

# Ocena osiągniętych efektów kształcenia przez studentów kierunku Położnictwo na studiach niestacjonarnych II stopnia

Evaluation of learning outcomes achieved by Midwifery students at part-time studies of the second degree

Mariusz Panczyk, Jarosława Belowska, Aleksander Zarzeka, Joanna Gotlib

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

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. +48-225-720490, fax. +48-225-720491  
E-mail: mariusz.panczyk@wum.edu.pl

## STRESZCZENIE

### OCENA OSIĄGANYCH EFEKTÓW KSZTAŁCENIA PRZEZ STUDENTÓW KIERUNKU POŁOŻNICTWO NA STUDIACH NIESTACJONARNYCH II STOPNIA

**Cel.** Analiza spójności w ocenie osiągnięć studentów z przedmiotów kończących się egzaminem i objętych programem kształcenia na kierunku Położnictwo na studiach niestacjonarnych II stopnia w Warszawskim Uniwersytecie Medycznym (WUM).

**Materiał i metody.** Dane egzaminacyjne 231 studentów kierunku Położnictwo, którzy podjęli studia niestacjonarne II stopnia w WUM w latach 2007-2012. Analizie retrospektywnej poddano wyniki kształcenia z ośmiu przedmiotów, które kończyły się egzaminem. Do określenia trendów w ocenianiu studentów w kolejnych latach użyto nieparametrycznego testu ANOVA rang Kruskala-Wallisa. Szacowanie zgodności oceniania dla kolejnych przedmiotów egzaminacyjnych określono za pomocą współczynnika Kendalla. Ocenę wewnętrznej trafności pomiaru edukacyjnego określano na podstawie wyników analizy interkorelacji oraz wielowymiarowej analizy regresji.

**Wyniki.** Analiza wewnętrznej spójności oceniania studentów w kolejnych latach wykazała, że dla zdecydowanej większości z badanych przedmiotów nie zaobserwowano jednolitości w ocenianiu osiągnięć studentów (test ANOVA rang Kruskala-Wallisa,  $p < 0,005$ ). Dla kolejnych zakresów kształcenia stwierdzono niedostateczny poziom zgodności pomiaru edukacyjnego (współczynnik Kendalla = 0,11). Oszacowanie trafności w analizie interkorelacji wskazuje, że dla ośmiu badanych przedmiotów istnieją pozytywne zależności między wynikami kształcenia studentów z poszczególnych obszarów kształcenia. Potwierdzenie powyższych zależności uzyskano także w analizie regresji.

**Wnioski.** Niski stopień zgodności oceniania osiągnięć studentów w kolejnych latach świadczy o niedostatecznej spójności systemu pomiaru efektów kształcenia. W przyszłości ważnym elementem ewaluacji systemu oceniania powinna być analiza prognostyczna, w której można by uwzględnić zmienne zależne związane z losami absolwentów.

## Słowa kluczowe:

położnictwo, pomiar edukacyjny, szkolnictwo wyższe

## ABSTRACT

### EVALUATION OF LEARNING OUTCOMES ACHIEVED BY MIDWIFERY STUDENTS AT PART-TIME STUDIES OF THE SECOND DEGREE

**Aim.** An analysis of consistence in evaluating students' achievements in subjects completed with an exam and included in the curriculum at the department of midwifery of part-time studies at Medical University of Warsaw (MUW).

**Materials and methods.** The authors used examination data of 231 subjects, part-time midwifery students of the second degree at MUW between the years 2007-2012. A retrospective analysis was used for analyzing students' achievements and learning results in eight subjects completed with an exam.

In order to determine the trends in evaluating students in consecutive years, a non-parametric ANOVA Kruskal-Wallis rank test was used. Assessing the subsequent compliance of evaluation of exam subjects was performed using Kendall coefficient. Evaluation of internal validity of educational measurement was established on the basis of inter-correlation analysis and multi-dimensional regression analysis.

**Results.** An analysis of internal compliance in evaluating students in subsequent years showed that for the vast majority of the studied subjects there was no consistency in evaluating students (Kruskal-Wallis rank ANOVA test,  $p < 0.005$ ). For subsequent ranges of education, an insufficient level of educational measurement was found (Kendall coefficient = 0.11). Evaluating accuracy in the analysis of inter-correlation shows that for the eight studied subjects there are positive dependencies between the results of teaching

students in individual areas of education. Confirmation of the above dependencies was also obtained in the analysis of regression. **Conclusions.** Low level of compliance in evaluating students in subsequent years proves that there is insufficient cohesion within the system of measuring the learning outcomes. Prognosis analysis which would include the dependent variables connected with the future of graduates should be an important element of educational system evaluation in the future.

**Key words:** midwifery, educational measurement, graduate education

## WPROWADZENIE

Funkcjonujący w Polsce system przewiduje konieczność podnoszenia kwalifikacji przez czynne zawodowo położne. Dlatego też, wszystkie aktywne zawodowo położne są zobowiązane do uczestniczenia w różnych formach kształcenia ustawicznego (podyplomowego) [1]. Zgodnie z koncepcją dwustopniowych studiów wyższych, kształcenie zakończone uzyskaniem dyplomu magistra ma w założeniu znacząco wpływać na podwyższenie kwalifikacji zawodowych absolwenta, co ma skutkować lepszym wykorzystaniem kompetencji specjalisty [2]. Ponadto, w większości krajów Europy Zachodniej i USA, zdobycie tytułu magistra jest postrzegane jako jeden z ważniejszych etapów rozwoju zawodowego, niezbędnego dla osiągnięcia profesjonalizmu w danej specjalności [3, 4]. Jest to także ważny moment w karierze prowadzącej do wyższych szczebli zawodowych, szczególnie w zakresie organizacji i zarządzania [4]. Ponadto, według *American Association of Colleges of Nursing* absolwenci studiów II stopnia są naturalnymi kandydatami na liderów zespołów [5].

