THE SCALE FOR TENDENCIES OF USING ALTERNATIVE ASSESSMENT APPROACHES: RELIABILITY AND VALIDITY STUDIES

Sevilay Kilmen

İlker Kösterelioğlu

Meltem Kösterelioğlu

Abstract

In this research, the purpose was to develop "The Scale for Tendencies of Using Alternative Assessment Approaches" and determine validity and reliability of the scale in a group of teachers. At the development stage of the scale, a scale has been prepared by reviewing the related literature and interviewing with teachers. The construct validity of scale was tested by Principal Component Analysis. The results showed that the scale had three factors. Item total correlations were also conducted, together with independent samples test in order to determine the meaningful difference between the top and the bottom %27 groups. The reliabilities of the subscales were analyzed using Cronbach Alpha. The results all showed that the scale is a valid and reliable instrument and have good psychometric properties.

Key words: alternative assessment, developing scale, validity, reliability.

Introduction

Assessment is the inseparable parts of education process. In assessment process, teachers are using traditional and alternative assessment approaches. In Turkey, Primary education program was changed in 2004 adapting the program to constructivist approach. Alternative assessment methods were added to traditional assessment and evaluation methods. But, researches were showed that teachers didn't have enough competence and knowledge to use these methods. Moreover, they were using these methods rarely (Ercan and Altun 2005; Güven; 2008; Kaplan, 2007; Kilmen and Demirtaşlı, 2009; Kutlu, 2005; Şekel 2007; Şenel Çoruhlu, Er Nas and Çepni, 2008; Özsevgeç, 2007).

According to Şenel Çoruhlu, Er Nas and Çepni (2009), teachers started to use alternative assessment with little theoretical and practical knowledge about alternative assessment techniques and teachers had problems because of lack of information and skills about techniques such as; composing and evaluating of portfolios. However, teachers hadn't got enough competence about alternative assessment techniques and they started to adapt traditional techniques to the new education program. In a research conducted by Adanalı and Doğanay (2010), research findings indicated that fifth grade teachers tended to use both alternative and traditional measurement and evaluation tools in the social studies. However, they declared that traditional measurement and evaluation methods have been used more than alternative methods. Moreover, results showed that shortage of time, inappropriate school environment and opportunities, equipment and material deficiencies, lack of family interest were among the problems have been encountered for teachers and students.

In some researches, teachers viewed themselves as more adequate for the traditional measurement method and they frequently experienced problems in the implementation assessment procedures. Most teachers also reported that they did not have enough time to perform most of student assessment work prescribed in the curriculum. Other problem reported by teachers is lack of knowledge about alternative assessment techniques and insufficient time (Çelikkaya, Karakuş and Öztürk Demirbaş, 2010; Gelbal and Kelecioğlu, 2007; Kartallıoğlu, 2005; Korkmaz, 2006)

In Turkish measurement and evaluation literature, there are some scales about alternative assessment (Tekindal, 1997; Erdoğdu, 2010; Çalışkan and Yazıcı 2013; Aktaş and Aktaş, 2012). But they are attitudes scales about alternative assessment. Lacking of tendencies of using alternative assessment approaches scale was noticed. So it was decided to develop scale about tendencies of using alternative assessment approaches. In this research, the purpose was to develop "The Scale for Tendencies of Using Alternative Assessment Approaches" (TUAAA) and determine validity and reliability of the scale in a group of teachers. Thus, it was aimed at developing a scale for identifying the teachers' tendencies of using alternative assessment approaches.

In this article, the development process of the scale is explained, possible explanations of the findings are discussed and interpretation of the use the results are provided.

Research Group

Participants of this study were 322 teachers working in Bolu. The data was collected in the 2012-2013 fall semester. Demographic information of the sample was given in Table 1.

Table 1. Demographic Information of Teachers.

