

Rzetelność i trafność oceny osiągniętych efektów kształcenia przez studentów kierunku ratownictwo medyczne w Warszawskim Uniwersytecie Medycznym

Reliability and validity of assessment of educational outcomes obtained by students of Medical Rescue at Medical University of Warsaw

Mariusz Panczyk¹, Grzegorz Stachacz¹, Robert Gałązkowski², Joanna Gotlib¹

¹ZAKŁAD DYDAKTYKI I EFEKTÓW KSZTAŁCENIA, WYDZIAŁ NAUKI O ZDROWIU, WARSZAWSKI UNIWERSYTET MEDYCZNY

²ZAKŁAD RATOWNICTWA MEDYCZNEGO, WYDZIAŁ NAUKI O ZDROWIU, WARSZAWSKI UNIWERSYTET MEDYCZNY

STRESZCZENIE

Wstęp: Dla zachowania wysokiego stopnia obiektywizmu w zakresie informacji dotyczącej osiągniętych przez studentów efektów kształcenia konieczne jest, aby system oceniania spełniał kryteria odpowiedniego poziomu rzetelności i trafności.

Cel pracy: Analiza rzetelności i trafności systemu oceny osiągnięć edukacyjnych studentów z przedmiotów kończących się egzaminem i objętych programem nauczania na kierunku ratownictwo medyczne na Warszawskim Uniwersytecie Medycznym (WUM).

Materiał i metody: Badanie retrospektywne obejmujące grupę 421 studentów z ośmiu kolejnych pełnych cykli kształcenia. Zebrano szczegółowe wyniki dotyczące uzyskanych ocen z czternastu przedmiotów kończących się egzaminem w całym toku studiów. Oceniono rzetelność (współczynnik alfa-Cronbacha) oraz trafność kryterialną (korelacja rang Spearmana). Wewnętrzna zgodność oceniania była oszacowana z wykorzystaniem metody regresji wielorakiej.

Wyniki: Poziom rzetelność oceniania dla zakresu ogólnouniwersyteckiego, przedklinicznego i klinicznego wyniosł odpowiednio, alfa: 0,42, 0,53 i 0,70. Wykazano, że najsilniejsze pozytywne zależności między wynikami kształcenia przedklinicznego względem klinicznego istnieją dla przedmiotu anatomia ($r \approx 0,30$). Jedynie w przypadku Farmakologii stwierdzono, że osiągnięcia studentów z tego obszaru są istotnie skorelowane ze wszystkimi przedmiotami kształcenia klinicznego. Udział wyników kształcenia z poszczególnych obszarów kształcenia klinicznego na końcową ocenę ze studiów był zróżnicowany (β regresji od 0,04 do 0,11). Najsilniejszy wpływ na wyniki końcowe zaobserwowano w przypadku Farmakologii. Natomiast najsłabszy wpływ na końcowe wyniki studenta zaobserwowano w przypadku Chirurgii ($\beta = 0,04$).

Wnioski: 1. Dostateczna rzetelność systemu oceny osiągniętych efektów kształcenia na kierunku ratownictwo medyczne świadczy o dobrej precyzji i powtarzalności oceniania. 2. Stwierdzony niski stopień trafności był determinowany niezachowaniem adekwatności oceny osiągniętych efektów kształcenia dla niektórych przedmiotów z grupy nauk klinicznych. 3. Konieczne jest w przyszłości oszacowanie trafności prognostycznej i diagnostycznej stosowanych metod oceny kompetencji studentów ratownictwa medycznego na WUM.

Słowa kluczowe: ewaluacja systemu oceniania, system oceniania, ocena osiągnięć, diagnoza edukacyjna, pomiar edukacyjny

ABSTRACT

Introduction: In the interest of preservation of high degree of objectivity of information about students' educational outcomes, a system of assessment needs to meet criteria of appropriate reliability and validity.

The aim: Analysis of reliability and validity of the system of assessment of students' educational outcomes for courses followed by an examination and covered by a curriculum in Medical Rescue at Medical University of Warsaw (MUW).

Materials and methods: A retrospective study enrolling a group of 421 students of eight subsequent full education cycles. Detailed data concerning grades for fourteen courses followed by an examination in the entire course of studies were collected. Reliability (Cronbach's alpha coefficient) and criteria validity (Spearman's rank correlation) were assessed. Internal consistency was estimated using a multiple regression model.

Results: The levels of assessment reliability for the general university, pre-clinical, and clinical scopes amounted to alpha: 0.42, 0.53, and 0.70, respectively. The strongest positive correlations between the results of pre-clinical and clinical trainings were found for the Anatomy course ($r \approx 0.30$). Only in the case of the Pharmacology course it was found that students' achievements in this field were significantly correlated with all other courses of clinical training. The influence of educational outcomes in particular areas of clinical training on the final grade for the entire course of studies was diverse (β regression between 0.04 and 0.11). While the Pharmacology course had the strongest impact on final results, the Surgery course had the least influence on students' final grades ($\beta = 0.04$).

