International Journal of Progressive Education, Volume 16 Number 2, 2020

The Effect of Chance Success on Equalization Error in Test Equation Based on Classical Test Theory*

Duygu Koçak [i]
Alanya Alaaddin Keykubat University

Abstract

The aim of this study was to determine the effect of chance success on test equalization. For this purpose, artificially generated 500 and 1000 sample size data sets were synchronized using linear equalization and equal percentage equalization methods. In the data which were produced as a simulative, a total of four cases were created with no chance success, and three different levels (20%, %25, %33) of chance success and the default chance success were corrected by the correction formula. In the simulated data, four different scenarios have been created that do not include chance success and contain three different success rates (20%, 25%, 33%). Accordingly, the test equalization was performed by using linear equalization and equipercentile equalization methods under two different sample sizes and four different chance success conditions. Weighted mean square error of equating methods was found for each situation, and the method with the lowest weighted mean square error was accepted as the most suitable equating method. At the end of the study, it was found out that; while linear equating is the most suitable method for equating test points with chance success; equipercentile equating is the most suitable method for equating test points without chance success.

Keywords: Test Equating, Linear Equating, Equipercentile Equating, Single Group Design, Chance Cucces.

DOI: 10.29329/ijpe.2020.241.15

----

* This study is an extended version or the oral presentation of the 4th International Symposium on Educational Sciences.

[i] **Duygu Koçak,** Assist. Prof. Dr., Faculty of Educatıon, Alanya Alaaddin Keykubat University

Correspondence: duygu.kocak@alanya.edu.tr

INTRODUCTION

Nowadays, exams are applied for many purposes, especially in the recruitment of students and staff to various institutions. Different and more various forms are developed to ensure that the validity and reliability of the tests do not fall below a certain level and to protect the confidentiality of tests, which is applied as regularly (Turgut, 1971). The questions in different forms were prepared with similar content and statistical characteristics. At the end of the application, it is seen that there is a difference between item difficulty levels of the tests. In other words, although the structure measured in the tests is the same, the difficulty of the substances and, therefore, the difficulty of tests differ. This prevents direct comparison of the scores obtained from the tests. The comparison of scores from different forms of tests that measure the same feature is of great importance in education (Tsai, 1997).

In order to make this comparison, it is necessary to establish a statistical relationship between the scores obtained from different forms of the same test, and this statistical relationship is called test equating. Felan (2002) defines test equalization as establishing a statistical relationship between the scores obtained from two forms measuring the same structure. Angoff (1971) describes the test equalization as the conversion of the unit system of one test into the unit system of another test that measures the same property. One of the aims of test equalization is to prevent bias among individuals taking different forms, and another is to report scores from different forms on the same scale and maintain the meaning of the reported scores (Barnard, 1996). Test scores are equalized in order to examine the change of an individual's ability and knowledge level over the years.

The test equalization can be used to test scores obtained from different test forms are equating observe the development of individuals, and compare the performance of individuals (Bozdağ, 2007; Crocker & Algina, 1986). Depending on the difficulty level of the forms to be equalized and the skill distribution of the applied group, horizontal equalization can be performed between groups with similar ability distribution and between tests with similar difficulty levels, and two different equalizations can be made, namely vertical equalization between different skill groups and tests with different difficulty levels. In addition, different equating methods can be used for equalization as depending on the theories based on the development of the tests (Crocker & Algina, 1986; Felan, 2002). Parallelism, symmetry, and independence from the group must be ensured in order to make the equalization. (Angoff, 1971; Gulliksen, 1967; Hambelton, 1985; Kelecioğlu, 1993; Şahhüseyinoğlu, 2005; Woldbeck, 1998). Parallelism is achieved when the test scores that are obtained from two different test forms are equal; hence, the forms must be one-dimensional, and the forms must be measured the same structure (Woldbeck, 1998). The symmetry is that the conversion between the unit systems of the two forms can be achieved by a single equation and that this transformation can be done by a single formula for both two-way tests (Angoff, 1971; Felan, 2002; Tanguma, 2000). The fact that the scores obtained as a result of equalization is independent of the group from which the conversion is made is expressed as independence from the group (Angoff, 1971; Felan, 2002; Kelecioğlu, 1993). The reliability, mean difficulties and variances of both forms should be the same (Angoff, 1971; Crocker & Algina, 1986; Kelecioğlu, 1993; Şahhüseyinoğlu, 2005; Tanguma, 2000; Turgut, 1971).

