

Araştırma Makalesi– Research Paper

**ESTABLISHING A MEASUREMENT AND EVALUATION LABORATORY IN A
NURSING FACULTY: A DESCRIPTIVE STUDY**

**HEMŞİRELİK FAKÜLTESİ ÖLÇME VE DEĞERLENDİRME LABORATUVARI ALT
YAPI ÇALIŞMASI: TANIMLAYICI ARAŞTIRMA**

**Fatma ORGUN¹, Fahriye VATAN¹, Nilay OZKUTUK¹, Hale SEZER², Meltem
DURŞUN ENGIN¹**

Özet

Bu çalışma hemşirelik fakültesi ölçme ve değerlendirme laboratuvarı kurularak eğitim programında uygulanan sınavların kalitesinin değerlendirilmesi amacıyla yapılmıştır. Tanımlayıcı kesitsel tipteki çalışmanın verileri, hemşirelik mesleki derslerinin toplam 42 adet sınavı ve 2671 test maddesinden oluşmaktadır. Araştırma, bir devlet üniversitesinin hemşirelik fakültesinde 26.06.2014-28.06.2018 tarihleri arasında yürütülmüştür. İlk aşamada ölçme değerlendirme laboratuvarı oluşturuldu. İkinci aşamada eğitici eğitimleri gerçekleştirildi. Üçüncü aşamada hemşirelik mesleki sınavlarının madde güçlük ve madde ayırtedicilik indeksleri belirlendi. Verilerin analizinde sayı ve yüzde dağılımları ve ortalamaları alındı. Hemşirelik derslerinin tüm test maddelerinin 2015-2016 öğretim yılında %40.49'u ve 2016-2017 öğretim yılında %38.90'ı çok kolay düzeyde (0.81-1.00) madde güçlük indeksine sahip olduğu ve 2015-2016 öğretim yılında %46.83'ü ve 2016-2017 öğretim yılında %43.56'sı çok düşük düzeyde (0.19 ve altı) madde ayırt edicilik indeksine sahip olduğu belirlenmiştir. Sonuç olarak hemşirelik derslerinin sınavlarının büyük bir kısmının kolay ve tüm sınavların ayırtıcılığının çok düşük düzeyde olduğu belirlendi. Gelecekteki araştırmaların hemşirelik mesleki bilgilerini ölçen çoktan seçmeli soruların madde analizlerini ve madde yazım hatalarını belirlemeye yönelik yapılması önerilir.

Anahtar Kelimeler: Hemşirelik, Çoktan Seçmeli Sorular, Madde Ayırt Edicilik İndeksi, Madde Güçlük İndeksi, Psikometrik

Abstract

Aim of this study, to establish a measurement and evaluation laboratory of a nursing faculty and evaluate the quality of tests applied as part of the education curriculum as a Descriptive cross-sectional design. A total of 42 tests related to nursing vocational courses and 2,671 test items. The research was conducted in a state university nursing faculty between June 26, 2014 and June 28, 2018. In the first stage, a measurement and evaluation laboratory was established. In the second stage, trainers were trained. In the third stage, the item difficulty and item discrimination indexes of nursing professional tests were determined. Number and percentage distributions and averages were obtained in the analysis of the data. Of all the test items of vocational nursing courses, 40.49% in the 2015-2016 academic year and 38.90% in the 2016-2017 academic year were very easy according to the item difficulty index (0.81-1.00), and 46.83% and 43.56%, respectively had very low item discrimination indexes (0.19 and below). Most of the tests applied in nursing courses were easy, and the discrimination of all tests was very low. Further studies are recommended to investigate the item analysis of multiple-choice questions that measure nursing professional knowledge and item spelling errors.

Keywords: Nursing, Multiple-Choice Questions, Item Discrimination Index, Item Difficulty Index, Psychometrics

Geliş Tarihi (Received Date): 08.06.2021, Kabul Tarihi (Accepted Date):24.08.2021, Basım Tarihi (Published Date): 30.09.2021 ¹ Ege Üniversitesi, Hemşirelik Fakültesi, Hemşirelik Bölümü, İzmir². İzmir Bakırçay Üniversitesi, Sağlık Bilimleri Fakültesi, Hemşirelik Bölümü, İzmir. **E-mail:** haleacan@yahoo.com **ORCID ID's:** F.O.; <https://orcid.org/0000.0002.2351.7227>, F.V.; <https://orcid.org/0000.0002.8044.0908>, N.Ö.; <https://orcid.org/0000.0003.1405.4600>, H.S.; <https://orcid.org/0000.0003.4199.7727>, M.D.E.; <https://orcid.org/0000.0002.7284.040X>.



