

Jakość testowego egzaminu wstępnego na kierunek położnictwo studia drugiego stopnia – 5-letnia analiza retrospektywna

Quality test entrance examination for full-time second-level in Midwifery faculty – 5-year retrospective analysis

Mariusz Panczyk, Jarosława Belowska, Henryk Rebandel, Joanna Gotlib

Zakład Dydaktyki i Efektów Kształcenia, Wydział Nauki o Zdrowiu, Warszawski Uniwersytet Medyczny

CORRESPONDING AUTHOR/AUTOR DO KORESPONDENCJI:

Mariusz Panczyk
Zakład Dydaktyki i Efektów Kształcenia
Wydział Nauki o Zdrowiu, Warszawski Uniwersytet Medyczny
ul. Żwirki i Wigury 61, 02-091 Warszawa
tel. +48 22 572 04 90, Fax: +48 22 572 04 91
e-mail: mariusz.panczyk@wum.edu.pl

STRESZCZENIE

JAKOŚĆ TESTOWEGO EGZAMINU WSTĘPNEGO NA KIERUNEK POŁOŻNICTWO STUDIA DRUGIEGO STOPNIA – 5-LETNIA ANALIZA RETROSPEKTYWNA

Cel. Ocena jakości testowego egzaminu wstępnego służącego do kwalifikacji kandydatów na studia stacjonarne drugiego stopnia na kierunku położnictwo na Warszawskim Uniwersytecie Medycznym.

Materiał i metodyka. Analizie poddano dane dla pięciu roczników kandydatów ($N = 335$, 100% kobiet), którzy przystąpili do egzaminu wstępnego na kierunku położnictwo w okresie 2009/10 – 2013/14. Zebrane dane dotyczyły wyników egzaminu testowego na studia (sumaryczna punktacja oraz w podziale na bloki tematyczne). Wyznaczono parametry statystyki opisowej, oszacowano łatwość i moc różnicującą pytań oraz oceniono rzetelność testów (alfa-Cronbacha). Wyznaczono ponadto indeksy dyskryminacyjne dla bloków tematycznych testu. Wszystkie obliczenia wykonano w pakiecie STATISTICA 12.5.

Wyniki. Najbardziej optymalny pod względem łatwości i mocy różnicującej dobór pytań zanotowano dla testu z roku 2009/10. Dla żadnej edycji egzaminu nie stwierdzono spełnienia minimalnych wymagań w zakresie wewnętrznej zgodności pomiaru. Najlepsze wyniki dotyczące rzetelności pomiaru zaobserwowano dla testu w roku 2009/10. Analiza mocy dyskryminacyjnej pytań w blokach tematycznych wskazywała na zróżnicowanie tego parametru jakościowego. Bardzo wysoki indeks dyskryminacyjny zanotowano dla bloku pytań z zakresu „opieki specjalistycznej” w roku 2009/10 (0,535), a najniższy w egzaminie w roku 2013/14 (0,150).

Podsumowanie. 1. Konieczna jest poprawa jakości testowych pytań egzaminacyjnych w zakresie trudności oraz mocy różnicującej. 2. Należy monitorować jakość kolejnych edycji egzaminu wstępnego w szczególności w odniesieniu do kontroli rzetelności tego narzędzia kwalifikacyjnego.

Słowa kluczowe:

położnictwo, ocena wiadomości, szkolnictwo wyższe, kryteria przyjęć do szkoły

ABSTRACT

QUALITY TEST ENTRANCE EXAMINATION FOR FULL-TIME SECOND-LEVEL IN MIDWIFERY FACULTY – 5-YEAR RETROSPECTIVE ANALYSIS

Aim. Quality assessment of the test entrance examination used for selection of candidates for full-time second-level studies in Midwifery at the Medical University of Warsaw.

Material and methods. Five grade-years of candidates ($N = 335$, 100% of women) who took the entrance exam for Midwifery between 2009/10 and 2013/14. Data collected was related to results of the test examination for studies (total score and scores divided into thematic blocks). Parameters of descriptive statistics were set. Ease and differentiating power of questions and reliability of tests were assessed (alpha-Cronbach). Discriminatory indexes for thematic blocks of the test were set. Calculations were performed in STATISTICA 12.5.

