HOW VALID ARE OPEN UNIVERSITY EXAM PASSING STANDARDS? CONSISTENCY OF CLASSIFICATION BY CUT-OFF POINTS

Dr. Hakan BARAN
ORCID: 0000-0002-1495-8477
Ministry of National Education
Eskisehir, TURKIYE

Dr. Murat AKYILDIZ
ORCID: 0000-0001-5069-0132
Open Education Faculty
Anadolu University
Eskisehir, TURKIYE

Received: 03/07/2024 **Accepted:** 16/11/2024

ABSTRACT

Evaluation decisions regarding students' success in Open Education faculties such as pass/fail based on cut-off scores affect the quality of these systems. The qualification of Open Education students to obtain a bachelor's or associate's degree is determined by their passing grade. The purpose of this study was to investigate whether the minimum passing scores used in Open Education examinations differ from the currently used minimum passing scores according to different standard-setting methods and the classification consistency of the cut-off scores obtained by these methods with the currently used cut-off scores and with each other. The participants consisted of 15 experts, consisting of textbook authors and lecturers of the Basic Disaster Knowledge Course this course. The results showed that in the final and midterm examinations, the percentage of successful students according to the cutoff points identified using the Angoff and Nedelsky methods was significantly lower than the percentage of successful students according to the current cut-off scores of the Open Education Examinations. The standards to be determined based on academic principles were different from the ones identified by administrative decisions. Further, the pass/fail decisions based on academic principles differed from those based on administrative decisions.

Keywords: Assessment and evaluation in distance education, assessment and evaluation in open education, standard-setting, cut-off score, the Angoff method, the Nedelsky method.

INTRODUCTION

Given the rapid expansion of open and distance education systems, ensuring the accuracy and fairness of assessment processes has become a critical issue. While existing research on assessment standards has focused largely on traditional educational settings, there is a gap in the literature regarding the implementation of standardized methods in open education systems. This study addresses this gap by examining the validity of passing standards based on academic principles rather than administrative decisions.

The programmes in Open Education Faculties aim to provide students with certain knowledge, skills, and qualifications. The fact that the examinations for the evaluation of students are prepared to measure the skills aimed at being gained in the programmes will ensure that the programmes provide qualified outputs. To respond to these outputs, the assessment processes and properties of the programmes that train human resources for institutions and business life should have some qualitative characteristics.

Assessment is the process of comparing measurement results with a criterion or set of criteria and reaching a judgement (Alkin and King, 2017). Decisions such as pass-fail, which is a result of the evaluation process, depend on the validity of the measurement results reflecting the real value of the measured characteristic, as well as the appropriateness of the criterion for the purpose of the evaluation and the accuracy of the

applications in the comparison stage with the criterion (Bloom, 1968). The realisation of the measurement process with minimum error is important in terms of measurement results that reflect the real value of the measured knowledge, skill and acquisition. The appropriateness of criterion is directly related to the accuracy of the decision to be made. The criterion is the cut-off score or performance level used to decide as a result of the measurement. It plays an important role in standardising the decisions to be made. If the criterion is not determined by the purpose of the assessment, undesirable results may be obtained. If the criterion set for an open education exam is lower than it should be, the candidates accepted to the programme will have difficulty showing the required performance and will fail the programme. This may lead to suspicions that education programmes are inadequate.

Of the two types of criteria used in education absolute criteria and relative criteria are preferred depending on the purpose of the assessment. If the aim is to test the presence of critical knowledge, skills or behaviours that should be at a minimum level of individuals to be assessed, it is appropriate to use absolute criteria. In assessments using absolute criteria, the minimum level to be considered sufficient should be determined based on the difficulty level of the test.

The Educational Testing Service (2008) explains standard setting as follows: Standard setting is the methodology used to define levels of achievement or proficiency and the corresponding cut-off scores. Cizek (1993) defined standard-setting as a legitimate and appropriate rule or procedure that assigns numbers to distinguish differences in performance. A cutoff score is simply a score used to categorise students below the cut-off score into one level and students above the cut-off score into the next and higher levels. Cutoff scores divide the distribution of test takers' test performance into two or more categories. For instance, in the context of licensure and certification testing programmes, it is often the case that only a single cut-off score is required, the application of which results in the creation of two categories of performance: pass/fail and award/reject (Cizek & Bunch, 2007). This process is not an administrative decision, but a clearly defined, systematic, academic, and scientific process. The standard is the conceptual aspect of the desired level of competence, while the cutoff score (passing score) is the operational aspect (Kane, 1994). The word standard refers to the minimum level of knowledge and skills for the relevant performance categories. Therefore, the standard is the answer to the question 'How adequate?' If the standards are not set appropriately, the assessment results may be skeptical. Therefore, standard setting is a fundamental element of the test development process (Educational Testing Service, 2008). Various standard-setting methods, including criterion-referenced and test-centred methods, have been proposed for written tests consisting of multiple-choice questions (Cizek & Bunch, 2007).