W ramach studiów II stopnia na kierunku Położnictwo, proces kształcenia zorganizowany jest w formie wieloprzedmiotowych modułów łączonych, obejmujących opiekę specjalistyczną w położnictwie, neonatologii i ginekologii, poświęconych węższymi i bardziej specjalistycznym obszarom tematycznym niż w przypadku studiów I stopnia. Ponadto, studia II stopnia podnoszą jakość uzyskanej dotąd wiedzy praktycznej, wzbogacając ją o elementy z zakresu administrowania i nadzoru nad opieką położniczo-ginekologiczną i pielęgniarstwa. Zajęcia praktyczne, trwające nie mniej niż 160 godzin, realizują efekty kształcenia związane z naukami w zakresie opieki specjalistycznej oraz nauk społecznych. Po ukończeniu studiów II stopnia absolwenci uzyskują zaawansowaną wiedzę z zakresu położnictwa oraz posiadają umiejętności współpracy z członkami zespołu w zakresie doskonalenia jakości opieki, podnoszenia jakości opieki położniczo-ginekologicznej; wdrażania nowych metod, technik i rozwiązań organizacyjnych oraz wyznaczania obszarów podejmowania badań [6].

Standardy kształcenia dla kierunku Położnictwo zawierają wykaz niezbędnych efektów kształcenia, które student tego kierunku musi osiągnąć w trakcie studiów [6]. W oparciu o powyższe ramy, każda uczelnia prowadząca kształcenie na kierunku Położnictwo powinna zbudować odpowiedni program studiów wraz ze szczegółowym opisem efektów kształcenia zawartym w sylabusach. Niezbędnym elementem tak zaplanowanego kształcenia jest odpowiedni system oceny osiągnięć studentów, którego prawidłowe działanie jest pochodną dwóch czynników. Musi on sprawdzać efekty kształcenia dla przedmiotu, ustalone przez autora sylabusu, a po drugie powinien także wynikać ze standardów kształcenia dla

danego poziomu studiów, co oznacza wymóg budowania systemu oceniania tak, aby sprawdzał, czy student posiadał pożądane na danym poziomie kompetencje [7]. Jeżeli traktować studia II stopnia jako kolejny etap na drodze do profesjonalizmu w zawodzie, to wymagania i zakres oczekiwanych osiągnięć absolwenta z dyplomem magistra powinien być znacząco wyższy, niż to ma miejsce w przypadku studiów I stopnia [8]. Biorąc pod uwagę powyższe założenia należy stawiać wysokie wymagania w odniesieniu do jakości metod pomiaru edukacyjnego, które są stosowane w procesie oceny kompetencji studentów. Spójność systemu oceniania szacowana rzetelnością oraz trafnością pomiaru to dwa podstawowe czynniki określające jakość pomiaru edukacyjnego [9]. Dbłość o wysokie standardy kształcenia i efektywne nauczanie realizowane poprzez stałą kontrolę i ewaluację jest kluczowym zadaniem dla uczelni wyższej kształcącej w tak odpowiedzialnym zawodzie jak Położnictwo.

## CEL PRACY

Celem badań jest analiza spójności w ocenie osiągnięć studentów z przedmiotów kończących się egzaminem i objętych programem kształcenia na kierunku Położnictwo na studiach niestacjonarnych II stopnia w Warszawskim Uniwersytecie Medycznym (WUM).

## MATERIAŁ I METODY

Do badania zakwalifikowano dane egzaminacyjne 231 studentów kierunku Położnictwo, którzy podjęli studia niestacjonarne II stopnia w WUM w latach 2007-2012. Grupę badaną stanowiły wyłącznie kobiety ze średnią wieku wynoszącą  $33,1 \pm 9,36$  lat. Wszystkie analizowane roczniki były porównywalne pod względem liczebności oraz struktury wiekowej. Charakterystykę badanej grupy przedstawiono w Tabeli 1.

Tab. 1. Charakterystyka badanej grupy studentów studiów niestacjonarnych II stopnia kierunku Położnictwo w latach 2007-2012

Rocznik	N	Procent	p*	Średnia wieku	p**	SD	p***
2007/08	49	21,21	0,096	34,0	0,162	8,31	0,165
2008/09	49	21,21		34,4		9,26	
2009/10	35	15,15		33,8		9,25	
2010/11	36	15,58		30,8		9,37	
2011/12	33	14,30		33,7		9,97	
2012/13	29	12,55		30,2		10,48	
Ogółem	231	100,00	----	33,1	----	9,36	----

\* test zgodności  $\chi^2 = 9,338$

\*\* nieparametryczny test ANOVA rang Kruskala-Wallis H = 7,900

\*\*\* test jednorodności wariancji Levena F = 1,586

Analizie retrospektywnej poddano wyniki kształcenia z ośmiu przedmiotów, które kończyły się egzaminem: (I) Kształcenie w zakresie diagnostyki ultrasonograficznej w położnictwie i ginekologii; (II) Kształcenie w zakresie dydaktyki medycznej; (III) Kształcenie w zakresie nowoczesnych technik diagnostycznych; (IV) Kształcenie w zakresie opieki specjalistycznej w położnictwie, neonatologii i ginekologii; (V) Kształcenie w zakresie wybranych zagadnień z pielęgniarstwa specjalistycznego; (VI) Kształcenie w zakresie geriatry; (VII) Kształcenie w zakresie intensywnego nadzoru położniczego; (VIII) Kształcenie w zakresie onkologii ginekologicznej. Zakres treści programowych oraz zakładane efekty kształcenia były zgodne z wymogami aktualnie obowiązujących *Standardów kształcenia dla kierunku Położnictwo*.

