European Journal of Education Studies
ISSN: 2501 - 1111
ISSN-L: 2501 - 1111
Available on-line at: www.oapub.org/edu

doi: 10.5281/zenodo.1101229 Volume 3 | Issue 12 | 2017

TEST EQUATING STUDY CONCERNING TO ALES (ACADEMIC PERSONNEL AND POSTGRADUATE EDUCATION ENTRANCE EXAM) SCORES OBTAINED AT DIFFERENT TIMES IN A YEAR[i]

Ceren Mutluer[1ii],
Zekeriya Nartgün[2]
[1]Lecturer, Measurement and Evaluation in Education Department, Education Faculty, Bolu, Turkey
[2]Prof. Dr, Measurement and Evaluation in Education Department, Education Faculty, Bolu, Turkey

Abstract:
The purpose of this to equate ALES (Academic Personnel and Postgraduate Education Entrance Exam) scores with linear equating and equipercentile equating methods in the periods of 2011 spring and autumn conducted by SPSS, to determine the most suitable one of these two methods for research and to propose the most suitable one thanks to the findings in case of a similar study. The study population was 21860 people participating in ALES made in the periods of both 2011 spring and 2011 autumn. The sample of the study was 2186 individuals, respectively selected %10 of population randomly via computer. In the study, the internal consistency was calculated without eliminating items by using KR-21 formula because data were raw scores. Analyzes were carried out because of the fact that internal consistencies were high without eliminating agents of heights. Linear and equipercentile methods applied separately to the raw scores of sizes. Error quantities were measured by WMSE formula to determine which equating method was more appropriate for research and when the data of research findings were analyzed it was found that WMSE coefficient of equipercentile method was lower. As a result of this study equipercentile equating was found more

[i] This paper is based on the first author's master thesis titled as "*Test Equating Study Concerning to ALES (Academic Personnel and Postgraduate Education Entrance Exam) Scores Obtained at Different Times in a Year*" under the supervision of the second author.
[ii] Correspondence: email cmutluer@yandex.com, nartgun@yahoo.com

appropriate on the equating study concerning to the field of quantitative, verbal, equiponderant.

Keywords: test equating, single group design, linear equating, equipercentile equating

1. Introduction

Every year tests such as Academic Personnel and Postgraduate Education Entrance Exam (ALES), Civil Servant Selection Test (KPSS), Civil Servants' Language Profiency Test (KPDS), Interuniversity Council Language Proficiency Test (UDS), Language Proficiency Test (YDS), Vertical Transfer Test (DGS), Student Selection Test (OSS), Postgraduate Placement Test (LYS) and High Education Entrance Test (YGS) are centrically carried out by The Assessment, Selection and Placement Centre (OSYM) in Turkey. Some tests such as YDS, ALES, KPDS and UDS are carried twice a year by the OSYM and their scores are valid for a few years. Different forms of tests are applied because of some security precautions for the entrance examination of college/university or other institutions for selecting the personnel. It is not possible that all the features of the test to be the same (such as the reliability of the test, the difficulty level/strength of the test items, the variant of the test...) eventhough the developers of the test for OSYM try to prepare equivalent tests. So that reason, it is necessary to have these equating to compare the points taken from different forms. The method to make these points transformed each other and compared is test equating. Angoff (1971) defined test equating as to transform the unity system of a form to another unity system of a form. This definition is the most clearly explained definition of the term. According to Kolen and Brennan (2004), however, test equating is a statistical process which organizes the differences between the forms that have gathered form these forms. However, Braun and Holland (1982) defined test equating as a quantitative regulation made to enable the interchangeably use of the scores collected form the forms which have different difficulty level.

1.1 Necessary Conditions for Test Equating

It is necessary to actualize five important conditions to equate the points in X and Y forms in most researches. These conditions are symmetry, equality, unidimensionality independency form the group and equal reliability values (Braun & Holland, 1982; Dorans & Holland, 2000; Hambleton & Swaminathan, 1985; Kolen & Brennan, 2004; Lord, 1980). Dorans and Holland (2000) explains the symmetry emphasized mostly for test equating as to get Y point from X point the same as how X point gotten from Y

point through equating. So this transformation must be possible to be exchanged backward. It is expresses that in the case of close statistics results in mean points, standard deviation, variance points taken from X and Y test forms by means of test equating, test forms are able to be converted each other or equal to each other (Lord, 1980). However this equality condition can be exceed by stability feature of Item Response Theory (IRT). According to Dorans and Holland (2000) it is important for test equating to examine the same behaviors, the same features and the same capabilities. Independence from the group means that equating relationship is being independent from the group which enforces equating. If this has been supplied the same relationship can be found both for girls and for boys (Hambleton & Swaminathan, 1985; Kolen & Brennan, 2004). The equality reliability points mean to make the points measured in a purified way from errors. If the two tests the reliabilities of which are not equal made to be equated; it seems that the test points containing more errors equating to test points containing more errors so this is also a kind of status that contradicts with the necessity of reliable test equating.

2. Equating Designs

The process of collecting data for equating is called equating designs. Equating designs must be objective and economical (Thorndike, 1982). There are various designs for test equating. Data collecting designs have an important role in the success of test equating. In test scores equating designs vary according to the organization of tests and groups (Angoff, 1982; Penfield & Camilli, 2007). In test equating "Random Group Design", "Single Group Design" and "Balanced Group Design" represents the equating mechanism created without using anchor while "Common-Item Nonequivalent Groups Design" represents the equating mechanism created by using anchor.

In this study, we tried to equate the scores of those who took both ALES 2011 spring and autumn tests, so single group design was found appropriate. In single group design in order to be equated, the two test forms are imposed to people in the same group (Kolen & Brennan, 2004). Thus, people randomly chosen from the group shown as P are imposed to both X and Y form. This is probably the first type of the equating pattern for the control of the who took the test's skills and the first time when same people are used with different tests (Dorans & Holland, 2000; Kolen & Brennan, 1995).

2.1 Test Equating Methods

Test equating methods are separate in two according to the theory based on the methods of test equating. These are test equating methods based on Classical Test

Theory and Item Response Theory. This study involves only the methods of Classical Test Theory.

2.2 Test Equating Methods Based on Classical Test Theory

Crocker and Algina (1986) enumerated the units to choose an equating method as supplying the hypothesis, applicability and consistency. The test equating methods based on Classical Test Theory could be discussed in three categories. These are 'mean equating', 'linear equating', 'equipercentile equating' (Kolen, 2007).