1. INTRODUCTION

The effects of the rapidly developing modern information and technology age have resulted in changes in the content of courses in the curriculum, teaching methods, and measurement and evaluation techniques (Gelbal & Kelecioglu, 2007). Evaluation, which has an important place in the education process, is critical for the success of any education program and used as data in many decisions (Brady, 2005, pp. 240; Dönder, Elaldı, & Özkaya, 2012, pp. 955; Semerci, 2007, pp. 130; Yeşilyurt, 2012, pp. 378) One of the most commonly used tools in the evaluation process is multiple-choice testing (Tarrant & Ware, 2012, pp. 3). In nursing, multiple-choice questions (MCQs) are mostly utilized as a student evaluation method (Tarrant, Knierim, Hayes, & Ware, 2006, pp. 356). Poorly structured MCQs can cause various problems in the evaluation of student competence (Brady, 2005, pp. 240).

In a multiple-choice test, each question is presented with options that can be an answer to that question, and students are expected to find and mark the accurate option (Tarrant et al., 2006, pp. 356). MCQs with a single correct answer include a root, a correct/best answer, and distractors (Tarrant & Ware, 2012, pp. 99). MCQs allow educators to efficiently evaluate a large number of candidates and measure a wide range of content and learning goals (Brady, 2005, pp. 239; Tarrant & Ware, 2012, pp. 99). Multiple-choice testing can be used in all types of evaluations to assess the student's ability to remember, interpret and analyze information (Brady, 2005, pp. 240; Dascalu, Enache, Mavru, & Zegan, 2015, pp. 23). Performance-based assessment methods are used to evaluate the practice skills of nursing students. If MCQs are properly structured, they can not only meet all the psychometric properties of the test (reliability, validity, objectivity, fairness, and practicality) but also help evaluate the high-level cognitive processes of Bloom's taxonomy (Case & Swanson, 2001, pp. 32; Dell & Wantuch, 2017, pp. 138; Hingorjo & Jaleel, 2012, pp. 143; Macerata, Costa, & Lages e Silva, 2018, pp. 147).

Although MCQs seem easy to prepare, they are actually difficult and contain more structural errors (Brady, 2005, pp.240). Preparing good test items that assess higher cognitive learning levels is time-consuming (Clifton & Schriener, 2010, pp.13). While preparing these tests, care should be taken to ensure that they do not contain any clues (Brady, 2005, pp. 240). If there are incorrectly written items in the test, this will cause errors in both student evaluation and the evaluation of the faculty education program (Brady, 2005, pp. 240). Test items developed without following accepted test preparation guidelines may lead to the misinterpretation and affect student performance (Clifton & Schriener, 2010, pp.13). The most common errors in test items are irregular length of options, negative questions, multiple correct answers, illogical options, and grammatical errors (Brady, 2005, pp. 240). A poorly written test item can confuse a student who actually knows the correct answer and reduce his/her score. On the other hand, a misstructured test item can reward a respondent that is not knowledgeable about the content of the item by inadvertently by providing logical clues that indicate the correct



answer (Case & Swanson, 2001, pp.33; Dell & Wantuch, 2017, pp. 138; Tarrant & Ware, 2012, pp. 99).

It has been argued that in nursing education, many test questions are prepared to evaluate low-level cognitive processes, and they are not appropriately related to learning goals (Nedeau-Cayo, Laughlin, Rus, & Hall, 2013, pp. 53; Nemec & Welch, 2016, pp. 161). Examined test item defects have led to the conclusion that systematic errors are present, which reduce the validity of evaluation (Downing, 2005, pp. 134; Nemec & Welch, 2016, pp. 161). Defective test items negatively affect the quality of the test. Tests with defective items cannot properly reflect the difference between successful and borderline students (D'Sa & Visbal-Dionaldo, 2017, pp. 110; Tarrant, Ware, & Mohammed, 2009, pp. 2).

