The Effect of Using Relative and Absolute Criteria to Decide Students' Passing or Failing a Course

Ayfer Sayin

Correspondence: Ayfer Sayin, Department of Measurement and Assessment, Gazi University, Ankara, Turkey.

Abstract

In the formation education that is carried out within the scope of undergraduate and non-thesis graduate programs within the same university, different criteria are used to evaluate students' success. In this study, classification accuracy of letter grades that are generated to evaluate students' success using relative and absolute criteria and decisions for students' passing or failing a course were examined. Within the scope of this study, it was also intended to determine the cut-off point required for students to pass a course. In this regard, midterm and final grades of a total of 141 students. First, correct classification percentages of the letter grades that the students scored with absolute and relative evaluations were calculated. Then, classification percentages for decisions regarding passing or failing a course were examined. Then, a cut-off point that decisions for students' passing or failing a course will be based on was determined using cluster analysis. At the end of the study, it was determined that the relative criterion provided a more accurate classification than the absolute criterion in determining letter grades, whereas the absolute criterion provided a more accurate classification in decisions for students' passing or failing courses. The relative criterion was found to have provided more benefits to students than the absolute criterion both in terms of letter grading and decision making process for students' passing a course. It was determined that the cut-off points required to decide students' passing a course is more parallel to the absolute criterion.

Keywords: absolute assessment, relative assessment, discrimination analysis, cluster analysis

1. Introduction

1.1 Introduce the Problem

Measurement is a process of separating experimental units by numbers according to their identified properties in a way to preserve relationships in the areas of behavior and to define their properties (Lord and Novick, 1968). According to Turgut (1992), it is a process of observing a property and displaying the result of the observation with numbers or other symbols. Evaluation, which is a judgment or decision-making process, is the process of reaching a decision for the results by comparing obtained measurement results with a criterion (Tekin, 2004). In this regard, measurement process is a prerequisite of the evaluation process; false or inaccurate measurement results may lead to inaccurate evaluations. On the other hand, measurement results make sense with the process of evaluation.

It is necessary to reach a decision by comparing measurement results with a criterion in order for them to make sense. In decision-making process, the criterion becomes important as well as measurement results.

Criterion is a measurement which could be used to determine the accuracy of a decision. In psychological testing, criteria typically represent measures of the outcomes that specific treatments or decision are designed to produce (Murphy and Davidshofer, 1991).

There are two kinds of criteria, absolute criterion and relative criterion, and decisions made based on these criteria are referred to as absolute and relative assessments. Glaser and Klaus (1962) used the term "criterion referenced measurement" as an indicator of individual success. Then this term gained recognition in the field of measurement as "criterion-referenced measures" (Glaser, 1963) and based on success standards of students rather than norms. The major reason for using a norm-referenced tests is to classify students. This tests are designed to highlight achievement differences between and among students to produce a dependable rank order of students across a continuum of achievement from high achievers to low achievers (Stiggins, 1994).

Relative criterion is the criterion used in making decisions for distribution of the measurement results within the group.

Mandernach (2003) reported that the relative criterion is effective in revealing differences between students and continuously provides standard distribution of scores.

Absolute evaluation is determined as a ratio of behaviors for which success can be considered sufficient in a particular scope. In this case the criterion is a pre-determined absolute value or an absolute threshold value that is independent of the group and is the same for everyone. Relative evaluation is an evaluation based on the criterion obtained from the results after measurement. The criterion used in this evaluation is of a norm quality obtained from the group from which measurement is taken (Atılgan, Yurdakul, Öğretmen, 2012).

The differences and similarities between criterion-referenced and norm-referenced assessments are purpose, design of assessment, unit of score interpretation, and score presentation. Criterion referenced tests reflect the progress of development of individual students (Bond, 1995, Huitt, 1996, Lok, McNaught & Young, 2015).

The differences and similarities between criterion-referenced and norm-referenced assessments are summarised in Table 1.