Angoff (1971) and Thorndike (1982) stated that the scores of the test with more errors would not be equal to the scores of the test with fewer errors, and the forms would not equalize significantly if the reliability was not high and similar. As can be seen from rules of equalization, many conditions must be met for equalization. Without these assumptions, equalization will be meaningless, and the equalization error will increase. The concept of error in test equalization is explained by the difference between the ability level of the individual and the predicted ability level for the test that he did not take. In the less error equalization, the ability levels obtained by different tests are expected to be equal (Cook & Eignor, 1991). In order to determine the error amount of the points obtained by the equalization methods, the raw score and the equalized scores corresponding to the raw scores are compared. For this comparison, the Weighted Mean Squares Error (WMSE) is used (Skaggs & Lissitz, 1986):

$$WMSE = \sum_{i=1}^{k-1} f_i (X_E - X_{crit})2 / \sum_{i=1}^{k} f_i S^2 Y$$

k: Number of items in the Y test.

S^2Y: variance of Y test.

X_{crit}: the raw score of i in the Y test

X_E: The score obtained by the equalization methods equal to the raw score in the X test.

f_i: The frequency of the raw score i in the Y test.

Equalization errors are divided into two as random equalization error and systematic equalization error. The random equalization errors may be caused by the test statistics such as standard deviation from the sample, mean or percentage order (Felan, 2002; Kolen, 1988). The systematic equalization error is mainly due to the deterioration of the equalization conditions (Kolen, 1988; Zeng, 1991). Test features, item features, and group features directly affect equalization errors. The most commonly used item type is the multiple-choice item in large-scale test applications. This item type has many advantages, which make it preferable. The major disadvantage of this type of item that negatively affects the psychometric features of the item and the test is that it contains chance success. The chance success is that the responder who takes the test finds the correct answer in the multiple-choice test by guessing (Turgut, 1971). Depending on the number of options of the multiple-choice item, the item includes chance success in different proportions. Item scores and test scores; hence, the psychometric properties of the test affect chance success (Araz, 2001; Çelen, 2002; Telli, 1993; Şahhüseyinoğlu, 1998). The validity and reliability increase in the tests that chance success is eliminated because it is predicted that corrected scores are estimated better than uncorrected scores in measuring an individual's ability (Çelen, 2002). In the test that chance success was eliminated, the average decreases, and the standard deviation increases (Koçak, 2013). Considering that the random error is caused by item and test parameters and the chance success is an item parameter, and it has an effect on the equalization error in the process of test equalization. It is thought that the chance success will have an effect on the equalization error in the test equalization process.

Significance of the Study and Research Questions

Relevant studies reveal that the properties of the equalized tests and items affect test equalization. Bozdağ (2007) states that a 20% chance of success will increase test equalization error. Considering that, the option number can be different in the multiple-choice item that is used in applied tests, and as a result of that, the proportion of chance success is different. It is thought that equalization errors can be different on the different proportion of the chance success. According to this, how the equalization error is affected by different levels of chance success and determining under which conditions the equalization method with lower error will be conducted will guide the researchers in the applications. Bozdağ (2007) considered only a 20% chance of success in the study. There are different chance successes in a test, depending on the number of choices. Therefore, considering other rates of chance success will increase the accuracy of the decisions to be made. This study differs from other studies that it deals with all percentages of chance success. In light of these discussions, the aim of this study is to determine the effect of different chance success levels on equalization error. For this purpose, the answers to the following questions were sought:

1. How does the equalization error obtained by equal percentage equalization method change according to sample size and chance success rate?
2. How does the equalization error obtained by linear equalization method change according to sample size and chance success rate?

METHODOLOGY

In this study, the effect of chance success on equalization error was determined by using simulated data by using linear and equal percentage equalization methods.

Data Generation

In this study, data generation and test equalization were based on Classical Test Theory. A total of 4 data sets of 25 items are scored in two categories, 500 and 1000 test lengths were artificially generated by the Monte Carlo simulation method. The R program (2011) "psych" package was used to generate the data. The data produced are scored in two categories (1-0), and the data are in the form of multiple-choice items. The test equalizations were made between data that have the same sample size as 1000-1000 and 500-500.