W celu określenia trendów w ocenianiu studentów w kolejnych latach użyto nieparametrycznego testu ANOVA rang Kruskala-Wallisa oraz testu homogenności wariancji Levena. Natomiast do oszacowania zgodności oceniania poszczególnych studentów dla kolejnych przedmiotów egzaminacyjnych wykorzystano nieparametryczny test ANOVA Friedmana dla porównania zmiennych zależnych ze współczynnikiem zgodności Kendalla. Analizę rzetelności oceniania określono za pomocą współczynnika  $\alpha$ -Cronbacha (współczynnik Kudara-Richardsona dla testu składającego się z pozycji dwukategorialnych) [10]. Ocenę interkorelacji dla ośmiu analizowanych przedmiotów przeprowadzono z użyciem nieparametrycznego współczynnika korelacji rang Spearmana. Wewnętrzna zgodność oceniania została także oszacowana z wykorzystaniem metody regresji wielorakiej dla dziewięciu zmiennych niezależnych (objaśniających): ocena dla ośmiu przedmiotów kończących się egzaminem oraz zmienna dychotomiczna dla cyklu kształcenia według starego i nowego układu przedmiotów (odpowiednio roczniki 2007/08-2008/2009 oraz 2009/10-2012/13). Oceniono parametry funkcji regresji wraz z oceną błędów standardowych oraz wyznaczono współczynnik standaryzowany  $\beta$  w celu określenia stopnia wpływu zmiennych niezależnych na średnią ocen ze studiów II stopnia (zmienna zależna, objaśniana) [11].

Do obliczeń wykorzystano pakiet statystyczny STATISTICA wersja 12 (StatSoft, Inc.) zgodnie z licencją WUM. Dla wszystkich analiz jako domyślny (a priori) poziom istotności przyjęto  $\alpha = 0,05$ .

## WYNIKI

Analiza wewnętrznej spójności oceniania studentów w kolejnych latach wykazała, że dla zdecydowanej większości badanych przedmiotów nie zaobserwowano jednolitości w ocenianiu osiągnięć studentów. Wysoki stopień zróżnicowania ocen studentów przejawiający się brakiem jednorodności wariancji wykazano dla sześciu przedmiotów (test Levena,  $p < 0,05$ ). Ponadto, zaobserwowano statystycznie istotną różnicę w ocenianiu studentów dla siedmiu przedmiotów w następujących po sobie rocznikach (test ANOVA rang Kruskala-Wallisa,  $p < 0,005$ ). Zestawienie wyników analizy stopnia zróżnicowania oceniania studentów przedstawiono w Tabeli 2.

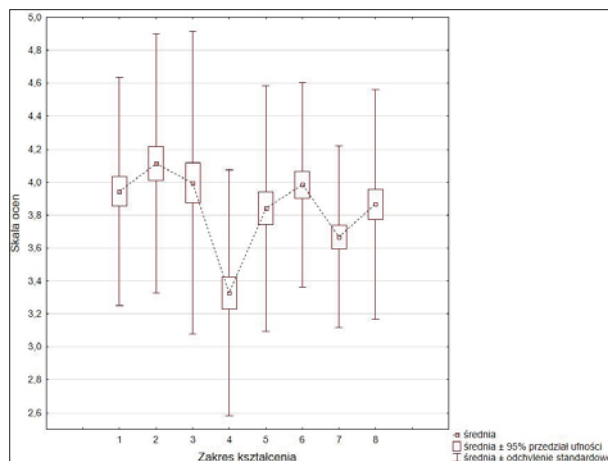
■ Tab. 2. Wyniki analizy stopnia zróżnicowania oceniania studentów z poszczególnych przedmiotów kończących się egzaminem na kierunku Położnictwo w latach 2007-2012

Zakres kształcenia	Średnia ocen	p*	SD	p**
1. Kształcenie w zakresie diagnostyki ultrasonograficznej w położnictwie i ginekologii	4,0	< 0,00001	0,673	< 0,00001
2. Kształcenie w zakresie dydaktyki medycznej	4,1	< 0,00001	0,788	0,02
3. Kształcenie w zakresie nowoczesnych technik diagnostycznych	4,0	< 0,00001	0,891	< 0,00001
4. Kształcenie w zakresie opieki specjalistycznej w położnictwie, neonatologii i ginekologii	3,4	< 0,00001	0,740	< 0,00001
5. Kształcenie w zakresie wybranych zagadnień z pielęgniarstwa specjalistycznego	3,9	< 0,00001	0,735	0,01
6. Kształcenie w zakresie geriatry	4,0	0,45	0,611	< 0,00001
7. Kształcenie w zakresie intensywnego nadzoru położniczego	3,7	< 0,00001	0,555	0,71
8. Kształcenie w zakresie onkologii ginekologicznej	3,9	0,003	0,687	0,75

\* poziom istotności dla nieparametrycznego testu ANOVA rang Kruskala-Wallisa

\*\* poziom istotności dla testu jednorodności wariancji Levena

Łączna analiza porównawcza średnich ocen uzyskanych przez studentów w latach 2007-2012 z poszczególnych przedmiotów wskazuje, że najgorzej ocenianym obszarem było *Kształcenie w zakresie opieki specjalistycznej w położnictwie, neonatologii i ginekologii* (średnia ocen  $3,4 \pm 0,74$ ), natomiast najlepiej *Kształcenie w zakresie dydaktyki medycznej* (średnia ocen  $4,1 \pm 0,79$ ). Dla pozostałych przedmiotów średnie ocen studentów mieściły się w przedziale od 3,7 do 4,0.