Table 1. Comparison between criterion-referenced and norm-referenced assessments

Comparasion Criterion	Criterion-referenced assessment	Norm-referenced assessment
Purpose	To determine whether each student has achieved specific skills or concepts. To find out how much students know before instruction begins and after it has finished.	To rank each student with respect to the achievement of others in broad areas of knowledge. To discriminate between high and low achievers.
Content	Measures specific skills which make up a designated curriculum. These skills are identified by teachers and curriculum experts. Each skill is expressed as an instructional objective.	Measures broad skill areas sampled from a variety of textbooks, syllabi, and the judgments of curriculum experts.
Item Characteristics	Each skill is tested by at least four items in order to obtain an adequate sample of student performance and to minimize the effect of guessing. The items which test any given skill are parallel in difficulty.	Each skill is usually tested by less than four items. Items vary in difficulty. Items are selected that discriminate between high and low achievers.
Score Interpretation	Each individual is compared with a preset standard for acceptable achievement. The performance of other examinees is irrelevant. A student's score is usually expressed as a percentage. Student achievement is reported for individual skills.	Each individual is compared with other examinees and assigned a score--usually expressed as a percentile, a grade equivalent score, or a stanine. Student achievement is reported for broad skill areas, although some norm-referenced tests do report student achievement for individual skills.
Design of assessment tasks	Align with content and expected outcomes	Discriminates high and low achievers
Score Presentation	Grades linked to criteria	Grades, derived from raw scores, usually presented in a bell curve and often coarse grained into letter grades

There are situations where absolute and relative criteria can be used separately. Especially in situations where there is a quota and an order is required, it is necessary to use relative criterion. However, absolute criteria should be used in situations where a specific skill and competency is required. In the process of evaluation of student success in higher education, some universities use relative criterion while some universities use absolute criterion. There is not a consensus in this regard yet. In addition, similar to the case which is the subject of this study, while some departments and courses in a university use absolute criterion, some other departments in the same university use relative criterion to evaluate student success. Therefore, it seems important to examine accuracy of the decisions made based on different criteria.

In absolute evaluation, cut-off points are determined to evaluate student success according to their performance standards at the same time. However, there are also different opinions as to what these cut-off points should be. In education, especially in the process of evaluation of large-scale successes, politicians, educators, content specialists, assessment professionals and other decision makers take part (McClarty, Way, Porter, Beimers & Miles, 2013). The length of the test, validity and reliability of the test and changes in the decision-making process affect changes in cut-off points. Reid and Dodds (2013) reported in their study that validity of cut-off points used to evaluate student attributes should be determined based on actual student data. As a result of border point and regression line methods, it was determined that higher cut-off points than previously set cut-off points are required.

1.2 Explore Importance of the Problem

In Gazi University, the letter grades for graduate students' passing courses are determined using relative criterion. However, absolute criterion is used to determine success in formation education of the same course, even the course instructed by the same academician most of the time. In this direction, it was aimed to investigate the effect of evaluations based on relative and absolute criteria, which vary even within the same university, on success grades and letter grades of students. Within the scope of the study, also passing grades which are required as fail-pass grades based on students' midterm and final grades were determined.

2. Method

In this section, research design, research group, data collection methods and data analysis are presented.

2.1 Research Model

Correlational research is an example of what is sometimes called asociational research. In associational research, the relationships among two or more variables are studied without ant attemp to influence them (Fraenkel & Wallen, 2009; Vanderstoep & Johnston, 2009). In this study, accuracy of the cut-off point determined to decide whether or not formation students pass a course. In this regard, the validity of the classification of units included in the tests applied to the students was examined, and then calculation of a cut-off point required for accurate classifications was carried out.

2.2 Participants

In this study, the research group included a total of 141 students, receiving formation education in two separate classes in Gazi University in the academic year 2015-2016. The students were seniors and receiving formation education at the same time. In other words, there were final year students of the university. 51.8% (n=73) of the students in the research group were studying Turkish philology and 48.2% (n=68) of the students were studying physical education.