Test equalization was performed under four different conditions: In the first case, it was assumed that the tests do not have chance success, in the second case it was assumed that the tests contain 20% chance success, in the third case it was assumed that the tests contain 25 % chance success, in the last case it was assumed that the tests contain 33% chance success, The correction formula was used to eliminate the chance success on the tests that contain the chance success. Accordingly, in the fiction where the questions in the test are considered to have three options, two wrong answers are deleted correct answers. In the fiction, where the questions have four options, three wrong answers delete one correct answer are deleted. In the fiction where the questions have five choices, four wrong answers delete one correct answer.

Data Analysis

In order to make the equalization, the forms that are to be equalized must measure the same structure and be one-dimensional (Angoff, 1971; Felan, 2002; Gulliksen, 1967; Tanguma, 2000; Thorndike, 1982; Woldbeck, 1998) and the reliability, mean of difficulties and variances of both forms should be the same (Angoff, 1971; Crocker & Algina, 1986; Kelecioğlu, 1993; Şahhüseyinoğlu, 2005; Tanguma, 2000; Turgut, 1971). Besides, the correlation between the forms that are equalized must be high (Dorans, 2000; Masse, Allen, Wilson & Williams, 2006). Table 1 provides statistics on these conditions.

Table 1 Factor Structure of Data Produced.

Test	Factors	Eigenvalue	Explained Variance Ratio	Explained Total Variance Ratio
A1	1	16,531	61,064	61,064
	2	0,917	3,387	64,451
	3	0,911	3,365	67,816
A2	1	15,437	54,126	54,126
	2	0,978	3,429	57,555
	3	0,902	3,162	60,717
B1	1	18,001	55,025	55,025
	2	0,980	2,995	58,02
	3	0,955	2,919	60,939
B2	1	16,208	59,808	59,808
	2	0,991	3,656	63,464
	3	0,970	3,346	66,81

Table 1 contains the results of the factor analysis of the produced data. For the factor analysis of the data that are scored in two categories, the R program "polycor" package was used. The tests appear to consist of a single and dominant dimension.

Table 2 Values for Correlation Between Tests to be Equalized

scores	r_{A1-A2}	r_{B1-B2}
Chance success was not eliminated on the test scores	0,78	0,81
20% chance success was eliminated on the test scores	0,79	0,82
25% chance success was eliminated on the test scores	0,79	0,81
33% chance success was eliminated on the test scores	0,79	0,80

The correlation between the forms that are equalized must be high to perform the equalization. When Table 2 is examined, it is seen that the correlations between the equalized tests are high.

Table 3 Comparison of The Test Difficulties.

	N	Test	$\bar{P}$	t	P
Chance success was not eliminated on the test scores	500	A1 A2	0,579 0,576	0,089	0,00
	1000	B1 B2	0,525 0,528	0,135	0,00
20% chance success was eliminated on the test scores	500	A1 A2	0,456 0,453	0,095	0,00
	1000	B1 B2	0,413 0,414	0,045	0,00
25% chance success was eliminated on the test scores	500	A1 A2	0,426 0,424	0,064	0,00
	1000	B1 B2	0,386 0,388	0,095	0,00
33% chance success was eliminated on the test scores	500	A1 A2	0,380 0,378	0,066	0,00
	1000	B1 B2	0,333 0,339	0,300	0,00

Table 3 shows whether there is a significant difference between the difficulties of equalized tests. Whether the difficulties of the tests were equal as examined by the ratio test for independent groups. When comparing difficulties, data having the same sample size were compared among themselves, and there is no significant difference between the difficulties of the tests.

Table 4 Comparison of The Test Reliability.

	N	Test	KR-20	Z_r	Z
Chance success was not eliminated on the test scores	500	A1	0,812	1,125	0,253
		A2	0,804	1,109	
	1000	B1	0,865	1,312	0,931
		B2	0,875	1,353	
20% chance success was eliminated on the test scores	500	A1	0,831	1,191	0,253
		A2	0,826	1,175	
	1000	B1	0,872	1,341	0,886
		B2	0,881	1,380	
25% chance success was eliminated on the test scores	500	A1	0,853	1,267	0,301
		A2	0,848	1,248	
	1000	B1	0,889	1,417	0,886
		B2	0,897	1,456	
33% chance success was eliminated on the test scores	500	A1	0,886	1,403	0,682
		A2	0,895	1,446	
	1000	B1	0,917	1,569	0,454
		B2	0,920	1,589	

First, the KR-20 internal consistency coefficient of each test was calculated to examine whether the reliability of the equalized tests was equal. The internal consistency coefficients obtained from equaled tests were transformed from Fishers to Z. It was investigated whether there was a significant difference between the two reliability coefficients. When the statistics in Table 4 are examined, it is seen that there is no significant difference between the reliability of the tests.