- 1 – Kształcenie w zakresie diagnostyki ultrasonograficznej w położnictwie i ginekologii;
- 2 – Kształcenie w zakresie dydaktyki medycznej;
- 3 – Kształcenie w zakresie nowoczesnych technik diagnostycznych;
- 4 – Kształcenie w zakresie opieki specjalistycznej w położnictwie, neonatologii i ginekologii;
- 5 – Kształcenie w zakresie wybranych zagadnień z pielęgniarstwa specjalistycznego;
- 6 – Kształcenie w zakresie geriatry;
- 7 – Kształcenie w zakresie intensywnego nadzoru położniczego;
- 8 – Kształcenie w zakresie onkologii ginekologicznej.

■ Ryc. 1. Struktura oceniania studentów z poszczególnych przedmiotów kończących się egzaminem na kierunku Położnictwo w latach 2007-2012 (wartość statystyki ANOVA Friedmana  $\chi^2(7,218) = 172,697$ ;  $p < 0,000001$ ; współczynnik zgodności Kendalla = 0,11)

Porównując oceny uzyskane przez danego studenta z kolejnych przedmiotów obejmujących kształcenie w zakresie Położnictwa stwierdzono, że zgodność pomiaru edukacyjnego kształtuje się na niskim poziomie (współczynnik zgodności Kendalla = 0,11). Ponadto, analiza osiągnięć studentów z różnych przedmiotów, wykonana za pomocą nieparametrycznego testu ANOVA Friedmana (porównanie dla ośmiu grup powiązanych) wskazuje na istotne statycznie odchylenie w spójność oceniania ( $p < 0,000001$ ). Najczęściej stosowanym sposobem oceny wewnętrznej zgodności wyników pomiaru dla co najmniej dwóch czynników jest współczynnik rzetelności  $\alpha$ -Cronbacha. Analiza wykazała średni poziom rzetelności oceniania dla wszystkich przedmiotów – współczynnik  $\alpha = 0,65$ , przy czym wskaźnik rzetelność w kolejnych latach oscylował w granicach od 0,60 do 0,75. Stopień zróżnicowania osiągnięć studentów mierzonych średnią ocen dla przedmiotów objętych kształceniem na kierunku Położnictwo w latach 2007-12 przedstawiono na Rycinie 1.

Oszacowanie trafności w analizie interkorelacji wskazuje, że dla ośmiu badanych przedmiotów istnieją pozytywne zależności między wynikami kształcenia studentów z poszczególnych obszarów kształcenia w Położnictwie. Współczynnik korelacji rang Spearmana kształtował się w przedziale od 0,13 do 0,53. Jedynie w przypadku przedmiotu *Kształcenie w zakresie intensywnego nadzoru położniczego* wykazano brak istotnie statystycznych interkorelacji z wynikami studentów uzyskanymi z czterech przedmiotów. Z kolei najsilniejszą korelację zaobserwowano dla pary przedmiotów *Kształcenie w zakresie diagnostyki ultrasonograficznej w położnictwie i ginekologii* – *Kształcenie w zakresie nowoczesnych technik diagnostycznych* ( $r = 0,53$ ). Zestawienie wyników analizy korelacyjnej przedstawiono w Tabeli 3.

■ Tab. 3. Wyniki analizy interkorelacji dla ocen studentów z poszczególnych przedmiotów kończących się egzaminem na kierunku Położnictwo w latach 2007-2012. Wszystkie współczynniki korelacji istotne ( $p < 0,05$ ) z wyjątkiem oznaczonych (\*)

Zakres kształcenia	Korelacja porządku rang Spearmana							
	1	2	3	4	5	6	7	8
1. Kształcenie w zakresie diagnostyki ultrasonograficznej w położnictwie i ginekologii	1,00	0,35	0,53	0,07*	0,22	0,15	0,07*	0,17
2. Kształcenie w zakresie dydaktyki medycznej		1,00	0,30	0,11*	0,18	0,22	0,22	0,20
3. Kształcenie w zakresie nowoczesnych technik diagnostycznych			1,00	0,16	0,28	0,17	0,08*	0,07*
4. Kształcenie w zakresie opieki specjalistycznej w położnictwie, neonatologii i ginekologii				1,00	0,13	0,15	0,23	0,20
5. Kształcenie w zakresie wybranych zagadnień z pielęgniarstwa specjalistycznego					1,00	0,17	0,01*	0,27
6. Kształcenie w zakresie geriatrii						1,00	0,11*	0,25
7. Kształcenie w zakresie intensywnego nadzoru położniczego							1,00	0,25
8. Kształcenie w zakresie onkologii ginekologicznej								1,00

W celu dokładnego oszacowania trafności wewnętrznej przeprowadzono analizę regresji z użyciem dziewięciu zmiennych niezależnych. W modelu regresji uwzględniono zmienną dychotomiczną odnoszącą się do zmiany układu przedmiotów jaki nastąpił na przełomie roczników 2008/09 a 2009/10. Jak przedstawiono w Tabeli 4 zaproponowane równanie funkcji regresji w dostatecznym stopniu wyjaśnia zmienność wyników mierzonych średnią ocen uzyskaną przez studentów z toku studiów II stopnia. Nieistotny natomiast był wpływ zmiany układu przedmiotów w cyklu kształcenia. Udział wyników kształcenia z poszczególnych obszarów tematycznych na końcową ocenę ze studiów był porównywalny (współczynnik  $\beta$  w przedziale od 0,11 do 0,14).