2.3 Data Collection

The research data were collected through measurement and evaluation midterm and final exams developed by the researcher. Each achievement test consisted of 25 multiple choice questions each with 5 possible answers.

During the development process of the achievement tests, steps suggested by Crocker and Algina (1986) were followed. Whether or not students passed a course would be determined by looking at their test results conducted to determine their levels. Forming a table of specifications, weights and levels of each subject were determined. In this regard, it was decided to include 25 questions in each exam. Earlier, an experimental application was carried out and first drafts of the exams were prepared by selecting items with an item discrimination index over 0.20. The generated draft exams were examined prior to actual examinations by a specialist of the subject field and a measurement and evaluation specialist and some options were revised. After giving the test items to the formation students constituting the research group, difficulty index and discrimination index of the items were recalculated.

The difficulty index of the 25 items included in the midterm exam was determined to vary between 0.32 and 0.96. Students who took the midterm exam correctly answered an average of 18.75 questions. It was determined that the discrimination indexes of the items varied between 0.21 and 0.63. The average discrimination index of the items was calculated as 0.36. In accordance with the students' answers, the KR-20 reliability coefficient was calculated as 0.69.

It was determined that the difficulty indexes of the 25 items included in the final exam varied between 0.35 and 0.89. It was also determined that the students correctly answered an average of 15.81 questions. The discrimination indexes of the items were found to vary between 0.19 and 0.59. It was determined that the average discrimination index of the items was 0.32. In accordance with the students' answers, the KR-20 reliability coefficient was calculated as 0.73.

2.4 Data Analysis

Discrimination analysis and cluster analysis were used in analyzing the data in accordance with the research problems. It was determined that the sample size was large enough for discrimination analysis and cluster analysis, which are multivariate statistical techniques (n=141). It was determined that there were no missing data in the data set. It was determined that the z scores of the midterm exam scores varied between -2.81 and 1.63; and the final exam scores varied between 2.23 and 2.33. Skewness and kurtosis coefficients were calculated to investigate normality assumption. The coefficients were calculated as -0.624, -0.217 for the midterm exam scores; and as 0.010, -0.602 for the final exam scores. The correlation coefficient between the midterm and final exams were calculated as 0.626. In other words, there are no multiple connection problems between the results. It was determined that Box's M test calculated for homoscedasticity assumption could not be provided and the variances of the variables were not homogeneous. Tabachnick and Fidell (2007) reported that the differences of the sample sizes between the groups could disrupt the homoscedasticity assumption.

Following the fulfillment of assumptions by the data set, the passing grades were calculated by taking 40% of the midterm grades and 60% of the final grades in accordance with the calculation weights of the university. As a result of calculations, the letter grades were calculated using absolute and relative criteria. The values that were based on in calculating letter grades were taken from the university's data IT system and shown in Table 2.

Table 2. Absolute and relative evaluation letter grade intervals

Letter Grades	Absolute criteria	Relative criteria
AA	90-100	85-100
BA	85-89	78-84
BB	80-84	70-77
CB	75-79	63-69
CC	70-74	56-62
DC	65-69	49-55
DD	60-64	42-48
FD	**50-59**	**34-41**
FF	**49 and below**	**0-33**

As shown in Table 2, the letter grades the students will score show differences depending on absolute and relative criteria. It is seen that students scoring below 60 according to the absolute evaluation; and below 42 according to the relative evaluation are supposed to fail the course. The students who participated in the study were classified initially according to their letter grades and then according to their "pass-fail" situation. Following the classification, discrimination analysis was performed for each evaluation and accuracy levels of the classifications were determined. Then, it was intended to determine cut-off points required for 100% correct classifications by performing cluster analysis.

The results were compared and reported.

3. Results

3.1 How Well Do Midterm and Final Grades of Formation Students Discriminate Students with Different Letter Grades and What Is the Percentage of Accurate Classification of the Students?