Table 5 Comparison of The Test Means and Variances.

	N	Test	X	t	p	S^2	F	p
Chance success was not eliminated on the test scores	500	A1 A2	14,475 14,402	0,219	0,099	28,398 27,457	1,23	0,159
	1000	B1 B2	13,145 13,214	0,263	0,105	33,907 35,988	1,11	0,191
20% chance success was eliminated on the test scores	500	A1 A2	11,400 11,331	0,159	0,122	47,527 46,076	1,22	0,180
	1000	B1 B2	10,336 10,371	0,103	0,101	56,055 59,259	1,10	0,244
25% chance success was eliminated on the test scores	500	A1 A2	10,674 10,616	0,129	0,123	51,279 49,702	1,18	0,205
	1000	B1 B2	9,668 9,703	0,099	0,082	60,497 63,968	1,10	0,280
33% chance success was eliminated on the test scores	500	A1 A2	9,508 9,459	0,101	0,089	59,089 57,289	1,20	0,217
	1000	B1 B2	8,627 8,673	0,122	0,101	69,622 71,876	1,08	0,292

The average and variance of the tests to be equalized should be equal. The difference between means was tested by the t-test, and the difference between variances was tested by the F test. When the values in Table 5 are examined, it is seen that there is no significant difference between the means and variances of the tests to be equalized.

After testing whether the equalization conditions were fulfilled in the generated data, equalized scores were obtained using equalization methods. Then, the mean error frames for each equalization method and condition were calculated, and the equalization errors were compared.

FINDINGS

In the following, firstly, the results obtained from equalization using the equal percentage equalization method, and then the results obtained from equalization using the linear equalization method are presented. Equalization was performed using the translation formula that is suggested by Livingston (2004) because the scores of the equalized tests have not coincided with the same percentage order in the equal percentage equalization method.