Conclusions: 1. Sufficient reliability of the system of assessment of educational outcomes in Medical Rescue showed good precision and repeatability of assessment. 2. A low level of validity was caused by a failure to keep the appropriateness of the assessment of educational outcomes in several clinical courses. 3. Prognostic and diagnostic validity of methods used for evaluation of competencies of Medical Rescue students at WMU needs to be assessed in the future.

Key words: performance evaluation system, performance assessment, achievement assessment, educational diagnosis, educational measurement

WSTĘP

W nieobowiązującej już Ustawie o Państwowym Ratownictwie Medycznym (PRM) z dnia 25 lipca 2001 roku (Dz.U. 2001 nr 113 poz. 1207 z późn. zm) w art. 30 ust. 3 podano, że „zawód ratownika medycznego może wykonywać (m.in.) osoba, która uzyskała w kraju dyplom ukończenia wyższych studiów zawodowych w zakresie ratownictwa medycznego” [1]. Generalnie zapis ten został utrzymany w art. 10 ust. 4 nowej Ustawy o Państwowym Ratownictwie Medycznym z dnia 8 września 2006 roku (Dz.U. 2006 Nr 191 poz. 1410) [2]. Takie uwarunkowania prawne zawodu przyczyniły się do tego, że już od początku ubiegłej dekady uruchamiano na wielu wyższych uczelniach 3-letnie studia licencjackie na nowym samodzielny kierunku ratownictwo medyczne lub rzadziej jako specjalność na kierunku zdrowie publiczne, jak i powoływano w nich jednostki organizacyjne odpowiedzialne za kształcenie i badania naukowe w zakresie ratownictwa. W roku akademickim 2001/02 także na Akademii Medycznej w Warszawie, obecnie Warszawski Uniwersytet Medyczny (WUM) na Wydziale Nauk o Zdrowiu stworzono możliwość kształcenia ratowników medycznych w ramach specjalność kierunku zdrowie publiczne. Następnie od roku 2004/05 kształcenie odbywa się już w ramach samodzielnego kierunku studiów. W chwili obecnej studia tego rodzaju prowadzą wszystkie uniwersytety medyczne w Polsce [3].

Wspomniana wyżej Ustawa o PRM z 2001 roku w zakresie wymagań programowych dotyczących kształcenia ratowników na poziomie wyższych studiów zawodowych odsyłała do przepisów wykonawczych, które jednak Minister Zdrowia nigdy nie ogłosił. Kształcenie kierunkowe prowadzone przez szkoły wyższe na mocy Ustawy z dnia 27 lipca 2005 r. Prawo o szkolnictwie wyższym (Dz.U. 2005 nr 164 poz. 1365) zostało określone przepisami Rozporządzenia Ministra Nauki i Szkolnictwa Wyższego z dnia 12 lipca 2007 roku w sprawie standardów kształcenia dla poszczególnych kierunków oraz poziomów kształcenia, a także trybu tworzenia i warunków, jakie musi spełniać uczelnia, by prowadzić studia międzykierunkowe oraz makrokierunki [4]. Rozporządzenie zdefiniowało wymagania ogólne, kwalifikacje absolwenta oraz ramowe treści kształcenia wraz z zasadami przeprowadzania praktyk zawodowych. W załączniku nr 88 do niniejszego Rozporządzenia, określono standardy kształcenia w ramach studiów I stopnia dla kierunku ratownictwo medyczne. Ustalono, że studia te nie mogą trwać krócej niż 6 semestrów, a liczba godzin zajęć nie powinna być mniejsza niż 3800, wliczając w to 320 godzin obowiązkowych praktyk. Program powinien obejmować również dwanaście tygodni praktyk zawodowych realizowanych w trakcie ostatniego roku studiów po czterdzieści godzin tygodniowo [5]. W ramach programu kształcenia wyróżniono grupę treści podstawowych (450 godzin w tym m.in. socjologia, psychologia, anatomia, fizjologia, patofizjologia, farmakologia i toksykologia) oraz grupę treści kierunkowych (1605 godzin w tym m.in. medycyna katastrof, medyczne czynności ratunkowe, intensywna terapia, choroby wewnętrzne, chirurgia, pediatria, neurologia, traumatologia narządów ruchu oraz położnictwo i ginekologia).