Results. The most optimal selection of questions in terms of ease and differentiating power was recorded for the 2009/10 test. None of the editions of the exam met the minimum requirements for the internal consistency of measurement. The best results concerning the reliability of the measurement was observed for the 2009/10 test. Analysis of differentiating power of questions in thematic blocks indicates the diversity of this quality parameter. Very high differentiating index was recorded for the block of questions on specialist care in 2009/10 (0.535), and the lowest in 2013/14 exam (0.150).

Conclusions. 1. It is necessary to improve the quality of exam questions in terms of difficulty and differentiating power. 2. Quality of subsequent editions of the admission examination should be monitored, in particular with regard to the control of the reliability of this qualifying tool.

Key words:

midwifery, educational measurement, graduate education, school admission criteria

WSTĘP

Od połowy lat 90. w krajach wysoko rozwiniętych obserwuje się stały wzrost liczby pielęgniarek i położnych podejmujących kształcenia na poziomie studiów magisterskich (*master's/postgraduate level*), co ma związek między innymi z rosnącym w tych krajach zapotrzebowaniem na kadre o wysokich kwalifikacjach w zakresie kompetencji klinicznych [1]. Rozwój kierunków studiów na poziomie podyplomowym jest więc odpowiedzią na wzrost poziomu specjalizacji w zawodzie pielęgniarki i położnej. Wpływ ma na to także pojawienie się nowych obszarów zawodowych związanych z koniecznością czynnego udziału przedstawicieli tych zawodów w procesie podejmowania decyzji oraz zarządzania w jednostkach ochrony zdrowia [2]. Podobne przemiany obserwujemy także w Polsce, co przekłada się na konieczność kształcenia wykwalifikowanej kadry pielęgniarek i położnych [3]. Wobec wysokich wymagań stawianych absolwentom kierunków medycznych, należy dążyć do jak najwyższego poziomu kształcenia w akredytowanych instytucjach akademickich. Jednym z istotnych czynników ograniczających efektywność nauczania jest niski poziom początkowego potencjału edukacyjnego wśród osób rozpoczynających studia [4]. Niedostateczne przygotowanie kandydatów do podjęcia studiów drugiego stopnia zwiększa niepewność w zakresie poziomu kompetencji wyjściowych wśród absolwentów kierunku [5]. Aby zminimalizować ryzyko „porażki edukacyjnej” należy stosować odpowiednią strategię rekrutacyjną podczas naboru na studia [6]. Jakość stosowanych metod oceny wiedzy i umiejętności kandydata powinna być podstawą dobrze zaplanowanej polityki rekrutacyjnej każdej uczelni.

W roku akademickim 2000/01 Warszawski Uniwersytet Medyczny (WUM) dawniej Akademia Medyczna, wprowadziła po raz pierwszy studia pierwszego stopnia na kierunku położnictwo, a od 2003/04 także drugiego stopnia. W związku z dwustopniowym podziałem studiów, uczelnia stanęła przed koniecznością zorganizowania od podstaw systemu rekrutacji kandydatów na studia położnicze drugiego stopnia. Brak wcześniejszych doświadczeń dotyczących prowadzenia tego typu rekrutacji, stanowił poważną trudność w doborze odpowiednich pod względem jakości kryteriów kwalifikacyjnych.

W Polsce w ramach naboru na studia drugiego stopnia zwykle zastosuje się jedno kryterium kwalifikacyjne: średnią ocen z toku studiów pierwszego stopnia lub wyniki testowego egzaminu wstępnego zbudowanego na bazie pytań wielokrotnego wyboru (MCQs, *Multiple-Choice Questions*). Obecnie, spośród dwunastu publicznych uniwersytetów jedynie cztery uniwersytety medyczne stosują w ramach rekrutacji na studia drugiego stopnia na kierunek położnictwo niestandardowy testowy egzamin wstępny. Ponadto, WUM od roku akademickiego 2009/10 w kwalifikacji kandydatów na ten kierunek studiów, jako jedyna uczelnia, stosuje trzy kryteria rekrutacyjne: średnią ocen, wyniki egzaminu testowego oraz dodatkową punktację za działalność naukową (autorstwo publikacji, udział w konferencjach i kołach naukowych) [7]. Testowy egzamin, jako bardziej obiektywne narzędzie, zastąpił stoso-

waną wcześniej kierowaną rozmową kwalifikacyjną. Obecnie 75% możliwej do uzyskania przez kandydata punktacji liczonej w rankingu rekrutacyjnym stanowią wyniki testu egzaminacyjnego. Duży wpływ tego kryterium kwalifikacyjnego na końcową punktację oznacza, że jego wartość dyskryminacyjna jest obecnie decydująca przy wyborze grupy najlepszych kandydatów. W związku z tym konieczne jest doskonalenie jakości egzaminu testowego, szczególnie w sytuacji braku standaryzowanego narzędzia służącego do oceny kompetencji wejściowych kandydatów na studia.

