisting of 18 items, works under a single factor. The consistency benefaction indexes obtained at the end of the confirmatory factor analysis shows that the model-data consistency is high. The KR-20 reliability coefficient has been calculated as.72 whereas the reliability coefficient related to the first half of the test has been calculated as .52 and reliability coefficient related to the second half has been calculated as .56.

Keywords: Level determination exam; mathematics test; validity; reliability; exploratory factor analysis; confirmatory factor analysis.

1. Introduction

Decisions on education and psychology are reached based on the results obtained from the testing tools, and the two primary problems of the calculations made are the extent the test can measure the variable which it aims to measure and to what extent is the test usable in order to make correct decisions (Atılgan, Doğan ve Kan, 2007).

Validity and reliability are the most necessary characteristics of the testing tools, which are prepared in order to meet important aims like selection and placement, need to possess in order for them to meet the above stated aims. The degree of test validity and reliability depends on the quality of the test items (Özgüven, 2004).

Tests are often used for educational or psychological decision making (i.e., placement of students in programs). Thus appropriate decision making is highly dependent upon validity of the instruements used. Validity refers to the degree to which an instrument or method measures what it is intended to measure. Therefore, validity has to do with exlusively measuring of the intended construct (Turgut, 1990). A person who uses the testing tools desires to make

* Duygu Anıl. Tel.: +90-242-505-8134873; fax: +90-312-399 20 77
E-mail address: aduygu@hacettepe.edu.tr

1877-0428 © 2010 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.sbspro.2010.03.863

decisions on the individual or indivuals based on the information obtained from the tool. Thus, test developers must first clearly define the aim, the decision or the judgement the testing tool will be used for and then they must organise empiric studies for this aim and must put forward proof of validity based on the content, answering process, result and the structure of the test as well as the relationship it has with other variables (Şencan, 2005; Urbina, 2004).

The concept of validity in the test standards report published in 1974 has been defined under three categories. These are criterion validity which consists of the later joined regression and concordance / immediate validity and scope and structural validity (American Education Research Association, American Psychology Association, National Council for Educational Evaluations, 1997; Urbina, 2004). This classification related to the validity within the test standards published in 1974 has been accepted until today.

Within this classification, structrural validity is the degree of stating structure or the characteristic the test intends to evaluate (Anastasi, 1988). Psychological characteristics are more abstract or latent rather than solid and observable and these are called "construct" (Nunnaly ve Bernstein, 1994). These latent characteristics are called "construct" or "factor" (Kline, 2005). One of the most used techniques in order to find proof related to the construct validity of the tests is factor analysis. Factor analysis can be defined as a multi variable statistic which aims to find and discover a lower amount of new variables (factors, size) that are conceptually meaningful by bringing together a large number of variables related to eachother (Büyüköztürk, 2002; 2007). Factor analysis divides into two as exploratory factor analysis and confirmatory factor analysis which are studied under the structural equation modeling.

According to Jöreskog and Sörbom (1993), exploratory factor analysis is highly useful in test development or in the first phases of the efforts to gain experience for test development. A large number of studies contain both the exploratory and the confirmatory factor analysis as it contains the variables related to the known and unknown situations. The desire is to confirm or refuse the assumptions with confirmatory techniques after the testing of the assumptions with exploratory techniques.

Reliability refers to the degree of sensitivity of the testing instrument or method that can evaluate the variable it tests; in other words, it refers to the degree to which measurement results are free of random errors (Turgut, 1990). Anastasi (1988) has listed the determining reliability methods under four headings; Test and retest method, alternative (equivalent) form development method, halving the test method and using the Kuder Richardson and Cronbach Alfa formulas method.

The Kuder Richardson method is a method used in order to measure the reliability of the tests as dual grading (1 for a correct response and 0 for an incorrect response) and can be held in multiple choice tests and true-false tests (Erkus, 2003). Because dual grading can be conducted in the success tests, it is frequently used in order to determine the internal consistency of the success tests. KR-20 and KR-21 are the two types of the Kuder Richardson method.

While the KR-20 internal consistency reliability is used when the item difficulty indexes of each item taking place in the test is known or can be measured, the KR-21 internal consistency reliability is developed to be used when item difficulty indexes are unknown or can not be measured. The KR-20 reliability coefficient of the tests which affect the lives of the students, like the entrance examimations, shouldn't be lower than .80 (Şencan, 2005).

