Journal of Institute of Social Sciences
Karabuk University

Year : 2018
Volume : 8
Issue : 2

A Statistical Approach to Improve the Standard Setting Process: Cluster Analysis

Ceylan Gündeğer*, Nuri Doğan[a]

[a]Haccettepe University, Faculty of Education, Ankara.

Abstract

This study aims to use k-means cluster analysis to improve the standard setting process, to determine the passing score in two samples by three methods, to examine the validity of the cluster analysis results using an external criteria, and to compare the cluster analysis results with those produced by Angoff, Yes/No, and Ebel test centered methods. In this study, the main sample consisted of 305 students and the validation sample consisted of 179 students. The data set is composed of the students' responses to a 20-item achievement test. On the other hand, the number of judges determining the passing score according to the test centered standard setting methods was 17. A moderate correlation was found between the external criteria and the results of cluster analysis for the validation sample. Medium and highly significant relationships were observed between the different statistical methods for determining the passing score. According to the study results, in order to achieve the highest relationship with the test centered standard setting methods' results, the following methods could be proposed respectively: determine the passing score based on the range comprising the lowest score of the first cluster and the highest score of the second cluster; logistic regression and the average score of successful cluster.

Keywords: *Cluster analysis, k-means, passing score, standard setting, logistic regression*

Standart Belirleme Sürecini Güçlendirmede İstatistiksel Bir Yaklaşım: Kümeleme Analizi

Öz

Bu araştırmada standart belirlemede k-ortalama kümeleme analizinin kullanılması; iki farklı örneklemde üç yöntemle kesme puanı belirlenmesi; dışsal kriter kullanarak sonuçların geçerliğinin incelenmesi; kümeleme analizi sonuçlarıyla Angoff, Yes/No ve Ebel yöntemlerinden elde edilen sonuçların ilişkisinin belirlenmesi amaçlanmıştır. Bu çalışmada ana örneklem (N=305) ve geçerleme örneklemi (N=179) olmak üzere iki örneklem kullanılmıştır. Veri seti 20 maddelik başarı testine verilen öğrenci cevaplarından oluşmaktadır. Test merkezli yöntemlerle kesme puanı belirleyen uzman sayısı ise 17'dir. Bunun yanında geçerleme örnekleminde dışsal kriter ile kümeleme analizi sonuçları arasında 0,01 hata ile orta düzeyde ilişki olduğu görülmüştür. Kesme puanı belirleme yöntemleri arasında orta ve üstü düzeyde manidar ilişkiler olduğu görülmüştür. Çalışma sonuçlarına göre, test merkezli standart belirleme yöntemlerinin sonuçlarıyla en yüksek ilişkiyi elde etmek için sırasıyla şu yöntemler önerilebilir: başarılılar kümesine atanan en düşük puan ile başarısızlar kümesine atanan en yüksek puanın oluşturduğu ranjın ortancasının belirlenmesi, lojistik regresyon ve başarılılar kümesine atanan bireylerin toplam puanlarının ortalaması.

Anahtar kelimeler: *Kümeleme analizi, k-ortalama, geçme puanı, standart belirleme, lojistik regresyon*

**Yazışma adresi. Email: cgundeger@gmail.com*

1. INTRODUCTION

Various tests have been used to determine students' achievement, to recruit individuals, and to select the students in Turkey. These tests help to identify whether the students have the desired qualities or not and to determine important decisions about these students. Success in these tests can be evaluated by an absolute criteria which is determined by test centered (as Angoff, Yes/No, Ebel) or performance based (as Borderline-group method) standard setting methods. Jaeger (1989) argued that more than thirty standard setting methods were described in the literature during the past few years and some studies were conducted to compare these methods in terms of their passing scores or their correspondencewith each other.

Standard setting methods aim to determine the passing score (standard, cut-off score or cut-off point) that separates the successful students and the unsuccessful students. In passing score determining or standard setting studies we can use some statistical approaches directly as cluster analysis which is used to reduce the number of variables (Hess, Subhiyah&Giordano, 2007). In this regard, cluster analysis can be used "alone" to determine the passing score when we have experts/judges to validate the analysis results based on the external criteria (Sireci, 2001). In standard setting process, the cluster analysis aims to cluster the students (or persons) with respect to the students' raw data and as it is data-driven it does not require judges or experts of the matter, as do most other standard setting methods.