CEL

Ocena jakości testowego egzaminu wstępnego służącego do kwalifikacji kandydatów na studia stacjonarne drugiego stopnia na kierunek położnictwo na Wydziale Nauki o Zdrowiu WUM.

MATERIAŁ I METODY

Na potrzeby badania retrospektywnego zgromadzono dane kwalifikacyjne kandydatów (N = 335, 100% kobiet), którzy przystąpili do egzaminu wstępnego na studia stacjonarne drugiego stopnia na kierunek położnictwo na Wydziale Nauki o Zdrowiu WUM w latach 2009/10 – 2013/14. Dane dotyczące wieku oraz ukończonej uczelni na studiach pierwszego stopnia były pozyskane na podstawie informacji złożonych w formularzach aplikacyjnych wypełnianych przez kandydatów na studia. Charakterystykę grupy kandydatów przedstawiono w Tabeli 1.

■ Tab. 1. Charakterystyka grupy kandydatów, którzy przystąpili do egzaminu wstępnego na studia stacjonarne drugiego stopnia na kierunek położnictwo na Warszawskim Uniwersytecie Medycznym w latach 2009/10 – 2013/14.

Rok rekrutacji	N	Wiek		Uczelnia na studiach I stopnia		
		Średnia	SD	WUM MUW	Inny uniwersytet	Wyższa szkoła zawodowa
2009/10	62	22,8	2,77	56	3	3
2010/11	57	23,6	4,12	52	4	1
2011/12	62	22,8	2,75	54	5	3
2012/13	65	23,3	3,72	51	13	1
2013/14	89	23,5	3,94	76	13	0
łącznie	335	23,2	3,53	289	38	8

SD – odchylenie standardowe; WUM/MUW – Warszawski Uniwersytet Medyczny

Zgodnie ze stanowiskiem Komisji Bioetycznej WUM, badania retrospektywne, ankietowe i inne działania nieinwazyjne nie wymagają zgody powyższej Komisji¹. Autorzy pracy uzyskali zgodę Lokalnego Administratora Ochrony Danych Osobowych na przetwarzanie danych osobowych kandydatów na studia prowadzone przez WUM.

Dane dotyczące kwalifikacji na studia drugiego stopnia były pobierane z Uczelnianego Systemu Rekrutacyjnego i dotyczyły szczegółowych wyników uzyskanych przez

1 Szczegółowe informacje oraz wzory dokumentów Komisji Bioetycznej Warszawskiego Uniwersytetu Medycznego dostępne pod adresem: <https://komisja-bioetyczna.wum.edu.pl/content/szczegolowe-informacje-oraz-wzory-dokumentow> (data dostępu: 12.11.2015).

kandydata z testowego egzaminu wstępnego w podziale na pięć kategorii tematycznych wyodrębnionych zgodnie z planem testu (odsetek pytań w teście): nauki podstawowe (6%), nauki społeczne (11%), nauki w zakresie podstaw opieki położniczej (13%), nauki w zakresie opieki specjalistycznej (12%), inne zagadnienia (8%).

Dane surowe zostały wstępnie opracowane w programie TESTY wersja 7 (Testy komputerowe, Copyright © 1994-2014 by Sławomir Zalewski). Wczytane do bazy danych informacje po przetworzeniu w programie Microsoft Excel 2013 (Microsoft Corporation) zostały wyeksportowane do pakietu statystycznego STATISTICA wersja 12.5 (StatSoft, Inc.), gdzie zostały poddane dalszej analizie. Wszystkie programy użyto zgodnie z licencją WUM.