There are also examinations like the Student Selection Exam (SSE), Entrance Exam for Academic Staff and Post-Graduation Education (EASP), Entrance Examination for Specialty in Medicine (ESM) that are applied in order to choose and place students to the educaiton programs in Turkey. In order to choose and place students to the secondary schools, Secondary Schools Student Selection and Placement Exam (SPE) has been applied from the 1997-1998 fiscal year to the 2007-2008 fiscal year.

As a continuing of the teaching programs development and application studies conducted graded by the Ministry of National Education since 2004, transfer to secondary school system has been re-structured and as an appendage, the Level Determination Exam (LDE) has been brought to practice as of the 2007-2008 fiscal year.

The Level Determination Exam has been first applied in 2008 to the sixth and seventh grades. These exams are central system exams where the achievement levels of the students on the goals stated on the teaching program of the year is measured in the sixth, seventh and eighth grades of the primary education.

In the Level Determination Exams, the content of the questions are prepared based on the education programs and the education and training provided at school. It has been decided that the questions are to be from the year the students are studying and that it shouldn't contain the previous years (www.oges.meb.gov.tr/docs/64_soru.pdf).

Different then the Secondary Schools Student Selection and Placement Exam, Level Determination Exam has a Foreign Language sub-test in addition to the Turkish, Mathematics, Science and Technology, Social Sciences sub tests. The number of questions taking place in the Level Determination Exam sub tests for all of the tests in the 6th, 7th and 8th grades are respectively like the following: Turkish (19,21,23); Mathematics (16, 18, 20); Science and Technology (16, 18, 20); Social Sciences (16, 18, 20) and Foreign Language (13, 15, 17). The total number of questions are 80 for the 6th grades, 90 for the 7th grades and100 for the 8th grades (Celik, 2008).

Instead of being perceived as a new exam system in terms of its characteristics, Level Determinatioın Exams should be perceived as one of the elements of the new system that is the Transfer to Secondary School System (TSS). In the new system that has been established with the TSS, the continuing of the student motivation for three years and paying attention to all of the lessons in a suitable way according to the primary education spirit is aimed. TSS consits of three elements which are the Level Determination Exam (LDE), End of Year Success Score (EYSS) and Attitude Score (AS). The End of Year Success Score (EYSS) show the scores the students get from all of the lessons at school whereas the Attitude Score (AS) shows the scores the students get from the evaluation of their adaptation to the school rules and attitude qualities like the productive study (LDE Manual, 2009). According to this, %70 of the scores obtained from the Level Determination Exams applied at the end of the 6,7 and 8th grades, %25 of the end of year success scores of the students and %5 of the attitude scores are added up and a “class score” is calculated. A testing and evaluation based on multi dimension performance indicators is aimed (LDE Manual, 2009) with this application. However, upon the stay of execution decision of the Council of State 8th Chamber on March 2, 2009 for the %5 effect of the attitude scores to the class scores, the Ministry of National Education has put into practice the decision of excluding the %5 attitude score on the calculation of the class scores. According to this application, a new method of calculation or evaluation related to the Level Determination Exams has not been developed but the calculations have been made by excluding the attitude scores in the evaluation (http://www.meb.gov.tr/duyurular/duyuruayrinti.asp?ID=2994, http://ceylanpinar.meb.gov.tr/duyuru/duyurular.php?announcements=84).

“Placement to Secondary School Score” is obtained by taking %25 of the above stated class scores of the 6th grades, %35 of the 7th grades and %40 of the 8th grades. With this score the students are placed to science high schools, anatolian highschools, social sciences highschools, anatolia technical highschools, anatolia vocational highschools, anatolian teacher highschools, anatolian islamic divinity students high-school, anatolian agriculture and agriculture vocational highschool, anatolia land registry and cadastre vocaitonal highschools (LDE Manual, 2009).

In accordance with these discussions, examining the validity and the reliability of the Level Determination Exam-2008 7th grade Mathematics sub test constitute the problem of this study. Answers to the questions on the level of the psychometric characteristics (item difficulty and discrimination strength indexes) of the items and then the psychometric characteristics of the test (structural validity, internal consistency -KR-20 and two halves reliability) have been searched.

2. Method

2.1 Research Model

The general aim of this research is to study the validity and reliability of the 2008 Level Determination Exam 7th grade Mathematics sub test. It is in a descriptive research quality aiming to put forward the situation current in the research.