Mimo że ratownictwo medyczne jest jednym z najważniejszych filarów współczesnego systemu ochrony zdrowia, to w przeciwieństwie do innych specjalności z obszaru nauk o zdrowiu, jak pielęgniarstwo i położnictwo, zawód ratownika medycznego

nie jest obecnie objęty standardami kształcenia. Obowiązujące Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 9 maja 2012 r. w sprawie standardów kształcenia dla kierunków studiów (Dz.U. 2012 poz. 631) nie wymienia ratownictwa medycznego [6]. Związku z tym każda uczelnia prowadząca kształcenie na tym kierunku powinna zbudować autorski program studiów wraz ze szczegółowym opisem efektów kształcenia zawartym w sylabusach. Niezbędnym elementem tak zaplanowanego kształcenia jest odpowiedni system oceny osiągnięć studentów, którego prawidłowe funkcjonowanie 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]. Brak jest jednak, jak wspomniano, obowiązujących ministerialnych standardów kształcenia dla zawodu ratownika medycznego. Związku z tym programy kształcenia są budowane na podstawie zasad zaczerpniętych z istniejących dobrych wzorców anglo-amerykańskich. Bazując na zachodnich standardach, obowiązujące obecnie programy studiów mają zapewniać zdobycie wiedzy oraz wykształcenie niezbędnych umiejętności składających się na złożone kompetencje ratownika medycznego. Osoba kończąca studia powinna posługiwać się wiedzą specjalistyczną i umiejętnościami w zakresie nauk medycznych, szczególnie w obszarze medycznych czynności ratunkowych, które są wykonywane w stanach zagrożenia życia u dzieci i dorosłych, niezależnie od przyczyny. Ponadto, absolwent powinien posiadać ogólną wiedzę w zakresie nauk społecznych. Absolwent powinien umieć samodzielnie wykonywać medyczne czynności ratunkowe, opanowując umiejętność oceny stanu pacjenta celem ustalenia właściwego postępowania, jego ułożenia w właściwej pozycji w odniesieniu do rodzaju schorzenia czy też odniesionych obrażeń [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]. Dbałość o wysokie standardy kształcenia i efektywne nauczanie realizowane poprzez stałą kontrolę i ewaluację powinna być kluczowym zadaniem dla uczelni wyższej kształcącej w tak odpowiedzialnym zawodzie jak ratownik medyczny.

CEL

Celem niniejszej pracy jest prezentacja wyników analizy zgodności oceniania mierzonych rzetelnością i trafnością w odniesieniu do osiągnięć studentów z przedmiotów kończących się egzaminem i objętych programem nauczania na kierunku ratownictwo medyczne na studiach I stopnia na Warszawskim Uniwersytecie Medycznym (WUM, dawniej Akademia Medyczna w Warszawie) w ośmiu kolejnych cyklach kształcenia.

MATERIAŁ I METODY

Do badania zakwalifikowano dane egzaminacyjne 421 studentów kierunku ratownictwo medyczne, którzy rozpoczęli naukę na studiach I stopnia w WUM w latach 2004–2011 i ukończyli

kształcenie, uzyskując tytuł licencjata w zawodzie. Grupa badana pod względem liczebności poszczególnych roczników charakteryzowała się jednorodnością. Natomiast pod względem struktury wiekowej oraz płci, analizowane grupy studentów wykazywały się istotnymi różnicami. Zestawienie danych dotyczących charakterystyki badanej grupy przedstawiono w tabeli I.

Analizie retrospektywnej poddano wyniki kształcenia z pięciu przedmiotów kształcenia ogólnouniwersyteckiego, czterech z zakresu nauk przedklinicznych oraz czternastu klinicznych. Zmiany w programie studiów w odniesieniu do różnych przedmiotów kończących się egzaminem na kierunku ratownictwo medyczne dla cykli kształcenia rozpoczynających się w okresie 2004–2011 przedstawiono w tabeli II.

W celu 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 dla poszczególnych zakresów kształcenia określono za pomocą współczynnika α -Cronbacha (współczynnik Kudara-Richardsona dla testu składającego się z pozycji dwukategorialnych) oraz wyznaczono błąd standardowy pomiaru (SEM). Jednowymiarową analizę korelacji dla analizowanych przedmiotów przeprowadzono z użyciem nieparametrycznego współczynnika korelacji rang Spearmana. Interkorelacje wyznaczono dla wyników kształcenia przedklinicznego vs. ocenianie z zakresu klinicznego. Ponadto, zgodność pomiaru edukacyjnego analizowano dla układu ocen uzyskanych przez studentów z przedmiotów klinicznych. Wewnętrzna zgodność oceniania została także oszacowana z wykorzystaniem metody regresji wielorakiej dla czternastu zmiennych niezależnych (objaśniających), wśród nich wiek i płeć studenta oraz oceny z dwunastu przedmiotów kończących się egzaminem. Oceniono parametry funkcji regresji wraz z oceną błędów standardowych oraz wyznaczono współczynnik standaryzowany β w celu określenia stopnia wpływu zmiennych niezależnych na średnią ocen ze studiów (zmienna zależna, objaśniana).

Do obliczeń wykorzystano pakiet statystyczny STATISTICA wersja 12 (StatSoft, Inc.) zgodnie z licencją WUM. Dla wszyst-

kich analiz jako domyślny (*a priori*) poziom istotności przyjęto $\alpha=0,05$.