Oszacowano parametry statystyki opisowej oraz oceniono zgodność rozkładu poszczególnych wyników egzaminu z rozkładem normalnym (test W Shapiro-Wilka przy domyślnym poziomie istotności $\alpha = 0,05$). W celu ewaluacji jakości egzaminu testowego wyznaczono poziom łatwości oraz moc różnicującą pytań egzaminacyjnych [8]. Ponadto wykorzystano trzy metody szacowania rzetelności zestawów testowych: (1) analizę właściwości statystycznych pozycji testowych (ocena zgodności wewnętrznej testu, współczynnik alfa-Cronbacha), (2) analizę korelacji między dwoma równoległymi połówkami (rzetelność połówkowa, współczynnik Guttmana) oraz (3) analizę związku pozycji testowych z ogólnym wynikiem dla danego bloku tematycznego testu (moc dyskryminacyjna) [9].

WYNIKI

Wyniki analizy dla poszczególnych edycji egzaminu wstępnego pod kątem charakteru rozkładu tej zmiennej, wskazywały na symetryczny (niewielka lewoskośność) i zbliżony od rozkładu normalnego zestaw danych (test W Shapiro-Wilka, $p > 0,05$). Zmienność wyników uzyskanych przez kandydatów w kolejnych latach była podobna, co odzwierciedlały wartości współczynników zmienności oraz zakres punktacji. Wąskie przedziały ufności dla średnich świadczyły o dużej precyzji szacowanych parametrów dla badanej grupy. Punkt odcięcia przyjęty/nieprzyjęty był zlokalizowany w centrum rozkładu na poziomie wartości średnich wyników. Szczegóły dotyczące wyliczonych parametrów statystyki opisowej zostały przedstawione w Tabeli 2.

Oceniając jakość poszczególnych kategorii łatwości pytań testowych zaobserwowano pewne różnice w poszczególnych wersjach egzaminu w kolejnych latach. Najbardziej optymalny pod względem doboru pytań był test zastosowany w roku 2009/10. Średnia łatwość pytań testowych wyniosła 0,5 przy medianie 0,4, ponadto zakres łatwości rozciągał się w tym przypadku w całej skali (0,0; 1,0>). Najmniej korzystny dobór pytań pod względem łatwości zanotowano w przypadku naboru w roku 2012/13. Szczegóły dotyczące analizy łatwości poszczególnych zestawów testowych przedstawiono na Rycinie 1.

Pod względem charakterystyki parametrów mocy różnicującej, poszczególne zestawy egzaminacyjne również różniły się między sobą. Podobnie jak w przypadku poziomu trudności także w zakresie różnicowania najle-

■ Tab. 2. Parametry statystyki opisowej dla wyników egzaminu uzyskanych przez kandydatów podczas naboru na studia na kierunku położnictwo na Warszawskim Uniwersytecie Medycznym w latach 2009/10 – 2013/14

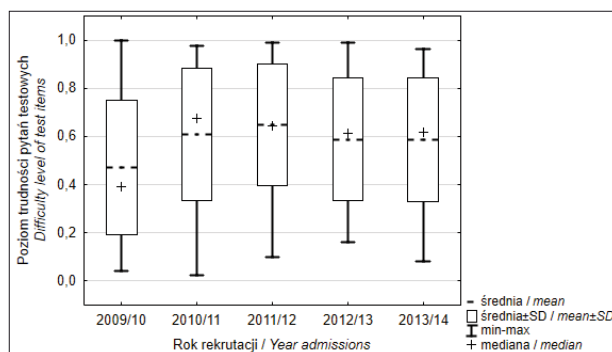
Parametry statystyki opisowej	Rok rekrutacji				
	2009/10	2010/11	2011/12	2012/13	2013/14
Procent przyjętych	69,4%	56,1%	53,2%	50,0%	42,7%
Próg punktowy	23,0	30,0	32,0	29,0	30,0
Rozkład normalny*	$p > 0,05$	$p > 0,05$	$p > 0,05$	$p > 0,05$	$p > 0,05$
Średnia	24,7	32,4	30,5	29,3	29,7
-95% CI	23,8	31,4	29,4	28,2	28,9
+95% CI	25,7	33,4	31,6	30,5	30,6
SD	3,90	3,93	4,21	4,46	4,20
Mediana	25,0	32,0	31,0	29,0	30,0
Minimum	16,0	22,0	20,0	21,0	19,0
Maksimum	32,0	42,0	40,0	40,0	40,0
CV	15,8%	12,1%	13,8%	15,2%	14,1%