2.2 Population and Sample

962991 7th grade students entering the Level Determination Exam first conducted in 2008 constitute the population of the research. 5000 students selected from this population random constitute the sample of the research. Although the data related to all of the population is in hand, because the data create difficulties in the analysis, the need for choosing a sample has been felt necessary and it has been limited to 5000 students.

2.3 Analysis of the Data

In order to determine the item level psychometric characteristics of the 2008 Level Determination Exam 7th grade Mathematics sub test, item difficulty and item discrimination strength have been calculated. Double series correlation coefficient has been used in the calculation of the item discrimination strength indexes. In order to determine the structural validity under the validity studies, basic component analysis based on the tetrachoric correlation coefficient has been applied. Whether the factor structure obtained is confirmed or not has been tested with a confirmatory factor analysis. KR-20 and two halves reliability (reliability with the test halving method) have been calculated under the reliability studies.

3. Results (Findings)

The determination of the psychometric characteristics of a test in a classic Test Theory is studied in two groups; the psychometric characteristics of the items and the test. First, the items and then the psychometric characteristics related to the test have been put forward in the presentation of the findings in this part.

3. 1 2008 Findings and Comments Related to the Psychometric Characteristics of the Level Determination Exam Mathematics Sub Test Items

Item difficulty and item discrimination strength indexes of the Mathematics sub test items have been presented in Table 1

Table 1. Item Difficulty and Item Discrimination Strength Indexes of the 2008 Level Determination Exam Mathematics Sub Test Items

Items	Item difficulty indexes (pj)	Item discrimination stregth indexes
1	.59	.42
2	.38	.51
3	.38	.59
4	.33	.51
5	.30	.26
6	.70	.23
7	.47	.50
8	.35	.41
9	.66	.49
10	.14	.08
11	.27	.18
12	.43	.67
13	.37	.53
14	.48	.58
15	.22	.29
16	.52	.63
17	.37	.46
18	.33	.40

When Table 1 is studied; it is seen that item difficulty indexes of the 18th item value taking place in the Mathematics sub test varies between .14 and .70. It is seen that the hardest item for the student group is the 10th item whereas the easiest item is the 6th. It is seen that the values are between .08and .67 when the item discrimination strength indexes, which are the indicators on how successful a test item is in distinguishing the students who know and the students who do not know, are examined. When the fact that the items giving a correlation of .30 and above between the test score and the item score are the most contributing items evaluating the quality of the intended measuring, it is seen that the 10th and 11th items are not distinguishing items. It is also seen that the 5th, 6th and the 15th items also give a correlation value below .30. If these findings were the findings obtained at the end of a trial application conducted in order to develop a test, then the 10th and 11th items would have been excluded from the final form and the 5th, 6th and the 15th items would have been corrected and included in the final form. The 12th item is the item with the highest discrimination in the test.

3.2 Findings and Comments Related to the Psychometric Characteristics of the 2008 Level Determination Exam Mathematics Sub Test

3.2.1 Findings and Comments Related to the Validity of the Mathematics Sub Test

In this part, first the findings related to the results of the exploratory factor analysis applied in order to test the structural validity of the Mathematics sub test and then the results related to the confirmatory factor analysis have been put forward. Before moving on to the exploratory factor analysis application, Kaiser-Meyer-Olkin (KMO) value which tests the appropriateness of the data to factorizing and Barlett Globosity Test results has been studied. The KMO value related to the data has been found as .88 and the Barlett Globosity Test has come out to be meaningful. As Büyüköztürk (2007) states, because the KMO being higher than 0.60 and the Barlett test being meaningful is accepted as an indicator of the data being suitable for factorizing, factor analysis application has continued. It has been determined that mathematics sub test items take a factor load value varying between -.62 and .86 under one factor at the end of the exploratory factor conducted based on the basic component analysis factorizing technique and the tetrachoric correlation matrix. It has been determined that the 12th item of the Mathematics sub test consisting of 18 items works under a single factor but the 5,6,8,10,11 and 15th items give a negative and low factor load values below this factor. When table 1 is studied once again, it can be seen that, apart from the 8th item, the item discrimination strength indexes of the above stated five items are also low. Moreover, it has been determined that the variance rate announced related to the Mathematics sub test is %26.9.