For the students receiving formation training in Gazi University, 2015-2016 Fall Semester, the midterm and final exam grades and consequently passing grades from them were obtained. Then, letter grades were calculated by absolute evaluation and relative evaluation. The descriptive statistics calculated for the results are shown in Table 3.

Table 3. Descriptive statistics calculated for letter grades determined through absolute and relative evaluations

Letter Grades	Score	Absolute			Relative		
		N	$\bar{X}$	SS	N	$\bar{X}$	SS
AA	Midterm	7	97,14	3,8	15	93,07	6,32
	Final	7	89,14	3,8	15	86,13	4,98
	Passing grade	7	92,34	1,72	15	88,91	3,61
BA	Midterm	8	89,50	6,02	21	86,67	6,61
	Final	8	83,50	4,5	21	76,19	4,81
	Passing grade	8	85,90	1,2	21	80,38	1,99
BB	Midterm	13	88,62	5,12	34	80,94	8,12
	Final	13	76,92	4,66	34	67,65	5,06
	Passing grade	13	81,60	1,46	34	72,96	2,35
CB	Midterm	19	83,37	6,29	22	76,73	8,7
	Final	19	72,42	4,79	22	60,00	6,05
	Passing grade	19	76,80	1,53	22	66,69	1,92
CC	Midterm	23	79,83	9,06	23	66,96	8,46
	Final	23	66,26	5,09	23	54,43	5,88
	Passing grade	23	71,69	1,67	23	59,44	1,97
DC	Midterm	19	76,63	8,77	16	57,50	8,99
	Final	19	60,84	5,75	16	47,25	6,4
	Passing grade	19	67,16	1,61	16	51,35	2,05
DD	Midterm	14	70,00	8,84	5	50,40	10,43
	Final	14	56,29	5,08	5	42,40	4,56
	Passing grade	14	61,77	1,41	5	45,60	2,04
FD	Midterm	26	61,54	9,61	5	40,80	5,93
	Final	26	49,85	6,32	5	37,60	3,58
	Passing grade	26	54,52	3,43	5	38,88	1,34
FF	Midterm	12	47,33	10,76	0	---	---
	Final	12	40,67	5,61	0	---	---
	Passing grade	12	43,33	4,35	0	---	---

Considering the information contained in Table 3, it is observed that the number of students with high grades of AA, BA and BB is more in the relative evaluation than the absolute evaluation. Similarly, it was determined that the number of students with low grades of DD, FD and FF is more than the number of students classified with the absolute evaluation.

The values calculated to determine to what extend the midterm and final exams taken by the formation students discriminate the letter grades of the students is shown in Table 4.

Table 4. Discrimination statistics calculated for the students' midterm and final exams

Assessment	Function	Eigenvalue	Cumulative %	Canonical Correlation	Wilks' Lambda	sig.
Criterion referenced	1	32,715	99,8	0,985	0,028	0,000
	2	0,052	100,0	0,222	0,951	0,447
Norm referenced	1	34,042	99,8	0,986	0,027	0,000
	2	0,070	100,0	0,256	0,934	0,164

Two discrimination functions established for discrimination analysis are shown in Table 4. The canonical correlation coefficient of the first function in the absolute evaluation is 0.985; the canonical correlation coefficient of the second function was calculated as 0.222. The first function formed in the absolute evaluation was determined to be highly effective in discriminating the groups. The same is also true for discrimination functions established for the relative evaluation. When the calculated Wilks' Lambda values were examined, it was determined that the discrimination power of the first function established for both evaluation methods was higher, while that of the second function was not significant in discrimination ($p > 0.05$). When the structure matrix coefficient was examined, it was determined that the contribution of final grades (0.439) to the determination of letter grades was higher than that of midterm grades (0.279) in the absolute evaluation. In the relative evaluation, on the other hand, it was determined that the final grades (0.418) had higher structure matrix coefficient than the midterm grades (0.285).