WYNIKI

Porównując oceny uzyskane przez danego studenta z kolejnych przedmiotów obejmujących kształcenie kliniczne w zakresie ratownictwa medycznego stwierdzono, że zgodność pomiaru edukacyjnego kształtuje się na średnim poziomie. Współczynnik zgodności Kendalla wyniósł 0,34, przy górnej granicy wartości 0,47. Ponadto, analiza osiągnięć studentów z różnych przedmiotów, wykonana za pomocą nieparametrycznego testu ANOVA Friedmana (porównanie dla czternastu grup powiązanych) wskazuje na istotne statycznie odchylenie w spójność oceniania ($p < 0,000001$). Łączna analiza porównawcza średnich ocen uzyskanych przez studentów z poszczególnych roczników dla przedmiotów klinicznych wskazuje, że najgorzej ocenianym obszarem była *Farmakologia* (średnia ocen $3,00 \pm 0,805$). W dolnej granicy skali ocen znalazły się także trzy inne przedmioty: *Traumatologia narządów ruchu* ($3,40 \pm 0,960$), *Neurologia* ($3,44 \pm 0,716$) oraz *Medyczne czynności ratunkowe - zaawansowane czynności ratunkowe ALS* ($3,49 \pm 1,143$). Ponadto, ostatni z wymienionych przedmiotów charakteryzował się największym stopniem zróżnicowania ocen. Natomiast najlepiej ocenianym zakresem kształcenia klinicznego były *Medyczne czynności ratunkowe - techniki zabiegów medycznych* ($4,52 \pm 0,615$) oraz *Intensywna terapia i zasady leczenia ostrego bólu* ($4,41 \pm 0,563$). Dla pozostałych przedmiotów średnie ocen studentów mieściły się w przedziale od 3,64 do 4,31. Stopień zróżnicowania osiągnięć studentów mierzonych średnią ocen dla przedmiotów objętych kształceniem klinicznym na kierunku ratownictwo medyczne w latach 2004–2011 przedstawiono na rycinie 1.

Najczęściej stosowanym sposobem oceny wewnętrznej zgodności wyników pomiaru dla co najmniej dwóch czynników jest współczynnik rzetelności α -Cronbacha. Analiza wykazała, że poziom rzetelność oceniania dla zakresu ogólnouniwersyteckiego, przedklinicznego i klinicznego wyniósł odpowiednio: 0,42, 0,53 i 0,70 (przy SEM równym: 1,54, 1,36 i 2,35). Dla przedmiotów z obszaru klinicznego ogólną rzetelność pomiaru edukacyjnego

Tabela I. Charakterystyka badanej grupy studentów, którzy rozpoczęli i ukończyli naukę na studiach I stopnia na kierunku ratownictwo medyczne na Warszawskim Uniwersytecie Medycznym (dawniej Akademia Medyczna w Warszawie).

Rok rozpoczęcia studiów	N	P*	N kobiet (%)	P**	Średnia wieku \pm SD	P***
2004/05	54	0,75	32 (59,3%)	0,02	19,7 \pm 1,70	0,001
2005/06	61		40 (65,6%)		19,1 \pm 1,47	
2006/07	55		35 (63,6%)		20,1 \pm 1,79	
2007/08	49		29 (59,2%)		19,9 \pm 1,53	
2008/09	57		28 (49,1%)		20,2 \pm 1,61	
2009/10	48		27 (56,3%)		20,1 \pm 1,60	
2010/11	43		29 (67,4%)		19,7 \pm 3,70	
2011/12	54		19 (35,2%)		20,0 \pm 1,60	
Ogółem	421	----	239 (56,8%)	----	19,9 \pm 1,92	----

* test zgodności $\chi^2=4,29$

** test χ^2 Pearsona =16,85

*** test ANOVA rang Kruskala-Wallisa $H=30,04$

SD – odchylenie standardowe

Tabela II. Wykaz przedmiotów kończących się egzaminem i obowiązujących w cyklu kształcenia dla kolejnych roczników rozpoczynających naukę na kierunku ratownictwo medyczne na Warszawskim Uniwersytecie Medycznym (dawniej Akademia Medyczna w Warszawie).

	04/05	05/06	06/07	07/08	08/09	09/10	10/11	11/12
Zakres kształcenia ogólnouniwersyteckiego								
Etyka	X	X	X	X	X	X	X	X
Psychologia	X	X	X	X	X	X	X	X
Język obcy	X	X	X	X	X	X	X	X
Biostatystyka		X	X	X	X	X	X	X
Zajęcia sprawnościowo-ruchowe				X	X	X	X	
Zakres kształcenia przedklinicznego								
Anatomia	X	X	X	X	X	X	X	X
Fizjologia	X	X	X	X	X	X	X	X
Patofizjologia	X	X	X	X	X	X	X	X
Mikrobiologia	X	X	X	X	X	X	X	X
Zakres kształcenia klinicznego								
Medycyna katastrof	X	X	X	X	X	X	X	X
Pierwsza pomoc BLS						X	X	X
Farmakologia	X	X	X	X	X	X	X	X
Toksykologia kliniczna	X	X	X	X	X	X	X	
Intensywna terapia i zasady leczenia ostrego bólu	X	X	X	X	X	X	X	X
Medycyna ratunkowa – ratownictwo specjalistyczne		X	X	X	X			
Medyczne czynności ratunkowe - techniki zabiegów medycznych	X	X	X	X	X	X	X	X
Medyczne czynności ratunkowe – zaawansowane czynności ratunkowe ALS				X	X	X	X	X
Kardiologia			X	X	X	X	X	X
Choroby wewnętrzne – interna	X	X	X	X	X	X	X	X
Chirurgia	X	X	X	X	X	X	X	X
Neurologia	X	X	X	X	X			X
Pediatrya	X	X	X	X	X	X	X	X
Położnictwo i ginekologia	X	X	X	X	X	X	X	X
Traumatologia narządów ruchu	X	X	X	X	X	X	X	X