After the exploratory factor analysis, a confirmatory factor analysis related to the 2008 Level Determination Test Mathematics sub test has been applied and the diagram related to the above stated analysis has been presented in Figure 1.

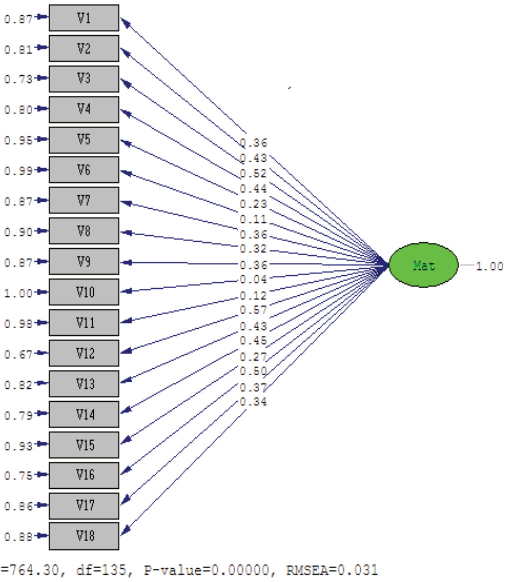

Figure 1. Diagram on the Confirmatory Factor Analysis Related to the 2008 Level Determination Exam Mathematics Sub Test

The most frequently used among the indexes calculated related to the confirmatory factor analysis and model-data consistency are x-square (χ2), χ2/sd, GFI, AGFI, RMR and RMSEA. The calculated χ2/sd rate being lower than 5, the GFI and AGFI values being higher than .90 and the RMR and RMSEA values being lower than .05 show a perfect model-data consistency (Jöreskog and Sörbom, 1993; Marsh and Hocevar, 1988). The GFI being higher than 0.85, AGFI being higher than 0.80, RMR and RMSEA values being lower than 0.10 are accepted as acceptable sub limits for the model-data consistency (Anderson & Gerbing, 1984; Cole, 1987; Marsh, Balla & McDonald, 1988).

The consistency benefaction indexes obtained as a result of the confirmatory factor analysis related to the 2008 Level Determination Exam Mathematics sub test are; χ2= 764.30, sd=135, P<.05, χ2/sd=5.66, CFI=.06, NFI=.95, AGFI=.98, IFI=.96, SRMR=.027, RMSEA=.031.The consistency index being high (AGFI=.98, CFI=.96, NFI=.95, IFI=.96, SRMR=.027 and RMSEA=.031) shows that the model-data consistency is high.

3. 2.2 Findings and Comments Related to the Reliability of the Mathematics Sub Test

a. Internal consistency (KR-20)

In order to estimate the reliability of the Mathematics sub test according to the Classical Test Theory, the KR-20 reliability coefficient has been calculated by means of the item variances and test variances taking place in the test. The Reliability coefficient value of the test has been calculated as .72. Although this value is not so high, when the fact that the coefficient obtained has internal consistency is taken into account, it could be said that the scale makes consistent estimations within itself and that it reliable in terms of internal consistency.

b.Split-Half Method

It has been calculated that the reliability coefficient related to the first half of the Mathematics sub test is .56 and the reliability coefficient related to the second half is .56. The Spearman Brown reliability coefficient value of the whole of the test has been calculated as 73. Apart from that, the Guttman internal consistency coefficient value which is one of the test halving methods has been calculated as .73. With the fact that the reliability coefficient calculated with the internal consistency method and the test halving method having close values, it could be stated that it is at an acceptable level especially when the number of items in the test is taken into account.

4. Conclusion and Recommendation

First, the examination of the reliability and the validity of the 2008 Level Determination Exam 7th Grade Mathematics sub test is aimed in this study. With this aim, it has been seen that the item difficulty index of the 18 items taking place in the Mathematics sub test varies between .14 and .70, item discrimination strength indexes vary between .08 and .67 and that the five items don’t have the desired discrimination level. When the items with a low difficulty level are examined it has been determined that it has been developed in order to test the high level skills. This situation is consistent with the 2000 and 2001 OKS, PISA and TIMMS results (Kutlu and Karakaya, 2007, MEB, 2003b; MEB, 2005). Secondly, at the end of the exploratory factor analysis, it has been determined that mathematics sub test has a single factor analysis and that the 6 items have a negative or a low factor load values. Apart form the 8th item, this result is consistent with the discrimination strength indexes. It has been determined that the announced variance rate related to the mathematics sub test is %26.9. This situation is an indicator that about one third of the test can be explained by this factor and the remaining can be explained by other variables. Thirdly, the construct obtained at the exploratory factor analysis has been examined with confirmatory factor analysis and it has been determined that it has high consistency indexes. Fourthly and finally, the KR-20 and two halves reliability coefficient have been calculated respectively and measured as .73 and .72. These results are at the acceptable limit level. When the fact the reliability coefficient value is affected by the number of indexes making up the test is taken into account, it could be stated that the increasing number of items within the test will contribute to the reliability coefficient.

