Mariusz Panczyk, Henryk Rebandel, Jarosława Belowska, Aleksander Zarzeka, Joanna Gotlib

Division of Teaching and Outcomes of Education, Faculty of Health Sciences, Medical University of Warsaw

Entrance exam for candidates for master's nursing studies — a 5-year retrospective analysis

Egzamin wstępny w kwalifikacji kandydatów na studia II stopnia na kierunek pielęgniarstwo — 5-letnia analiza retrospektywna

ABSTRACT

Aim of the study. Analyze admission criteria and assess the quality of the entrance exam for the MA program in Nursing at the Medical University of Warsaw (MUW) between 2009 and 2013.

Material and methods. The study used the admission data of candidates for a Master's degree program in Nursing in 2009–2013 (n = 2,257; median of age 23 years). The reliability level was established by estimating Cronbach's alpha (reliability coefficient). ROC curve analysis for two categories (admitted/not admitted) was performed to estimate the selective capacity of the admission criteria.

Results. An analysis of particular editions of the entrance exam demonstrated its insufficient reliability (α coefficient ranging between 0.429–0.559). The discriminatory power for the adopted admission criteria was diverse. The point score obtained in the exam has a significantly greater selection power than the criterion based on grade point average (ROC curves, P < 0.00001). A high discriminatory power of the test questions regarding "Clinical Nursing" was also observed.

Conclusions. The role of GPA and scoring scientific activity needs to be reduced and, at the same time, the importance of the results of the exam needs to be increased. Moreover, improvements in the quality parameters of the exam test questions are needed, so as to achieve a better discriminative ability of this selective tool.

Nursing Topics 2015; 23 (3): 324–331

Key words: school admission criteria; educational measurement; nursing education

STRESZCZENIE

Cel pracy. Analiza kryteriów selekcji kandydatów oraz ocena jakości testowego egzaminu wstępnego na studia pielęgniarskie II stopnia na Warszawskim Uniwersytecie Medycznym (WUM) w latach 2009–2013.

Materiał i metody. W badaniu wykorzystano dane rekrutacyjne kandydatów na studia II stopnia na kierunku Pielęgniarstwo z lat 2009–2013 (n = 2257, średnia wieku 23 lata). Jakość egzaminu testowego oszacowano poprzez ocenę stopnia rzetelności (współczynnik α-Cronbacha). Zdolności selekcyjną kryteriów rekrutacyjnych wykonano za pomocą analizy krzywych ROC dla dwóch kategorii przyjęty/nieprzyjęty.

Wyniki. Analiza poszczególnych edycji wstępnego egzaminu testowego wskazuje na jego niedostateczną rzetelność (α w przedziale 0,429–0,559). Dla stosowanych kryteriów selekcji wykazano zróżnicowaną moc dyskryminacyjną. Wynik punktowy z egzaminu miał istotnie większą moc selekcji niż kryterium oparte na średniej ocen (krzywe ROC, p < 0,00001). Zaobserwowano także wysoką zdolność dyskryminacyjną pytań testowych z zakresu „Pielęgniarstwa klinicznego".

Wnioski. Konieczne jest zmniejszenie roli selekcyjnej średniej ocen oraz punktowanych osiągnięć naukowych w procesie rekrutacji przy równoczesnym zwiększeniu znaczenia wyników z egzaminu testowego. Ponadto, należy dążyć do poprawy parametrów jakościowych testowych pytań egzaminacyjnych, tak aby uzyskać lepszą niż dotychczas zdolność dyskryminacyjną dla tego narzędzia selekcyjnego.

Introduction

In the academic year 2000/2001, the Medical University of Warsaw (MUW) (former Academy of Medicine) Faculty of Health Sciences, introduced a Bachelor's program in Nursing for the first time in its history, and a Master's program was introduced in 2003/2004. In practice, this followed the strategy stipulated in the Bologna Declaration [1] and the provisions resulting from Directive 2005/36/EC of The European Parliament and of The Council of 7 September 2005 on the recognition of professional qualifications [2]. Therefore, the University had to face the necessity to develop a bottom-up admission system for Master's degree candidates. The absence of any previous experience regarding this kind of recruitment posed a considerable difficulty in choosing the criteria of selection of the best candidates.