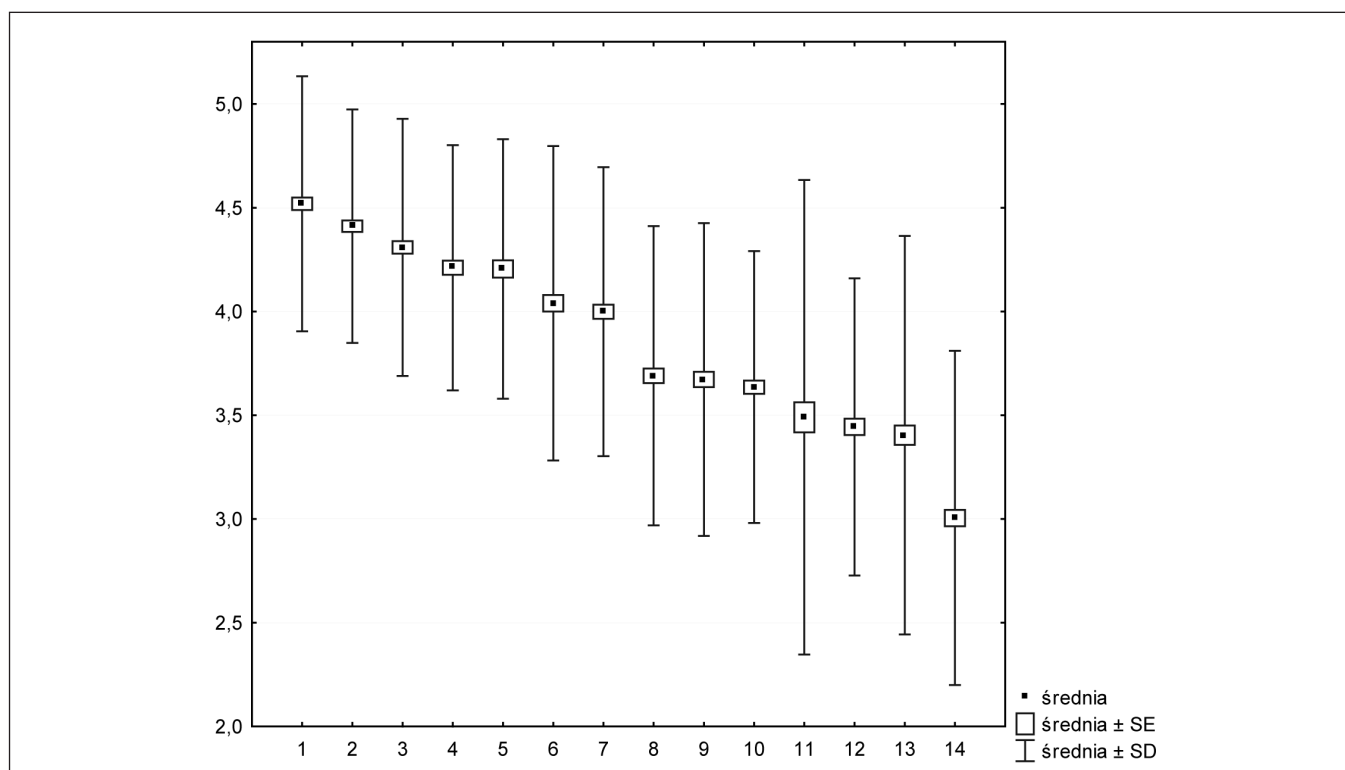
obniżały wyniki studentów z egzaminów z *medycyny katastrof* oraz *medycyny ratunkowej – ratownictwa specjalistycznego*.

Oszacowanie trafności w analizie interkorelacji wskazuje, że najsilniejsze pozytywne zależności między wynikami kształcenia przedklinicznego względem klinicznego istnieją dla przedmiotu *anatomia*. Dla tego przedmiotu współczynnik korelacji rang Spearmana osiągał w niektórych przypadkach wartości powyżej 0,30. Mniejszą siłę korelacji zaobserwowano w przypadku *mikrobiologii*, gdzie tylko jedna zależność przekroczyła próg 0,30. Szczegółowe zestawienie wyników analizy korelacyjnej dla zależności: wyniki kształcenia przedklinicznego vs. wyniki dla zakresu klinicznego, przedstawiono w tabeli III.