■ Tab. 4. Parametry statystyczne modelu regresji dla zmiennych objaśniających: ocena dla ośmiu przedmiotów kończących się egzaminem i zmienna dychotomiczna dla cyklu kształcenia według starego i nowego układu przedmiotów oraz zmiennej objaśnianej: średnia ocen ze studiów II stopnia na kierunku Położnictwo w latach 2007-2012 ( $R = 0,988$ ; wartość statystyki  $F(9,208) = 985,16$ ; istotność statystyczna modelu  $p < 0,00001$ ; błąd standardowy estymacji = 0,06012)

Zmienne objaśniające	b	SE dla b	$\beta$	SE dla $\beta$	t	p
Wyraz wolny			0,04	0,0472	0,9320	0,352
Układ przedmiotów w cyklu kształcenia (0 – stary = 2007/08-2008/09; 1 – nowy = 2009/10-2012/13)	-0,02	0,0132	-0,02	0,0104	-1,696	0,091
1. Kształcenie w zakresie diagnostyki ultrasonograficznej w położnictwie i ginekologii	0,21	0,0128	0,12	0,0074	16,768	< 0,0001
2. Kształcenie w zakresie dydaktyki medycznej	0,25	0,0134	0,12	0,0066	18,681	< 0,0001
3. Kształcenie w zakresie nowoczesnych technik diagnostycznych	0,26	0,0133	0,11	0,0058	19,766	< 0,0001
4. Kształcenie w zakresie opieki specjalistycznej w położnictwie, neonatologii i ginekologii	0,26	0,0112	0,14	0,0059	23,482	< 0,0001
5. Kształcenie w zakresie wybranych zagadnień z pielęgniarstwa specjalistycznego	0,23	0,0115	0,12	0,0061	20,434	< 0,0001
6. Kształcenie w zakresie geriatrii	0,18	0,0112	0,11	0,0071	15,814	< 0,0001
7. Kształcenie w zakresie intensywnego nadzoru położniczego	0,19	0,0113	0,14	0,0079	17,076	< 0,0001
8. Kształcenie w zakresie onkologii ginekologicznej	0,21	0,0119	0,12	0,0067	18,048	< 0,0001

b – współczynnik regresji; SE – błąd standardowy;  $\beta$  – współczynnik standaryzowany beta; t – wartość statystyki dla istotności zmiennej niezależnej, p – poziom istotności statystycznej dla zmiennej niezależnej

## DYSKUSJA

Problematyka oceniania w edukacji medycznej oraz kształceniu specjalistów nauk o zdrowiu jest przedmiotem wielu badań i budzi szerokie zainteresowanie środowiska akademickiego. Fakt ten nie jest zaskakujący, ponieważ ocena studentów jest postrzegana jako jeden z ważniejszych elementów całego systemu edukacyjnego. Z jednej strony określa ona stopień osiągnięcia przez studenta zakładanych efektów kształcenia, a z drugiej może być także miarą jakości procesu edukacyjnego [12].

Bez względu na cel jakemu ma służyć ocena, zawsze wiąże się ona z mniej lub bardziej systematycznym gromadzeniem danych z obserwacji, które prowadzą do postawienia wniosków na temat cech i właściwości ocenianego studenta [12]. Aby proces ten był wysoce obiektywnym źródłem informacji na temat osiągniętych efektów kształcenia musi on spełniać pewne kryteria, określane jako cechy diagnozy edukacyjnej. Do właściwości pomiaru edukacyjnego, które pozwalają na jego ocenę i optymalizację należą: (I) niezależność sytuacji pomiarowej, (II) obiektywizm punktowania, (III) rzetelność, (IV) trafność oraz (V) obiektywizm pomiaru [9, 13, 14]. Jakość pomiaru zależy od zastosowanych metod oceny osiągniętych efektów kształcenia, ale także od czynników zewnętrznych, które w istotny sposób mogą wpływać na jego właściwości. Ocenę wpływu różnego rodzaju czynników kulturowych i osobistych na uzyskane przez studentów wyniki pomiaru w niniejszym badaniu.

Bezstronność pomiaru, czyli niezależność sytuacji pomiarowej oznacza, stworzenie wszystkim studentom jednakowych (sprawiedliwych) warunków do ceny ich osiągnięć. Natomiast stronniczość oceny, która często jest źródłem błędów systematycznych pomiaru, prowadzi do niezaskarżenia wysokich lub niskich wyników uzyskiwanych przez określoną grupę studentów. Jednakowe traktowanie wszystkich uczniów w kolejnych latach oznacza niezależną ocenę ich osiągnięć, bez względu na to jakie wyniki osiągnęli w poprzednich cyklach kształcenia, jakie ukończyli szkoły/uczelnie, czy też w jakiej grupie studenckiej się uczyli. Istotnym elementem bezstronności pomiaru jest stworzenie odpowiednich warunków podczas egzaminowania oraz dobór takich metod oceny osiągnięć studentów, które zapewniałyby porównywalny stopień niezależności sytuacji pomiarowej w kolejnych latach [15]. Jak pokazały wyniki analizy spójności oceniania studentów z poszczególnych przedmiotów w kolejnych rocznikach, możemy postawić tezę o wysokim stopniu zróżnicowania ocen (brak jednorodności wariancji). Ponadto, zaobserwowano statystycznie istotne różnice w ocenach studentów z danego przedmiotu dla następujących po sobie roczników. Powyższe wyniki wskazują, że istnieją czynniki, które wpływają na stopień zróżnicowania ocen studentów. Jest mało prawdopodobne, że istotne różnice w strukturze ocen mierzonych średnią/medianą i wariancją są spowodowane skokową zmianą ocenianych osiągnięć. Jedną z przyczyn może być duży stopień zróżnicowania grupy studentów podejmujących naukę w trybie niestacjonarnym. Wysoka, jak na studia II stopnia, średnia wieku wskazuje, że wśród studentów są osoby, które skończyły studia pomostowe, a następnie podjęły dalsze kształcenie. Jednak grupa ta staje się z roku na rok coraz mniej liczna, więc liczebnie zaczynają przeważać studenci młodszy wiekiem i stażem zawodowym. Taki stan rzeczy może mieć wpływ na obniżającą się średnią ocen z poszczególnych przedmiotach, co obserwujemy w ostatnich latach. Otwarte pozostaje więc pytanie, czy brak powtarzalności w ocenie osiągnięć studentów ma związek ze stronniczością oceny, czy też decydujące znaczenie ma tutaj niewłaściwy dobór metod pomiaru edukacyjnego lub ich niska jakość, a może zmienność grupy studiujących?