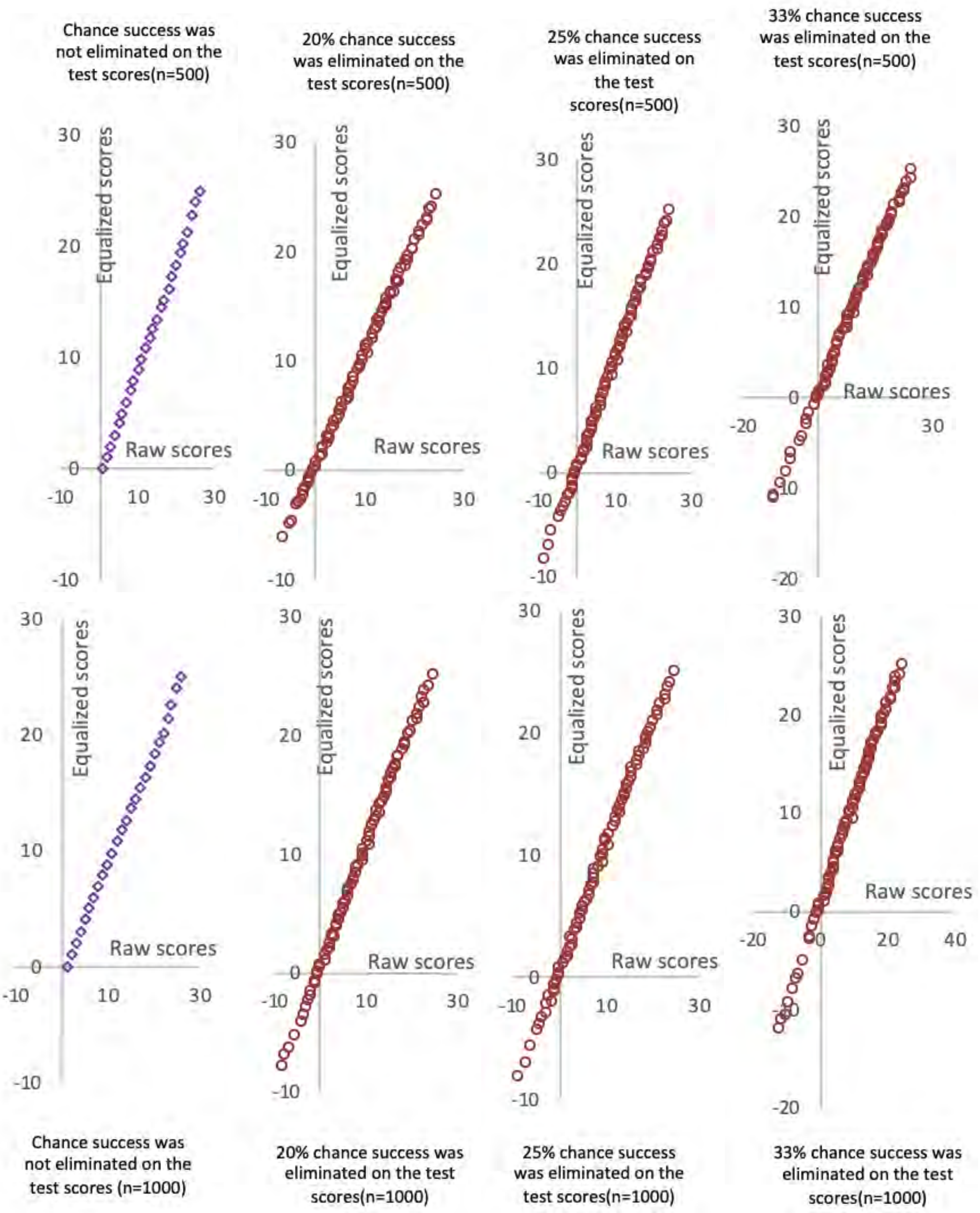

Figure 1. **Graphs of equalization with equal percentage equalization.**

Figure 1 shows the distribution of raw scores and equalized scores equalized by the equal percentage equalization method. In the tests in which chance success was eliminated, equalization was made with higher errors at low skill levels. Accordingly, it can be stated that there is a relationship between ability level and elimination error. Livingston (2004) and Taguma (2000) state that the level of individuals' ability affect the equalization error. The fact that the equalized score and raw score pairs in the presented graph are frequent indicates that the equalization error is low. Accordingly, as the chance success rate increases, the graph becomes more frequent except for low skill levels. The equalization error is, therefore, reduced.

The minimum values of the scores that chance success was eliminated have a lower value than the scores that chance success was not eliminated because the correction formula that was used to eliminate the chance success has an algorithm that reduces the individual's total test score. Therefore, after applying the correction formula, the scores that chance success was eliminated can be negative. According to this, it can be said that equalization was realized with higher error in sub talent groups in the scores that chance success was eliminated. However, the error is lower as talent distribution becomes more frequent in data where chance success is eliminated. Şahhüseyinoğlu (2005) states that there is less equalization error in the tests that chance success is eliminated.