In mean equating when X and Y are considered to be the two different forms of a test, the difference between the medians of X and Y forms is regarded as fixed. Equated scores point to the same success level. Mean equating can be expresses as equating 4 and 5 (Kolen, 2007):

X - Mean (X) = Y - Mean (Y)
X = Y - Mean (Y) + Mean (X)

It is suitable to use linear equating test when the level of capability is the same in groups. The standard deviation and mean points of the groups taking the X and Y forms are calculated. The points of the groups are equated by using the equating below (Angoff, 1971):

(Y - Mean (Y)) / (Standart Deviation (Y)) = (X - Mean (X)) / (Standart Deviation (X))

In equipercentile equating, these groups can be selected when the two forms which are thought equal to the save percentiles class, taking the similar test group (Angoff, 1982). In the equating method the distribution of the form X to the distribution of form Y by calculating the percentile of the two tests.

In equipercentile equating, if the two forms considered as equal are equal to the same percentage range among the people who took the test in similar groups, these groups could be chosen for transformation (Angoff, 1982). In this equating method, Range of form X is equated to range of form Y by calculating the percentage range of the two tests. When the distribution of the test points is different, the precision of equating is supplied by equating to equipercentiles equating (Angoff, 1971). It is stated that the points taken from X and Y forms measuring the same feature by the equal reliability rank can be equivalent to the equipercentile sequence is acceptable.

ALES is a kind of examination that gives the points concerned with the election of candidates, send to abroad for postgraduate education or in the entrance of

postgraduate education in Turkey. ALES is done two times in a year that in spring and autumn period. It is validity until to three years from the first taken (OSYM, 2012). When the reliability of the test is taken into consideration, the same questions aren't asked. In that case, the transformation of the points and comparisons of them can be realized by test equating methods.

When the reached literature is examined, it is seen that the studies intended for test equating are too limited. Also, this study is thought to be important for adding a new perspective to the test equating.

The aim of this study is to determine the most appropriate method between ALES centrically done by OSYM is equated with equipercentile equating method and linear equating method belong to Classical Test Theory.

In line with this objective the below mentioned questions were sought to be answered:

1. According to linear equating method what are the equating scores of the sub dimension of the 2011 ALES spring and autumn?
2. According to equipercentile equating method what are the equating scores of the sub dimension of the 2011 ALES spring and autumn?
3. Which equating method used for scores of the sub dimension of the 2011 ALES spring and autumn is the most proper?

3. Material and Methods

3.1 Research Design

The results from 2011 ALEs spring and autumn are equated according to linear and equipercentile equating methods in Classical Test Theory and the errors gotten from equating method compared with each other. So this study is a kind of basic study.

3.2 Population and Sample

The research population consists of the 21,860 candidates who took both of the 2011 spring and autumn ALESs. The research samples, however, consist of the 2,186 candidates chosen by the random sampling method.

3.3 Data Collecting Instruments and Collecting the Data

The data in this study is consisted of the scores which belong to quantitative, verbal and equiponderant fields obtained from Quantitative-1, Quantitative-2, Verbal-1 and Verbal-2 tests of the participants who took both ALESs in 2011 spring and autumn terms. The population of the research is 21860 candidates who participated both of the

ALES exams in spring and autumn period. The sample is 2186 candidates who are chosen randomly from the population.

3.4 Data Analysis and Meeting to the Equating Conditions

Data analysis was made in four phases. In the first stage, descriptive statistics of the ALES scores was calculated. In the second stage, the equating conditions were tested whether they were ensured or not. In the third stage, the scores were equated by using linear and equipercentile equating methods. In the fourth stage, the mistaken amounts of the each equating method were calculated.

3.4.1 Descriptive Statistics

In the first stage of the data analysis, the descriptive statistics of the test scores of the 2011 spring and autumn ALES were calculated and given on Table 1.

Table 1: Descriptive Statistics for the 2011 Spring and Autumn ALES Test Results

	2011 Spring ALES Results			**2011 Autumn ALES Results**		
Statistics	**Numerical**	**Equiponderant**	**Verbal**	**Numerical**	**Equiponderant**	**Verbal**
K	150	150	150	150	150	150
N	2186	2186	2186	2186	2186	2186
Mean	64,15	63,73	63,21	65,33	65,43	65,39
Median	64,43	60,94	52,95	66,04	62,84	56,27
Mod	71	65	65	65	64	65
Standard Deviation	11,87	12,22	10,43	12,78	9,80	12,81
Variance	141,30	149,35	108,71	163,301	164,00	96,036
Skewness	,005	-,012	-,021	,050	-,057	,064
Kurtosis	-,553	-,312	,157	-,704	-,513	,250
KR-21	,89	,90	,86	,89	,89	,81

According to the descriptive analysis result, the ordinary distribution table created by the comparison of the mod, median and mean values of the ALES dimensions, stem-and-leaf plot and the results of the Shapiro-Wilk test, it was seen that each dimension seemed to have an almost normal distribution. After examining the skewness and kurtosis coefficients of each dimension, the results were found to be close to "0". KR-20 reliability coefficient could not be calculated because of the data gotten from OSYM

matrix of 1-0. Instead of this each of raw dimension's reliability coefficients is measured and this value is changing from 0,81 to 0,90.

3.5 Testing the Equating Conditions

Test conditions whether they providing the equating conditions or not were tested to equate the test forms in each before equating the test results with the research methods. It was stated that single dimensionality, equal reliability and having similar strength levels in researches on equating conditions must be obtained (Angoff, 1982; Crocker & Algina, 1986; Dorans & Holland, 2000; Kolen & Brennan, 2004).

3.5.1 Unidimensionality in the Tests to be Equated

Tests which evaluate same skill or feature can be equated (Lord, 1980). Also, It is not enough to measure the same feature of the two tests, and in the meantime measuring the single feature namely being unidimensional are necessity (Woldbeck, 1998). The data obtained for the research are not the data created in the 1-0 matrix but are the raw scores which correspond to quantitative, verbal and equiponderant dimensions of the Quantitative-1, Quantitative-2, Verbal-1 and Verbal-2 sub tests of the ALES. For this reason, whether each sub test ensures unidimensionality could not be examined by using factor analysis. Instead, KR-21 value was calculated in order to identify whether unidimensionality, which is an equating condition, was provided for each test. The KR-21 reliability coefficient is an indicator of the test's internal consistency (Baykul, 2000; Tekin, 2008). Based on this information, it was tried to be identified by examining the size of the KR-21 reliability coefficient whether the unidimensionality was provided for each sub dimension.