However, only few nurse educators are sufficiently prepared and knowledgeable on how to develop high-quality multiple-choice tests. Educators often develop test items themselves or rely on question banks as the source of questions, both of which may result in lower than optimal test quality. Thus, there may be significant deficiencies in the tests prepared by course instructors (Dell & Wantuch, 2017, pp. 138; Tarrant & Ware, 2012, pp. 99). Without proper training, most novice test writers develop low-quality test items that only measure the recall ability or insignificant content (Tarrant & Ware, 2012, pp.100). In addition, spelling mistakes in an item can result in evaluating the student's ability to understand what they have read rather than the content (Dell & Wantuch, 2017, pp. 138).

Educators should use well-designed and valid tools for evaluation (Race & Brown, 2001, pp. 24). Among the difficulties encountered during the preparation of objective tests is ensuring that what is intended to be measured is actually measured and the measurement is consistent (Brady, 2005, pp. 240). It is recommended that the prepared multiple-choice tests be reviewed critically by other instructors, and feedback should be received from colleagues (Brady, 2005, pp. 241; Quinn, 2000, pp.32; Race & Brown, 2001, pp. 24).

Carefully structured tests should be used by educators to ensure that the evaluation process is effective and reliable (D'Sa & Visbal-Dionaldo, 2017, pp. 110; Mahjabeen et al., 2018, pp. 311). The test can be made more effective by not repeatedly using items that have been found to receive weak and ineffective responses from students (D'Sa & Visbal-Dionaldo, 2017, pp. 110 ; Tarrant et al., 2009, pp. 2). In this context, the current study aimed to establish a measurement and evaluation laboratory of a nursing faculty and to evaluate the quality of tests applied as part of the education curriculum.

2. METHODS

This research had a cross-sectional descriptive design that included the review of test items administered to students in the 2015-2016 and 2016-2017 academic years in a nursing faculty of a state university. This research was carried out in accordance with STROBE



Statement. The study sought an answer to the following question: “Do tests used in nursing courses effectively evaluate students?”. The sub-problems of the research were as follows: “What are the mean scores of the nursing course tests for each academic year?”, “What are the test difficulty indexes of the nursing course tests in each academic years/are they within acceptable limits?”, and “What are item discrimination indexes of the nursing course tests in each academic year/are they within acceptable limits?”.

The research was conducted in the nursing faculty between June 26, 2014 and June 28, 2018. The software was purchased between these dates, and the trainings continued intermittently on January 16.01.2015, February 3, 2016 and June 16, 2017. All tests applied within the scope of vocational nursing courses in the nursing faculty in the 2015-2016 and 2016-2017 academic years (30 modules, six finals and six make-up tests) were included in the research sample. The education program of the nursing faculty offers integrated education, and vocational courses consist of modules. Nursing courses consist of a combination of these modules: six modules in the first grade, four modules in the second grade, and five modules in the third grade. Each module involves one MCQ test. As a data collection tool, we used the software ‘Nursing Faculty Measurement and Evaluation System (NFMES)’. NFMES software is based on a web-based architecture and can run on any web browser on the client side through an independent platform. On the server side, there is a web server and a relational database server. Measurement and evaluation system pages were prepared using the dynamic query language ASPX. In terms of security, the use of the program is limited to the local network, and necessary precautions were taken by preventing remote access.

2.1 Data collection stages

Stage 1. Establishment of laboratory procurement of NFMES software: this study was funded as a research project. By this project budget, NFMES software, a server and computer suitable for this software were purchased.

Stage 2. Training for software use: In order to use NFMES effectively, training sessions were held on January 16.01.2015, February 3, 2016 and June 16, 2017 to help lecturers prepare MCQs targeting learning goals. In addition, eight workshops each lasting four hours were organized on January 17, 2017 and February 9, 10, 14, 16 and 17, 2017 to increase the quality of MCQs to be written by lecturers.

Stage 3. Assessment of tests: Transforming tests into raw data using an optical reader and evaluating them through the system: The optical forms of the nursing vocational course tests were read with an optical reader in the evaluation process. The test evaluation was undertaken by the researchers using the student assessment system (NFMES), and the test difficulty index, item difficulty and item discrimination indexes and scores of each test were determined. When calculating item difficulty indexes through the software, the following reference ranges were used: ‘very difficult’ (0.00-0.20), ‘difficult’ (0.21-0.40), ‘moderate’



(0.41-0.60), ‘easy’ (0.61-0.80), and ‘very easy’ (0.81-1.00) (Atılğan, 2011, pp. 92; Linn & Miller, 2005, pp. 123; Nartgün et al., 2016, pp.33). Item discrimination indexes were calculated with the same software by comparing 27% lower and upper groups and applying the reference ranges of ‘very low-poor’ (0.19 and below), ‘low-needs further work’ (0.20-0.29), ‘fair to good discriminative ability’ (0.30-0.39), and ‘very good-high discriminative ability’ (0.40 and above) (Atılğan, 2011, pp. 92; Linn & Miller, 2005, pp. 123; Nartgün et al., 2016, pp. 33).