This paper aims to determine the passing score in two samples which are the main (study) sample and the validation sample by three statistical standard setting methods which are the results of the k-means cluster analysis and mentioned here as the average method (AM), the range median method (RMM) and Logistic Regression (LR). Moreover this study examines the validity of these statistical standard setting methods using an external criteria which is the school grades of the students, to compare these statistical methods with the test centered standard setting methods (Angoff, Yes/No and Ebel methods).

Standard setting can be defined as the process of determining the passing score. The passing score indicates the proper performance point that can distinguish the students with the desired level of performance from the students at an insufficient level. Essentially, standard setting is essential to determine the criteria, based on which decisions will be made regarding students' performance. In particular, there are several standard setting methods in the literature. As the present study focuses on the test centered (Angoff, Yes/No and Ebel) and the statistical (mentioned above as AM, RMM and LR) standard setting methods, these methods are briefly outlined below.

1.1. Angoff Method

This test centered standard setting method is used to determine the possibility of whether a student with minimum competency, at the passing–failing limit (calling borderline group), answers the item correctly. The judges in the Angoff method consider the items as a whole. The easier the questions are, the more the probability of

the answering item correctly increases. If the probability is considered zero for a multiple choice test item, then the chance score (1/the number of the choices) for that test item should be considered as a passing score. The judges' minimum passing score (MPS) is calculated by adding the percentages assigned to the individual items (Livingston&Zieky, 1982).

1.2. Yes/No Method

In Yes/No method, the judges are asked to examine the items in the test and determine whether a student at the borderline group can correctly answer the item or not. If the judge believes that the student at the borderline group can correctly answer the item, the item is assigned score 1, if not then it is assigned score 0, and then the MPS is calculated for each judge (Zieky, 2001).

1.3. Ebel Method

Unlike the other two test centered standard setting methods above, the judges in the Ebel method should provide two different decisions for each item. One of the decisions is estimating the difficulty and the other is estimating the relevance level for each item. These two characteristics are classified as "Hard, Medium, Easy" and "Necessary, Important, Acceptable, Negotiable," respectively (Ebel, 1972). In Ebel method, the judges are first requested to place the items according to the 3×4 classification table by considering only the structure of the test. Then, the percentage of the correct answer to this classification should be determined by considering only the students in the borderline group. Once these two steps are completed, the score for each cell is determined by multiplying the number of items in each cell with the percentage of the cell; moreover, the identified scores are added to determine the MPS of the judges (Livingston&Zieky, 1982).

The performance of these test centered standard setting methods depends on the validity of the judges' decisions, which are subjective (Downing, Tekian&Yudkowsky, 2006). The student pass–fail or successful–unsuccessful status are based on the passing score obtained by these test centered standard setting methods, which may result different passing scores for the same test.

1.4. What is Cluster Analysis?

Cluster analysis methods reduce a dataset by grouping/clustering samples in homogeneous groups based on certain predetermined specifications (Alpar, 2011). Cluster analysis classifies objects or variables based on owned qualities (such as similarity and distance). The obtained clusters are homogeneous within themselves and heterogeneous against each other. In this context, the method is descriptive and the results are valid only for that dataset; they cannot be generalized; therefore, there is no estimation purpose in cluster analysis. The analysis reveals only the instant case.

Cluster analysis methods can be divided into two categories: hierarchical and nonhierarchical methods. Hierarchical methods are based on the assumption that all observations belong in a cluster, and they generate a similarity matrix to reduce the cluster number. Observations are placed in 1, 2, 3, …, (n − 2), (n − 1), n clusters

according to their similarity distance. Sireci (2001) stated that hierarchical cluster analysis is not suitable for large datasets and the individuals usually remain in the cluster to which they are first assigned. On the other hand, nonhierarchical cluster analysis methods are used when there is a prior information about the number of clusters. The most preferred method is the k-means method, in which the researcher defines k clusters (for example two clusters for successful and unsuccessful students in the standard setting process) at the beginning of the analysis. At this point, the first assignment is arbitrary. Each added observation is assigned to the cluster that is closer to the cluster center and the mean is recalculated to determine the new cluster center. This iterative process continues until all of the observations are processed. The most important feature of the method is that interpretation is based on theoretical knowledge.

1.5. Cluster Analysis in Standard Setting Studies

Hierarchical cluster analysis and k-means clustering were primarily used in previous standard setting studies (Sireci, Robin&Patelis, 1999; Sireci, 1995 as cited in Sireci, 2001; Violato, Marini&Lee, 2003; Hess et al., 2007; Khalid, 2011). Considering the problems of using hierarchical cluster analysis, which were indicated by Sireci (2001) above, in this study, only k-means cluster analysis was used. This method maximizes the difference between the sets as well as minimizes the inter-individual differences in the same cluster. After data clustering, the passing scores for the test can be calculated through different statistical standard setting methods which are AM, RMM and LR.