* test Shapiro-Wilka (rozkładu normalnego dla $p > 0,05$)

CI – 95% przedział ufności dla średniej

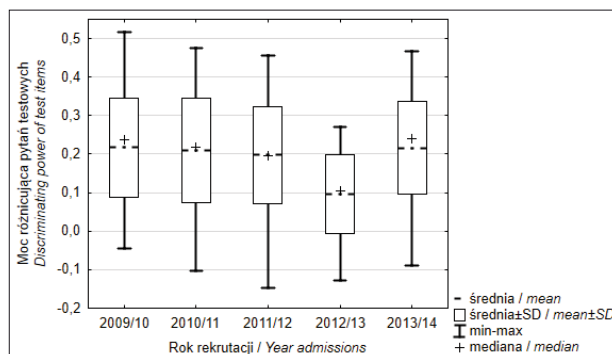
SD – odchylenie standardowe

CV – współczynnik zmienności



■ Ryc. 1. Porównanie poziomu trudności pytań tworzących test egzaminacyjny w kolejnych latach rekrutacji na kierunku położnictwo na Warszawskim Uniwersytecie Medycznym

piej dobrane zostały pytania w egzaminie w roku 2009/10 a najgorzej 2012/13. Ponadto, w każdej edycji egzaminu testowego zanotowano obecność pytań z ujemną mocą różnicującą, przy czym najwięcej pytań o niewłaściwej konstrukcji zaobserwowano w roku 2012/13. Szczegóły dotyczące analizy mocy różnicującej poszczególnych zestawów testowych przedstawiono na Rycinie 2.



■ Ryc. 2. Porównanie mocy różnicującej pytań tworzących test egzaminacyjny w kolejnych latach rekrutacji na kierunku położnictwo na Warszawskim Uniwersytecie Medycznym

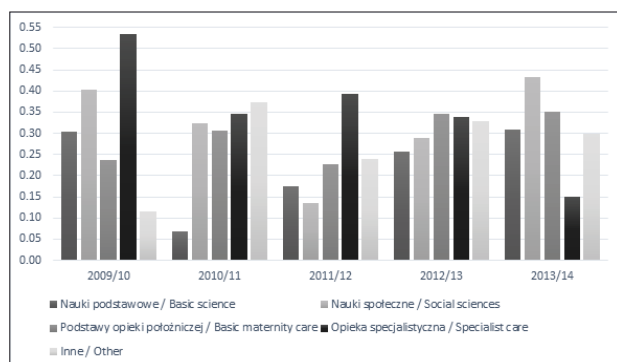
Wyniki analizy rzetelności poszczególnych testów wskazują na niedostateczną jakość wszystkich egzaminów.

Dla żadnej edycji nie stwierdzono spełnienia minimalnych wymagań w zakresie wewnętrznej zgodności pomiaru. Najlepsze wyniki w tym zakresie zanotowano dla testu w rekrutacji z roku 2009/10. Szczegółowe zestawienie wyników analiz rzetelności poszczególnych edycji egzaminu wstępnego przedstawiono w Tabeli 3.

■ Tab. 3. Wyniki analizy rzetelności poszczególnych edycji wstępnego egzaminu testowego

Rok rekrutacji	Zgodność wewnętrzna (alfa-Cronbacha)	Rzetelność połówkowa (współczynnik Guttmana)	Korelacja między połówkami
2009/10	0,595	0,728	0,572
2010/11	0,572	0,579	0,416
2011/12	0,495	0,530	0,361
2012/13	0,569	0,520	0,354
2013/14	0,586	0,585	0,414

Analizując moc dyskryminacyjną pytań w pięciu wyodrębnianych blokach tematycznych stwierdzono zróżnicowanie tego parametru jakościowego w kolejnych latach. Jak przedstawiono na Rycinie 3, bardzo wysoki indeks dyskryminacyjny zanotowano dla bloku pytań z zakresu *opieki specjalistycznej* w roku 2009/10 (0,535), a najniższy w egzaminie w roku 2013/14 (0,150). Dodatkowo, łącznie dla kategorii *nauki w zakresie podstaw opieki położniczej i nauki w zakresie opieki specjalistycznej*, które stanowiły połowę całkowitej liczby pytań w każdym teście, indeksy dyskryminacyjne zawsze przekraczały wartość 0,300.