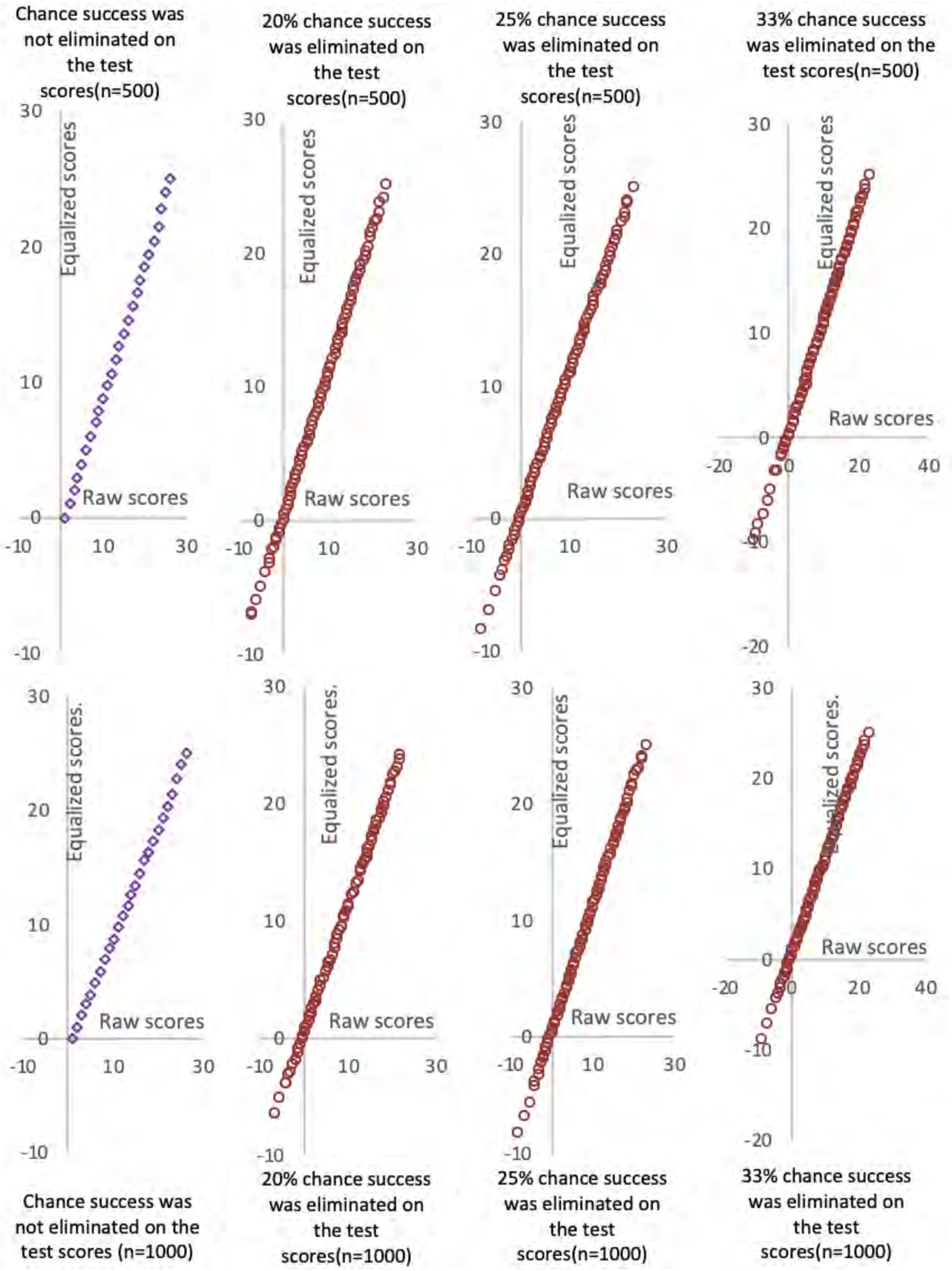

***Figure 2*. Graphs of equalization with linear equalization method**

Figure 2 shows the distribution of raw scores and equalized scores equalized by the linear equalization method. When the graphs are analyzed, it is seen that the equalized score distribution obtained by the equalization of the data that chance success was eliminated is more linear. The more linear the distribution, the less the error of the equalization error. Accordingly, as the eliminated chance success rate increases, equalization error decreases. Şahhüseyinoğlu (2005) states that there is less equalization error in the tests that chance success was eliminated.

Table 6 Means of Error Squares Obtained Using Equal Percentage and Linear Equation Method.

Tests	Means of Error Squares			
	Linear Equating		Equipercentile Equating	
	N=500	N=1000	N=500	N=1000
Chance success was not eliminated on the test scores	0,007	0,006	0,009	0,007
20% chance success was eliminated on the test scores	0,0069	0,0055	0,0087	0,0068
25% chance success was eliminated on the test scores	0,0065	0,0051	0,0074	0,0063
33% chance success was eliminated on the test scores	0,006	0,0048	0,0071	0,0051

Table 6 presents the equalization errors (Weighted Error Squares Mean) obtained from the equalization using the linear equalization method and the equal percentage equalization method. It is concluded that the equalization method, which gives the lowest error in all conditions, is the linear equalization method. As the sample size increases, it is seen that equalization error decreases both in linear equalization method and equal percentage equalization method. In the literature, it is stated that equalization error decreases with increasing sample size (Kim & Cohen, 2002; Lee & Ban, 2010; Tsai, 1997).

It is seen that equalization error decreases when chance success is eliminated in each sample size. The lowest equalization error was obtained from the equalization, which chance success was eliminated from the points of the test in the case of the highest chance achievement (33%). Accordingly, it can be stated that chance success increases the equalization error, and therefore the scores are eliminated from chance success can be equalized with lower error. Şahhüseyinoğlu (2005) states that there is less equalization error in the tests in which chance success is eliminated. The results obtained support this.

The research findings point to three main points: first, the equalization error decreases as the sample size increases. Second, the scores eliminated from the chance success are equalized with lower error, and as a result, chance success increases the equalization error. The third, linear equalization method, performs equalization with lower error than the equal percentage equalization method.