The first method is to consider the average score based only on the students assigned to the successful cluster (Average Method: AM). The second one is to consider the median of the range formed by the lowest score assigned to the successful cluster and the highest score assigned to the unsuccessful cluster (Range Median Method: RMM). Finally, the third method is to equate the logistic regression (LR) equation to 0.50 and calculate the equation value of x (Sireci, 2001).

2. LITERATURE REVİEW

Few studies compared the statistical standard setting methods (mentioned above as AM, RMM and LR) and the test centered standard setting methods (as Angoff, Yes/No or Ebel) for determining the passing scores with each other. For example, Sireci (1995) performed cluster analysis on a general educational development test (GED) having 50 multiple-choice and one open-ended items with samples belonging to two different years (cited in Sireci, 2001). This cluster analysis result indicated that, the students were divided into five clusters. The passing score was calculated by considering the median of the students' grades in the middle cluster. In conclusion, the passing score of GED was found to be slightly below the passing score calculated using cluster analysis. A validation study with an external criteria was conducted using the year end marks of the students. The point-biserial correlation coefficient was found 0.38 according to the passing scores with student grades. Sireci (1995) revealed two important results with this study. The first result is that, if the cluster analysis between samples changes, the solution is not tenable. In large samples,

this comparison can be made by drawing numerous samples like bootstrap method. The same results are obtained for various samples in this study. The second result of these study was that the analysis results must be validated with an external criteria. The correlation coefficient of the study was positive and high. Nevertheless, note that even if the coefficient was low, the analysis results could be supported (cited in Sireci, 2001).

Sireci et al. (1999) conducted another study to determine passing score using cluster analysis on 7th grade math test results. The original standard of this test was set by a revised Angoff method, and the students were placed on three levels. The aim of this study was to validate the passing scores using an external criteria, i.e., the year-end achievement levels. For this purpose, the researchers first applied hierarchical cluster analysis and k-mean cluster analysis to all of the samples. The C index and the stability between the samples were considered in determining the number of possible clusters. The results were validated with the student grades. Based on the analysis results, three clusters were identified, the stability of the clusters among the samples was tested, and the relationship with the student grades was found 0.69.

Violato et al. (2003) conducted a key study of, the passing scores that were obtained from Nedelsky and Ebel test centered standard setting methods were compared with k-means cluster analysis results. The correlation coefficient was 0.81 for Nedelsky method and 0.93 for Ebel method.

Hess et al. (2007) used the Angoff method with 15 experts; k-means clustering that was applied to all samples comprising an a priori sample, a validation sample, and the combination of these two-samples. Moreover LR method was used to determine the passing scores and the classification accuracy was assessed using Cohen's Kappa statistics. The study concluded that a high level of significant correspondence was found among cluster analysis results with the Angoff method in the priori sample; furthermore, this correspondencedecreased slightly in the validation sample. When the entire sample was considered, there was still a high correspondence; however, the study used no external criteria.

Finally, Khalid (2011) applied both hierarchical and k-means clustering to determine passing score on a 60-item achievement test and compared the results of the two analyses. The most important deficiency is the absence of an external criteria in this study. The results of the study showed that both methods assigned the students into five clusters. The differences between the total scores of individuals were tested with one-way variance analysis and significant differences were found between the clusters. According to the results, thirty-six students were placed in the first cluster and thirty-four students in the second cluster by both methods. However, none of the students in the third, fourth, and fifth cluster were common. In other words, both clustering methods could assign students in to first two clusters. Overall, 42% of the students were assigned to a common cluster by the two methods, whereas 58% of students were placed in different clusters.

There are no published studies comparing all three passing score determination methods by cluster analysis set forth by Sireci (2001). In this study, it was implemented k-means cluster analysis on two samples including a main sample of 305 students and a validation sample of 179 students to determine the passing score by three statistical standard setting methods (*AM, RMM, LR*) based on the results of the k-means cluster analysis. Moreover, the research aims that the correspondence of the determined passing scores with Angoff, Yes/No and Ebel test centered standard setting methods. In particular, this study aims to find an answer for the following research questions:

- What is the relationship between the passing scores determined by the statistical standard setting methods (*AM, RMM, LR*) based on k-means cluster analysis of the main sample and the validation sample?
- How is the correspondence of the passing scores obtained in the main sample in terms of classifying the students as successful/unsuccessful with Angoff, Yes/No, and Ebel test centered standard setting methods?