Z bezstronnością pomiaru bezpośrednio wiąże się dokładność punktowania (obiektywizm punktowania) rozumiana jako adekwatność danej skali pomiarowej do ocenianych właściwości. W praktyce osiągnięcie wysokiej dokładności punktowania z jednej strony zależy od sposobu formułowania zadań egzaminacyjnych oraz jakości instrukcji oceniania opisanego w kluczu (przyczyny konstrukcyjne), a z drugiej jest on pochodną kompetencji, doświadczenia zawodowego i osobowości egzaminatora (przyczyny osobowe) [16]. Najczęstszym źródłem niezgodności w odniesieniu do wyników pomiaru jest nadmierna surowość/łagodność egzaminatora, a także skłonność do skrajnych ocen lub nadmiernego uśredniania [9]. Jak wskazują wyniki analizy dotyczącej ocen z poszczególnych przedmiotów, w przypadku *Kształcenia w zakresie opieki specjalistycznej w położnictwie, neonatologii i ginekologii* studenci uzyskiwali wyraźnie niższe noty w porównaniu z pozostałymi przedmiotami (surowy egzaminator). Ponadto, dla wszystkich analizowanych przedmiotów zaobserwowano szeroki zakres zmienności wyników (mierzonych wielkością odchylenia standardowego), co w naturalny sposób odzwierciedla zróżnicowanie osiągnięć w badanej grupie studentów. Skrajne wartości nie muszą oczywiście świadczyć o braku obiektywizmu punktowania, ale mogą wynikać z faktu, że *Kształcenie w zakresie opieki specjalistycznej* wymaga złożonych i trudnych w opanowaniu kompetencji (bardzo niska średnia ocen studentów). Zróżnicowany poziom oczekiwań względem studentów dla poszczególnych przedmiotów może w pewnym stopniu uzasadniać wynik analizy osiągnięć studentów, który wskazują na istotne statycznie odchylenie w spójność oceniania. Jednakże, zgodność oceny była różna w kolejnych latach. Należałoby oczekiwać, że niezależnie od rocznika, dobry student w porównaniu z mniej zdolnym będzie miał istotnie wyższe wyniki ze wszystkich przedmiotów. Miarą takiej zgodności ocen jest współczynnik Tau Kendalla, który opiera się na różnicy między prawdopodobieństwem tego, że dwie zmienne układają się w tym samym porządku w obrębie obserwowanych danych, a prawdopodobieństwem, że ich uporządkowanie się różni [17]. Wartość tego współczynnika była bardzo różna dla kolejnych roczników (od 0,22 do 0,42) co świadczy o niezachowaniu we wszystkich przypadkach zasady obiektywizmu punktowania w ocenie studentów w kolejnych rocznikach. Błędy w zakresie dokładności punktowania mogą mieć związek z nadmiernym usztywnieniem kryteriów oceny w sytuacji, w której przyznawana punktacja jest wyprowadzona nie z celów kształcenia, ale w prosty sposób wynika jedynie z poprawności rozwiązania danego zdania, która nie koniecznie mieści się w zakresie wiedzy i umiejętności przewidzianych dla mierzonego efektu kształcenia. Problem ten będzie szczególnie istotny przy ocenie rzetelności i trafności pomiaru [14].

Rzetelność pomiaru to powtarzalność uzyskanych wyników w określonych warunkach. Jeżeli wyniki pomiaru są takie same lub bardzo podobne w ustalonych okolicznościach (np. w sytuacji egzaminowania) to taki pomiar edukacyjny można uznać za rzetelny. Niedostateczna rzetelność stosowanej procedury oceniania przyczynia się do niskiego stopnia zaufania, skoro w podob-

nych okolicznościach poszczególne wyniki różnią się znacząco między sobą. Do oceny czy dany pomiar jest rzetelny można wykorzystać różne metody analityczne. Najczęściej stosowana jest w praktyce ocena wewnętrznej spójności wyników pomiaru poprzez oszacowanie średnich wariancji ocen dla wszystkich egzaminów (*współczynnik  $\alpha$ -Cronbacha*) [10]. Jak pokazują przedstawione wyniki badania rzetelności, stosowane w latach 2007-2012 ocenianie osiągnięć studentów charakteryzowało się niedostatecznym poziomem rzetelności ( $\alpha = 0,65$ ). Jednakże, analiza poszczególnych roczników wskazuje, że w kolejnych latach poziom wewnętrznej zgodności był zróżnicowany ( $\alpha$  od 0,60 do 0,75). Zły dobór metod egzaminacyjnych, a w szczególności nieprawidłowa konstrukcja zadań sprawdzających, które są podstawą oceniania, przyczyniają się do spadku rzetelności pomiaru. Student może nie mieć możliwości pokazania pełni swoich osiągnięć z danej dziedziny, jeżeli egzamin znacząco zawęża zakres treści, które służą ocenie danych efektów kształcenia. Inną przyczyną niskiej wartości współczynnika  $\alpha$  ( $< 0,7$ ) może być duży udział błędów losowych w wynikach danego pomiaru [18]. Przy wartości  $\alpha = 0,65$  błędy losowe stanowią 35% zmienności uzyskanych wyników, a pomiar w takich warunkach może być stosowany właściwie jedynie przy porównywaniu międzygrupowym, a nie do różnicowania indywidualnego [19].