Table 2: KR-21 Reliability Coefficients of ALES Lower Dimensions of 2011 Spring and 2011 Autumn

	2011 Spring ALES Results			**2011 Autumn ALES Results**		
Internal Consistency	**Numerical**	**Equiponderant**	**Verbal**	**Numerical**	**Equiponderant**	**Verbal**
KR-21	0,89	0,90	0,86	0,89	0,89	0,81

According to the values on Table 2, internal consistency of each sub dimension is high and items in each sub dimension evaluate the same feature and thus unidimensionality, which is the first condition of equating, can be said to be provided.

3.5.2 Testing the Difference between the Reliability Coefficients of the Tests to be Equated

Another prior condition for test equating is the equivalence of the reliabilities. Zr values must be acquired out of the KR-21 reliability coefficients of the tests by examining the table of the Transformation of Pearson's R to Fischer's Zr (Akhun et al., 1984; Hovardaoğlu, 1995). The obtained Zr and Z values are shown on Table 3.

Table 3: Comparison of the Reliability of 2011 Spring and Autumn ALES Dimensions

Dimensions of ALES	KR-21	Zr	Z
Spring Numerical	0,89	1,422	0,000
Autumn Numerical	0,89	1,422	
Spring Equiponderant	0,90	1,472	0,158
Autumn Equiponderant	0,89	1,422	
Spring Verbal	0,86	1,293	0,155
Autumn Verbal	0,81	1,127	

3.5.3 Testing the Difference Between the Means and the Variances of the Tests to be Equated

Whether there is a significant difference between the means of the tests was tested by the related sampling t test. The Levene's test was used for testing whether the variances were equal. The arithmetic mean of the 2011 spring and autumn ALES dimensions and the results of t test which tested the significance of the difference and the results of the Levene's test which tested variances and the differences between the variances are shown on Table 4.

Table 4: 2011 Spring and Autumn Means and Variances of ALES Dimensions Comparison Results

		t test		Levene's	Test	
Dimensions	$\bar{X}$	**t**	**p**	S_x^2	**F**	**p**
Spring Numerical	64,15	0,56	,29	141,300	1,86	,41
Autumn Numerical	65,33			163,301		
Spring Verbal	63,21	0,78	,38	108,658	2,85	,22
Autumn Verbal	65,39			96,035		
Spring Equiponderant	63,73	0,63	,36	149,346	1,46	,52
Autumn Equiponderant	65,43			164,004		

$p<0,05$*

In the aftermath of the several analysis, it can be pointed out that the quantitative, verbl and equiponderant sub dimensions of the 2011 spring and autumn ALESs evaluated

just one and same structure and that they had equal reliability level and in terms of their means and variances there was not a significant difference between them.

3.6 Equating Methods

Lord (1980) divided the test forms into two parts as "old form" and "new form". According to the equating methods, old form must be equated to new form. In this test equating study, the 2011 spring term test was regarded as the old form and the 2011 autumn term test was regarded as the new form. While using the equating methods the 2011 spring data were equated to 2011 autumn data.

Under the research, these two forms were managed to be equalted by using linear and equipercentile equating methods and by calculating WMSE values as evaluation criteria, the most appropriate equating method was chosen.

4. Findings and Comment

4.1 What Are the Equated Scores Belong to ALES Examination's Subdimensions of the 2011 Spring and Autumn Period According to Linear Equating Method?

The new form of the test needed to be equated to the old form (Lord, 1980). In this research, the test conducted in 2011 spring term ALES is old from while the autumn term test is the new form.

4.1.1 Linear Equating Findings of the Quantitative Dimension

The raw scores and the equated scores of the autumn term scores of the 2011 quantitative dimensions are shown in the Figure 4.1 below.

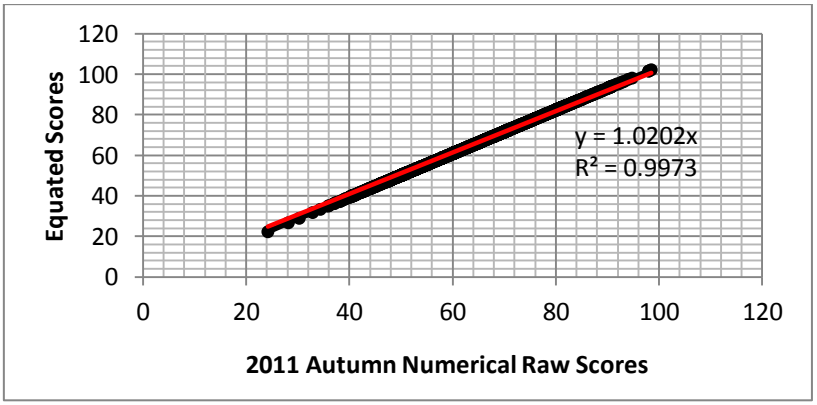

Figure 1: The Equating of the 2011 Spring and Fall Term ALESs Numerical Raw Scores According to the Linear Equating Method

When the Figure 1 is examined, the relation between the raw scores and the scores acquired by using the linear equating method is linear. The equating of this relation was found to be Y=1,0202X. And a positively high relation is seen between the raw scores and the equated scores. The raw scores and the scores acquired by the equating corresponding to these scores are shown in Table 5.

Table 5: Raw Scores for the 2011 Autumn Numerical Dimension and Equated Scores Obtained by Linear Equating Corresponding to These Scores

Raw Score	Equated Score	Difference	Raw Score	Equated Score	Difference	Raw Score	Equated Score	Difference
0	-	-	34	33,35	1,05	68	70,02	1,88
1	-	-	35	35,01	0,94	69	70,86	-1,56
2	-	-	36	36,35	0,84	70	72,15	-1,65
3	-	-	37	37,56	0,76	71	73,04	-1,72
4	-	-	38	38,65	0,68	72	74,15	-1,79
5	-	-	39	39,91	0,6	73	75,21	-1,87
6	-	-	40	40,91	0,53	74	76,83	-1,94
7	-	-	41	41,98	0,45	75	77,38	-2,02
8	-	-	42	42,99	0,38	76	78,4	-2,09
9	-	-	43	44,01	0,31	77	79,52	-2,17
10	-	-	44	45,09	0,23	78	80,58	-2,24
11	-	-	45	46,17	0,16	79	81,67	-2,32
12	-	-	46	47,31	0,08	80	82,9	-2,4
13	-	-	47	48,31	0,01	81	83,83	-2,47
14	-	-	48	49,4	-0,07	82	84,92	-2,55
15	-	-	49	50,66	-0,15	83	85,97	-2,62
16	-	-	50	51,55	-0,22	84	87,04	-2,69
17	-	-	51	52,65	-0,29	85	88,13	-2,77
18	-	-	52	53,72	-0,37	86	89,15	-2,84
19	-	-	53	54,78	-0,44	87	90,16	-2,92
20	-	-	54	55,89	-0,52	88	91,39	-3,02
21	-	-	55	56,9	-0,59	89	92,36	-3,06
22	-	-	56	57,99	-0,67	90	93,62	-3,15
23	-	-	57	59,1	-0,74	91	94,62	-3,22
24	22,39	1,82	58	60,16	-0,82	92	95,84	-3,31
25	-	-	59	61,41	-0,9	93	97	-3,39
26	-	-	60	62,31	-0,97	94	98,28	-3,48
27	-	-	61	63,42	-1,04	95	-	-
28	26,58	1,53	62	64,49	-1,12	96	-	-
29	-	-	63	65,55	-1,19	97	-	-