The observed and predicted percentages for the letter grades of the students receiving formation education, calculated by absolute and relative evaluations, are shown in Table 5.

Table 5. As a result of Discrimination Analysis, Classification Decisions (%) - Letter Grades

Original	Letter Grades	Predicted Group Memebership (%)								
		AA	BA	BB	CB	CC	DC	DD	FD	FF
Absolute	AA	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
	BA	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
	BB	0,0	15,4	84,6	0,0	0,0	0,0	0,0	0,0	0,0
	CB	0,0	0,0	10,5	89,5	0,0	0,0	0,0	0,0	0,0
	CC	0,0	0,0	0,0	0,0	91,3	8,7	0,0	0,0	0,0
	DC	0,0	0,0	0,0	0,0	5,3	94,7	0,0	0,0	0,0
	DD	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0
	FD	0,0	0,0	0,0	0,0	0,0	0,0	11,5	88,5	0,0
	FF	0,0	0,0	0,0	0,0	0,0	0,0	0,0	8,3	91,7
Relative	AA	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
	BA	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
	BB	0,0	5,9	88,2	5,9	0,0	0,0	0,0	0,0	0,0
	CB	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0
	CC	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0
	DC	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0
	DD	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0
	FD	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
	FF	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0

When the data in Table 5, are examined, it is seen that all of the students with AA, BA and DD letter grades with absolute evaluation were classified with 100% accuracy. However, 15.4% (n=2) of the students with BB letter grade was predicted to have BA letter grade. In the absolute evaluation, it was determined that the original and predicted group members were classified with 92.2% accuracy.

Examining Table 5, it is determined that all of the students with AA, BA, CB, CC, DC, DD, FD and FF letter grades with relative evaluation method were classified with 100% accuracy. However, it was determined that 5.9% (n=2) of the students with BB letter grade were supposed to have BA; and 5.9% (n=2) of the students with BA letter grade were supposed to have CB letter grade. It was determined that, the original and predicted group members were classified with 97.2% accuracy in the relative evaluation.

3.2 How Well Do Midterm and Final Grades of Formation Students Discriminate Students Passing or Failing the Course And What Is the Percentage of Accurate Classification of the Students?

Based on the midterm and final exam grades of the students receiving formation education in Gazi University, during 2015-2016, Fall Semester, it was decided whether the students passed or failed the courses. These decisions were made based on the absolute criterion. Within the scope of the study, also pass-fail classification of students was made through relative evaluation.   The descriptive statistics are shown in Table 6.

Table 6. Descriptive statistics calculated for the pass-fail decisions determined by relative and absolute evaluations

Assessment	Group	Score	N	$\overline{X}$	SS
Absolute	students passed the course	Midterm	103	57,05	11,90
		Final	103	46,95	7,42
		Passing grade	103	50,99	6,43
	students failed the course	Midterm	38	81,59	10,33
		Final	38	69,28	10,69
		Passing grade	38	74,21	8,67
Relative	students passed the course	Midterm	136	40,80	5,93
		Final	136	37,60	3,58
		Passing grade	136	38,88	1,34
	students failed the course	Midterm	5	76,24	14,05
		Final	5	64,21	13,35
		Passing grade	5	69,02	12,12
All student		Midterm	141	74,98	15,32
		Final	141	63,26	14,03
		Passing grade	141	67,95	13,15

When the information contained in Table 6 are examined, it is seen that the midterm grades of all students, students who passed or failed the course, are higher than their final grades and passing grades. When Table 5 is examined, it was determined that 103 students passed the course based on the absolute evaluation; and 136 students passed the course based on the relative evaluation.

Based on the absolute and relative evaluations, pass-fail decisions for the students were classified. The values calculated for this classification derived from the midterm and final exams are shown in Table 7.