Oprócz badań skupiających się na zagadnieniu rzetelności, które odnoszą się do pytania „jak się mierzy?“, istotne w budowaniu dobrych narzędzi oceniających jest określenie trafności pomiaru, który odpowiada na pytanie „co się mierzy?“. Trafność w tym zakresie należy rozumieć, jako użyteczność danej metody w ocenie konkretnego zestawu cech i właściwości zdającego [20]. Nie istnieje dokładna metoda pomiaru trafności, a jedynie pewna jej pośrednia ocena, co zwykle opiera się na zastosowaniu jednej z pięciu koncepcji, według których można określić trafność pomiaru: programowa, wewnętrzna, diagnostyczna, prognostyczna i teoretyczna [9, 14, 18]. W przypadku egzaminów przewidzianych programem kształcenia na kierunku Położnictwo, istotna w zakresie trafności programowej jest zgodność treści zadań egzaminacyjnych z celami kształcenia dla poszczególnych przedmiotów. Walidacja tego parametru wymaga więc analizy programu kształcenia, co nie jest przedmiotem niniejszego opracowania. Natomiast, analiza trafności wewnętrznej dotyczy stopnia, w jakim poszczególne komponenty, czyli części składające się na całość pomiaru, wchodzą ze sobą w interkorelacje [21]. Do oceny trafności wewnętrznej konieczne jest wyznaczenie korelatów (*cross-correlates*) między poszczególnymi przedmiotami. Wyniki analizy interkorelacji są w dużym stopniu spójne z wynikami oceny zgodności punktowania szacowanej wartością współczynnika Tau Kendalla oraz pomiarami rzetelności za pomocą współczynnika  $\alpha$ -Cronbacha. Przedstawiona w niniejszej pracy ocena trafności w analizie interkorelacji wskazuje, że dla zdecydowanej większości przedmiotów istnieją pozytywne zależności między wynikami uzyskanymi przez studentów z kolejnych egzaminów. Jednakże, siła korelacji dla różnych par przedmiotów nie była zbyt duża (*r-Spearmana* od 0,13 do 0,53). Wyjątkiem

są wyniki studentów uzyskane z Kształcenia w zakresie intensywnego nadzoru położniczego. Dla tego przedmiotu nie zaobserwowano istotnych interkorelacji z czterema innymi zakresami kształcenia. Brak takich korelatów międzyprzedmiotowych pozwala postawić tezę, że uzyskane przez studentów wyniki z *Kształcenia w zakresie intensywnego nadzoru położniczego*, odzwierciedlają ocenę różnych treści i umiejętności zdających niż ma to miejsce w przypadku innych przedmiotów. Przypuszczenie to, chociaż mało prawdopodobne, wymaga pogłębionej analizy trafności pomiaru w oparciu o szczegółowe efekty kształcenia dla przedmiotów, dla których brak jest pozytywnych korelatów. Bardziej prawdopodobne jest że ocena osiągnięć studentów dla *Kształcenia w zakresie intensywnego nadzoru położniczego* została uzyskana w wyniku złe zaplanowanego egzaminu lub świadczy o niepoprawnym przeprowadzeniu procesu kształcenia. W innym przypadku należałoby podważyć trafność oceny dla czterech zakresów kształcenia, dla których nie wykazano pozytywnych interkorelacji z tym przedmiotem.

*Analiza trafności* pomiaru edukacyjnego ma za zadanie zapobiegać nadużyciom w interpretacji wyników pomiaru [22]. Jeżeli student uzyskał wysoką średnią ocen z toku studiów, to wartość takiej oceny jest tylko wtedy istotna, jeśli odzwierciedla ona rzeczywiste jego osiągnięcia w szczególności w odniesieniu do wymagań programowych. Dlatego też jednym z ważnych aspektów analizy trafności pomiaru jest ocena predykcji, która odnosi się do pomiaru zdolności prognostycznej danych wyników do przewidywania przyszłych losów studentów, np. osiągnięcie sukcesu w trakcie studiów czy późniejszy status zawodowy absolwenta. Z oczywistych względów określenie trafności prognostycznej w analizowanym przypadku nie jest możliwe. Nie dysponujemy bowiem szczegółowymi danymi o przyszłych losach studentów i absolwentów w zakresie ich aktywności zawodowej po zakończeniu studiów.

Ze względu na fakt, że przyjmowanie w badaniach trafności, jako wskaźnika oceny, współczynnika korelacji rang Spearmana jest bardzo uproszczonym modelem, znacznie bardziej wiarygodne wyniki można uzyskać z wielowymiarowej analizy z użyciem metody regresji wielorakiej. Zaproponowany model regresji uwzględniał dziewięć zmiennych niezależnych, w tym osiem obejmujące wyniki uzyskane przez studenta z poszczególnych przedmiotów oraz jednej zmiennej dychotomicznej odnoszącej się do zmiany układu przedmiotów jaka nastąpiła na przełomie roczników 2008/09 a 2009/10. Wyniki analizy regresji prezentowane w niniejszej pracy, potwierdzają obserwacje uzyskane z jednowymiarowej analizy korelacji. Wyznaczone standaryzowane współczynniki  $\beta$  dla ośmiu przedmiotów, wskazują na ich bardzo zbliżony udział w średniej ocen z toku studiów. Uzyskanie przez studenta wyższej oceny o 0,5 z danego przedmiotu przyczynia się do wzrostu średniej ocen na koniec studiów o wartość od 0,11 do 0,14. Nie zaobserwowano natomiast wpływu zmiany organizacji kształcenia na końcowe wyniki studentów (brak wpływu zmiennej dychotomicznej), co oznacza że stosowane metody pomiaru edukacyjnego funkcjonowały podobnie w obu okresach.