DISCUSSION

In this study, it is aimed to determine the equalization errors to be obtained from linear and equal percentage equalization methods in the artificially generated data at different levels of chance success. Classical test theory is used in the research. Although the use of Item Response Theory has been increasing in recent years, Classical Test Theory is still prevalent, especially in classroom measurement and evaluation activities. More than one test is administered throughout the semester to monitor the progress of students' in-class achievement. Test equalization is used to compare the scores obtained from these tests. For this reason, in this study, two basic equalization methods that can be used in the comparison of these tests, which are mostly developed based on Classical Test Theory, are discussed. In this study, it is aimed to determine the effect of chance success on test equalization error by considering the widespread use of multiple-choice items and the effects of chance success on psychometric features of the test. It is thought that the results of the research will guide teachers, researchers, and test developers in practice.

As a result of the research, it is seen that the linear equalization method makes equalization with fewer errors under all different sample sizes and chance success conditions. According to this, the linear equalization method is more successful than the equal percentage equalization method. The linear equalization method was found to be the method with the least error squares in all chance success rates when equalizing the data. Skaggs and Lissitz (1986) indicate that there is no statistical test to determine the significance of error squares means and that the values for this error may be meaningful in practice if the values are 0.05 or greater. It is seen that the mean values of error squares for both methods are less than 0.05 and close to each other. It is thought that equalization errors take close values because the distributions of data are similar (Felan, 2002). Since the error of linear equalization is smaller, it can be said that linear equalization is a more appropriate method for equalization when the chance success was not eliminated. This finding is consistent with Budescu's (1987) findings.

Angoff (1971) and Woldbeck (1998) stated that the distribution of scores should be frequent and tense for the equal percentage equalization method. Thus, each point that is a raw score in the score distribution will correspond to one point that is an equalized point in the other score distribution. It can be said that the equalization error of equal percentage method decreases because the score range of the equalized scores decreases and becomes frequent after the chance success was eliminated. The results obtained are in agreement with the results of Bozdağ (2007). The very low equalization errors obtained by both methods can be related to the similarity of the distributions of data that are equalized data. Under conditions where chance success is eliminated, the equalization error is higher at low skill levels. Wordbeck (1998) states that the frequency of the score distribution points to the skill range and that there is greater equalization error at low ability levels. The findings of the study support this.

In both equalized methods, as the sample size increased, the equalization error decreased. Accordingly, it can be stated that the sample size reduces the equalization error. In the literature, it is stated that equalization error decreases with increasing sample size (Kim & Cohen, 2002; Lee & Ban, 2010; Tsai, 1997). Zimmerman and Williams (2003) state that chance success negatively affects reliability in small samples. As a result of the research, more equalization errors were obtained when the sample size was 500 under similar conditions. It can be stated that this is due to the fact that chance success affects the reliability of the tests negatively.

Suggestions for applications and future research

In the study, it was obtained that the equalization error decreased with increasing sample size. Accordingly, it is recommended to use data to be obtained from a large sample as possible as for equalization. In multiple-choice items, it is seen that equalization error is higher when chance success is not eliminated. Therefore, in the case of using multiple-choice items, the chance success was eliminated by correcting formula that can provide error-free equalization. In this study, the distribution characteristics of the equalized data are similar, and in this condition, the linear equalization method has equalized with lower errors. Therefore, if the distributions of data to be equalized are similar, the linear equalization method is recommended.

In this study, the effect of chance success on equalization error was investigated. The effect of features such as item difficulty distribution and item discrimination distribution on test equalization can be investigated in future researches. In this study, the sample sizes were determined as 500 and 1000. Equalization errors can be examined in smaller samples and different test length. In addition to the effect of the features of the test and the items on the equalization error, the equalization error can be examined depending on the cognitive level measured by the test and the items. In the study, only the chance success variable was manipulated from the item features, and other features of the items could be considered in future researches. It has been shown that chance success affects equalization error in tests developed and equalized on the basis of Classical Test Theory; a similar study can be performed with Item Response Theory.

REFERENCES

Angoff, W. H. (1971). Scales, Norms and Equivalent Scores. Thorndike, R. L. (Ed.) *Educational Measurement*, 2nd ed., Washington, D. C. American Council on Education.

Araz, G. (2001). *Aynı davranışı ölçmeye yönelik kısa cevaplı, üç ve beş Seçenekli çoktan seçmeli testlerin madde ve test özelliklerini şans başarısı ile birlikte incelenmesi*. (Unpublished master's thesis). Hacettepe University, Ankara.

Barnard, J. J. (1996). *In search for equity in educational measurement: Traditional versus modern equating methods.* Paper presented at ASEESA's National Conference, Pretoria South Africa.