In test-centred methods, experts make judgements about the test items in relation to the measured features. When studies on standard-setting methods in the related literature are examined, it is seen that test-centred methods are used much more often (Park et al., 2018; Shulruf et al., 2016; Yim & Shin, 2020). The Angoff and Nedelsky methods, which are test-centred methods, are frequently preferred in these studies (Chang, 1996). This is because of the relative simplicity of the Angoff and Nedelsky methods compared to the complexity of other test-centred methods (Cizek, 2001). The Angoff method is the most widely used test-centred method. Experts analyse each test item and estimate the probability that a competent person with a minimum level of proficiency will answer the item correctly in this method (Cizek & Bunch, 2007). The Angoff method, which is widely applied in licensing or achievement tests, is easy to understand because it is much simpler than other methods, and is considered to strike the best balance between technical suitability and practicality (Angoff, 1984; Berk, 1986). However, the Nedelsky method is used only in multiple-choice tests. Experts are asked to analyse each distractor in a test question throughout the application of this method. According to the basic assumption of the Nedelsky method, a student at the minimum proficiency level can randomly select the correct option from the remaining options by eliminating the options that they know are incorrect while answering the test question (Livinston & Zieky, 1982).

Open and distance education services are developing and spreading rapidly. When standard setting methods are not used in the process, the cut-off scores are determined by the test administrators or according to the administrative decisions of institutions in open education systems. Inthe research setting of this study, 30% of the midterm exam grades and 70% of the final exam grades were used to calculate the passing grade for each course. Accordingly, the final passing grade was determined to be 30 out of 100 by administrative decision. The Standards for Educational and Psychological Testing (the American Educational Research Association,

the American Psychological Association, and the American Council on Measurement in Education, 1999) recommend the following robustness criteria:

> *"When proposed score interpretations include one or more cut-off scores, the rationale and procedures used to determine the cut-off scores should be clearly documented. Sufficient precision in the regions of the score scales on which cut-off scores are determined is a prerequisite for reliable categorisation of test takers into categories" (p. 59).*

It is important to examine how the standards set based on academic principles with certain justifications and procedures differ from the standards set through administrative decisions. This research will present the differences stemming from the individuals who actually need to make a pass-fail decision and the ones based on administrative decisions. This research presents a contribution to the literature since this difference has not been examined before in open and distance education systems.

Literature on open and distance education assessment has predominantly focused on areas such as learner engagement, course design, and technological challenges (Park et al., 2018; Yim & Shin, 2020). However, there is a distinct lack of research examining the determination of cut-off scores for assessments within open education contexts. While studies on assessment in open education systems are available, these typically focus on general evaluation processes or the effectiveness of different assessment types (Gikandi et al., 2011; Johnson & Aragon, 2003). Very few studies have explored the methodological rigor involved in setting cut-off scores, especially in relation to academic standards. For example, Gikandi and Morrow (2015) discuss automated assessment systems, but they do not address the process of establishing cut-off points for pass/fail decisions. Similarly, Shulruf et al. (2016) focus on technological challenges in open and distance education without delving into the specifics of assessment standards. This gap becomes even more evident when considering that most open education systems rely on administrative decisions to determine pass/fail cut-off points, rather than using systematic, research-backed methods. This research, by contrast, aims to fill this gap by examining the use of academically grounded standard-setting methods, such as the Angoff and Nedelsky methods, to ensure fairness and consistency in pass/fail decisions in open education settings.

PURPOSE AND RESEARCH QUESTIONS

Evaluation decisions such as pass/fail based on the cut-off scores determined by standard-setting methods affect the number of students in different open education systems and the quality of open education systems in the standard setting processes. The eligibility of open-education students to receive undergraduate and associate degree diplomas was determined according to the passing grade. In this context, the main purpose of this study was to examine whether the passing score used in open education exams according to different standard-setting methods differs from the passing score currently used and whether the cut-off scores determined by these methods are consistent with the currently used cut-off scores In line with the purpose, answers to the following questions were sought.