Data were evaluated using the Statistical Package for Social Sciences (SPSS) v. 21.0. Number and percentage distributions and averages were taken in the analysis of the data.

2.2 Ethical Approval

Approval was obtained from the Scientific Ethics Committee of the university (approval date: 31.12.2013, number: 2013-51) and the dean of the nursing faculty in order to conduct the study.

3. RESULTS

The research was conducted to establish an evaluation and measurement laboratory of a nursing faculty and evaluate the quality of tests applied in the education program of vocational nursing courses. A total of 42 exams and 2,671 test items were examined using the established evaluation and measurement system of the faculty.

3.1 Distribution of Findings Concerning the First Sub-problem

The findings related to the first sub-problem of the research, “What are the mean scores of the nursing course tests for each academic year?”, are shown in Tables 1 and 2. It was determined that for the 2015-2016 academic year, the lowest mean score was obtained from the make-up test of the first-grade vocational nursing course (61 ± 6.96) and the highest from the fifth module test the first-grade vocational nursing course (89 ± 6.57) (Table 1). The mean score of all tests applied during the vocational nursing courses was 67.90 ± 8.91 for the 2015-2016 and 67.81 ± 9.37 for the 2016-2017 academic year.

3.2 Distribution of Findings Concerning the Second Sub-problem

Tables 1 and 2 present the findings related to the second sub-problem of the research, “What are the test difficulty indexes of the nursing course tests in each academic years/are they within acceptable limits?”. The mean test difficulty index of all vocational nursing course tests was 0.68 for the 2015-2016 academic year and 0.66 for the 2016-2017 academic year, and accordingly all tests were classified as easy.

According to the test difficulty index averages of the tests in the 2015 academic year, 16 of the tests (final, make-up and module) were evaluated as easy, two (module) were very



easy, and three them (make-up and module) were of moderate difficulty (Table 1). The test difficulty index averages of the tests in the 2016 academic year indicated that 16 of the tests (final, make-up and module) were easy, three (make-up and module) had moderate difficulty, one (module) was very easy, and one (module) was difficult (Table 2).

3.3 Distribution of Findings Concerning the Third Sub-problem

The findings concerning the third sub-problem of the research, “What are item discrimination indexes of the nursing course tests in each academic year/are they within acceptable limits?”, are given Tables 1 and 2. Of all the test items of vocational nursing courses, 40.49% in the 2015-2016 academic year and 38.90% in the 2016-2017 academic year were very easy according to the item difficulty index (0.81-1.00), and 46.83% in the 2015-2016 academic year and 43.56% in the 2016-2017 academic had very low item discrimination indexes (0.19 and below).

An acceptable item difficulty index was observed in two tests in the 2015-2016 academic year (third module test in second grade and second module test in third grade) and four tests in the 2016-2017 academic year (make-up, second module and third module tests in first grade). In terms of the item discrimination index, it was determined that all the tests in both academic years were below the acceptable lower limit (Tables 1, 2).



Table 1. Distribution of The Difficulty and Discrimination Indexes of The Tests Applied in Vocational Nursing Courses in the 2015-2016 Academic Year