30	28,92	1,36	64	66,62	-1,27	98	101,92	-3,73
31	-	-	65	67,68	-1,34	99	-	-
32	-	-	66	68,76	-1,42	100	-	-
33	31,73	1,17	67	69,85	-1,49			

The scores that can be obtained from the quantitative dimension of the ALES starting from 1 to 100 are shown in raw score section in Table 5 while the equated scores range from 22,39 to 101,92. The raw scores and the equated scores in the consecutive values between 48 and 53 and between 41 and 47 are close to each other. This result is linked to the difference which is between -0,5 and 0,5. Also when the autumn scores are equated to the spring form, it is seen that excluding the scores corresponding to "-" in the score distribution in Table 5, the equated scores of other scores such as 24 and 47 had higher values and those who obtained scores between 48 and 98 had higher equated scores in the aftermath of linear equating.

According to the data in the table, it is seen that the equated raw score of 98 exceeded the 0-100 raw score interval and had a value of 101,92. The exceeding of the equated scores out of the raw score interval is the characteristic of the linear equating.

4.1.2 Linear Equating Findings of the Verbal Dimension

The raw scores and the equated scores of the autumn term scores of the 2011 verbal dimensions are shown in the Figure 2 below.

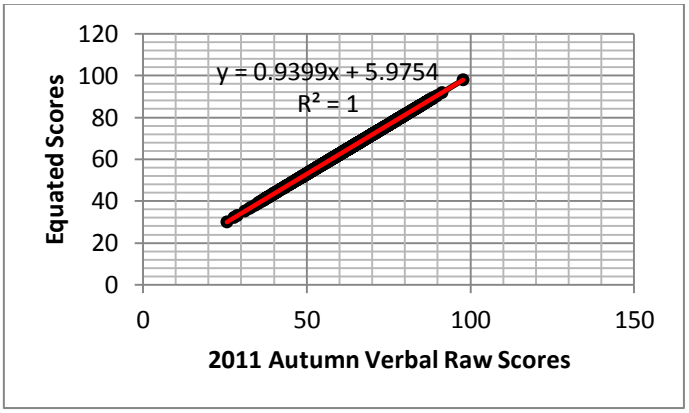

Figure 2: The Equating of the Verbal Raw Scores of the 2011 Spring and Fall Term ALESs By Using Linear Equating Method

When the raw scores in the Figure 2 and the scores obtained by the linear equating method are analyzed, it is found out that they were distributed on a linear line. The relation between data groups in the graphic is linear. The equating of this relation is Y=0,9399X+ 5,9754. When the graphic is analyzed, it can be said that there is a positively

high relation between the raw scores and the equated scores. The raw scores of the verbal dimension and the equated new values of these scores are shown in Table 6.

Table 6: Raw Scores for the 2011 Autumn Verbal Dimension and Equated Scores Obtained by Linear Equating Corresponding to These Scores

Raw Score	Equated Score	Difference	Raw Score	Equated Score	Difference	Raw Score	Equated Score	Difference
0	-	-	34	38,05	-3,92	68	70,36	-1,86
1	-	-	35	39,19	-3,85	69	71,35	-1,8
2	-	-	36	40,21	-3,79	70	72,19	-1,74
3	-	-	37	41,14	-3,73	71	73,2	-1,68
4	-	-	38	42,04	-3,67	72	74,11	-1,62
5	-	-	39	43,15	-3,6	73	75,03	-1,56
6	-	-	40	43,96	-3,55	74	76,02	-1,5
7	-	-	41	44,96	-3,48	75	76,94	-1,44
8	-	-	42	45,77	-3,43	76	77,81	-1,38
9	-	-	43	46,78	-3,37	77	78,83	-1,32
10	-	-	44	47,75	-3,3	78	79,82	-1,25
11	-	-	45	48,77	-3,24	79	80,62	-1,2
12	-	-	46	49,77	-3,18	80	81,55	-1,14
13	-	-	47	50,65	-3,12	81	82,58	-1,08
14	-	-	48	51,57	-3,06	82	83,56	-1,01
15	-	-	49	52,56	-3	83	84,44	-0,96
16	-	-	50	53,45	-2,94	84	85,4	-0,9
17	-	-	51	54,32	-2,88	85	86,35	-0,84
18	-	-	52	55,35	-2,82	86	87,31	-0,77
19	-	-	53	56,21	-2,76	87	88,09	-0,72
20	-	-	54	57,22	-2,7	88	88,97	-0,67
21	-	-	55	58,15	-2,64	89	90,24	-0,59
22	-	-	56	59,12	-2,58	90	-	-
23	-	-	57	60,01	-2,52	91	91,71	-0,49
24	-	-	58	60,98	-2,46	92	-	-
25	30,07	-4,44	59	61,92	-2,4	93	-	-
26	-	-	60	62,83	-2,34	94	-	-
27	32	-4,31	61	63,8	-2,28	95	-	-
28	33,04	-4,24	62	64,74	-2,22	96	-	-
29			63	65,61	-2,16	97	97,87	-0,1
30			64	66,63	-2,1	98	-	-
31	35,22	-4,11	65	67,54	-2,04	99	-	-
32	36,75	-4,06	66	68,45	-1,98	100	-	-
33			67	69,44	-1,92			

When the Table 6 above is examined, it is seen that all raw scores regarded as exceptional are lower than the equated scores. Kolen and Brennan (2004) links this situation to the regulation, which is made by the linear equating, of the difficulty level through the score scale between the forms. According to the research it can be said that the difficulty level of the 2011 spring and autumn term ALES's verbal scores vary through the score scale. It was monitored that the raw scores between 0 and 100 scale of the verbal scores did not exceed 0-100 scale after the equating.