For the first time in 2009, mini-interviews at the MUW were replaced by a test examination. Multiple-choice question (MCQ) exam became the main selection criterion (a maximum of 50 points) but the possibility of granting additional points for scientific activity was also preserved (a maximum of 8 additional points). GPA (without considering the diploma examination grade) still constituted the third important criterion.

All state higher education schools that are at the rank of a university and which educate at the nursing faculty in the second degree, carry out their recruitment process following their individual rules. There is no standardized tool that would evaluate the competences of candidates and which could be used by different universities. When building up their strategy of selecting the best candidates, universities usually base their choice on one or two selection criteria (the result of a test exam and/or GPA), which allows to create a score placement ranking (Table 1).

Test exam carried out at the Faculty of Health Sciences at MUW verifies the scope of knowledge and skills of candidates in the following four areas: *basic science, health sciences, primary health care, and clinical nursing*. Five editions of this exam took place between 2009-2013 and the importance of scores obtained on the MCQ exam had been increasing compared to the remaining admission criteria. A substantial influence of this exam on the final score means that its discriminatory value has been crucial in selecting the best candidates. Therefore it has been necessary to improve its quality so that the exam meets high standards imposed on this kind of tools. Due to the fact that a Master's program is offered to a smaller number of students than a much more common Bachelor's program, searching for an effective selection tool to find the best candidates who, more likely than not, shall achieve professional success on higher levels of nursing management seems to be justified [3].

Aim

The aim of the present study results was to assess the quality of the entrance test exam and critically analyze the selection criteria for candidates for the Master's program in Nursing at the MUW in the period of 2009–2013.

Material and methods

The present study is a 5-year retrospective surveillance analysis. The study used the admission data of candidates for a Master's degree program in Nursing in 2009–2013 (n = 2,257; median of age = 23 years), including the following: MCQ exam score and GPA in the Bachelor's program. See Table 2 for detailed characteristics of the study group.

Each MCQ exam comprised 50 questions in the "best answer from a list of possible answers" format, with four possible answers for each question in 2009–2010 and 2012 and five option items in 2011 and 2013. Three varieties of tests that differed only in the order of questions were prepared for each edition of the exam.

Normal distribution parameters of particular exam results were assessed using the Shapiro-Wilk test and data were screened for outliers using Grubbs test. The assessment of the quality of test exam was conducted with establishing the item easiness and substitute discriminating power of particular questions and distractors, and Cronbach's coefficient was estimated to determine the reliability level of the test [4]. Pearson's r linear correlation coefficient was used to determine the strength of correlation between the test exam score and GPA in Bachelor's program.

Receiver Operating Characteristics (ROC) curve analysis for two categories (admitted/not admitted)

Table 1. Candidates' selection criteria for full-time studies of the second degree at Polish universities which provide education at the Nursing Faulty

Tabela 1. Kryteria selekcji kandydatów na studia stacjonarne drugiego stopnia w polskich uniwersytetach prowadzących kształcenie na kierunku pielęgniarstwo

Public University	Admission criteria		
	Entrance exam	GPA	Scoring additional
Jagiellonian University Medical College	•		
Medical University of Bialystok		•	
Medical University of Gdansk	•		
Medical University of Lublin	•		
Medical University of Lodz	•		
Medical University of Warsaw	•	•	•[4]
Nicolaus Copernicus University Collegium Medicum in Bydgoszcz		•[1]	
Poznan University of Medical Science	•		
Pomeranian Medical University		•	
Wroclaw Medical University		•	•[4]
Faculty of Medical Sciences University of Warmia and Mazury in Olsztyn		•[2]	
Faculty of Health Sciences Jan Kochanowski University in Kielce		•[3]	
Faculty of Medicine University of Rzeszow	•		

GPA — final grade point average of Bachelor's degree program. [1]final evaluation on undergraduate diploma; [2]GPA + grade of diploma thesis + grade of diploma examination; [3]GPA + final evaluation on undergraduate diploma; [4]scientific activity (publications, conferences / symposia, student's scientific circles, Erasmus)