■ Ryc. 3. Porównanie wartości indeksów dyskryminacyjnych pytań testowych w poszczególnych blokach tematycznych wyodrębnionych zgodnie z planem testu

OMÓWIENIE

Wśród dostępnych w światowej literaturze publikacji dotyczących naboru kandydatów na studia drugiego stopnia (*master's/postgraduate level*) brak jest takich, które odnosiłyby się ściśle do rekrutacji na kierunek położnictwo. Opublikowano jednak wyniki amerykańskich badań, które są związane z problematyką przyjęć na kierunek pielęgniarstwo [10-13]. Dotyczą one w głównej mierze oceny trafności predykcyjnej dwóch kryteriów kwalifikacyjnych: średniej ocen (GPA, *grade point average*) wyliczonej z okresu kształcenia dyplomowego (*undergraduate/baccalaureate nursing*) oraz wyników standaryzowanego wielomodułowego testu GRE (*Graduate Record*

Examinations) podsumowującego studia pierwszego stopnia [14]. Stosowany w USA przez wiele uczelni wyższych test GRE jest prowadzony metodą komputerową (*computer-based test*) i składa się z trzech podstawowych domen oceniających różne kompetencje absolwenta studiów dyplomowych: *analytical writing* (GRE-AW), *verbal reasoning* (GRE-V), *quantitative reasoning* (GRE-Q). Mimo pozytywnych danych dotyczących testu GRE [11], istnieją także krytyczne oceny, które podważają sensowność stosowania tego typu egzaminu w procesie naboru na studia drugiego stopnia na kierunek pielęgniarstwo [13]. Należy podkreślić, że ponieważ w Polsce nie ma tradycji stosowania standaryzowanych testów w procesie naboru na studia drugiego stopnia, trudno jest odnosić wyniki Autorów niniejszego opracowania do ustaleń uzyskanych przez cytowanych powyżej badaczy amerykańskich. Jednakże, opublikowane wcześniej przez Panczyk i wsp. wyniki analiz dotyczących rzetelności i trafności egzaminu wstępnego na studia drugiego stopnia na kierunek pielęgniarstwo na WUM, pozwalają zakładać że w polskich warunkach może być to dobre narzędzie selekcyjne o dostatecznych właściwościach różnicujących i predykcyjnych [15-17].

Ważną właściwością dobrych narzędzi selekcyjnych jest ich rzetelność, rozumiana jako powtarzalność uzyskanych wyników w określonych warunkach [9,18]. Niedostateczna rzetelność stosowanej procedury kwalifikacyjnej przyczynia się do niskiego stopnia zaufania, skoro w podobnych okolicznościach poszczególne wyniki różnią się znacząco między sobą. Jak pokazują przedstawione wyniki badania rzetelności za pomocą współczynnika alfa-Cronbacha, wewnętrzna zgodność pomiaru we wszystkich analizowanych rocznikach nie była zadowalająca ($\alpha < 0,600$). Zgodnie z powszechnie przyjętymi kryteriami dobre wyniki selekcji, czyli ocenę indywidualnych różnic między zdającymi można uzyskać, kiedy pomiar ma rzetelność większą niż 0,800 [19]. Przy wartości $\alpha=0,500$ błędy losowe stanowią aż połowę zmienności uzyskanych wyników, a pomiar w takich warunkach może być stosowany właściwie jedynie przy porównywaniu międzygrupowym, a nie do różnicowania indywidualnego [20]. Także uwzględniając ocenę rzetelności za pomocą współczynnika Guttmana, w której mierzy się korelację między dwoma połówkami testu, uzyskano wartości mniejsze niż 0,600 (wyjątek rocznik 2009/10). Należy podkreślić, że osiągnięcie wysokich wartości współczynnika rzetelności, większych niż 0,900 jest ważne w przypadku konieczności rozróżnień dokonywanych wewnątrz ogólnego wyniku pomiaru, np. kiedy chcemy porównać osiągnięcia zdających w różnych zakresach tematycznych, co może mieć znaczenie przy ocenie szczególnych kompetencji oczekiwanych u kandydatów na studia.