3. METHODOLOGY

The objective of this descriptive study was to set the passing scores based on the statistical standard setting approaches (AM, RMM, LR) with k-means cluster analysis and to investigate the validity of the these methods' results. Moreover, the passing scores obtained by k-means cluster analysis were compared with the passing scores calculated with Angoff, Yes/No and Ebel test centered standard setting methods. In this section, study populations, the measuring instrument and procedure are explained.

3.1. Study Populations

In this research two study groups were used: *(i)* The examinees and *(ii)* the judges (it is known as experts of the area, here as classroom teachers, because the achievement test is for fourth grade students):

The examinees consisted of 4th grade students who have participated in instruction on the sub-learning area of division and fractions. In this study, the examinee study group has two samples which are the main and validation sample. The validation sample of the study includes 179 students, whose school grades were used as an external criteria, and the main sample includes 305 students. Descriptive statistics were calculated for the two samples and the difference of the sample averages was tested by t-test for independent samples. The samples were found to be similar in terms of descriptive statistics, exhibited a normal distribution, and there was no significant difference between their means. Accordingly, the validation sample was considered to be representative of the main sample.

The judge (expert) group consisted of the 4th grade classroom teachers who are the experts of the study area and includes 17 experts. They made their decisions about the test items in respect to Angoff, Yes/No and Ebel test centered standard setting methods. The passing scores were obtained from their decisions.

3.2. Data Collection Tools

The Achievement Test: The research dataset includes students' answers, which were given to a 20-item multiple-choice test prepared for the sub-learning area of division and fractions for the 4th grade in the 2010–2011 academic year. These were analyzed in terms of validation and reliability by the researcher (item discrimination values range from 0.45 to 0.80; KR-20 internal consistency coefficient 0.84). Moreover, judge opinions were considered on the three test centered standard setting methods for the same test (Gündeğer, 2012).

3.3. Procedure

The study data were collected in the 2010–2011 academic year. As a first step of data collection, the students took the test. The students' responses were transferred to excel. In the second step, the judges made their opinions with respect to the Angoff, Yes/No and Ebel test centered standard setting methods and the passing scores were obtained from their decisions (Gündeğer, 2012).

The second and main part of this study is to set the passing score with the statistical standard setting methods (AM, RMM and LR) based on k-means cluster analysis results. Cluster analysis aims to reduce the dataset by reducing the number of variables and placing the student samples into clusters. In this study, the k-means cluster analysis was conducted to the raw data which consisted of the students' responses to the 20-item achievement test with two clusters as successful students cluster and unsuccessful students cluster. After the clustering process using the results of the cluster analysis, Sireci (2001) stated that there are three different statistical methods to determine the passing score. The first method is the average method (*AM*) which is to take the average of the total scores of the students in the successful cluster after the cluster analysis. The second method (*RMM*) is to take the median of the range formed by the lowest score assigned to the successful cluster and the highest score assigned to the unsuccessful cluster. The third method (*LR*) is to solve the equation obtained by logistic regression, in which the dependent variable is determined as a group membership by cluster analysis and the independent variable is determined as the total score by putting the a (constant) and b (slope) parameters into their places in the probability function below, and finding x for P = 0.50.

$$P = \frac{1}{1 + e^{-(a+bx)}} \qquad \text{Equation 1}$$

In data analysis, cluster analysis was performed on the students' responses in the two samples which are the main sample and the validation sample; and then the passing scores were determined with the three statistical methods mentioned above as AM, RMM and LR according to the k-means cluster analysis results. The relationship between the cluster analysis results for the validation samples and students' school grades were analyzed by Cramer's V; a significant relationship with 0.53 was found ($p < 0.05$). Accordingly, a meaningful and moderate correspondencewas found between the school grades and the results of the k-means cluster analysis. The classification

correspondenceof the passing scores with the results of the test centered standard setting methods was investigated by Cohen's Kappa statistics.

Kappa is a statistic developed by Cohen (1960) to measure of agreement between the two individuals when two binary (1 and 0) variables are attempts by two individuals to measure the same thing. If the Kappa is less than .20 it is poor agreement; Kappa between .20-.40 is fair agreement; Kappa between .40-.60 is moderate agreement; Kappa between .60-.80 is good agreement and Kappa above .80 is very good agreement (Şencan, 2005).