Table 7. Discrimination statistics calculated for the students' midterm and final grades

Assessment	Function	Eigenvalue	Cumulative %	Canonical Correlation	Wilks' Lambda	sig.
Criterion referenced	1	1,646	100,0	0,789	0,378	0,000
Norm referenced	1	0,243	100,0	0,442	0,804	0,000

When Table 7 is examined, as the students were divided into two classes as passed or failed, one function was established in the discrimination analysis. It is seen that the Wilks' Lambda values of the function established as a result of the absolute evaluation were significant and moderate level. In the discrimination function where the formation students were classified as "passed-failed", the canonical correlation was calculated as 0.789. It was determined that the grades in accordance with the discrimination function (wl=0.378; p<0.05) were effective in discrimination of the students.

In the process of decision-making regarding whether students pass or fail a course, the structure matrix coefficient for the grades was calculated as 0.793 for midterm grades and 0.784 for final grades to see the effects of midterm and final grades. It was determined that the function established as a result of the relative evaluation had less effect in discrimination of the students. In other words, the grades were found to be moderately effective in discrimination of the students (wl=0.807; p<0.05). In the decisions made as a result of the relative evaluation, the structure matrix coefficient was calculated as 0.964 for midterm grades and as 0.762 for final grades. Although weight of the final grades in passing grades was 60%, the midterm grades were found to be more effective in both absolute and relative evaluations in students' passing or failing a course.

The observed and predicted values and percentages for pass-fail conditions of the students receiving formation education, calculated by absolute and relative evaluations, are shown in Table 8.

Table 8. As a result of Discrimination Analysis, Classification Decisions (%) – Pass-Fail Decisions

Original	Result	Predicted Group Memebership (%)	
		Kaldı	Geçti
Absolute	**Fail**	100,0	0,0
	Pass	6,8	93,2
Relative	**Fail**	100,0	0,0
	Pass	10,3	89,7

When Table 8 is examined, it is seen within the scope of "60 passing grade" (based on the absolute criterion) that the students' midterm and final exam grades accurately classified 93.2% of the students who passed the course, while they accurately classified all of the students (n=38, 100%) who failed the course. It was determined that 6.8% (n=7) of the students were misclassified. Based on the absolute evaluation, the students were found to be classified with 95.0% accuracy.

In Table 8, it was determined within the scope of "passing grade based on the mean and standard deviation values of the class" (based on the relative criterion) that the students' midterm and final exam grades accurately classified all of the students (n=5; 100%), who passed the course, while they accurately classified 89.7% (n=122) of the students who failed the course. Based on the relative evaluation, the students were found to be classified with 90.1% accuracy.

3.3 What Is the Cut-Off Point of the Midterm and Final Grades of Formation Students, Which is Required to Classify Students Passing or Failing a Course?

Cluster analysis was performed with the aim to determine which students were misclassified and to make a 100% accurate decision in classification of passing-failing courses. Using the initial cluster centers of the best kept (detained) class solution as non-random starting points in a repeated k-means set, a k-means cluster analysis was calculated. At the end of the cluster analysis, F values were determined to be significant for midterm grades ($F_{(1.139)}$=184.746; p<0.05) and final grades ($F_{(1.139)}$=158.599; p<0.05). It was determined that the midterm grades of the students explained 60.98% of the variance of the students who failed the course, and 84.48% of the variance of the students who passed the course. The final grades, on the other hand, were found to explain 50.88% of the variance of the students who failed the course, and 71.67% of the variance of the students who passed the course. The proportion of the variance that the grades explain was found to be over 50% of all classifications.

The students were divided into two groups following the cluster analysis, and 55 students took part in the first group and 85 students took part in the second group. The descriptive statistics calculated for the grades of the students included in the groups are shown in Table 9.

Table 9. Descriptive statistics calculated for the midterm, final and passing grade of the students included in the groups that were determined as a result of cluster analysis.