Rzetelność i trafność to dwie hierarchicznie powiązane ze sobą cechy, dlatego też zawsze dla danego pomiaru należy obie właściwości rozpatrywać łącznie. Stosunkowa dobra trafności oznacza, że w większości przypadków wyniki uzyskiwane przez studentów odzwierciedlają rzeczywiste właściwości i cechy zdających. Jednakże niedostateczna rzetelność pomiaru przyczynia się do mniej dokładnej oceny wiedzy i umiejętności zdających, a co za tym idzie słabszego różnicowania indywidualnego. Ponadto, sytuacja taka skutkuje gorszą precyzją w ocenie kompetencji szczególnie tych studentów, dla których średnia ocen z toku studiów lokuje się w pobliżu dolnej granicy skali (średnia ocen  $\approx 3,0$ ). Ze względu na fakt, że są to najslabsi studenci, to istnieje duże ryzyko niedostatecznego poziomu kompetencji reprezentowanych przez tę grupę absolwentów kierunku Położnictwo.

## WNIOSKI

Każda z trzech strategii analitycznych – szacowanie współczynnika korelacji  $r$ -Spearmana, Tau Kendalla oraz  $\alpha$ -Cronbacha, mimo że ma różne podstawy logiczne i formuły obliczeniowe, prowadzi do podobnych wniosków. Niski stopień zgodności oceniania osiągnięć studentów w kolejnych latach świadczy o niedostatecznej spójności systemu pomiaru efektów kształcenia na kierunku Położnictwo. W przyszłości ważnym elementem ewaluacji systemu oceniania powinna być analiza prognostyczna, w której można by uwzględnić zmienne zależne związane z losami absolwentów. Krytyczna analiza obecnie stosowanych rozwiązań pozwala na zaplanowanie szczegółowej ewaluacji stosowanych kryteriów oceniania w celu zdiagnozowania słabych punktów w obszarze nauczania oraz oszacowania rzetelności i trafności stosowanych metod oceny kompetencji studentów Położnictwa w WUM.

## PIŚMIENNICTWO

1. Rozporządzenie Ministra Zdrowia z dnia 29 października 2003 r. w sprawie wykazu dziedzin pielęgniarstwa oraz dziedzin mających zastosowanie w ochronie zdrowia, w których może być prowadzona specjalizacja i kursy kwalifikacyjne, oraz ramowych programów specjalizacji dla pielęgniarzek i położnych (Dz.U. 2003 nr 197 poz. 1922).
2. Dobrowolska V. Ocena przygotowania do zawodu w opinii pielęgniarzek i pielęgniarzy zatrudnionych w oddziałach szpitalnych. *Piel Pol.* 2010;35 (1): 7-13.
3. Drennan J. Masters in nursing degrees: an evaluation of management and leadership outcomes using a retrospective pre-test design. *J Nurs Manag.* 2012;20 (1): 102-12.
4. Begley CM, Oboyle C, Carroll M, et al. Educating advanced midwife practitioners: a collaborative venture. *J Nurs Manag.* 2007;15 (6): 574-84.
5. The Essentials of Master's Education in Nursing: American Association of Colleges of Nursing 2011. Available from: <http://www.aacn.nche.edu/education-resources/MastersEssentials11.pdf> (dostęp 06-02-2015).
6. 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 poz. 631).
7. Duszyński M. Efekty kształcenia w Polsce: perspektywa brytyjska. *Nauka.* 2011;1 : 137-44.
8. Harden RM, Crosby JR, Davis MH. AMEE Guide No. 14: Outcome-based education: Part 1 – An introduction to outcome-based education. *Med Teach.* 1999;21 (1): 7-14.
9. Niemierko B. Diagnostyka edukacyjna. Warszawa: Wydawnictwo Naukowe PWN; 2009.
10. Feldt LS. A test of hypothesis that Cronbachs alpha or Kuder-Richardson coefficient 20 is same for 2 tests. *Psychometrika.* 1969;34 (3): 363.
11. Nathans LL, Oswald FL, Nimon K. Interpreting multiple linear regression: A guidebook of variable importance. *Pract Assess Res Eval.* 2012;17 (9): 2.
12. Schuwirth LW, van der Vleuten CP. General overview of the theories used in assessment: AMEE Guide No. 57. *Med Teach.* 2011;33 (10): 783-97.
13. Norman GR, Vleuten C, Newble DI. International handbook of research in medical education. Dordrecht: Kluwer Academic Publishers; 2002.
14. Niemierko B. Pomiar wyników kształcenia. Warszawa: Wydawnictwo Szkolne i Pedagogiczne; 1999.
15. Rowley J. Measuring quality in higher education. *Qual High Educ.* 1996;2 (3): 237-55.
16. Tam M. Measuring Quality and Performance in Higher Education. *Qual High Educ.* 2001;7 (1): 47-54.
17. Kendall MG. A new measure of rank correlation. *Biometrika.* 1938: 81-93.
18. Niemierko B. Testy osiągnięć szkolnych. Podstawowe pojęcia i techniki obliczeniowe. 1st ed. Warszawa: Wydawnictwo Szkolne i Pedagogiczne; 1975.
19. Guilford JP. Psychometric methods. 2nd ed. New York: McGraw-Hill; 1954.
20. Goodwin LD. Changing conceptions of measurement validity: an update on the new standards. *J Nurs Educ.* 2002;41 (3): 100-6.
21. Meagher DG, Pan T, Wegner R, Olson AT, Overgaard SL, Mehle JJ. PCAT Reliability and Validity. 3rd ed. San Antonio: Pearson Executive Office; 2012.
22. Kubielski W. Podstawy pomiaru, konstruowania i ewaluacji testu dydaktycznego. Warszawa: Wydawnictwo Wyższej Szkoły Pedagogicznej TWP; 2006.

Praca przyjęta do druku: 07.02.2015

Praca zaakceptowana do druku: 07.05.2015