		1-5	6-10	11-15	16 years more	Total
Classroom teache	Male	2	20	28	19	69
	Female	16	36	30	4	86
	Total	18	56	58	23	155
Branch teachers	Male	21	47	14	5	87
	Female	46	30	3	1	80
	Total	67	77	17	6	167
Total	Male	23	67	42	24	156
	Female	62	66	33	5	166
	Total	85	133	75	29	322

When Table 1 examined, of the participants, 156 were male while 166 were female. 155 of them were classroom teachers; 167 of them were branch teachers. Most of them had 6-10 years seniority.

Development Process of the Scale

In this study, the purpose was to develop "The Scale for Tendencies of Using Alternative Assessment Approaches" and determine validity and reliability of the scale in a group of teachers. While determining the validity and reliability of the test, factor analysis, item difficulty, item discrimination, item total correlation were taken into consideration. In this part, the findings of the study were described and interpreted in the framework of the purpose of the study.

In the development process of scale, related literature is examined. In Turkish measurement and evaluation literature, there were some scales about attitudes towards to measurement and evaluation or alternative assessment (Tekindal, 1997; Erdoğdu, 2010; Çalışkan and Yazıcı 201; Aktaş and Aktaş, 2012) but lacking of tendencies of using alternative assessment approaches scale was noticed. So it was decided to develop scale abaut tendencies of using alternative assessment approaches. The instrument was constructed after reviewing relevant research articles in the field. Also, 21 teachers' opinions and comments were taken in order to determine the scale statements of the scale with a semi structured interview form. each interview was recorded and finished 40 minutes. After interview was analyzed, item pool was generated 38 items on a five point Likert type items such as "strongly agree", "agree", "undecided", "disagree" and "strongly disagree". For the content validity, the draft scale was given to three experts in measurement and evaluation for taking their opinions about whether the selected items were valid items for assessing teachers' tendencies of using alternative assessment approaches. After the expert reviewing process, 6 items were deleted because; the experts founded them unclear or not suitable. Final form of the scale with 32 items was administered to 322 teachers to determine item quality, construct validity and reliability.

Findings

The data gained from the scale were analyzed with SPSS 20.0 program. For item analysis, p value and t tests between items' means of upper and lower 27% points of the scale were calculated in order to determine item validity. p value and t values were given in Table 2.

Table 2. t and p values of upper and lower groups.

İtem number	t	p	İtem number	t	p	İtem number	t	p
I1	,932	,353	I12	3,749	,000	I23	-10,560	,000
I2	-9,673	,000	I13	-11,314	,000	I24	-13,015	,000
I3	3,392	,001	I14	4,642	,000	I25	-5,607	,000
I4	-11,954	,000	I15	-10,799	,000	I26	3,797	,000
I5	1,985	,055	I16	-9,344	,000	I27	3,581	,000
I6	-12,194	,000	I17	3,676	,000	I28	-13,251	,000
I7	-13,907	,000	I18	-12,511	,000	I29	2,311	,022
I8	3,092	,002	I19	-16,705	,000	I30	-13,679	,000
I9	-12,926	,000	I20	-13,584	,000	I31	-15,331	,000
I10	1,906	,058	I21	-11,126	,000	I32	-15,006	,000
I11	-11,165	,000	I22	7,144	,000			

When the figure was revised, the t test results showed significant differences between each item's means of upper 27% and lower 27% points except from items 1, 5 and 10. These items were removed from the scale. For item discriminations, item total correlations were calculated. Item total correlations were given in Table 3.

Table 3. Item total correlations.

İtem number	r	İtem number	r	İtem number	r
I2	,457	I14	,570	I24	,694
I3	,357	I15	,716	I25	,179
I4	,559	I16	,467	I26	,567
I6	,607	I17	,550	I27	,584
I7	,711	I18	,625	I28	,747
I8	,443	I19	,738	I29	,424
I9	,630	I20	,739	I30	,711
I11	,646	I21	,617	I31	,693
I12	,522	I22	,639	I32	,595
I13	,576	I23	,653		

As seen from the Table 3, item total correlations ranged from ,357 to ,747 except item 25. According to Büyüköztürk (2011), when item total correlations are 0,30 and higher, it means that item can discriminate individuals. Generally items had very high item discriminations. It means that, items discriminates high and low tendencies toward of using alternative assessment approaches. Item 25 had very low item discrimination. So, this item removed from the scale.