Table 2. Characteristics of the study group as divided into subgroups with respect to gender, mode of Master's program and type of university they graduated from with Bachelor's degree

Tabela 2. Charakterystyka grupy badanej w podzielone na podgrupy ze względu na płeć oraz ukończoną uczelnię na studiach pierwszego stopnia

	2009		2010		2011		2012		2013		Total	
	n	%	n	%	n	%	n	%	n	%	n	%
Total	541	–	581	–	492	–	346	–	297	–	2,257	–
Women	520	96.1	558	96.0	465	94.5	331	95.7	287	96.6	2,161	95.7
Men	21	3.9	23	4.0	27	5.5	15	4.3	10	3.4	96	4.3
MUW graduates	347	64.1	418	71.9	314	63.8	195	56.4	160	53.8	1,434	63.5
Graduates from other medical universities	25	4.7	20	3.5	38	7.7	49	14.2	45	15.2	177	7.9
Graduates from higher vocational schools	169	31.2	143	24.6	140	28.5	102	29.4	92	31.0	646	28.6

was performed to estimate the selective capacity of the question sets and a modified U-statistic as proposed by Hanley and Hajian-Tilaki was conducted to compare the Area Under the Curve (AUC) [5, 6].

Calculations were made in a statistical set of STATISTICA 12.5 (StatSoft©, Inc.) according to the MUW license. For all analyses, the relevance level assumed *a priori* was $\alpha = 0.05$.

Results

The analysis of the results of particular editions of the MCQ exam with regard to the nature of the distri-

Table 3. Characteristics of particular editions of the entrance exam.

Tabela 3. Charakterystyka poszczególnych edycji wstępnego egzaminu testowego

	2009	2010	2011	2012	2013
Number of questions	50	50	50	50	50
Number of options	4	4	5	4	5
Normal distribution*	$P < 0.05$	$P < 0.05$	$P < 0.05$	$P > 0.05$	$P > 0.05$
Data outliers**	$P > 0.05$	$P > 0.05$	$P > 0.05$	$P = 0.05$	$P > 0.05$
Mean	30.2	25.6	24.3	31.5	25.7
SD	4.33	4.59	4.16	4.44	4.71
Median	30.0	25.0	25.0	31.0	26.0
Q1–Q3	27.0–33.0	23.0–29.0	21.0–27.0	29.0–34.0	22.0–29.0
CV	14.3%	17.9%	17.1%	14.1%	18.4%
Skewness	0.046	0.109	–0,050	–0.112	0.064
Kurtosis	0.178	–0.005	0.022	0.327	–0.274
Index of test difficulty	0.62	0.52	0.49	0.67	0.52
Average power discriminating	0.205	0.203	0.184	0.190	0.207
Cronbach's α	0.558	0.546	0.429	0.446	0.559
Number of questions with negative correlation	7	9	10	7	8
Cronbach's α after optimization	0.587	0.604	0.523	0.534	0.622

*Shapiro-Wilk test (for $P < 0.05$ distribution is not normal); **Grubbs test for outliers (for $P < 0.05$ presence of outliers); Q_1 — first quartile; Q_3 — third quartile; SD — standard deviation; CV — coefficient of variation

bution of this variable demonstrated non-symmetric (skew) data diverging, in some cases, from the normal distribution (kurtosis $\neq 0$ and Shapiro-Wilk test, $P < 0.05$). Each edition of the MCQ exam differed with respect to the scope of variability of the results, which was reflected in different values of coefficients of variation, the range of results, and the scope of scores obtained by candidates in particular years. Narrow confidence intervals for the mean and standard deviation indicated high precision of parameters used for the study population. See Table 3 for a list of all results.

The analysis of particular exam editions demonstrated insufficient reliability of the question set. In subsequent years, α-coefficient ranged between 0.429–0.559. A detailed evaluation of the test scores allowed the selection of test questions with a negative correlation, i.e. significantly decreasing the reliability of the exam. With these questions excluded, the overall reliability of tests increased. After optimization, the α-coefficient ranged between 0.523–0.622 (Table 3).