4.1.3 Linear Equating Findings of the Equiponderant Dimension

The scores of the equiponderant dimension of the 2011 autumn term ALES were sought to be equated by using the linear equating method. The raw scores of the equiponderant dimension of the 2011 autumn term the equated scores are shown in Figure 3.

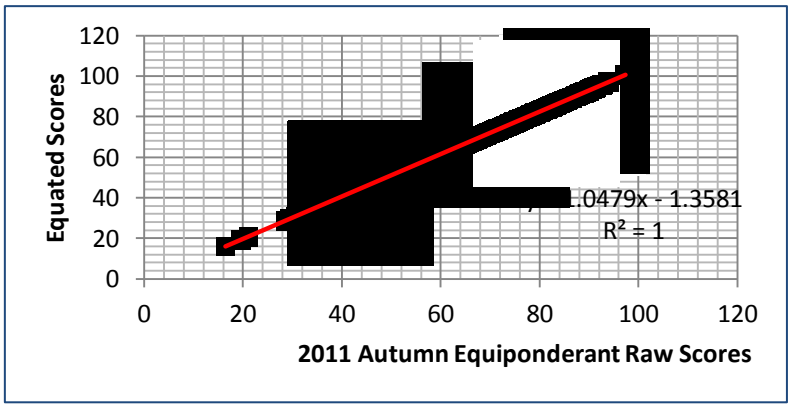

Figure 3: The Equating of the Equiponderant Raw Scores of the 2011 Spring and Fall ALESs According to the Linear Equating Method

When the raw scores and the scores obtained by using the linear equating method in Figure 3 are examined, it is determined that they are distributed in a linear line. The relation between the data groups in the graphic is linear. The equating of this relation is Y= 1,0479X-1,3581. When the graphic is analyzed, it is seen that the there is a positive relation between the raw scores and the equated scores. The equiponderant raw scores and the new values acquired by equalizing these scores are shown in Table 7.

Table 7: Raw Scores for the 2011 Autumn Equiponderant and Equated Scores Obtained by Linear Equating Corresponding to These Scores

Raw Score	Equated Score	Difference	Raw Score	Equated Score	Difference	Raw Score	Equated Score	Difference
0			34	34,65	-0,29	68	78,27	-2,28
1	-	-	35	35,73	-0,34	69	71,51	-1,97
2	-	-	36	36,83	-0,39	70	72,5	-2,02
3	-	-	37	37,99	-0,44	71	73,58	-2,07
4	-	-	38	38,9	-0,48	72	74,65	-2,41
5	-	-	39	39,96	-0,53	73	75,62	-2,16
6	-	-	40	41,1	-0,58	74	76,71	-2,21
7	-	-	41	42,09	-0,63	75	77,72	-2,26
8	-	-	42	43,18	-0,68	76	78,82	-2,31
9	-	-	43	44,3	-0,73	77	79,8	-2,4
10	-	-	44	45,3	-0,78	78	80,95	-2,41
11	-	-	45	46,32	-0,82	79	82,05	-2,46
12	-	-	46	47,3	-0,87	80	83,03	-2,5
13	-	-	47	48,53	-0,92	81	83,98	-2,55
14	-	-	48	49,55	-0,97	82	85,04	-2,59
15	-	-	49	50,49	-1,01	83	86,18	-2,65
16	15,93	0,07	50	51,58	-1,06	84	87,14	-2,69
17	-	-	51	52,64	-1,11	85	88,6	-2,74
18	-	-	52	53,64	-1,16	86	89,16	-2,93
19	19,18	-0,18	53	54,73	-1,21	87	90,21	-2,96
20	-	-	54	55,78	-1,26	88	91,47	-2,99
21	20,72	0,28	55	56,83	-1,3	89	92,42	-2,93
22	-	-	56	57,9	-1,35	90	93,38	-3,08
23	-	-	57	58,95	-1,4	91	94,26	-3,11
24	-	-	58	59,95	-1,45	92	95,5	-3,07
25	-	-	59	61	-1,49	93	97,03	-3,14
26	-	-	60	62,07	-1,54	94	97,29	-3,15
27	-	-	61	63,15	-1,59	95	-	-
28	28,59	-0,59	62	64,1	-1,64	96	-	-
29	29,44	-0,44	63	65,2	-1,69	97	100,61	-3,31
30	30,83	-0,83	64	66,16	-1,73	98	-	-
31	31,75	-0,16	65	67,27	-1,78	99	-	-
32	32,59	-0,19	66	68,33	-1,83	100	-	-
33	34,02	-0,26	67	69,35	-1,88			

The equated scores exceeded the 0-100 score scale. This is a characteristic of linear equating. The equated scores of 21, 19, 30 and 16 are seen to be lower than the raw

scores in the table. Normally, the equated scores are seen to be higher than the raw scores. This, in linear equating, is clarified by the raw and equated scores of the 2011 spring term test in which the difference of the difficulty levels of the test forms varied through the score scale.

In light of the comments made for the sub problem situations by using the linear equating method, it is seen that the linear equating and the raw scores do not coincide with each other. Livingston (2004) describes this as a characteristic of linear equating. Kolen and Brennan (1995), on the other hand, state that the exceeding of the equated scores acquired by linear equating of the tests, which are graded upon the number of true answers, is an expected outcome in linear equating and offer two ways to deal with this situation. The first way is to allow equated high and low scores that exceed the raw score limit and the second way is to remove the highest and the lowest equated scores from the equating process.

4.2 What are the Equated Scores of Sub Dimensions of the 2011 Spring and Autumn Terms ALESs According to the Equipercentile Equating Method?

2011 spring term ALES was chosen as a reference form for the solution of the quantitative dimension data of this sub problem. Under the research, equipercentile equating method was conducted to the reference and the new form single group mechanism. Before equalizing the tests, cumulative percentages of the raw scores were calculated and their percentage arrays were identified. Since the raw scores of both sub tests rarely coincide with the same percentage array, shift formula, which is offered by Livingston, was conducted.

4.2.1 The Equipercentile Equating Findings of the Quantitative Dimension

The graphic of the raw scores and equal scores corresponding of these scores is shown in Figure 4.