Bozdağ, S. (2007). *Şans başarısının test eşitlemeye etkisi*. (Unpublished master's thesis). Mersin University, Mersin.

Budescu, D. V. (1987). Selecting an equating method: Linear or equipercentile?. *Journal of Educational Statistics, 12*(1), 33-43. https://doi.org/10.3102/10769986012001033.

Cook,L.L., & Eignor, D.R. (1991). NCME instructional module on IRT equating methods. *Educational Measurment:Issuse and Practice,10*(3),37-45.

Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. CBS Collage Publishing, New York.

Çelen, Ü. (2002). *Şans başarısı için düzeltme formülü kullanılacağına ilişkin yönergenin testin psikometrik özelliklerine etkisinin Araştırılması*. (Unpublished master's thesis) Ankara University, Ankara.

Dorans, N.J. (2000). Research Notes: Distinctions Among Classes of Linkages. *The College Board, Office of Research and Development*.

Felan, G.D. (2002). *Test equating: Mean, linear, equipercentile and Item Response Theory*. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Austin.

Gulliksen, H. (1967). *Theory of Mental Tests*. New York: John Wiley & Sons

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Kluwer Academic Publishers Group, Boston

Kelecioğlu, H. (1993). *Öğrenci seçme sınavı puanlarının eşitlenmesi üzerine bir çalışma*. (Unpublished doctoral thesis). Hacettepe University, Ankara.

Kim S. H., & Cohen A. S. (2002). A comparison of linking and concurrent calibration under the graded response model. *Applied Psychological Measurement, 26*, 25-41. https://doi.org/10.1177/0146621602026001002.

Koçak, D. (2013). *Farklı yönergelerle verilen çoktan seçmeli testlerde yanıtlama davranışlarının incelenmesi*. (Unpublished master's thesis). Ankara University, Ankara.

Kolen, M. J. (1988). An NCME instructional module on traditional equating methodology. *Educational Measurement: Issues and Practice. 7*(4), 29-36. https://doi.org/10.1111/j.1745-3992.1988.tb00843.x.

Lee, W.C., & Ban, J.C. (2010). A comparison of IRT linking procedures. *Applied Measurement in Education, 23*(1), 23-48. https://doi.org/10.1080/08957340903423537.

Livingston, S. A.(2004). *Equating Test Scores (Without IRT).* Educational Testing Service.

Masse, L. C., Allen, D., Wilson, M., & Williams, G. (2006). Introducing equating methodologies to compare test scores from two different self-regulation scales. *Health Education Research 21*(1), 110-120. https://doi.org/10.1093/her/cyl088.

Skagg, G. & Lissitz R. W. (1986). An Exploration of the Robustness of Four Test Equating Models. *Applied Psychological Measurement. 10*, 303-317. https://doi.org/10.1177/014662168601000308.

Şahhüseyinoğlu, D. (1998). *Sayısal yetenek testlerinde seçenek sayısının test ve madde istatistikleri üzerindeki etkisinin şans başarısı ile birlikte incelenmesi*. (Unpublished master's thesis). Hacettepe University, Ankara.

Şahhüseyinoğlu, D. (2005). *İngilizce yeterlik sınavı puanlarının üç farklı eşitleme yöntemine göre karşılaştırılması*. (Unpublished doctoral thesis). Hacettepe University, Ankara.

Tanguma, J. (2000). *Equating test scores using linear method*. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Dallas.

Telli, A. (1993). *Şans başarısının madde türlerindeki madde ve test istatistiklerine etkisi.* (Unpublished mater's thesis). Hacettepe University, Ankara.

Thorndike, R.L. (1982). *Aplied Psychometrics*. Houghton Mifflin Company, Boston.

Tsai, T.H. (1997). *Estimating minumum sample sizes in random groups equating*. Poster presented at the Annual Meeting of the National Council on measurement in Education, Chicago.

Turgut, F. (1971). Ş*ans Başarısının Test Puvanlarına Etkisi*, ODTÜ Yayınları, Ankara.

Woldbeck, T. (1998). *Basic concepts in modern methods of test equating.* Paper presented at the Annual Meeting of the Southwest Psychological Association, New Orleans.

Zeng, L.(1991). Standard errors of linear equating for the single group design. *ACT Research Report Series* (4).

Zimmerman D. W., & Williams R. H. (2003). A new look at the influence of guessing on the reliability of multiple-choice tests. *Applied Psychological Measurement, 27*, 357-371. https://doi.org/10.1177/0146621603254799.