1. What are the cut-off scores determined by the Angoff and Nedelsky methods in open education exams?
2. What are the internal consistencies between experts using the Angoff and Nedelsky methods?
3. Is there a significant difference between the percentage of students who pass according to the cut-off scores used in the Open Education System and the percentage of students who pass according to the cutoff scores determined by the opinions of experts using the Angoff and Nedelsky standard setting methods?

METHODOLOGY

Study Group

The study group of the present research consisted of 15 experts, including 10 authors of the textbook used in the preparation of the exams of the 2018-2019 fall term of the Basic Disaster Knowledge course in the Open Education Faculty Emergency and Disaster Management Programme, and five experts teaching this course. Jeager (1989) stated that 15 raters would be sufficient for standard-setting processes. Similarly, Wu and Tzou (2015) revealed in their study that the number of experts should be at least ten.

While determining the course that was the subject of the research, the Emergency and Disaster Management Programme, in which the authors of each of the textbooks taught within the scope of the programme are relatively more than the other programmes, was selected. The criterion sampling method was then used to select the course from the programme. The selection criteria were as follows: i) the number of authors being more than the average number of authors of all textbooks of the program (6-8 authors), and ii) the maximum number of experts also being textbook authors. Accordingly, the Emergency and Disaster Management Programme textbooks and number of authors were identified. Six courses with more than (approximately) seven authors were identified. Among these six courses, The Basic Disaster Knowledge course, whose authors were all field experts, was selected as the course of focus in the present study. The number of students enrolled in the Faculty of Open Education Emergency and Disaster Management Programme in the 2018-2019 fall term was 8545, and 5485 students took the Basic Disaster Knowledge course and the exams of this course.

Data Collection Process

The process of consulting expert opinions started with a session in which all of the authors of the Basic Disaster Knowledge textbook (ten authors) participated. Five non-author experts were interviewed in the second session. In these sessions, the researcher first explained the Angoff and Nedelsky test-centred standard-setting methods to the participants. The experts were informed about the concepts needed in the standard-setting process, such as the cut-off score and the learner's minimum proficiency level. The experts were then asked to evaluate each question of the exams according to the framework of Angoff and Nedelsky test-centred standard-setting methods. In this context, an Expert Opinion Form prepared by the researchers was used to document expert opinions. The form consisted of midterm and final exam questions (a total of 40 exam questions, 20 midterm questions and 20 final exam questions) and the Angoff and Nedelsky standard setting theory explanations and examples. A sample of this form is presented in Figure 1.

Evaluation 1 **Angoff**	**Question**	**Evaluation 2** **Nedelsky**
Of the 100 students on the pass-fail border The student can answer correctly.	Question 1 What is the percentage expression of the ratio of the amount of water vapour in a mass of air or gas at a given temperature to the highest amount of water vapour that can be found in a mass of air or gas at that temperature? A) Air humidity B) Relative humidity C) CS Constant humidity D) Variable humidity E) Soluble moisture	A student on the pass-fail border A | B | C | D | E can eliminate the choices knowing that they are wrong.

Figure 1. A Sample Section of the Expert Opinion Form

Data Analysis

In the first evaluation using the Angoff method, the experts were asked to predict how many of the 100 students at the minimum proficiency level could correctly answer the first question of the mid-term and final exams, consisting of 20 questions each. The experts were then asked to make predictions about the remaining questions. The percentage for each expert evaluation were calculated and the arithmetic means of the percentage values were determined. Thus, the minimum passing scores (MGP) for the midterm and final examinations were determined. Subsequently, the arithmetic mean of the minimum passing scores was identified for each expert. Thus, midterm and final-term cut-off scores were determined for the Angoff method. According to the course passing regulation of the 2018-2019 academic year, 70% of the final cutoff score and 30% of the midterm cutoff score were summed, and the final cutoff score was calculated.

In the second evaluation using the Nedelsky method, the experts were asked to predict the number of distractors of the first question of the midterm and final exams of a learner at the minimum proficiency level could be eliminated by knowing that they were wrong. As a result of the evaluation, if four distractors

were eliminated, the probability of the learner answering the question correctly was 100%. Similarly, if the number of eliminated distractors was three, two, one, and zero the probability of the learner answering the question correctly was determined to be 50%, 33%, 25%, and 20%, respectively. Subsequently, the sum of the percentages was divided by the number of questions (20). Thus, the minimum passing scores for the midterm and final exams were determined. On the other hand, the arithmetic means of the minimum passing scores identified by each expert was calculated. Thus, the cut-off scores of the midterm and final exams were determined according to the Nedelsky method. According to the 2018-2019 academic year course passing regulations, 70% of the final exam cut-off score and 30% of the midterm exam cut-off scores were summed, and the final cut-off score was calculated. In addition, the Intraclass Correlation Coefficient statistic was used to calculate the internal consistency between experts.