To assess their selective capacity, particular exam questions were divided into four thematic areas: *basic science, health sciences, primary health care,* and *clinical nursing* (Table 4). The proportions of questions from particular thematic fields differed significantly in subsequent editions of the exam. A considerable tendency was observed towards increasing the proportion of exam tasks associated with *clinical nursing* (increase from 24% in 2009 to 66% in 2013). At the same time, the proportion of exam questions connected with *basic science* and *health sciences* decreased. It is also worth emphasizing that questions concerning *primary health care* were insufficiently represented in all exam editions.

Since the written exam is not the only selection criterion, a correlation between MCQ exam score and the grade point average in Bachelor's program was also analyzed. For the analysis of correlations between MCQ exam scores and GPA, candidates were grouped according to the criterion of school of graduation from a Bachelor's degree program. Candidates from other medical universities (not MUW) were excluded from the analysis of correlations because of their small number in subsequent admission years. Table 5 shows that in the case of MUW Bachelor graduates, there was a significant correlation between the results of the MCQ exams and GPA in all admission years (Pearson's r between 0.263 and 0.430). The correlations were also found

Table 4. Characteristics of particular categories of test questions in subsequent editions of the entrance exams

Tabela 4. Charakterystyka poszczególnych kategorii pytań wchodzących w skład egzaminu wstępnego

	Basic science	Health sciences	Primary health care	Clinical nursing
2009				
N	16 (32%)	21 (42%)	1 (2%)	12 (24%)
Cronbach's α	0.239	0.387	–	0.315
2010				
N	18 (36%)	10 (20%)	2 (4%)	20 (40%)
Cronbach's α	0.317	0.334	–	0.215
2011				
N	5 (10%)	9 (18%)	8 (16%)	28 (56%)
Cronbach's α	0.014	0.004	0.058	0.362
2012				
N	4 (8%)	8 (16%)	5 (10%)	33 (66%)
Cronbach's α	0.006	0.112	–0.061	0.326
2013				
N	6 (12%)	9 (18%)	2 (4%)	33 (66%)
Cronbach's α	0.168	0.256	–	0.478

Table 5. Results of analysis of correlation between GPA in Bachelor's program and total score for MCQ exam and score in the area of Clinical nursing

Tabela 5. Wyniki analizy korelacji między średnią ocen ze studiów pierwszego stopnia a punktacją za testowy egzamin oraz za pytania z kategorii Pielęgniarstwo kliniczne

|	Graduates of MUW	Graduates of Higher vocational schools
2009		
Total (n = 50)	0.303*	0.151
Clinical nursing (n = 12)	0.220*	0.128
2010		
Total (n = 50)	0.430*	0.206*
Clinical nursing (n = 20)	0,322*	0.115
2011		
Total (n = 50)	0.312*	0.178*
Clinical nursing (n = 28)	0.274*	0.148
2012		
Total (n = 50)	0.263*	0.316*
Clinical nursing (n = 33)	0.193*	0.265*
2013		
Total (n = 50)	0.365*	0.391*
Clinical nursing (n = 33)	0.354*	0.415*

*$P < 0.05$

in the analysis of a set of questions regarding *clinical nursing*, however, in this case, Pearson's r coefficients were of lower value (r ranging between 0.193 and 0.354). Significantly worse results of correlation analysis were obtained for Bachelor graduates from non-medical universities (higher vocational schools). Moreover, statistically significant correlations were higher among graduates from non-medical universities only in 2013, as compared to MUW graduates.

ROC curves were used to verify whether the selection criteria had sufficient discriminatory power. Criteria of significant specificity are most valuable for the selection of best candidates so that individuals with insufficient level of initial competences are not admitted to the course. The analysis of admission data of 2010 (Figure 1) showed that discriminatory power of particular selection criteria was different, which was expressed by a difference in areas under the curve (AUC). Due to obvious reasons, the total score for admission had the greatest discriminatory power (AUC = 0.924) when the MCQ exam scores had selection power at AUC = 0.8697. Evaluation of the ROC curves for GPA showed that this selection criterion worked the worst (AUC = 0.6418). An analysis of the remaining exam editions produced similar results. A modified *U*-statistic as proposed by Hanley et al. [6] showed that, in all the years that were analyzed, the discriminatory value of MCQ exam scores was significantly higher than selective capacity based on GPA (P < 0.00001). The analysis of the ROC curves might have also indicated that the discriminatory capacity of clinical test questions was high. For example, in 2010, as many as 40% of the questions were devoted to *clinical nursing*, but the discriminatory value of this group of exam tasks represented as much as 86.7% of the selection power of the entire exam set that comprised 50 questions.