Ponieważ ważnym czynnikiem, który ma wpływ na rzetelność pomiaru za pomocą testu jest jego prawidłowa konstrukcja, to istotna jest ocena jakości poszczególnych pytań tworzących pulę egzaminacyjną [21]. Porównując różne edycje egzaminu zaobserwowano zróżnicowany poziom ogólnej trudności od stosunkowo łatwego testu w 2011/12 roku do średniotrudnego w 2009/10 (średnia łatwość odpowiednio 0,65 i 0,47). Także pod względem

mocy różnicującej poszczególne egzaminy nie były jednorodne. Duża liczba pytań łatwych i aprobujących, czyli takich, które wymagają od zdającego jedynie potwierdzenia poprawności podanej informacji (szczególnie używanie opcji typu: „wszystkie z powyższych” lub „żaden z powyższych”) są intelektualnie mniej wymagające i przyczyniają się do uzyskania przez kandydatów przeciętnie wyższych wyników z takiego egzaminu [22,23]. Istotne jest więc dobranie takich pytań egzaminacyjnych, które wykazują znaczną moc różnicującą. Pytania takie mają zdolność odróżniania kandydatów, którzy dla danej mierzonej wartości uzyskują wyniki istotnie różne, co pozwala na dobrą selekcję. W literaturze pytania takie określane są jako zadania „rozrywające” („burst”), ponieważ powodują one spłaszczenie rozkładu wyników, a to znacznie ułatwia dyskryminację i ustalenie punktu odcięcia przyjęty/nieprzyjęty [24].

W przypadku każdego egzaminu wstępnego, który ma służyć ocenie kluczowych kompetencji istotnych dla efektywnego podjęcia studiów drugiego stopnia jest analiza zgodności treści pytań egzaminacyjnych ze standardami kształcenia w zakresie położnictwa [25]. Walidacja tego parametru wymagało oceny planu testu. Analizując założenia programowe dla studiów drugiego stopnia określono pewne istotne obszary wiedzy i umiejętności niezbędne do efektywnego podjęcia studiów na tym kierunku. Wydaje się że wyodrębnione obszary tematyczne wraz z przypisanymi im procentowymi udziałami w puli testowej są *dobrym prerekwizytem dla efektywnego studiowania i mogą one stanowić w dalszej perspektywie czasowej podstawę do badań trafności predykcyjnej* [6].

Ponieważ rzetelność i trafność to dwie hierarchicznie powiązane ze sobą cechy, to zawsze dla danego narzędzia pomiarowego należy obie właściwości rozpatrywać łącznie. Dobra trafność egzaminu testowego, oznacza że w większości przypadków wyniki uzyskiwane przez kandydatów odzwierciedlają rzeczywiste właściwości i cechy zdających. Natomiast dostateczna rzetelność pomiaru odpowiada za dokładność oceny wiedzy i umiejętności zdających, a co za tym idzie dobrego różnicowania indywidualnego kandydatów. Ponadto, lepsza precyzja w ocenie kompetencji tych kandydatów dla których wynik z testu egzaminacyjnego lokuje się w pobliżu punktu odcięcia przyjęty/nieprzyjęty w mniejszym stopniu zależy od błędów przypadkowych pomiaru. Ponieważ grupa tych kandydatów jest obciążona największym niepowodzeniem w trakcie studiowania, należy zwrócić szczególną uwagę na prawidłowy pod względem jakości dobór pytań testowych stanowiących pulę zadań egzaminacyjnych. Tylko dobra konstrukcja testu może zapewnić odpowiedni poziom rzetelności egzaminu wstępnego w kolejnych rocznikach rekrutacji.

PODSUMOWANIE

1. Konieczna jest poprawa jakości testowych pytań egzaminacyjnych w zakresie trudności oraz mocy różnicującej.
2. Należy systematycznie rozbudowywać bazę pytań testowych oraz monitorować jakość kolejnych edycji

egzaminu wstępnego w szczególności w odniesieniu do kontroli rzetelności tego narzędzia kwalifikacyjnego.

3. W klejonym etapie ewaluacji należy dokonać analizy dyskryminacyjnej w celu oceny parametrów różnicujących testu egzaminacyjnego.