Finally, we examined any significant difference between the percentages of students whowere considered as successful or unsuccessful according to the cut-off scores determined by the Angoff and Nedelsky methods and the percentages of students who were considered as successful or unsuccessful according to the cut-off scores of the open education examinations. For this purpose, the difference between two dependent percentages was tested and its significance was tested using the Z test. The significance level was set as 0.05. Fraenkel, Wallen, and Hyun (2012) interpreted the intraclass correlation coefficient as follows: "<0.70 incompatible"; "0.70-0.84 good"; "0.84-0.94 high"; "0.94-1 excellent".

FINDINGS

Findings Related to Cut-off Scores Determined by the Angoff and Nedelsky Methods in Open Education Exams

The arithmetic means of the minimum passing scores of the midterm and final examinations were calculated. The cutoff scores of the midterm and final exams according to the Angoff method are presented in Table 1.

Table 1. Minimum Passing Scores and Cutoff Scores for Midterm and Final Examination Identified Using the Angoff Method

Expertise	Final Examination Angoff	Midterm Exam Angoff
Expert 1	63.5	57
Expert 2	50.75	46.8
Expert 3	78.75	70
Expert 4	55	51.75
Expert 5	43.75	43.5
Expert 6	38	35
Expert 7	59.75	56.5
Expert 8	58,25	52.25
Expert 9	69	49.5
Expert 10	72.25	65.25
Expert 11	47.5	42.25
Expert 12	72.25	67
Expert 13	43.25	38.25
Expert 14	77.5	71.25
Expert 15	66.5	59.15
Cut Points	**59.733**	**53.697**

A midterm and final exam were administered within the sccope of the Basic Disaster Information course. 20 questions were asked during the exams. Each question had a score of 5 points. The examination reauslts varied between 0 and 100 points. The course passing grade was 30. The results presented in Table 1 showed that the difference in the experts' evaluations of the questions yielded the minimum passing scores to to differ from each other. Using the Angoff method, the cutoff scores for the final and midterm exam were set as 59.733 and 53.697, respectively.

The descriptive statistics of the final exam and midterm cut-off scores identified using the Angoff method are given in Table 2.

Table 2. Descriptive Statistics of Midterm and Final Examination Cutoff Scores Identified Using the Angoff Method

Descriptive Statistics	Final Examination Angoff	Midterm exam Angoff
N	15	15
Average	59.73	53.70
Hydrangea	59.75	52.25
Standard Deviation	13.07	11.42
Variance	170.78	130.51
Minimum	38.00	35.00
Maximum	78.75	71.25
Openness	40.75	36.25
Skewness Coefficient	-0.145	0.037
Kurtosis Coefficient	-1.204	-1.024

According to Table 2, the mean and median values of the cutoff scores of the midterm and final exams identified using the Angoff method were similar to each other. The skewness and kurtosis values did not differ significantly from zero. A range of -1.5 and +1.5 indicated that the data were normally distributed (Tabachnick & Fidell, 2013).

As a result of the evaluations of 15 experts according to the guidelines determined for the Nedelsky method, the MGPs and arithmetic means of these scores were calculated. The cutoff scores of the midterm and final exam scores according to the Nedelsky method are presented in Table 3.

Table 3. Final and Midterm Examination Cut-off Scores Identified Using the Nedelsky Method

Expertise	Final Examination Nedelsky	Midterm Nedelsky
Expert 1	37.8	41.6
Expert 2	40.8	42.95
Expert 3	61.2	39.05
Expert 4	42.7	38.9
Expert 5	35.3	37.7
Expert 6	41	36.2
Expert 7	41.15	35.35
Expert 8	40.75	33.3
Expert 9	48.2	45.3

Expert 10	50.85	36.5
Expert 11	41.55	35.25
Expert 12	39.5	36.4
Expert 13	44.8	39.75
Expert 14	47.35	44.8
Expert 15	48.5	43.75
Cut Points	44.096	39.120

Table 3 shows that the difference in the experts' evaluations of the questions yielded the minimum passing scores to differ from each other. Using the Nedelsky method, the cutoff scores for the midterm and final exams were 44.096 and 39.120, respectively. The cut-off score of the final exam was higher than that of the midterm exam.

Descriptive statistics of the cutoff scores of the midterm and final exams determined using the Nedelsky method are presented in Table 4.