References

American Education Research Association, American Psychology Association, Education Estimaitons International Council (1998). Testing Standards in Education and Psychology. (Tra. S. Hovardaoğlu and N. Sezgin). (First Edition). Ankara: Türk Psikologlar Derneği Yayınları. No 14.

Anastasi, A. (1988). Psychological Testing. (Sixth Edition). New York: Macmillan Publishing Company.

Anderson, J. C. and Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin,163 (3), 411–423.

Atılgan, H., Doğan, N. and Kan, A. (2007). Eğitimde Ölçme ve Değerlendirme (2. Edition). Ankara: Anı Yayıncılık.

Baykul, Y. (2000). Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması. (First Edition). Ankara: ÖSYM Yayınları.

Büyüköztürk, Ş. (2007). Sosyal Bilimler için Veri Analizi El Kitabı: İstatistik, Araştırma Deseni, SPSS Uygulamaları ve Yorum. (Seventh Edition). Ankara: PEGEM A Yayıncılık.

Büyüköztürk, Ş. (2002). Faktör Analizi: Temel Kavramlar ve Ölçek Geliştirmede Kullanımı. Kuram ve Uygulamada Eğitim Yönetimi. Volume 32, 470–483.

Cole, D.A. (1987). Utility of confirmatory factor analysis in test validation research. Journal of Consulting and Clinical
Psychology, 55, 1019-1031.

Cronbach, L. J. (1970). Essentials of Psychological Testing. New York: Harper & Row Publishers.

Çelik, H. (2008). SBS İle İlgili Basın Toplantısı.http://www.meb.gov.tr/haberler/haberayrinti.asp?ID= 1254.

Erkuş, A. (2003). Psikometri Üzerine Yazılar. (First Edition). Ankara: Türk Psikologlar Derneği Yayınları. No 24.

http://www.meb.gov.tr/duyurular/duyuruayrinti.asp?ID=2994

http://www.oges.meb.gov.tr/docs/64_soru.pdf).

http://ceylanpinar.meb.gov.tr/duyuru/duyurular.php?announcements=84).

Jöreskog, K. G. and Sörbom, D. (1993). Lirsel 8: Structural Equation Modeling with the Simplis

Command Language. Lincolnwood: Scientific Software International, Inc.

Kline, R. B. (2005). Principles and Practice of Structural Equation Modeling.(Second Edition). NY: Guilford Publications, Inc.

Marsh, H. W., Balla, J. R. & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103 (3), 391–410.

MEB. (2003b). 1999 TIMSS Uluslararası Matematik ve Fen Bilgisi Çalışması Ulusal Raporu, Ankara:

Eğitimi Araştırma Geliştirme Dairesi Başkanlığı (EARGED) Yayını.

MEB. (2005). PISA 2003 Projesi Ulusal Nihai Raporu, Ankara: Eğitimi Araştırma Geliştirme Dairesi Başkanlığı (EARGED) Yayını.

MEB (2008). Seviye Belirleme Sınavı (SBS) Kılavuzu. http://www.oges.meb.gov.tr/sbs/sbskilavuz. htm

Nunnally, J. & Bernstein, I. (1994). Psychometric theory. New York: McGraw Hill, 3rd ed.

Özgüven, İ. E. (2004). Psikolojik Testler. Ankara: PDREM yayınları.

Şencan, H. (2005). Sosyal ve Davranışsal Ölçümlerde Güvenilirlik ve Geçerlilik. (First edition). Ankara: Seçkin Yayınları.

Turgut, M. F. (1990). Eğitimde Ölçme ve Değerlendirme Metotları (7th edition. Ankara: Yargıcı Matbaası.

Urbina, S. (2004). Essentials of Psychological Testing. (First Edition). NJ: Wiley & Son