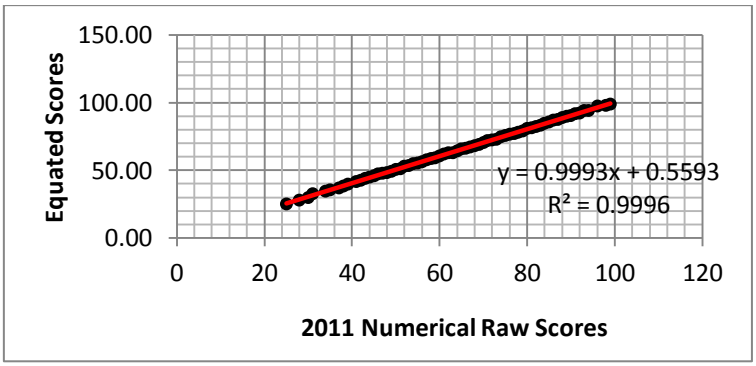

Figure 4: The Equating of the 2011 Spring and Fall Term ALESs Numerical Raw Scores According to the Equipercentile Equating Method

When the Figure 4 is analyzed, it is seen that there is a linear relation between the 2011 autumn term ALES's quantitative raw scores and the data acquired after the equipercentile equating method. This relation can be pointed out with the Y=0,9993X+0,5593 linear equating. When the graphic is analyzed, it is seen that there is a positively high relation between the raw scores and the scores obtained with the equipercentile equating.

Table 8: Numerical Dimension Raw Scores and Equated New Scores of the Scores

Raw Score	Equated Score	Difference	Raw Score	Equated Score	Difference	Raw Score	Equated Score	Difference
0	-	-	34	34.56	-0.56	68	68.31	-0.31
1	-	-	35	35.64	-0.64	69	69.14	-0.14
2	-	-	36	-	-	70	70.56	-0.56
3	-	-	37	37.00	0	71	71.88	-0.88
4	-	-	38	38.53	-0.53	72	72.55	-0.55
5	-	-	39	39.90	-0.9	73	73.03	-0.03
6	-	-	40	-	-	74	74.67	-0.67
7	-	-	41	41.82	-0.82	75	75.58	-0.58
8	-	-	42	42.40	-0.4	76	76.63	-0.63
9	-	-	43	43.93	-0.93	77	77.05	-0.05
10	-	-	44	44.92	-0.92	78	78.02	-0.02
11	-	-	45	45.85	-0.85	79	79.02	-0.02
12	-	-	46	47.11	-1.11	80	80.72	-0.72
13	-	-	47	47.87	-0.87	81	81.41	-0.41
14	-	-	48	48.03	-0.03	82	82.45	-0.45
15	-	-	49	49.15	-0.15	83	83.16	-0.16
16	-	-	50	50.62	-0.62	84	84.60	-0.6
17	-	-	51	51.05	-0.05	85	85.52	-0.52
18	-	-	52	52.98	-0.98	86	86.89	-0.89
19	-	-	53	53.51	-0.51	87	87.71	-0.71
20	-	-	54	54.98	-0.98	88	88.94	-0.94
21	-	-	55	55.34	-0.34	89	89.93	-0.93
22	-	-	56	56.50	-0.5	90	90.12	-0.12
23	-	-	57	57.49	-0.49	91	91.67	-0.67
24	-	-	58	58.40	-0.4	92	92.36	-0.36
25	25.00	0	59	59.26	-0.26	93	93.95	-0.95
26	-	-	60	60.56	-0.56	94	94.31	-0.31
27	-	-	61	61.93	-0.93	95	-	-
28	28.00	0	62	62.93	-0.93	96	97.44	-1.44
29	-	-	63	63.07	-0.07	97	-	-

30	30.00	0	64	64.22	-0.22	98	98.00	0.00
31	32.60	-1.6	65	65.51	-0.51	99	99.00	0.00
32	-	-	66	66.29	-0.29	100	-	-
33	-	-	67	67.36	-0.36			

It is seen in Table 8 that the equated scores got values between 25 and 99 and that no scores given in the range exceeded the raw score scale. Kolen and Brennan (1995) state that the problem of the equated scores' getting values out of the raw score scale in linear equating does not occur in the equipercentile equating. The obtained findings seem to be supporting this outcome.

The equated scores are always higher than raw scores in the table. Because the equipercentile equating organizes the difficulty levels of the test forms with the help of a curve, the score distribution of the equated scores' always getting higher values than the raw scores can be described as there is no difference with the difficulty levels of the items through the score distribution scale in the quantitative dimensions of the 2011 spring and autumn terms of ALES.

There is not a significant surge between the data through the line in Figure 4 which shows the raw and equated scores. That might be because there is not a noteworthy difference between the means of the equated scores in the quantitative dimensions of the 2011 spring and autumn ALESs, because their variances get close values and because there are small differences between the skewness and kurtosis coefficients.

4.2.2 The Equipercentile Equating Findings of the Verbal Dimension

The graphic of the raw scores and the equal scores corresponding of these scores is shown in Figure 5.

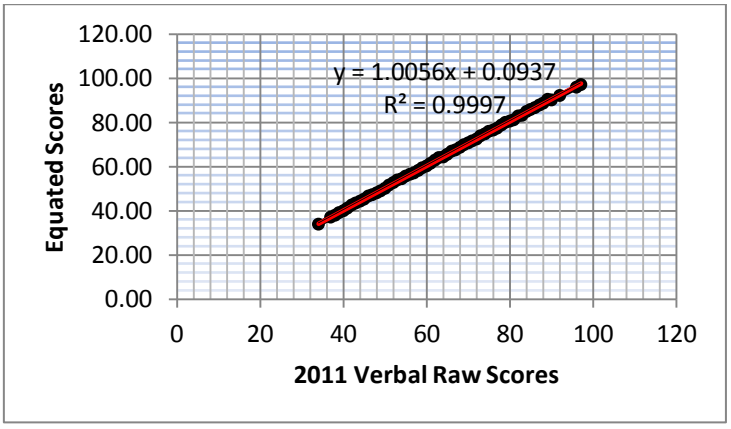

Figure 5: The Equating of the 2011 Spring and Fall Term ALESs Verbal Raw Scores According to the Equipercentile Equating Method

When the Figure 5 is analyzed, it is seen that there is a linear relation between the 2011 autumn term ALES's verbal raw scores and the data acquired after the equipercentile equating method. This relation can be pointed out with the Y=1,0056X+0,0937 linear equating. When the graphic is analyzed, it is seen that there is a positively high relation between the raw scores and the scores obtained with the equipercentile equating.