Table 4. Descriptive Statistics of Midterm and Final Examination Cutoff Scores Identified Using the Nedelsky Method

Descriptive Statistics	**Final Examination Nedelsky**	**Midterm Nedelsky**
N	15	15
Average	44.10	37.02
Hydrangea	41.55	38.90
Standard Deviation	6.39	3.79
Variance	40.88	14.36
Minimum	35.30	33.30
Maximum	61.20	45.30
Openness	25.90	12.00
Skewness Coefficient	1.364	0.327
Kurtosis Coefficient	2.600	-1.123

As shown in Table 4, the mean and median values of the cut-off scores of the midterm and final exams identified using the Nedelsky method were close to each other. The skewness and kurtosis values did not differ significantly from zero. A range of -1.5 and +1.5 indicated that the data were normally distributed (Tabachnick & Fidell, 2013). However, it should be noted that the kurtosis value was 2.6. According to Tabachnick & Fidell (2013), kurtosis values between 2.0 and 3.0 are considered indicative of a distribution with moderately increased leptokurtic characteristics, meaning the distribution has more pronounced extreme values compared to a normal distribution. This level of kurtosis is generally still acceptable for parametric statistical tests.

According to the 2018-2019 academic year course passing regulation, the cutoff scores for pass/fail decisions according to the Angoff and Nedelsky methods are presented in Table 5.

Table 5. Cut-off Score For Pass/Fail Decision According to the Angoff and Nedelsky Methods (Course Passing Regulations in 2018-2019 Academic Year)

	The Angoff Method	The Nedelsky Method
Final Exam Cut-off Score	59.733	44.096
Midterm Exam Cut-off Score	53.697	39.120
Cut-off score for pass/fail decision	**57.92**	**42.60**

According to Table 5, when the cutoff scores for the final exam were compared, the score figured out using the Nedelsky method was 44.096, while the score identified using the Angoff method was 59.733. More specifically, the cutoff score determined using the Angoff method was higher. Similarly, a comparison of the cutoff scores for the midterm exam showed that the score revealed using the Nedelsky method was 39.120, while the score revealed using the Angoff method was 53.697. Here, the cut-off score revealed using the Angoff method was higher. Accordingly, this situation was reflected in the cut-off score for the pass/fail decision. The cutoff score for the pass/fail decision figured out using the Nedelsky method was 42.60 points, it was 57.92 for the Angoff method.

Findings on the Internal Consistencies between Experts Using the Angoff and Nedelsky Methods

The Intraclass Correlation Coefficient statistic was used to identify the inter-expert agreement coefficient in the Angoff and Nedelsky methods as part of the reliability in identifying the cut-off scores. The results of the analyses regarding the agreement between expert decisions are presented in Table 6.

Table 6. The Concordance between Expert Decisions in Identifiying the Cut-off Score According to the Angoff and Nedelsky Methods

		In-class Correlation Coefficient	p
The Angoff Method	Final Exam	0.885*	<0.001
	Midterm Exam	0.887*	<0.001
The Nedelsky Method	Final Exam	0.929*	<0.001
	Midterm Exam	0.877*	<0.001

* $p<0.05$

As shown in Table 6, the intraclass correlation coefficients examined in determining the cut-off score according to the Angoff method were 0.885 and 0.887 for the final and midterm exams, respectively. Similarly, the intraclass correlation coefficients examined in determining the cutoff score according to the Nedelsky method were 0.929 and 0.877 for the final exam and midterm exam, respectively. In this respect, there was a high agreement between expert judgements.

Findings Related to the Significance of the Difference between the Percentages of Students Passing the Exams

The percentage of successful students according to the cutoff scores of the Angoff and Nedelsky methods in the final exam and the percentage of successful students according to the open education exam cutoff score are compared pairwise in Table 7.

Table 7. *Z Values Related to the Significant Difference Between the Percentages of Successful Students*

Examination	Compared to Cut-off Score	Cut-off Score	Percentage of Successful Students	Z
Final Exam	Open Education Exams Cut score	30	%53	**26.52***
	Cut-off Score Determined by the Angoff Method	59.733	%6,5	
	Open Education Exams Cut score	30	%53	**14.37***
	Cut-off Score Determined by the Nedelsky Method	44.096	%26	

*p<0.05

In Table 7 shows, the results of the Z value of the difference between the percentages of successful and unsuccessful students according to the cut-off score determined by identified using the Angoff method and the open education exams cut-off score for the final exam are seen. While the percentage of successful students according to the Angoff method was 6.5%, the percentage of successful students according to the open education exams cut-off score was calculated as 53%. Regarding the final exam, the percentage of successful students according to the cut-off score determined revealed using by the Angoff method was significantly lower than the percentage of successful students according to the open education exams cut-off score "Z=(26.52); p<0.05".