PIŚMIENNICTWO

1. Drennan J. Professional and academic destination of masters in nursing graduates: a national survey. *Nurse. Educ. Today.* 2008; 28(6): 751-9.
2. Raines CF, Tagliaireni ME. Career pathways in nursing: Entry points and academic progression. *Online. J. Issues. Nurs.* 2008; 13(3): 3.
3. Kostrzanowska Z, Zarzycka D. Struktura i zasoby pielęgniarstwa europejskiego. [w:] Wrońska I, Krajewska-Kulak E, red. *Wybrane zagadnienia z pielęgniarstwa europejskiego.* Lublin: Wydawnictwo Czelej; 2007, s. 27-47.
4. Muecke L. Pre- and post-admission criteria as predictors of academic success in an associate degree nursing program. *Ames: Iowa State University;* 2008.
5. Kuncel NR, Hezlett SA, Ones DS. A comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. *Psychol. Bull.* 2001; 127(1): 162-81.
6. Panczyk M, Zarzeka A, Belowska J, et al. Risk of attrition from Bachelor's degree of Nursing programme – predictive evaluation. *Piel.* XXI. 2015; 51(2): 66-72.
7. Uchwała nr 13/2008 z dnia 25 lutego 2008 roku w sprawie uchwalenia zasad i trybu postępowania rekrutacyjnego na pierwszy rok studiów na rok akademicki 2009/2010 w Akademii Medycznej w Warszawie.
8. Niemierko B. *Pomiar wyników kształcenia.* Warszawa: Wydawnictwo Szkolne i Pedagogiczne; 1999.
9. Jankowski K, Zajenkowski M. *Metody szacowania rzetelności pomiaru testem.* [w:] Fronczyk K, red. *Psychometria - podstawowe zagadnienia.* Warszawa: Vizja Press & IT; 2009, s. 84-110.
10. Rhodes ML, Bullough B, Fulton J. The Graduate Record Examination as an admission requirement for the graduate nursing program. *J. Prof. Nurs.* 1994; 10(5): 289-96.
11. Auerhahn C. *Predictors of success in master's level nurse practitioner programs.* New York: Columbia University; 1996.
12. Newton SE, Moore G. Undergraduate grade point average and graduate record examination scores: the experience of one graduate nursing program. *Nurs. Educ. Perspect.* 2007; 28(6): 327-31.
13. Katz JR, Chow C, Motzer SA, Woods SL. The graduate record examination: help or hindrance in nursing graduate school admissions? *J. Prof. Nurs.* 2009; 25(6): 369-72.
14. Megginson L. Exploration of nursing doctoral admissions and performance outcomes. *J. Nurs. Educ.* 2011; 50(9): 502-12.
15. Panczyk M, Rebandel H, Belowska J, et al. Entrance exam for candidates for master's nursing studies – a 5-year retrospective analysis. *Prob. Piel.* 2015; 23(3): in press.
16. Panczyk M, Gotlib J. Analiza dyskryminacyjna kryteriów rekrutacyjnych na studia II stopnia na kierunku Pielęgniarstwo na Warszawskim Uniwersytecie Medycznym. *Prob. Piel.* 2015; 23(1): 51-6.
17. Panczyk M, Ściegłińska B, Woynarowska-Sołdan M, Gotlib J. MSc Nursing Entrance Exams - admission to full-time and part-time programmes at medical university. *Prob. Piel.* 2015; 22(4): 477-4483.
18. Fleiss JL. *Reliability of Measurement. The Design and Analysis of Clinical Experiments.* New York: John Wiley & Sons; 1999.
19. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J. Pers. Assess.* 2003; 80(1): 99-103.
20. Guilford JP. *Psychometric methods.* 2nd ed. New York: McGraw-Hill; 1954.
21. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. *Med. Educ.* 2008; 42(2): 198-206.
22. Harasym PH, Leong EJ, Violato C, et al. Cuing effect of „all of the above” on the reliability and validity of multiple-choice test items. *Eval. Health. Prof.* 1998; 21(1): 120-33.
23. Boland RJ, Lester NA, Williams E. *Writing Multiple-Choice Questions.* *Academic Psychiatry.* 2010; 34(4): 310-6.
24. Downing SM. 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(2): 133-43.
25. Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 9 maja 2012 r. w sprawie standardów kształcenia dla kierunków studiów: lekarskiego, lekarsko-dentystycznego, farmacji, pielęgniarstwa i położnictwa (Dz.U. 2012 nr 0 poz. 631).

Praca przyjęta do druku: 16.11.2015

Praca zaakceptowana do druku: 31.12.2015