Discussion

The MUW has been carrying out admission processes for a Master's program in Nursing for 11 years. Admission strategies have varied significantly during this period. Beginning with the first admission process in 2003, GPA estimated for all subject exams taken during the Bachelor's program was a constant criterion. However, the importance of this criterion with reference to the total score has been changing. In 2003–2008, points were also awarded for a mini-interview that aimed to evaluate certain predispositions of candidates. Beginning with the 2006 admission process, additional points were introduced for involvement in scientific activity in a Bachelor's program (scholarships, honorable mentions, internship trips, participation in scientific conferences, and co-authorship of publications). A review of the
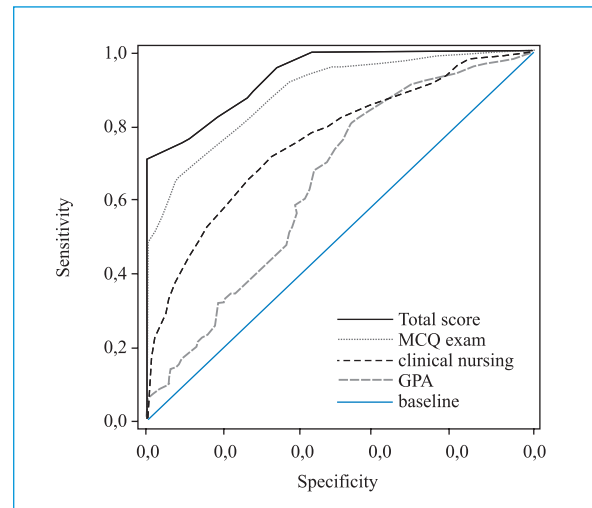
Figure 1. Example of analysis of discriminatory capacity of various selection criteria on the basis of ROC curve for 2010 exam

Rycina 1. Przykład analizy dyskryminacyjnej uzyskanej na podstawie krzywej ROC dla różnych kryteriów selekcyjnych w rekrutacji z 2010 roku

abovementioned admission practices showed that the selection of candidates had had several major defects: a) mini-interviews are very time-consuming and difficult in standardization, particularly if a university does not have all the necessary experience to conduct examinations in such a form [7]; b) a very robust list of scoring scientific activity contributes to a significant decrease in discriminatory capacity of the remaining selection criteria; c) attributing great importance to GPA is not justified in cases when grading systems in university-level schools offering Bachelor's programs are very diverse and if most of those schools have not developed good practices in this regard. The above remarks were also reflected in results published by other researchers concerned with these issues. McNelis et al. pointed out a good feature of a mini-interview [8]: it allows academic teachers to become acquainted with a candidate. This was compliant with the previous observations made by Trice and Foster [9]. However, as it was mentioned above, this selection criterion does not work well in all circumstances [8]. On the other hand, many authors have pointed to the problem of effectiveness of the selection based on GPA due to the heterogeneity of grading systems in various educational institutions; therefore, the selection of candidates based on GPA might be of poor reliability [8, 10]. Correlation analysis between GPA and MCQ exam score that was added to the set of criteria in 2009 confirmed the abovementioned observations. ROC curves analysis showed that the selective capacity of GPA was low,

and it was statistically insignificant in 2012–2013, falling far behind the scores in MCQ exam.

Recruitment team for a Master's degree program at MUW decided to introduce in 2009 a MCQ exam as the key selection criterion, which was justified by positive results of qualitative and quantitative analyses published in the literature. Positive results of correlation analyses and good predictive value of this criterion are important reasons supporting the thesis above since MCQ exam allows, with high probability, prediction of future career of students and graduates [10].