Similarly, according toas shown in Table 7, the percentage of successful students according to the Nedelsky method was 26%, and the percentage of successful students according to the cut-off score of open education exams was 53%. Regarding the final exam, the percentage of successful students according to the cut-off score determined figured out by using the Nedelsky method was significantly lower than the percentage of successful students according to the open education exams cut-off score [Z=(14.37); p<0.05]. The percentage of successful students according to the cut-off scores of the methods in the midterm exam and the percentage of successful students according to the open education exam cut-off score were compared pairwise in Table 8.

Table 8. Z Values Related to the Significant Difference Between the Percentages of Successful Students

Examination	Compared Cut-off Scores	Cut-off Score	Percentage of Successful Students	Z
Midterm Exam	Open Education Exams Cut score	30	64%	21.82*
	Cut-off Score Determined by the Angoff Method	53.697	32%	
	Open Education Exams Cut score	30	64%	9.19*
	Cut-off Score Determined by the Nedesky Method	39.120	50%	

*p<0.05

Table 8 showsthe results of the Z value of the difference between the percentages of successful and unsuccessful students according to the cut-off score determined by the Angoff method for the midterm exam and the open education exams cut-off score. The percentage of successful students according to the Angoff method was 32%, and the percentage of successful students according to the open education exams cut-off score was 64%. Regarding the midterm exam, the percentage of successful students according to the cut-off score determined by the Angoff method was significantly lower than the percentage of students considered successful according to the open education exams cut-off score "Z=(21.82); p<0.05".

Similarly, as shown in Table 8, the percentage of successful students according to the Nedelsky method was 50%, and the percentage of successful students according to the cut-off score of open education exams was 64%. Regarding the midterm exam, the percentage of successful students according to the cut-off score determined by the Nedelsky method was significantly lower than the percentage of successful students according to the open education exams cut-off score "Z=(9.19); p<0.05".

DISCUSSIONS AND CONCLUSIONS

The present study focused on examining whether the passing (cutoff) scores that can be used in open education examinations according to different standard-setting methods differ from the passing (cutoff) scores currently used. The classification consistencies of the cut-off scores determined by these methods with the cut-off scores that are currently used and with each other were revealed. The results showed the importance of the phase of determining the passing (cutoff) score used for pass/fail decisions, which concerns mass of students in the open education system and affects the quality of the open education system.

The eligibility of open education students to receive undergraduate and associate degree diplomas was determined according to the passing grade. There are many undergraduate and associate degree programmes and many courses under these programmes in the open education system. The difference in outputs and minimum qualifications of each program is an important variable for determining the passing (cut-off) grade. In addition, the difficulty or ease of the examination is an important variable in answering a sufficient number of questions correctly and getting the passing (cutoff) score. In this context, it is important to employ appropriate standard-setting methods to determine a standard in accordance with scientific qualifications and to make the right decisions regarding students.

As a result of the study, a higher cut-off score was found with the Angoff method than with the Nedelsky method. This result supports previous research, which indicated that the cut-off scores calculated using the Nedelsky method were lower than those calculated using the Angoff method (Chang, 1996; Demir, 2014; Tanriverdi, 2006; Tasdemir, 2013). However, making a generalization based solely on the fact that the cut-off score determined by the Nedelsky method is typically lower may be misleading. This is because such results are context-dependent and influenced by the difficulty level of the test and the specific requirements of each program (Chang, 1996). In this regard, decision makers may prefer the Angoff method to determine the cut-off score to increase the difficulty of passing exams based on their educational policies. The Nedelsky method could potentially make exams 'easier' in the sense that it allows for a more lenient scoring process. Since the method evaluates test items by eliminating obviously incorrect choices, it could lower the threshold for passing, making it easier for students to achieve the required score at the minimum proficiency level.

Another remarkable result of this study was that the agreement between expert judgements was high for both the Angoff and Nedelsky methods. Therefore, the reliability of the passing score determination process using the Angoff and Nedelsky methods in open education exams was also high. This result provides supportive evidence to previous research, which indicated that standard-setting methods can be used as a passing score determination process in different tests such as the Medical Licensing Examination, the Medical Performance Tests (Afrashteh, 2021; Park, 2022). In this regard, the use of cutoff scores determined by standard-setting methods is appropriate if the standard-setting process is carried out appropriately and carefully.