Table 9: Verbal Dimension Raw Scores and Equated New Scores of the Scores

Raw Score	Equated Score	Difference	Raw Score	Equated Score	Difference	Raw Score	Equated Score	Difference
0	-	-	34	34	0	68	68,74	-0,74
1	-	-	35	-	-	69	69,83	-0,83
2	-	-	36	-	-	70	70,78	-0,78
3	-	-	37	37	0	71	71,67	-0,67
4	-	-	38	38	0	72	72,5	-0,5
5	-	-	39	39	0	73	73,74	-0,74
6	-	-	40	40	0	74	74,69	-0,69
7	-	-	41	41,25	-0,25	75	75,8	-0,8
8	-	-	42	42,69	-0,69	76	76,6	-0,6
9	-	-	43	43,4	-0,4	77	77,24	-0,24
10	-	-	44	44,22	-0,22	78	78,92	-0,92
11	-	-	45	45,18	-0,18	79	79,8	-0,8
12	-	-	46	46,75	-0,75	80	80,46	-0,46
13	-	-	47	47,15	-0,15	81	81,06	-0,06
14	-	-	48	48,21	-0,21	82	82,77	-0,77
15	-	-	49	49	0	83	83,14	-0,14
16	-	-	50	50,19	-0,19	84	84,78	-0,78
17	-	-	51	51,59	-0,59	85	85,72	-0,72
18	-	-	52	52,63	-0,63	86	86,7	-0,7
19	-	-	53	53,75	-0,75	87	87,7	-0,7
20	-	-	54	54,44	-0,44	88	88,6	-0,6
21	-	-	55	55,66	-0,66	89	90,38	-1,38
22	-	-	56	56,58	-0,58	90	90	0
23	-	-	57	57,08	-0,08	91	-	-
24	-	-	58	58,32	-0,32	92	92	0
25	-	-	59	59,24	-0,24	93	-	-
26	-	-	60	60,3	-0,3	94	-	-
27	-	-	61	61,42	-0,42	95	-	-
28	-	-	62	62,78	-0,78	96	96	0
29	-	-	63	63,93	-0,93	97	97	0

30	-	-	64	64,34	-0,34	98	-	-
31	-	-	65	65,46	-0,46	99	-	-
32	-	-	66	66,77	-0,77	100	-	-
33	-	-	67	67,52	-0,52			

The equated scores are always higher than raw scores in Table 9. Because the equipercentile equating organizes the difficulty levels of the test forms with the help of a curve, the score distribution of the equated scores' always getting higher values than the raw scores can be described as there is no difference with the difficulty levels of the items through the score distribution scale in the verbal dimensions of the 2011 spring and autumn terms of ALES.

4.2.3 The Equipercentile Equating Findings of the Equiponderant Dimension

The equiponderant raw scores of the 2011 autumn term ALES were sought to be equated by using the equipercentile equating method. The graphic of the raw scores and the equal scores corresponding of these scores is shown in Figure 6.

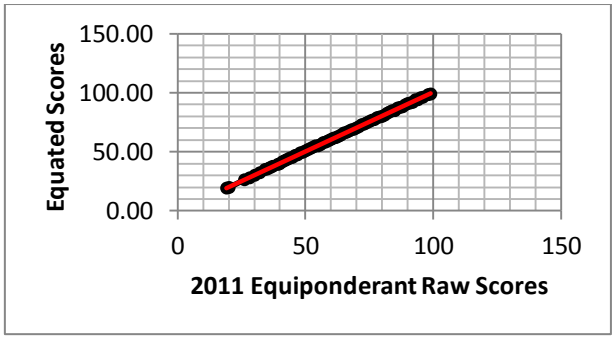

Figure 6: The Equating of the 2011 Spring and Fall Term ALESs Equiponderant Raw Scores According to the Equipercentile Equating

When the Figure 6 is analyzed, it is seen that there is a linear relation between the 2011 autumn term ALES's equiponderant raw scores and the data acquired after the equipercentile equating method. This relation can be pointed out with the Y=1,0017X+0,323 linear equating. When the graphic is analyzed, it is seen that there is a positively high relation between the raw scores and the scores obtained with the equipercentile equating.

Table 10: Equiponderant Dimension Raw Scores and Equated New Scores of the Scores

Raw Score	Equated Score	Difference	Raw Score	Equated Score	Difference	Raw Score	Equated Score	Difference
0			34	34,78	-0,78	68	68,51	-0,51
1	-	-	35	35,38	-0,38	69	69,09	-0,09
2	-	-	36	36,56	-0,56	70	70,05	-0,05
3	-	-	37	37,61	-0,61	71	71,05	-0,05
4	-	-	38	-	-	72	72,97	-0,97
5	-	-	39	39,32	-0,32	73	73,50	-0,50
6	-	-	40	40,00	0,00	74	74,76	-0,76
7	-	-	41	41,34	-0,34	75	75,52	-0,52
8	-	-	42	42,66	-0,66	76	76,52	-0,52
9	-	-	43	43,77	-0,77	77	77,23	-0,23
10	-	-	44	44,77	-0,77	78	78,98	-0,98
11	-	-	45	45,00	0,00	79	79,29	-0,29
12	-	-	46	46,63	-0,63	80	80,09	-0,09
13	-	-	47	47,63	-0,63	81	81,00	0,00
14	-	-	48	48,94	-0,94	82	82,71	-0,71
15	-	-	49	49,79	-0,79	83	83,82	-0,82
16	-	-	50	50,67	-0,67	84	84,95	-0,95
17	-	-	51	51,73	-0,73	85	85,00	0,00
18	-	-	52	52,79	-0,79	86	86,86	-0,86
19	19,00	0,00	53	53,73	-0,73	87	87,74	-0,74
20	20,00	0,00	54	54,82	-0,82	88	88,00	0,00
21	-	-	55	55,02	-0,02	89	89,78	-0,78
22	-	-	56	56,26	-0,26	90	90,83	-0,83
23	-	-	57	57,66	-0,66	91	91,55	-0,55
24	-	-	58	58,50	-0,50	92	92,00	0,00
25	-	-	59	59,45	-0,45	93	93,78	-0,78
26	26,00	0,00	60	60,37	-0,37	94	94,18	-0,18
27	-	-	61	61,24	-0,24	95	95,68	-0,68
28	28,00	0,00	62	62,09	-0,09	96	96,00	0,00
29	-	-	63	63,02	-0,02	97	-	-
30	30,01	-0,01	64	64,12	-0,12	98	98,36	-0,36
31	-	-	65	65,89	-0,89	99	99,00	0,00
32	32,01	-0,01	66	66,62	-0,62	100	-	-
33	-	-	67	67,28	-0,28			

The equated scores are always higher than raw scores in the table. Because the equipercentile equating organizes the difficulty levels of the test forms with the help of a curve, the score distribution of the equated scores' always getting higher values than

the raw scores can be described as there is no difference with the difficulty levels of the items through the score distribution scale in the equiponderant dimensions of the 2011 spring and autumn terms of ALES.

4.3 Among the Equating Methods Carried Out for the Scores of the Sub Dimensions of the 2011 Spring and Autumn Terms ALESs, Which One is the Most Appropriate?