W analizie interkorelacji między wynikami kształcenia z poszczególnych obszarów klinicznych wykazano, że w szczególności dla dwóch przedmiotów *medycyna katastrof* oraz *medycyna ratunkowa – ratownictwo specjalistyczne* zaobserwowano niski stopień zgodności z wynikami kształcenia dla pozostałych przedmiotów. Ponadto jedynie w przypadku *farmakologii* stwierdzono, że osiągnięcia studentów z tego obszaru są istotnie skorelowane ze wszystkimi przedmiotami

kształcenia klinicznego. Szczegółowe zestawienie wyników analizy interkorelacji dla przedmiotów klinicznych objętych programem kształcenia ratownictwa medycznego zestawiono w tabeli IV.

W celu dokładnego oszacowania trafności wewnętrznej przeprowadzono analizę regresji z użyciem *czternastu zmiennych niezależnych (objaśniających)*. Jak przedstawiono w tabeli V zaproponowane równanie funkcji regresji w dostatecznym stopniu wyjaśnia zmienność wyników mierzonych średnią ocen uzyskaną przez studentów podczas studiów I stopnia na kierunku ratownictwo medyczne. Nieistotny natomiast był wpływ płci oraz wieku studenta na uzyskane oceny. Udział wyników kształcenia z poszczególnych obszarów kształcenia klinicznego na końcową ocenę ze studiów był zróżnicowany (współczynnik β w przedziale od 0,04 do 0,11). Najsilniejszy wpływ na wyniki końcowe zaobserwowano w przypadku *Farmakologii*. Na każde podwyższenie oceny z tego przedmiotu o 0,5 rośnie prawdopodobieństwo uzyskania wyższej średniej ocen na zakończenie studiów o wartość 0,11. Analogicznie, najsłabszy wpływ na końcowe wyniki studenta zaobserwowano w przypadku *Chirurgii* ($\beta = 0,04$).



1 – intensywna terapia i zasady leczenia ostrego bólu, 2 – medyczne czynności ratunkowe – techniki zabiegów medycznych, 3 – choroby wewnętrzne - interna, 4 – kardiologia, 5 – medycyna ratunkowa – ratownictwo specjalistyczne, 6 – toksykologia kliniczna, 7 – chirurgia, 8 – medycyna katastrof, 9 – pediatria, 10 – położnictwo i ginekologia, 11 – medyczne czynności ratunkowe – zaawansowane czynności ratunkowe ALS, 12 – neurologia, 13 – traumatologia narządów ruchu, 14 – farmakologia SE – błąd standardowy dla średniej, SD – odchylenie standardowe

Ryc. 1. Struktura oceniania studentów z poszczególnych przedmiotów kształcenia klinicznego kończących się egzaminem na kierunku ratownictwo medyczne w Warszawskim Uniwersytecie Medycznym (wartość statystyki ANOVA Friedmana $\chi^2(13,106) = 467,78$; $p < 0,000001$; współczynnik zgodności Kendalla = 0,34).

Tabela III. Wyniki analizy korelacji między ocenami uzyskanymi przez studentów z przedmiotów przedklinicznych a wynikami z egzaminów dla przedmiotów klinicznych. Wartości korelacji wyrażono za pomocą współczynnika porządku rang Spearmana, wielkości oznaczone (*) są nieistotne statystycznie ($p > 0,05$).

Nauki przedkliniczne	Anatomia	Fizjologia	Patofizjologia	Mikrobiologia
Nauki kliniczne				
Farmakologia	0,34	0,20	0,21	0,29
Toksykologia kliniczna	0,26	0,26	0,12	0,18
Medycyna katastrof	-0,10	0,06*	0,13	0,23
Intensywna terapia i zasady leczenia ostrego bólu	0,22	0,16	0,12	0,11
Medycyna ratunkowa – ratownictwo specjalistyczne	0,40	0,10*	0,16	-0,18
Medyczne czynności ratunkowe – techniki zabiegów medycznych	0,24	0,26	0,13	0,14
Medyczne czynności ratunkowe – zaawansowane czynności ratunkowe ALS	0,21	0,20	0,07*	0,25
Kardiologia	0,07*	0,13	0,10*	0,18
Choroby wewnętrzne – Interna	-0,03*	-0,02*	0,11	0,17
Chirurgia	0,11	0,10*	0,04*	0,11
Neurologia	0,33	0,25	0,22	0,31
Pediatria	0,21	0,17	0,13	0,23
Położnictwo i ginekologia	0,24	0,19	0,18	0,23
Traumatologia narządów ruchu	0,45	0,33	-0,05*	0,09*