Grade	Test name	N	\bar{X}	SD	Test difficulty index Mean	Item difficulty index %					Item discrimination index %			
						Very difficult	Difficult	Moderately difficult	Easy	Very easy	Very low	Low	Moderate	Very high
First grade	FT1	100	70	7.79	0.7	3	11	17	25	44	58	18	17	7
	MUT1	100	61	6.96	0.6	14	17	21	26	22	35	32	-	33
	MT1	50	64	9.63	0.64	8	16	14	30	32	50	12	24	14
	MT2	50	72	8.73	0.71	6	8	16	20	50	52	18	16	14
	MT3	31	40	7.65	0.63	3.2	25.8	19.4	12.9	38.7	35.5	9.7	16.1	38.7
	MT4	50	74	7.99	0.74	-	12	14	26	48	62	12	14	12
	MT5	50	89	6.57	0.88	-	2	6	10	82	72	8	6	14
	MT6	40	69	10.9	0.68	-	7.5	27.5	32.5	32.5	45	20	12.5	22.5
Second grade	FT2	100	67	8.57	0.67	7	9	22	25	37	51	24	11	14
	MUT2	100	71	9.46	0.7	5	10	17	20	48	60	-	16	24
	MT1	50	77	8.44	0.77	-	8	14	24	54	56	14	18	12
	MT2	50	72	12.14	0.71	4	4	18	30	44	34	18	22	26
	MT3	50	69	9.7	0.68	4	14	20	20	42	40	26	14	20
	MT4	50	82	10.52	0.81	-	-	6	42	52	36	20	30	14
Third grade	FT3	100	69	0.69	0.69	6	10	17	25	42	56	27	13	4
	MUT3	100	55	8.86	0.54	14	20	20	22	24	49	25	-	26
	MT1	50	64	9.6	0.64	8	12	16	28	36	42	26	14	18
	MT2	50	60	10.99	0.6	4	20	26	26	24	34	22	22	22
	MT3	50	67	11.41	0.66	4	10	20	36	30	36	20	16	28
	MT4	50	63	12.37	0.62	6	12	26	34	22	28	22	18	32
	MT5	50	71	8.08	0.7	8	6	14	26	46	52	26	14	8
Total		1321	67.90	8.91	0.68	7.14	12.86	17.66	25.73	40.49	46.83	22.36	16.8	19.2

**FT: Final test, MUT: Make-up test, MT: Module test



Table 2. Distribution of The Difficulty and Discrimination Indexes of The Tests Applied in Vocational Nursing Courses in the 2016-2017 Academic Year

Grade	Tests	N	\bar{X}	SD	Test difficulty index Mean	Item difficulty index %					Item discrimination index %			
						Very difficult	Difficult	Moderately difficult	Easy	Very easy	Very low	Low	Moderate	Very high
First grade	FT1	100	67	7.93	0.67	6	10	21	24	39	53	25	16	6
	MUT1	100	58	10.82	0.56	8	13	36	24	19	46	28	26	-
	MT1	50	76	8.72	0.75	4	4	12	26	54	50	26	16	8
	MT2	50	59	9.54	0.59	8	18	24	20	30	32	44	24	-
	MT3	30	52	9.68	0.68	2.5	7.5	25	15	50	37.5	32.5	20	10
	MT4	40	71	10.03	0.71	-	10	20	36.7	33.3	26.7	23.3	20	30
	MT5	40	93	7.32	0.91	-	-	2.5	10	87.5	72.5	20	2.5	5
Second grade	MT6	40	67	11.08	0.66	5	12.5	15	32.5	35	40	20	15	25
	FT2	100	70	8.88	0.69	2	12	22	26	38	50	16	21	13
	MUT2	100	61	3.71	0.61	15	14	20	27	24	55	24	-	21
	MT1	50	67	7.66	0.67	4	22	12	16	46	62	22	8	8
	MT2	50	69	13.36	0.68	2	6	24	42	26	26	18	20	36
	MT3	50	75	8.67	0.75	4	6	14	18	58	56	14	14	16
Third grade	MT4	100	81	8.9	0.4	2	2	10	30	56	50	24	18	8
	FT3	100	66	9.4	0.66	3	17	23	26	31	38	28	18	16
	MUT3	100	59	8.01	0.6	6	21	21	26	26	50	24	10	16
	MT1	50	64	9.49	0.64	10	18	6	32	34	42	28	10	20
	MT2	50	66	10.68	0.66	2	14	20	32	32	30	34	22	14
	MT3	50	69	11.9	0.68	2	10	22	32	34	32	16	14	38
	MT4	50	67	10.85	0.66	2	10	24	34	30	32	26	20	22
MT5	50	67	10.21	0.66	2	12	26	26	34	34	30	22	14	
	Total	1350	67.81	9.37	0.66	4.36	12.61	19.02	26.44	38.90	43.56	24.90	16	18.35

**FT: Final test, MUT: Make up test, MT: Module test



4. DISCUSSION

The research was carried out to evaluate the quality of tests administered in a nursing faculty education program to seek an answer to the following question: “Do tests used in nursing courses effectively evaluate students?”. It was determined that the mean score of all tests in nursing courses was similar in both academic years (Tables 1, 2) and that the mean score of all the tests was over 100, and the threshold for passing a grade level (minimum 60) was achieved. Tests that measure knowledge in nursing education help determine students’ strengths and weaknesses (Mehmood et al., 2021, pp. 237). It can be stated that the strengths and weaknesses of the students could not be precisely determined in line with the mean scores obtained from the tests analyzed.