The present analysis of reliability conducted with the use of Cronbach's coefficient demonstrated that the test exams used between 2009–2013 did not meet the assumed reliability criteria. Cronbach's coefficient did not reach the value higher than 0.7 in any of the years that were analyzed [11]. Low quality of these exams was strongly influenced by an insufficient tool structure, i.e. inadequate content selection that, due to obvious reasons, constituted only a small part of all features and qualities of candidates we would like to evaluate. At the value of $\alpha = 0.5$, random errors constitute as much as half of the variability of results and a measurement in such conditions might be used only for intergroup comparisons and not for individual differentiation [12]. Low values of coefficient is a prompt to review the present exam tasks, consisting in a rejection of poorly differentiating components, increase in the domain coherence of a set, and extend the measurement scale [12].

Validity of a particular selection method may vary depending on how well a candidate knows the assessment tool. A phenomenon called "overall testing experience" consisting in taking an entrance exam repeatedly by one candidate after not passing the exam in the previous years may constitute a problem [13]. If we want to have a good candidate selection procedure, we follow the principle of choosing those who have possessed a minimum of knowledge and skills necessary for a certain profession and represent certain predispositions for this particular profession. Therefore, if we use the MCQ exam score as the main selection criterion, this tool should accurately assess those initial competences that were described as essential for taking a Master's program in Nursing.

Internal validity (known also as content validity) of a particular tool may only be achieved by developing a strict test content outlines consistent with the outcomes defined in the educational standards [14]. The analysis of individual thematic areas represented in the test plan shows an increasing share of questions in the area of clinical nursing. This tendency is connected with applying formal adjustment of a entry exam to the requirements stipulated in standards of education for the first degree of studies. This standard assumes that throughout these studies, at least 60% of time will be devoted to education in the area of *clinical nursing*. Such a big share of questions in this category in the pool of exam questions is to eventually determine the selection of such candidates who are well prepared considering their clinical skills and knowledge alike. Another thematic area presenting good discriminative parameters were questions related to *basic science*. Yet a low or very low level of selective ability was observed for *health sciences* and *primary health care*, which means that, especially for these two areas, it is necessary to improve quality. Implementation of the outlines in the subsequent editions of MCQ exam shall allow validation of assumptions and their potential improvement in the next years.

The issue of internal validity is also associated with "overfitting" of a test to a certain curriculum. A comparative analysis of MCQ exam scores of graduates from MUW and higher vocational schools (non-medical universities) showed that the scores obtained by the latter had been significantly worse. It is justified to pose a question whether the difference indeed resulted from the fact that MUW graduates were better candidates for the Master's program or whether they were excessively favored by the selection criteria. It is problematic to judge whether the selection tool developed by MUW lecturers was not overfitting to the training offered in a Bachelor's program in Nursing at MUW. The problem of "overfitting" was pointed out by Cronbach who argued that if the test content reflects the content of the classes taught e.g. on lectures rather than the assumed outcomes of education, then such a tool is not reliable [15]. In conclusion, it can be said that, while a significant increase in reliability of MCQ exam is possible, validity of this tool may be "overfitted" to a great extent if the exam is not standardized with a control group of non-MUW students.

As it was mentioned above, a development of good test content outlines constitutes a crucial component of a properly developed selection tool. When developing a complex tool such as an examination test, we assume that particular components are supposed to represent certain domains in certain proportions that can be changed during later evaluation. Apart from the above, selection of exam questions with reference to the level of difficulty is an important aspect that needs to be considered in test design. A large number of easy and approving questions that require only confirmation of correctness of given information, e.g., the use of "all of the above" (AOTA) and "none of the above" (NOTA) as options) is intellectually less demanding and contributes to obtaining better scores [16]. The distribution of variable of results is