Group	Score	N	Minimum	Maksimum	$\overline{X}$	SS
students passed the course	Midterm	56	32,00	88,00	61,36	12,40
	Final	56	32,00	64,00	50,21	8,18
	Passing grade	56	36,80	64,80	54,67	7,61
students failed the course	Midterm	85	60,00	100,00	83,95	9,23
	Final	85	48,00	96,00	71,86	9,79
	Passing grade	85	65,60	94,40	76,70	7,44

When the data contained in Table 9 are examined, it is seen that the passing grades of the students who failed the course varied between 36.80 and 64.80. The passing grades of the students who passed the course, on the other hand, were found to vary between 65.60 and 94.40. In other words, the cut-off point, which was required for accurate classification of the students, was determined as "65".

A discrimination analysis was performed on the students who were already classified as passed-failed by a cluster analysis and it was determined that all of the students who either passed (n=56; 100%) or failed (n=85; 100%) had been accurately classified.

4. Discussion

The study was conducted with the participation of a total of 141 students who received formation education in two different departments in Gazi University in the academic year 2015-2016. In the study, the success grades the students receiving formation education got from the measurement and evaluation course were determined based on absolute and

relative criteria. It was determined that receiving nine different letter grades based on the absolute criteria, the students were classified with 92.2% accuracy. The letter grades of the students were calculated in accordance with relative criteria and the students were found to be classified with 97.2% accuracy. It was determined that more students scored high success grades (AA, BA) with the relative criterion than the absolute criterion. Saral (2012) also evaluated same grades with relative and absolute criteria in Selçuk University. At the end of the study, it was reported that the number of students with high success grades (A1, A2) using relative evaluation was higher than the number of students evaluated with absolute evaluation.   Similarly, it was determined that the number of students with low grades (F, D) who were classified with absolute criterion was more than the number of students classified with relative criterion (Saral, 2012). Lok, McNaught & Young (2015) focused to the tension between criterion-referenced and norm-referenced assessment is examined in the context of curriculum planning and assessment in outcomes-based approaches to higher education. Cox & Vargas (1966) to determine to what extent two methods of item analysis - norm referenced and criterion referenced yield the same relative evaluation of test items. At the end of the study, it was found that the method of item analysis attempted in this study (pretest and post-test method) seems to produce results sufficiently different from traditional methods to warrant its consideration when criterion referenced tests are desired. Traditional item analysis procedures were deemed appropriate in the selection of norm referenced measures (Cox & Vargas, 1966).

In accordance with the regulations of the university, it is decided that the students with FD and FF letter grades should fail the course while the students with other grades should pass the course. In this regard, the students were classified as students who passed the course and students who failed the course, both by absolute and relative evaluations. It was determined that 95% of the classifications made by absolute evaluation were accurate, while 90.1% of the classifications made by relative evaluation were accurate. In the study conducted by Saral (2012) in Selçuk University, while the number of students who failed the course, based on the absolute criterion, was 16, the number of students who failed the course, based on the relative criterion, was 6. Kaya (2013) reported in his study that the grades of students evaluated by relative evaluation were higher than that of students evaluated by absolute evaluation.

In order to determine a cut-off point required for pass-fail classification of students, cluster analysis was performed. At the end of the cluster analysis, it was determined that the passing grades of the students varied between 36.80 and 64.80. The passing grades of the students who passed the course, on the other hand, were found to vary between 65.60 and 94.40. In other words, the cut-off point, which was required for accurate classification of the students, was determined as "65". Başokçu and Kelecioğlu (2014) noticed that: there are no students assumed as unsuccessful in spite of having necessary properties in criterion-referenced assessment although one out of every five students is assumed as a successful despite of not having necessary properties. And also It is determined that in assessment made by absolute criteria the inconsistency ratio especially for students about who fall decision is given is higher (Başokçu and Kelecioğlu, 2014). Its another study results that normative assessment results is more representative for students' achievement level in comparison with criterion-referenced assessment (Nartgün, 2007).

In the light of all this information, it was determined that a more accurate classification is obtained by absolute evaluation than relative evaluation in determining whether or not students pass or fail their courses. While students' passing grades are lower in relative evaluation, it was determined as a result of the cluster analysis that the passing grade should be 65. It is suggested that absolute evaluation should be used in determining whether or not students pass or fail their courses.