In the comparison of the nursing course tests of the 2015-2016 and 2016-2017 academic years, it was observed that most of the tests were easy and the discrimination ability of all tests was very low (Tables 1, 2). Although the high values of the test difficulty index of all the tests indicate that they were easy, contrary to what was expected, the low item discrimination index of the tests questions may have led students to get lower scores from the tests.

We determined that almost half the tests conducted in the 2015-2016 and 2017-2018 academic years had very low difficulty according to the item difficulty index (0.81-1.00) (Tables 1, 2). Basically, the item difficulty index is evaluated within the range between 0-100%, and a good item should have a p value varying between 50 and 60%, and an acceptable item 30-70% (Rush, Rankin, & White, 2016, pp. 8; Topal, Aybek, Kara Orhan, Büke, & Aybek, 2008, pp. 124). Accordingly, it can be stated that the item difficulty indexes of the items analyzed in the current study were not within acceptable limits.

The item difficulty index is evaluated based on the proportion of individuals who correctly answered that item. The higher the value of the item difficulty index, the easier it is to evaluate the item (Hingorjo & Jaleel, 2012, pp. 144). In the current study, it was determined that nearly half the tests in both academic years had very low item discrimination indexes (0.19 and below), and only six tests had acceptable item difficulty indexes (Tables 1, 2). In the literature, a value of 0.30 (moderate difficulty) is accepted as the lower limit for item discrimination (Erkuş, 2006, pp. 85; Sahin, Atay, Yagdi, & Aka, 2017, pp. 604). In terms of the item discrimination index, all the tests in both academic years were below the acceptable lower limit. The item discrimination index helps determine how well an item can distinguish between students who are knowledgeable and those that are not (Hingorjo & Jaleel, 2012, pp. 145). The low item discrimination indexes of all tests in our study show that they were not able to discriminate between good and low-performing students.

MCQs are efficient, objective, and easy to mark, and they can be used to test a large sampling of the curriculum (Brady, 2005, 239). However, it may take an hour to write a single well-structured test item (Morrison & Free, 2001, pp. 18), and it is stated that lecturers of the nursing faculty generally do not have sufficient time to construct and analyze a test (Hicks,



2011, pp. 267). It is emphasized that if a test with poorly structured items misdirects students, this will distort the proper evaluation of student competence (Brady, 2005, pp. 239). The item difficulty and discrimination indexes of the tests in the study not being at the desired level indicate that the competence of the students could not be effectively determined.

Reports on MCQ analyses used in the evaluation of students in nursing education suggest that there is more than one defect and violation in these items (Hicks, 2011, pp. 268). The guidelines for the preparation of MCQs contain various rules, such as being understandable in terms of the elements of the test items, not including very long sentences, complying with grammar rules, and structural and content integrity of the options (Hicks, 2011, pp. 268; Przymuszała, Piotrowska, Lipski, Marciniak, & Cerbin-Koczorowska, 2020, pp. 9). In addition to item analysis, it is also important to identify spelling mistakes in MCQs (Przymuszała et al., 2020, pp. 9). Preparing MCQs is a skill that can be improved. The creation of high-quality MCQs requires familiarity with the guidelines on these items and a willingness to change personal writing habits (Abdulghani et al., 2015, pp. 3; AlFaris et al., 2015, pp. 1307). Considering the results of our item analysis, training led to an improvement in the difficulties of the test items but did not change item discrimination. Gupta et al. (2020) indicated that the one-day training sessions did not improve MCQ writing skills (Gupta, Meena, Khan, Malhotra, & Singh, 2020, pp. 212). It is important that these tests, which allows for the evaluation of the cognitive field of nursing education, can distinguish between good and poorly performing students, and MCQs should have difficulty levels within acceptable limits.

5. CONCLUSION

This research was carried out to evaluate the quality of tests applied in a nursing faculty education program. In the comparison of the nursing course tests in both academic years, it was determined that most of the tests were easy and had very low discrimination ability. Educational interventions led to an improvement in the item difficulty status, but did not have an effect on item discrimination. In this respect, it is recommended that rather than the group teaching method, item preparation training should be planned and continued intermittently using test items that have been written by lecturers and analyzed in terms of difficulty and discrimination ability. For further studies, it is recommended to increase the number of item analyses of tests applied in nursing faculties, as well as those analyzing spelling mistakes in test items.