4. FINDINGS

The passing scores (on the scale of 100) calculated based on the k-means cluster analysis results with the three methods (*AM, RMM, LR*) are presented in Table 1. The passing score calculated using the average method (*AM)* for the main sample is 78.63 and for the validation sample is 77.65. The range median method (*RMM*) calculated the passing score as 60 in both samples, whereas the passing score calculated based on LR is 62.04 in the main sample and 59.47 in the validation sample. These differences in the calculated passing scores are attributed to the differences between the methods.

Table 1. The Passing Scores Calculated in the Main and the Validation Sample

Method	Sample	
	Main Sample	**Validation Sample**
Average Method (AM)	78.63	77.65
Range Median Method (RMM)	60	60
Logistic Regression Method (LR)	62.04	59.47

According to the Table 1, the AM provides the highest passing score in both samples, followed by LR method, and the RMM. However, standard setting methods often provide close values, although they are different. In this research, it was tested the correspondenceof the calculated passing scores with each other (see Table 2), which was turned into two artificial categories, i.e., successful or unsuccessful students, and was examined by Cohen's Kappa statistics.

According to the Table 2, a significant correspondence appears between the three methods at an error level of *0.01*. The correspondence in the passing score calculation methods is moderate between the AM and the RMM, moderately high between the AM and the LR method, and high between the RMM and the LR method. Thus, the highest correspondence among the methods in terms of the passing scores assigned to the students into successful or unsuccessful categories was found between the RMM and the LR methods.

Table 2. Correspondence among the Passing Scores*

	Average Method	**Range Median Method**
Average Method	1.00	
Range Median Method	.37**	1.00
LR Method	.50**	.80**

*The calculated coefficients are the same in both samples. **p < 0.01

The test centered standard setting methods were carried out in a previous study by Gündeğer (2012). In this study the passing scores (on the scale of 100) were calculated as 51.47, 57.94, and 47.12, respectively, based on the Angoff, Yes/No, and Ebel test centered standard setting methods. The correspondence of these passing scores with those calculated through k-means cluster analysis was investigated using Cohen's Kappa statistics (see Table 3).

Table 3. Correspondence with Test Centered Standard Setting Methods

Test Centered Methods	Average Method (AM)	Range Median Method (RMM)	LR Method
Angoff Method	.28*	.84*	.66*
Yes/No Method	.37*	1.00*	.80*
Ebel Method	.18*	.62*	.46*

*p < 0.01

Table 3 shows that all methods display meaningful relationships with each other at an error level of *0.01*. The method showing the highest correspondence with the Angoff, Yes/No, and Ebel methods is the range median method (RMM), followed by the LR method and the average method (AM). Moreover, the RMM gives a high correspondence with the three test centered standard setting methods.

The highest correspondence in the passing score determination based on k-means cluster analysis is given by the Yes/No method. The lowest correspondence value for this method is .37 calculated by the AM, and it indicates a moderate correspondence. The highest correspondence for the same method was calculated by the RMM providing a perfect correspondence. Moreover, Yes/No method shows a very high correspondence with LR method.

5. CONCLUSION & DISCUSSION

This study examined the relationship of three different passing scores identified according to k-means cluster analysis results in the main sample and the validation sample with each other and with the test centered standard setting methods. Also, the classification accuracy of the calculated passing scores was evaluated using an external criteria in the validation sample. The relationship between the cluster analysis results for the validation samples and the student school grades was analyzed by Cramer's V; a meaningful relationship (.53) was found between the student grades and the cluster analysis results. Indeed, there were only a few previous studies using an external criteria. Sireci (1995) proposed a sufficient relationship (.38) for validation in his study. In the present study, a moderate correspondence was found between the students' school grades and the results of the k-means cluster analysis. The passing scores obtained using the k-means cluster analysis were considered to be valid for both the validation and the main samples (cited in Sireci, 2001).

Furthermore, the highest passing score was given by AM for both samples, followed by LR, and RMM. However, the passing score determination methods may give similar values to each other even though different values were produced. When

the students were classified using the different method based on these calculated close values, significant moderate-to-high–level relationships were found among the three methods. The correspondence in the passing score calculation methods was moderate, i.e., between the AM and the RMM; moderately high, i.e., between AM and LR method; and high, i.e., between RM and LR methods. When the literature is examined, there is no study comparing AM, RMM and LR methods all together. Therefore this research focused on the aim of the comparing these cluster analysis methods. So the results of the study give an idea to the researchers about the three methods.