References

Atılgan, H., Yurdakul, H., & Oğretmen, T. (2012). A research on the relative and absolute evaluation for determination of students achievement. *Inonu University Journal of the Faculty of Education, 13*(2), 79-98.

Başokçu, T. O., & Kelecioglu, H. (2014). Comparison of Validity of Classification Made According to DINA Model with Criterion-Referenced and Normative Assessment. *e-International Journal of Educational Research, 5*(1), 1-19. http://dx.doi.org/10.19160/e-ijer.82227

Bond, L. (1995). *Norm- and criterion-referenced testing: The differences in purpose, content, and interpretation of results.* North Central Regional Educational Laboratoy, Oak Brook, IL.

Cox, & Vargas (1966). *A comparison of item selection techniques for norm-referenced and criterion-referenced tests.* Learning Research and Devolepment Center, University of Pittsburgh.

Crocker, L. M., & Algina, L. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.

Fraenkel, J. R., & Wallen, N. E. (2009). *How To Design And Evaluate Research In Education.* (7). New York: McGraw-Hill İnternational Edition.

Glaser, R. (1963). Instructional Technology and the Measurement of Learning Outcomes: Some Questions. *American Psychologist, 18*(8), 519–521. http://dx.doi.org/10.1037/h0049294

Glaser, R., & Klaus. D. J. (1962). Proficiency Measurement: Assessing Human Performance. *In Psychological Principles in System Development*, edited by R. M. Gagne, 419–474. New York, NY: Holt, Rinehart & Winston.

Huitt, W. (1996). Measurement and Evaluation: Criterion- versus Norm-referenced Testing. Educational Psychology Interactive, Valdosta, GA: Valdosta State University.

Kaya, U. (2013). *Students and instructors opinions about relative and absolute evaluation in higher education.* Fırat University, Master Thesis.

Lok, B., McNaught, C., & Young, K. (2015). Criterion-referenced and norm-referenced assessments: compatibility and complementarity, *Assessment & Evaluation in Higher Education, 41*(3), 450-465. http://dx.doi.org/10.1080/02602938.2015.1022136

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Welsley Publishing Company.

Mandernach, B. J. (2003). Effective Grading Strategies. *Park University Faculty Development Quick Tips*.

McClarty, K. L., Way, W. D., Porter, A. C., Beimers, J. N., & Miles, J. A. (2013). Evidence-Based Standard Setting: Establishing a Validity Framework for Cut Scores, *Educational Researcher*, *42*(2), 78–88. http://dx.doi.org/10.3102/0013189X12470855

Murphy, K. R., & Davidshofer, C. O. (1991). Psychological Testing: Principles and Applications. Prentice Hall.

Nartgün, Z. (2007). An Investigation of Criterion Referenced and Norm Referenced Evaluation Practic es Based on Same Scores Causing A Difference on Grades or No. *Ege Journal of Education, 8*(1), 19-40.

Reid, K. J., & Dodds, A. (2013). Comparing the borderline group and borderline regression approaches to setting Objective Structured Clinical Examination cut scores. *Journal of Contemporary Medical Education, 2*(1), 8-12. http://dx.doi.org/10.5455/jcme.20130921013523

Saral, D. G. (2012). *A Study on Comparison of Classification Validity of Decisions Taken About the Same Students Evaluated According to Norm and Criterion*, Hacettepe University, Master Thesis.

Stiggins, R. J. (1994). *Student-Centered Classroom Assessment*. New York: Merrill

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Allyn and Bacon.

Tekin, H. (2004). *Eğitimde Ölçme ve Değerlendirme*. Ankara. Yargı Yayınevi.

Turgut, F. (1992). *Eğitimde Ölçme ve Değerlendirme Metotları*. Ankara.

Vanderstoep, S. W., & Johnston, D. D. (2009). Research Methods for Everyday Life. USA: Jossey-Bass.