Funding: This study was financially supported by the Scientific Research Project Unit of Ege University [grant number: 14-HYO-001, Budget: 35373TL]



6. REFERENCES

- Abdulghani, H. M., Ahmad, F., Irshad, M., Khalil, M. S., Al-Shaikh, G. K., Syed, S., ... Haque, S. (2015). Faculty development programs improve the quality of Multiple Choice Questions items' writing. *Scientific Reports*, 5(9556), 1–6. <https://doi.org/10.1038/srep09556>
- AlFaris, E., Naeem, N., Irfan, F., Qureshi, R., Saad, H., Al Sadhan, R., ... Van der Vleuten, C. (2015). A One-Day Dental Faculty Workshop in Writing Multiple-Choice Questions: An Impact Evaluation. *Journal of Dental Education*, 79(11), 1305–1313.
- Atılğan, H. (2011). *Eğitimde Ölçme ve Değerlendirme* (4. Baskı). Ankara: Anı yayıncılık.
- Brady, A. M. (2005). Assessment of learning with multiple-choice questions. *Nurse Education in Practice*, 5(4), 238–242. <https://doi.org/10.1016/j.nepr.2004.12.005>
- Case, S. M., & Swanson, D. B. (2001). Constructing Written Test Questions For the Basic and Clinical Sciences. *National Board of Medical Examiner*, (June), 1–181. <https://doi.org/10.1111/j.1365-2990.2004.00612.x>
- Clifton, S. L., & Schriener, C. L. (2010). Assessing the quality of multiple-choice test items. *Nurse Educator*, 35(1), 12–16. <https://doi.org/10.1097/NNE.0b013e3181c41fa3>
- D'Sa, J. L., & Visbal-Dionaldo, M. L. (2017). Analysis of Multiple Choice Questions: Item Difficulty, Discrimination Index and Distractor Efficiency. *International Journal of Nursing Education*, 9(3), 109-114. <https://doi.org/10.5958/0974-9357.2017.00079.4>
- Dascalu, C. G., Enache, A. M., Mavru, R. B., & Zegan, G. (2015). Computer-based MCQ Assessment for Students in Dental Medicine—Advantages and Drawbacks. *Procedia - Social and Behavioral Sciences*, 187, 22–27. <https://doi.org/10.1016/j.sbspro.2015.03.005>
- Dell, K. A., & Wantuch, G. A. (2017). How-to-guide for writing multiple choice questions for the pharmacy instructor. *Currents in Pharmacy Teaching and Learning*, 9(1), 137–144. <https://doi.org/10.1016/j.cptl.2016.08.036>
- Dönder, A., Elaldı, Ş., & Özkaya, Ö. M. (2012). Views of Lecturers on Complementary Methods of Measurement And Evaluation That Can be Used In Basic English Teaching at University Level. *Turkish Studies*, 7(1), 953–968.
- Downing, S. M. (2005). The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement examinations in medical education. *Advances in Health Sciences Education*, 10(2), 133–143. <https://doi.org/10.1007/s10459-004-4019-5>



Erkuş, A. (2006). Sınıf öğretmenleri için ölçme ve değerlendirme: kavramlar ve uygulamalar. Ankara: Ekinoks Yayınları.

Gelbal, S., & Kelecioğlu, H. (2007). Teachers' Proficiency Perceptions of about The Measurement and Evaluation Techniques and The Problems They Confront. Hacettepe University Journal of Education, 33, 135–145.

Gupta, P., Meena, P., Khan, A., Malhotra, R., & Singh, T. (2020). Effect of Faculty Training on Quality of Multiple-Choice Questions. International Journal of Applied and Basic Medical Research, 10, 210–214. <https://doi.org/10.4103/ijabmr.IJABMR>

Hicks, N. A. (2011). Guidelines for identifying and revising culturally biased multiple-choice nursing examination items. Nurse Educator, 36(6), 266–270. <https://doi.org/10.1097/NNE.0b013e3182333fd2>

Hingorjo, M. R., & Jaleel, F. (2012). Analysis of one-best MCQs: The difficulty index, discrimination index and distractor efficiency. Journal of the Pakistan Medical Association, 62(2), 142–147. <https://doi.org/10.1080/13658810701731168>

Linn, R., & Miller, M. (2005). Measurement and assessment in Teaching (9th Editio). Upper Saddle River: NJ: Pearson Education.