Exploratory factor analyses were conducted to examine the construct validity. To aid in the interpretation of the components, because of the relationship between the factors, promax rotation was performed. Kaiser-Meyer-Olkin value was 0,931, exceeding the recommended value of 0,600 and Barlett's Test of Sphericity reached statistical significance ($\chi^2 = 2963,689$; p=0,00), supporting the factorability of the correlation matrix. So the data were adequate exploratory factor analyses (Field, 2009). Principal component analysis revealed a structure with items clustered into three factors. The three factor solution explained 55,825 per cent of the variance, with factor 1 contributing 40,843 per cent, factor 2 contributing 9,206 per cent, factor 3 contributing 5,775 per cent.

Table 4. Factor Structures and Loadings of TUAAA

Factor 1		Factor2		Factor3	
Item number	Foctor loadi	Item number	Factor load	Item number	Factor load
I26	,802	I16	,828	I6	,868
I27	,774	I19	,722	I4	,776
I29	,749	I18	,710	I7	,624
I17	,734	I21	,705	I9	,566
I8	,620	I20	,680	I2	,552
I14	,600	I32	,564		
I12	,579	I28	,469		
I22	,523				
Cr α	,856	Cr α	,872	Cr α	803

From table 4, after expletory factor analyses process, many items were it was seen distribution of 20 items to three factors. Factor loading of the items ranged from 0,469 to 0,868. Factor 1 includes eight items: 26, 27, 29, 17, 8, 14, 12, and 22. These items measure teachers' opinions towards requirement of the use of alternative assessment approaches. This factor was named as ''requirement. Factor 2 includes seven items: 16, 19, 18, 21, 20, and 32. These items measure teachers' opinions about the "importance" of the use of alternative assessment approaches. This factor was named as "importance". Factor 3 includes five items: 6, 4, 7, 9, and 2. These items measure teachers' opinions towards usefulness of the use of alternative assessment approaches. Also, it was seen that from Table 4, Cronbach alpha value of factor 1 is 0.856, factor 2 is 0.872, and factor 3 is 0.803. According to these results, it can be said that The Scale for Tendencies of Using Alternative Assessment Approaches is a reliable scale.

Result and Recommendations

In this research, The Scale for Tendencies of Using Alternative Assessment Approaches was developed. The Scale for Tendencies of Using Alternative Assessment Approaches was administered to 322 teachers. 156 were male while 166 were female. 155 of them were classroom teachers; 167 of them were branch teachers. The scale's item total correlations were computed. Results showed that each items' with total ranged from 357 to 747. According to factor analysis results, the scale consist three factors. Factor loading of the items ranged from 0,469 to 0,868. Factor 1 includes eight items: 26, 27, 29, 17, 8, 14, 12, and 22. These items measure teachers' opinions towards requirement of the use of alternative assessment approaches. This factor was named as ''requirement. Factor 2 includes seven items: 16, 19, 18, 21, 20, and 32. These items measure teachers' opinions about the "importance" of the use of alternative assessment approaches. This factor was named as "importance". Factor 3 includes five items: 6, 4, 7, 9, and 2. These items measure teachers' opinions towards usefulness of the use of alternative assessment approaches. Cronbach alpha values were calculated as 0.856 for factor 1, as 0.872 for factor 2 and as 0.803 for factor 3. The results of item total correlations and Cronbach alpha have shown that the scale was a highly reliable scale to measure tendencies of using alternative assessment approaches. Thus the general evaluation of these findings suggested that the scale has acceptable reliability and validity. Further research is needed to extend the results into new samples.

References
Adanalı, K., Doğanay, A. (2010). The Evaluation of Alternative Assessment Practices in Fifth Grade Social Studies Instruction. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 19* (1), 271-292.

Büyüköztürk, Ş. (2011). *Sosyal Bilimler İçin Veri Analizi El Kitabı*, Pegem Akademi Yayınları, Ankara.