Tabela IV. Wyniki analizy korelacji między ocenami uzyskanymi przez studentów z poszczególnych przedmiotów klinicznych. Wartości korelacji wyrażono za pomocą współczynnika porządku rang Spearmana, wielkości oznaczone (*) są nieistotne statystycznie ($p > 0,05$).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1		0,04*	0,18	0,11	-0,03*	0,08*	-0,37	0,01*	0,03*	0,06*	0,06*	0,10	0,06*	-0,12
2	0,04*		0,18	0,24	0,10*	-0,02*	0,10*	0,07*	0,18	0,16	0,31	0,12	0,10*	0,01*
3	0,18	0,18		0,18	0,08*	0,08*	-0,07*	0,05*	0,12	0,24	0,20	0,15	0,13	-0,04*
4	0,11	0,24	0,18		0,29	0,32	0,15	0,18	0,25	0,33	0,31	0,17	0,27	0,25
5	-0,03*	0,10*	0,08*	0,29		0,21	0,17	0,09*	0,21	0,27	0,22	0,17	0,20	0,30
6	0,08*	-0,02*	0,08*	0,32	0,21		0,02*	0,18	0,18	0,17	0,19	0,07*	0,30	0,31
7	-0,37	0,10*	-0,07*	0,15	0,17	0,02*		0,13*	0,10*	0,00*	0,06*	0,07*	0,16	0,20
8	0,01*	0,07*	0,05*	0,18	0,09*	0,18	0,13*		0,21	0,19	0,24	0,17	0,16	0,08*
9	0,03*	0,18	0,12	0,25	0,21	0,18	0,10*	0,21		0,38	0,30	0,22	0,21	0,28
10	0,06*	0,16	0,24	0,33	0,27	0,17	0,00*	0,19	0,38		0,25	0,16	0,19	0,18
11	0,06*	0,31	0,20	0,31	0,22	0,19	0,06*	0,24	0,30	0,25		0,28	0,32	0,18
12	0,10	0,12	0,15	0,17	0,17	0,07*	0,07*	0,17	0,22	0,16	0,28		0,29	0,17
13	0,06*	0,10*	0,13	0,27	0,20	0,30	0,16	0,16	0,21	0,19	0,32	0,29		0,26
14	-0,12	0,01*	-0,04*	0,25	0,30	0,31	0,20	0,08*	0,28	0,18	0,18	0,17	0,26	

1 – medycyna katastrof, 2 – kardiologia, 3 – choroby wewnętrzne – interna, 4 – farmakologia, 5 – intensywna terapia i zasady leczenia ostrego bólu, 6 – toksykologia kliniczna, 7 – medycyna ratunkowa – ratownictwo specjalistyczne, 8 – chirurgia, 9 – medyczne czynności ratunkowe – techniki zabiegów medycznych, 10 – medyczne czynności ratunkowe – zaawansowane czynności ratunkowe ALS, 11 – neurologia, 12 – pediatria, 13 – położnictwo i ginekologia, 14 – traumatologia narządów ruchu

Tabela V. Parametry statystyczne modelu regresji dla zmiennych objaśniających: ocena dla dwunastu przedmiotów klinicznych kończących się egzaminem, wiek i płeć studenta oraz zmiennej objaśnianej: średnia ocen ze studiów I stopnia na kierunku Ratownictwo medyczne ($R=0,911$; wartość statystyki $F(14,146)=50,709$; istotność statystyczna modelu $p < 0,00001$; błąd standardowy estymacji $=0,1359$).

Zmienne objaśniająca	b	SE dla b	β	SE dla β	t	p
Wyraz wolny			0,61	0,217	2,82	0,005
Płeć (0 - kobieta; 1 - mężczyzna)	-0,02	0,038	-0,01	0,024	-0,54	0,588
Wiek	-0,03	0,036	-0,01	0,007	-0,93	0,355
Farmakologia	0,27	0,041	0,11	0,016	6,72	< 0,0001
Toksykologia kliniczna	0,13	0,041	0,06	0,018	3,27	0,001
Medycyna katastrof	0,15	0,037	0,07	0,016	4,15	0,0001
Intensywna terapia i zasady leczenia ostrego bólu	0,14	0,038	0,08	0,022	3,62	0,0004
Medyczne czynności ratunkowe - techniki zabiegów medycznych	0,11	0,041	0,07	0,027	2,59	0,01
Kardiologia	0,12	0,038	0,06	0,019	3,03	0,003
Choroby wewnętrzne – Interna	0,10	0,039	0,06	0,022	2,63	0,009
Chirurgia	0,10	0,038	0,04	0,016	2,74	0,007
Pediatria	0,22	0,042	0,09	0,018	5,18	< 0,0001
Neurologia	0,19	0,044	0,09	0,020	4,37	< 0,0001
Traumatologia narządów ruchu	0,18	0,036	0,08	0,017	5,02	< 0,0001
Położnictwo i ginekologia	0,11	0,041	0,05	0,021	2,57	0,01