then negatively skewed and the scope of discriminatory capacity is considerably narrowed. Therefore, it is so important to choose exam questions with a significant differentiation power. These questions are capable of distinguishing those candidates who obtain significantly different results for a particular parameter, which allows for a good selection. In the literature, these tasks are referred to as "burst tasks" [17], since they cause the flatness of the distribution of results (kurtosis < 0), which facilitates discrimination and establishment of the "passed/failed" cut-off point. A detailed technical analysis of individual test questions may provide very useful information on a group undergoing examination and, being a complementary part of a content-related analysis, it provides a foundation for evaluation and development of a high-quality database of exam questions. The current 5 years of experience in preparation of MCQ exams have demonstrated the weaknesses of this exam: non-equivalence of particular editions and improper selection of questions with regard to the easiness and differentiation power. If the questions with optimum easiness have sufficient differentiation parameters, they should constitute the core of the pool of exam questions and this is what should be pursued in the next years. The development of a proper database of exam questions is a long-lasting process and database resources need to be successively updated due to the process of "ageing" of exam questions. This process consists in an increase of easiness and decrease of differentiation power in the subsequent exam editions comprising the database [18].

Conclusions

The present analysis is the first step to improve the efficiency of the admission system for candidates for a Master's degree program in Nursing at MUW. The present findings and critical remarks on the quality of the selection criteria may become the basis for the introduction of essential changes in the admission policy: a) improvement of the test exam tool as the primary task for the team developing the next editions of MCQ exams; b) revision of the importance of the percentage share of GPA in the total score so that high validity of the entire selection procedure is preserved and score in MCQ exams is not diminished; and c) monitoring of career development of students and graduates in order to gather data for predictive analysis of the selection methods for nursing candidates.

References

1. Satu K.U., Leena S., Mikko S., Riitta S., Helena L.K. Competence areas of nursing students in Europe. Nurse. Educ. Today 2013; 33: 625–632.
2. Directive 2005/36/EC of The European Parliament and of The Council of 7 September 2005 on the recognition of professional qualifications (Official Journal of the European Union L 255/41).
3. Creech C.J., Aplin-Kalisz C. Developing a selection method for graduate nursing students. J. Am. Acad. Nurse. Pract. 2011; 23: 404–409.
4. Feldt L.S. A test of hypothesis that Cronbachs alpha or Kuder-Richardson coefficent 20 is same for 2 tests. Psychometrika. 1969; 34: 363.
5. Wanvarie S., Sathapatayavongs B. Logistic regression analysis to predict Medical Licensing Examination of Thailand (MLET) Step1 success or failure. Ann. Acad. Med. Singapore. 2007; 36: 770–773.
6. Hanley J.A., Hajian-Tilaki K.O. Sampling variability of nonparametric estimates of the areas under receiver operating characteristic curves: an update. Acad. Radiol. 1997; 4: 49–58.
7. Ehrenfeld M., Tabak N. Value of admission interviews in selecting of undergraduate nursing students. J. Nurs. Manag. 2000; 8: 101–106.
8. McNelis A.M., Wellman D.S., Krothe J.S, Hrisomalos D.D., McElveen J.L., South R.J. Revision and evaluation of the indiana university school of nursing baccalaureate admission process. J. Prof. Nurs. 2010; 26: 188–195.
9. Trice L.B., Foster P.H. Improving nursing school diversity through use of a group admission interview. AORN J. 2008; 87: 522–532.
10. Chen S., Voyles D. HESI Admission Assessment Scores: Predicting Student Success. J. Prof. Nurs. 2013; 29 (2 supl. 1): S32–S37.
11. Nunnally J.C., Bernstein I.H. Psychometric theory. 3rd ed. McGraw-Hill, New York 1967.
12. Guilford J.P. Psychometric methods. 2nd ed. McGraw-Hill, New York 1954.
13. Terry W.S. Learning and memory: Basic principles, processes, and procedures. 3rd ed. Pearson Education, Boston 2006.
14. Ebel R.L. Must all tests be valid? Am. Psychol. 1961; 16: 640–647.
15. Cronbach L.J., Meehl P.E. Construct validity in psychological tests. Psychol. Bull. 1955; 52: 281.
16. Boland R.J., Lester N.A., Williams E. Writing Multiple-Choice Questions. Acad. Psychiatry. 2010; 34: 310–316.
17. Downing S.M. 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. Adv. Health. Sci. Educ. Theory. Pract. 2005; 10: 133–143.
18. Case S.M., Swanson D.B. Constructing Written Test Questions For the Basic and Clinical Sciences. 3rd ed. National Board of Medical Examiners, Philadelphia 2002.