In addition to the aforementioned results, the percentage of successful students according to the cut-off scores determined using the Angoff and Nedelsky methods in both midterm and final exams was significantly lower than the percentage of successful students according to the currently used cutoff score. This result indicates that the passing scores determined based on academic principles differ from those set by administrative decisions. The American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (2014) state that the level of performance required to pass a certification test should be based on the knowledge and skills necessary for acceptable performance in a profession. More specifically, passing scores should not be set to regulate the number or proportion of students who pass the test. In the same vein, Downing et al. (2003) stated that a rigorous and legally defensible standard-setting process should be used to support the validity of performance-based inferences in decision-making based on assessment and evaluation studies conducted within the scope of diploma or certificate-awarding programmes. In this context, the importance of using standard-setting processes to determine the cut-off scores of open education exams was revealed.

A fixed cut-off score is used for each programme in open education systems. When standard-setting methods are used, it is likely to result in different cut-off scores for each program. In many open education programs, a fixed passing score is used. However, applying standard-setting methods could result in different passing scores for each program, tailored to the specific requirements and difficulty levels of the respective courses. These systems have a comprehensive exam preparation process by the coordination of the Test Research Unit. The difficulty level of each of the questions can be determined by using the Angoff and Nedelsky test-centred methods in the exam preparation process Experts can use both methods to determine the probability of correct answers according to the students at the minimum proficiency level. Accordingly, exam questions can be selected from question banks to provide a fixed cutoff score to be used in all programmes. Thus, exam difficulty levels can be identified using scientific methods in the exam preparation process of the open education systems. In addition, an automation system can be established based on the Angoff and Nedelsky test-based methods. Thus, it can be ensured that passing scores are figured out in a short time based on the exam questions. In this way, passing score information can be obtained for all courses of the programs in a short time. The automation system to be established can provide flexible acquisition of passing score information according to the competencies of each program and minimum predicted student levels. In this context, coordinators of open education programs can be authorized to identify thepassing grade for each course.

BIODATA AND CONTACT ADDRESSES OF AUTHORS

Dr. Hakan BARAN has been serving as a middle school mathematics teacher under the Ministry of National Education since 2011. He holds a Ph.D. in Distance Education from Anadolu University and a master's degree in Educational Administration from Eskisehir Osmangazi University. His academic and research interests focus on assessment and evaluation in distance education, open education, online learning, learning design, and educational technology. He is particularly interested in measurement and evaluation processes in open and distance learning environments, with an emphasis on standard setting and cut score determination.

Hakan BARAN
Ministry of National Education
Address: Melahat Unugur Middle School, 26100, Eskisehir, Turkiye
Phone: +90 5374393636
E-mail: hakanbrn@gmail.com

Dr. Murat AKYILDIZ is an associate professor of distance education at Anadolu University, Open Education Faculty. After receiving his bachelor's degree in psychology, he completed his master's and Ph.D. in the field of measurement and evaluation. His primary research interest is the prediction of individuals' cognitive and affective characteristics using multidimensional and multiparameter logistic models. His other research interests include psychometrics, statistical analysis, methodology, and the philosophy of science. He is currently a faculty member at the Open Education Faculty of Anadolu University.

Murat AKYILDIZ
Department of Distance Education, Open Education Faculty
Address: Anadolu University, 26470, Eskisehir, Turkiye
Phone: +902223350580
Email: muratakyildiz@anadolu.edu.tr

REFERENCES

Afrashteh, M. Y. (2021). Comparison of the validity of bookmark and Angoff standard-setting methods in medical performance tests. *BMC Medical Education*, *21*(1), 1. https://doi.org/10.1186/s12909-020-02436-3

Alkin, M. C., & King, J. A. (2017). Definitions of evaluation use and misuse, evaluation influence, and factors affecting use. *American Journal of Evaluation, 38*(3), 434–450. https://doi.org/10.1177/1098214017717015

American Educational Research Association, American Psychological Association, & American Council on Measurement in Education (1999). *Standards for educational and psychological testing.* American Educational Research Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Angoff, W. H. (1984). *Scales, norms, and equivalent scores.* Princeton (NJ): Educational Testing Service.