Under the research, the quantitative, verbal and equiponderant raw scores of the 2011 autumn term ALES was equated to the quantitative, verbal and equiponderant raw scores of the 2011 spring term ALES. In order to find out which method is more appropriate for equalizing and which method fails, the WMSE coefficients, which gives the weighted error mean squares of both methods, were calculated.

Table 11: WMSE Results of Dimensions and Equating Methods

Equating Methods	Dimensions	WMSE
Linear Equating Method	Numerical	0,0292
Equipercentile Equating Method	Numerical	0,0273
Linear Equating Method	Verbal	0,114
Equipercentile Equating Method	Verbal	0,087
Linear Equating Method	Equiponderant	0,394
Equipercentile Equating Method	Equiponderant	0,283

When WMSE values in the table above are examined, it is seen that there are fewer mistakes in every dimension of the equipercentile method. However, the difference between the amounts of mistakes among the equating methods is not big. Still, equipercentile equating method can be said to be the most appropriate method for this research.

5. Result and Suggestions

The raw scores of the quantitative and equiponderant dimensions of the 2011 autumn term ALES were equated to the spring term ALES same year by using linear equating method and it was monitored that the equated scores exceeded the raw score scale. It was seen that the difference between the equated scores of the sample's raw scores were gradually increasing in the extreme values. This could be because the mean and variance values in the descriptive statistics are different than each other. Jaeger (1981) states that because of the nature of the linear equating method the score scale of the raw scores and the equated scores do not match exactly while Livingston (2004) describes this as a characteristic of linear equating.

The raw scores of the verbal dimension of the 2011 autumn term ALES were equated to the spring term ALES same year by using linear equating method. The equated scores did not exceed the raw score scale. And a positively high and linear relation was found between the raw scores and the equated scores concerning the quantitative, verbal and equiponderant dimensions in linear equating. The raw scores of the quantitative dimension of the 2011 autumn term ALES were equated to the spring term ALES same year by using equipercentile equating method. The equated scores exceeded the raw score scale. It was seen that the difference between the equated scores of the sample's raw scores were gradually increasing in the extreme values. This could be because the mean and variance values in the descriptive statistics are different than each other. The raw scores of the verbal and equiponderant dimensions of the 2011 autumn term ALES were equated to the spring term ALES same year by using equipercentile equating method. But the equated scores did not exceed the raw score scale in these dimensions. After the equipercentile equating method, the equiponderant raw scores always got higher values than the raw scores. This occurred because the difficulty levels of the items, which evaluated this dimension, did not have any differences through the score scale. Additionally, the mean and variance values in the descriptive statistics can be said to be different than each other. And a positively high and linear relation was found between the raw scores and the equated scores concerning the quantitative, verbal and equiponderant dimensions in linear equating. When the ALES scores of different terms were analyzed, it was seen that the score distribution of the dimensions in the research were similar in spring and autumn terms, the equated scores were similar and the equated mistake amount had as low values as possible.

When WMSE values are examined, it is seen that there are fewer mistakes in every dimension of the equipercentile method. Equipercentile equating method can be said to be the most appropriate method for this research.

5.1 Suggestions

If a similar research will ever be carried out, given the results of this research, it might be offered to use equipercentile equating method to equate the scores of the quantitative, verbal and equiponderant dimensions of the ALES. ALES is a very important test for people who are to be accepted to many institutions and many post-graduate programmes. Therefore, when attending the post-graduate programmes, the academicians from the assessment and evaluation departments of the universities can work together and contact OSYM in order to prepare an equivalence of the scores of the applicants that took tests such as ALES and YDS in different years and thus by

calculating the scores of the applicants with a common test form created with the statistics of the tests, there can be fairer assessment and evaluation. A different test equating study can be sought with a different equating pattern and an equating method in Item Response Theory. Equating of the scores acquired in different years can be made by choosing the most appropriate method according to the size of the sample and by including chance factor into the study it can be ensured by using the scores obtained in language proficiency tests like TOEFL, IELTS and tests such as KPDS, UDS and YDS carried out by OSYM.

References

1. Akhun, İ. (1984). İki Korelasyon Katsayısı Arasındaki Farkın Manidarlığının Test Edilmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, *17*(1).
2. Angoff, W. H. (1971). Scales, Norms and Equivalent Scores *Educational Measurement* (pp. 509-600): American Council on Education.
3. Angoff, W. H. (1982). Summary and Derivation of Equating Methods Used at ETS. *Test Equating*, *55*, 69.
4. Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması* ÖSYM Yayınları.
5. Braun, H. I., & Holland, P. W. (1982). *Observed Score Test Equating. A Mathematical Analysisof Same ETS Equating Procedure*: Academic Press
6. Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*: Holt, Rinehart and Winston.
7. Dorans, J. N., & Holland, P. W. (2000). Population in Variance and the Equitability of Tests: Basic Theory and the Linear Case. *Journal of Educational Measurement*, *37*, 281-306.
8. Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications.*: Kuluwer-Nijhoff Publisihing.
9. Hovardaoğlu, S. (1995). *Doğrusal Regresyon ve Korelasyona Giriş*: Habitoğlu Basım ve Yayın.
10. Jaeger, R. M. (1981). Some Exploratory Indices for Selection of a Test Equating Method. *Journal of Educational Measurement*, *18*(1), 23-38.
11. Kolen, M. J. (2007). Data Collection Designs and Linking Procedures *Linking and Aligning Scores and Scales* (pp. 31-55): Springer.
12. Kolen, M. J., & Brennan, R. L. (1995). Observed Score Equating Using the Random Groups Design *Test Equating* (pp. 28-65): Springer.

13. Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking*: Springer.
14. Livingston, S. A. (2004). Equating test scores (without IRT). *Educational testing service*.
15. Lord, M. F. (1980). *Application of Item Response Theory to Practical Testing Problems.*: Lawrence Erlbaum Associates Publishers.
16. OSYM. (2012). Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) Sonbahar Dönemi Kılavuzu.
17. Penfield, R., & Camilli, G. (2007). Test Fairness and Differential Item Functioning. *Handbook of Statistics: Psychometrics* (Vol. 26, pp. 125-167).
18. Tekin, H. (2008). *Eğitimde Ölçme ve Değerlendirme* (19 ed.): Yargı Yayınevi.
19. Thorndike, R. L. (1982). *Applied Psychometrics*: Houghton Mifflin Company.
20. Woldbeck, T. (1998). *Basic Concepts in Modern Methods of Test Equating*: Southwest Psychological Association.