Macerata, I. M., Costa, L. A., & Lages e Silva, R. L. (2018). Brutalists Walytopias: Operations for a savage clinic. Arquivos Brasileiros de Psicologia, 70(1), 146–160. <https://doi.org/10.1016/j.jtumed.2016.08.014>

Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., & Rizvi, M. (2018). Difficulty Index , Discrimination Index and Distractor Efficiency in Multiple Choice Questions. Annal of PIMS, 13(4), 310-315.

Mehmood, B., Sultan, S., Azhar, M., Shakoore, A., Khan, N. B., & Niazi, Z. (2021). Quality analysis of multiple choice questions. Community Dentistry, 40(4), 236–239.

Morrison, S., & Free, K. W. (2001). Writing multiple-choice test items that promote and measure critical thinking. The Journal of Nursing Education, 40(1), 17–24.

Nartgün, Z., Çetiner, B., Gömleksiz, M., Bıçak, B., Karaca, E., & Yurdabakan, İ. (2016). Eğitimde Ölçme ve Değerlendirme (4. Baskı; S. Erkan, ed.). Ankara: Nobel Akademik Yayıncılık.

Nedeau-Cayo, R., Laughlin, D., Rus, L., & Hall, J. (2013). Assessment of item-writing flaws in multiple-choice questions. Journal for Nurses in Professional Development, 29(2), 52–57. <https://doi.org/10.1097/NND.0b013e318286c2f1>



Nemec, E. C., & Welch, B. (2016). The impact of a faculty development seminar on the quality of multiple-choice questions. *Currents in Pharmacy Teaching and Learning*, 8(2), 160–163. <https://doi.org/10.1016/j.cptl.2015.12.008>

Przymuszała, P., Piotrowska, K., Lipski, D., Marciniak, R., & Cerbin-Koczorowska, M. (2020). Guidelines on Writing Multiple Choice Questions: A Well-Received and Effective Faculty Development Intervention. *SAGE Open*, 10(3), 1–12. <https://doi.org/10.1177/2158244020947432>

Quinn, F. (2000). *The Principles and Practice of Nurse Education* (fourth ed.). Cheltenham: Stanley Thorne (Publishers) Ltd.

Race, P., & Brown, S. (2001). *The Lecturers Toolkit* (second ed.). London.: Kogan Page.

Rush, B. R., Rankin, D. C., & White, B. J. (2016). The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. *BMC Medical Education*, 16(250), 1–10. <https://doi.org/10.1186/s12909-016-0773-3>

Sahin, H., Atay, Y., Yagdi, T., & Aka, S. A. (2017). Item analyses of 2016 written exam of Turkish Society of Cardiovascular Surgery Proficiency Board. *Turkish Journal of Thoracic and Cardiovascular Surgery*, 25(4), 600–607. <https://doi.org/10.5606/tgkdc.dergisi.2017.14513>

Semerci, Ç. (2007). A View to the New Primary School Curricula With the Metaphors Relating to “Curriculum Development.” *Cumhuriyet University Faculty of Literature Journal of Social Sciences*, 31(2), 125–140.

Tarrant, M., Knierim, A., Hayes, S. K., & Ware, J. (2006). The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. *Nurse Education in Practice*, 6(6), 354–363. <https://doi.org/10.1016/j.nepr.2006.07.002>

Tarrant, M., & Ware, J. (2012). A Framework for improving the quality of multiple-choice assessments. *Nurse Educator*, 37(3), 98–104. <https://doi.org/10.1097/NNE.0b013e31825041d0>

Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. *BMC Medical Education*, 9(1), 1–8. <https://doi.org/10.1186/1472-6920-9-40>

Topal, K., Aybek, H., Kara Orhan, C., Büke, A., & Aybek, Z. (2008). Item and Test Analysis of Multiple-Choice Exams Which Applied to The First Term Students Of Pamukkale University Medical Faculty In 2006-2007 Academic Year. *Pamukkale Medical Journal*, 1(3), 120–126.



Establishing A Measurement And Evaluation Laboratory In A Nursing Faculty: A

Descriptive Study

Orgun et al.

Yeşilyurt, E. (2012). Öğretmen Adaylarının Ölçme Ve Değerlendirme Alanına İlişkin Genel Yeterlik Algıları. Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 9(17), 377–395.