Behuniak, P., Archambault, F. X., & Gable, R. K. (1982). Angoff and Nedelsky standard-setting procedures: Implications for the validity of proficiency test score interpretation. *Educational and Psychological Measurement, 42*(1), 247-255. https://psycnet.apa.org/doi/10.1177/0013164482421031

Bejar, I. I. (2008). Standard-setting: what is it? Why is it important? *R&D Connections, 7*, 1-6. https://www.ets.org/Media/Research/pdf/RD_Connections7.pdf

Berk, R. A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research, 56*, 137-172. https://doi.org/10.3102/00346543056001137

Bloom, B. S. (1968). *Toward a theory of testing which includes measurement-evaluation-assessment (CSE Report 9)*. Los Angeles: University of California, Los Angeles, Centre for the Study of Evaluation. https://files.eric.ed.gov/fulltext/ED036878.pdf

Chang, L. (1996, April*). A comparison between the Nedelsky and Angoff standard-setting methods* [Paper presentation]. The Annual Meeting of the American Educational Research Association, New York. https://files.eric.ed.gov/fulltext/ED398247.pdf

Cizek, G. J. (2001). Conjectures on the rise and call of standard-setting: An introduction to context and practice. In G. J. Cizek & R. J. Sternberg (Eds.), *Setting performance standards: Concepts, methods and perspectives* (pp. 3-17). Mahwah, N. J.: Erlbaum.

Cizek, G. J., & Bunch, M. B. (2007). *Standard-setting: A guide to establishing and evaluating performance standards on tests.* Thousand Oaks, CA: Sage Publications.

Demir, O. (2014). *Comparison of cut-off scores determined by Angoff, Nedelsky and Ebel standard-setting methods*. [Master thesis, Abant Izzet Baysal University].

Downing, S. M., Lieska, N. G., & Raible, M. D. (2003). Establishing passing standards for classroom achievement tests in medical education: A comparative study of four methods. *Academic Medicine*, *78*(10), 85-87. http://doi.org/10.1097/00001888-200310001-00027

Educational Testing Service. (2008). *Standards for educational and psychological testing*. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education.* (8th ed.). New York: McGram-Hill Companies.

Gikandi, J. W., Morrow, D., & Davis, D. (2011). Online assessment in higher education: A review of the literature. Computers & Education, 57(4), 2333-2351. https://doi.org/10.1016/j.compedu.2011.06.004

Gikandi, J. W., & Morrow, D. (2015). A review of automated assessment systems in open and distance education. British Journal of Educational Technology, 46(2), 295-311. https://doi.org/10.1111/bjet.12183

Jeager, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), *Educational measurement* (pp. 485-514). New York: Macmillan.

Johnson, S. D., & Aragon, S. R. (2003). An instructional strategy framework for distance education. Educational Technology Research and Development, 51(3), 35-51. https://doi.org/10.1007/BF02504856

Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research, 63*(4), 425-461. https://doi.org/10.2307/1170678

Livingston, S. A., & Zieky, M. J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests.* New Jersey: Educational Testing Service.

Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement, 14*(1), 3:19. https://doi.org/10.1177/001316445401400101

Park, J., Ahn, D., Yim, M. K., & Lee, J. (2018). Comparison of standard-setting methods for the Korean Radiological Technologist Licensing Examination: Angoff, Ebel, bookmark, and Hofstee. *Journal of Educational Evaluation for Health Professions, 15*(32). https://doi.org/10.3352/jeehp.2018.15.32

Park, J. (2022). Possibility of using the yes/no Angoff method as a substitute for the percentage Angoff method for estimating the cutoff score of the Korean Medical Licensing Examination: A simulation study. *Journal of Educational Evaluation for Health Professions*, *19*(23). https://doi.org/10.3352/jeehp.2022.19.23

Shulruf, B., Poole, P., Wilkinson, T., Weller, J., & Jones, P. (2016). Insights into the Angoff method: Results from a simulation study. *BMC Medical Education*, *16*(134). https://doi.org/10.1186/s12909-016-0656-7

Shulruf, B., Tumen, S., & Byun, W. (2016). Technological challenges in open and distance education: An analysis of past research and future directions. Open Learning: The Journal of Open, Distance, and e-Learning, 31(3), 1-17. https://doi.org/10.1080/02680513.2016.1160639

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Tanriverdi, S. (2006). *The effect of standard-setting methods on passing scores.* [Master thesis, Hacettepe University].

Tasdemir, F. (2013). *Analysing the classification accuracy of a test with Angoff (1-0), Nedelsky and boundary value detection methods*. [Doctoral dissertation, Ankara University].

Wu, Y. F., & Tzou, H. (2015). A multivariate generalisability theory approach to standard-setting. *Applied Psychological Measurement*, *39*(7), 507-524. http://doi.org/10.1177/0146621615577972

Yim, M. K., & Shin, S. (2020). Using the Angoff method to set a standard on mock exams for the Korean Nursing Licensing Examination. *Journal of Educational Evaluation for Health Professions*, 17(14). https://doi.org/10.3352/jeehp.2020.17.14