Moreover, the method showing the highest correspondence with the Angoff, Yes/No, and Ebel methods was found the RMM, followed by LR, and AM. Indeed, RMM showed a high correspondence with the three test centered standard setting methods, whereas the highest correspondence was given by Yes/No method calculated by RMM. Yes/No method combined with RMM produced a perfect correspondence. Moreover, Yes/No method showed a very high correspondence with LR method. Similarly, in the literature there is no study comparing these three cluster analysis methods with Angoff, Yes/No and Ebel standard setting methods.

This is the first study in terms of comparing the three statistical standard setting methods based on k-means cluster analysis with the three test centered standard setting methods. Although Hess, Subhiyah&Giordano (2007) proposed that there is high relationship only between the Angoff and LR method, but did not validate their results with an external criteria. Moreover, cluster analysis produced results, which refer only to that dataset; therefore, there is no problem of generalization. Nevertheless, the literature supports the result stating that there is a moderate correspondence between the Angoff and LR methods.

Test centered standard setting methods such as Angoff, Yes/No, and others employed in standard setting are based on judge opinion, and these opinions are subjective, occasionally even arbitrary. This is the most criticized aspect of the test centered standard setting methods. In order to ensure the validity of the opinions, standard setters should support these decisions with statistical methods. If researchers are to use test centered standard setting methods to determine the standards or passing scores, then cluster analysis, in particular, the k-mean method, in terms of correspondence with large datasets can be suggested. The relationship of the passing scores obtained with judge opinion can be compared with the k-means cluster analysis results and these passing scores can be supported with an external criteria. In this respect, k-means cluster analysis may be used by researchers to validate their standard setting studies.

REFERENCES

Alpar, R. (2011). *Uygulamalı Çok Değişkenli İstatistiksel Yöntemler*. Ankara: Detay Yayıncılık

Downing, S. M., Tekian, A. & Yudkowsky, R. (2006). Research Methodology: Procedures forEstablishing Defensible Absolute Passing Scores on Performance Examinations in Health Professions Education. *Teaching and Learning in Medicine*, *18* (1), 50–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16354141

Ebel, R. L. (1972). *Essentials of Educational Measurement*. New Jersey: Prentice Hall.

Gündeğer, C. (2012). *Angoff, Yes/No ve Ebel Standart Belirleme Yöntemlerinin Karşılaştırılması* (Master's Thesis, Hacettepe Üniversitesi, Ankara, Turkey) Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Hess, B., Subhiyah, R. G. & Giordano, C. (2007). Convergence between cluster analysis and the Angoff method for setting minimum passing scores on credentialing examinations. *Evaluation and the Health Professions*, *30*, 362-375. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17986670

Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), *Educational measurement* (pp. 485–514). Washington, DC: American Council on Education.

Khalid, M. N. (2011). Cluster analysis–a standard setting technique in measurement and testing. *Journal of Applied Quantitative Methods, 6*, 46–58. Retrieved from http://www.jaqm.ro/issues/volume-6,issue 2/pdfs/6_khalid_shafiq_khan.pdf

Livingston, S. A. & Zieky, M. J. (1982). *Passing Scores (A Manual for Setting Standards of Performance on Educational and Occupational Tests).* New Jersey: Educational Testing Service.

Sireci, S. G. (2001). Standard setting using cluster analysis. In G. J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods and Perspectives* (pp. 339-354). Mahwah, NJ: Lawrence Erlbaum Associates.

Sireci, S. G., Robin, F. & Patelis, T. (1999). Using cluster analysis to facilitate standard setting. *Applied Measurement in Education*, *12*, 301-325. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/S15324818AME1203_5?journalCode=hame20#.VxX4EvmLTIU

Şencan, H. (2005). Sosyal ve Davranışsal Ölçümlerde Güvenirlik ve Geçerlik. Ankara: Seçkin Yayıncılık

Violato, C., Marini, A., & Lee, C. (2003). A validity study of expert judgment procedures for setting cutoff scores on high-stakes credentialing examinations using cluster analysis. *Evaluation and the Health Professions*, *26*, 59-72. Retrieved from http://www.internationalgme.org/Resources/Pubs/Validity%20Cutoff%20Scores%20-%20Violato.pdf

Zieky, M. J. (2001). So Much Has Changed: How the Setting of Cut-scores Has Evolved Since the 1980s. In G. J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods and Perspectives* (pp. 19–51). Mahwah, NJ: Lawrence Erlbaum Association