Çalışkan, H., & Yazıcı, K. (2013). Developing an attitude scale assessment and evaluation and analyzing of social studies teachers attitude levels in terms of several variables. *International Journal of Human Sciences , 10* (1), 398-415.

Çelikkaya, T., Karakuş, U. Öztürk Demirbaş, Ç. (2010). Utilization Levels of Teachers of Social Studies in Assessment –Evaluation Tools and the Problems They Met. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi, 11* (1), 57–76.

Erdoğdu, M. Y. (2010). Öğretmenlerin ölçme ve değerlendirmeye yönelik tutumlarının bazı değişkenler açısından incelenmesi. *International Conference on New Trends in Education and Their Implications*. Antalya: 11-13 November.

Field, A. (2009) *Discovering Statistics using SPSS*. (3rd ed.). London: Sage.

Gelbal, S. ve Kelecioğlu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 33*, 135-145.

Gül, E. ve Doğan, Ç. (2011). Online Değerlendirme Güvenilir midir?, *5th International Computer & InstructionalTechnologies Symposium*, 22-24 September 2011, Fırat University, Elazığ, Turkey.

Güven, S. (2008), Sınıf öğretmenlerinin yeni ilköğretim ders programlarının uygulanmasına ilişkin görüşleri, *Milli Eğitim Dergisi, 177*, 224–236.

Ercan, F. ve Altun, S. A., (2005). "İlköğretim Fen ve Teknoloji Dersi 4. ve 5. Sınıflar Öğretim Programına İlişkin Öğretmen Görüşleri", *Eğitimde Yansımalar: VII Yeni İlköğretim Programlarını Değerlendirme Sempozyumu*, Erciyes Eğitim Fakültesi, Erciyes.

Kaplan, S. (2007). *Sınıf Öğretmenlerinin Ölçme Ve Değerlendirme İlkelerinin Önem Ve Uygulama Düzeylerine İlişkin Görüşleri*. Ankara Üniversitesi Eğitim Bilimleri Enstitüsü. Yayımlanmamış yüksek lisans tezi.

Kartallıoğlu, F. (2005). *Yeni ilköğretim programlarının uygulandığı pilot okullardaki öğretmenlerin yeni program ve pilot çalışmalar hakkındaki görüşleri*. Yüksek Lisans Tezi, Abant İzzet baysal Üniversitesi Sosyal Bilimler Enstitüsü, Bolu.

Kilmen, S. ve Çıkrıkçı-Demirtaşlı, N. (2009). The Perceptions of primary school teachers about their application levels of measurement and evaluation principles. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*. *42* (2), 027-054.

Korkmaz, İ. (2006). *Yeni ilköğretim programının öğretmenler tarafından değerlendirilmesi*. Ulusal Sınıf Öğretmenliği Kongresi, Cilt II (s.249-260). Ankara: Kök Yayıncılık.

Kutlu, Ö., (2005). Yeni İlköğretim Programlarının 'Öğrenci Başarısındaki Gelişimi Değerlendirme, *Eğitimde Yansımalar: VIII Yeni İlköğretim Programlarını Değerlendirme Sempozyumu*, Erciyes Üniversitesi, Eğitim Fakültesi, Kayseri.

Özsevgeç, T. (2007). İ*lköğretim 5. Sınıf Kuvvet ve Hareket Ünitesine Yönelik 5E Modeline göre Geliştirilen Rehber Materyallerin Etkililiğinin Belirlenmesi*, Doktora Tezi, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü,Trabzon.
 Şenel Çoruhlu, T., Er Nas, S. & Çepni, S., (2008). Fen ve Teknoloji öğretmenleri için alternatif ölçme ve değerlendirme tekniklerine yönelik bir HİE programından yansımalar: Trabzon örneği, *Necatibey Eğitim Fakültesi Dergisi, 2* (2), 1-22.

Tekindal, S. (1997). Ölçme ve Değerlendirmeye İlişkin Tutum Aracı Geliştirme. Samsun: Cem Ofset.