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

DYSKUSJA

Konieczność budowania autorskiego programu kształcenia dla kierunku ratownictwo medyczne wymagało określenia pożądanych efektów kształcenia oraz systemu oceniania osiągnięć studentów przy zachowaniu ogólnych norm wyznaczonych przepisami prawa m.in. zapisanych w Krajowych Ramach Kwalifikacji. Autonomia uniwersytecka w tym zakresie równocześnie wiąże się ze szczególnie dużą odpowiedzialnością za jakość procesu dydaktycznego prowadzonego w oparciu o taki program kształcenia. Rozwojowi systemu kształcenia zorientowanego na efekty powinna towarzyszyć zmiana sposobu realizowania zajęć dydaktycznych. Kształcenie powinno być ukierunkowane nie na przekazywanie informacji, a na skuteczne pomaganie studentom w zdobywaniu wiedzy i umiejętności oraz na kształtowanie ich postaw [10, 11]. Dążeniu do spełnienia wysokich wymagań jakościowych w odniesieniu do prowadzonych zajęć powinno w równym stopniu dotyczyć systemu oceny kluczowych kompetencji, które są ważne dla zawodu ratownika medycznego. Problematyka pomiaru dydaktycznego w edukacji medycznej oraz kształceniu specjalistów nauk o zdrowiu od zawsze budziła szerokie zainteresowanie środowiska akademickiego. Wiąże się to z faktem, że 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, jeśli prowadzona jest przez niezależny zewnętrzny ośrodek (np. Centrum Egzaminów Medycznych), może być także miarą jakości procesu edukacyjnego [12, 13].

Pomiar edukacyjny zawsze wiąże się z mniej lub bardziej systematycznym gromadzeniem danych z obserwacji, które prowadzą do postawienia wniosków na temat cech i właściwości ocenianego studenta [12]. Dla zachowania wysokiej obiektywności takiego pomiaru konieczne jest spełnienie pewnych kryteriów, określanych jako cechy diagnozy edukacyjnej. Do właściwości pomiaru edukacyjnego, które pozwalają na jego ocenę i optymalizację należy: (I) niezależność sytuacji pomiarowej, (II) obiektywizm punktowania, (III) rzetelność, (IV) trafność oraz (V) obiektywizm pomiaru [9, 14–16]. Jakość pomiaru zależy od zastosowanych metod oceny osiąganych efektów kształcenia, ale także od czynników zewnętrznych (np. demograficznych czy kulturowych), które w istotny sposób mogą wpływać na jego właściwości.

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 niezasłużenie wysokich lub niski 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 [17–19].

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) [17]. 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 *farmakologii* studenci uzyskiwali wyraźnie niższe noty w porównaniu z pozostałymi przedmiotami (surowy egzaminator), podczas gdy dla *intensywnej terapii i zasad leczenia ostrego bólu* oraz *medycznych czynności ratunkowych – techniki zabiegów medycznych* oceny były skrajnie wysokie (nadmiernie łagodny egzaminator). Ponadto dla jednego z analizowanych przedmiotów (*medyczne czynności ratunkowe – zaawansowane czynności ratunkowe ALS*) zaobserwowano szeroki zakres zmienności wyników, co w naturalny sposób odzwierciedla różnicowanie osiągnięć w badanej grupie studentów. Skrajne wartości nie muszą oczywiście świadczyć o braku obiektywizmu punktowania, ale mogą wynikać z faktu, że dany przedmiot wymaga złożonych i trudnych w opanowaniu kompetencji (jak w przypadku *farmakologii* bardzo niska średnia ocen studentów) lub wręcz przeciwnie osiągnięcie efektów kształcenia w danym zakresie jest dla studentów stosunkowo łatwe (wysoka średnia ocen). 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 [20]. Średnia wartość tego współczynnika wyniosła 0,34 i była ona zróżnicowana w kolejnych cyklach kształcenia, 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 uszczywnieniem 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 [16].

Rzetelność pomiaru oznacza 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. W odniesieniu do badań społecznych, Earl Babbie wskazał warunek konieczny, który decyduje o rzetelności: bezstronność warunków pomiarowych oraz precyzja w punk-

towaniu [21]. Niedostateczna rzetelność stosowanej procedury oceniania przyczynia się do niskiego stopnia zaufania, skoro w podobnych okolicznościach poszczególne wyniki różnią się znacząco między sobą. Do oceny czy dany pomiar jest rzetelny można wykorzystać różne metody analityczne. Najczęściej stosowane jest w praktyce określenie stopnia korelacji wyników poszczególnych egzaminów lub ich części (np. *odd-even* lub *split-half reliability*) [22] albo ocena wewnętrznej spójności wyników pomiaru poprzez oszacowanie średnich wariancji ocen dla wszystkich egzaminów (α -Cronbacha – współczynnik 20 Kudera-Richardsona dla testu składającego się z pozycji dwukategorialnych) [23]. Jak pokazują przedstawione wyniki badania rzetelności za pomocą α -Cronbacha, stosowane w latach 2004–2011 ocenianie osiągnięć studentów w obszarze nauk klinicznych charakteryzowało się dostatecznym poziomem rzetelności (spełnione kryterium Jum Nunnally'ego $\alpha > 0,70$, chociaż niektórzy badacze dopuszczają wartość 0,60) [24]. Jednakże analiza poszczególnych roczników wskazuje, że w kolejnych latach poziom wewnętrznej zgodności był zróżnicowany (α od 0,64 do 0,73). 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. W szczególności problem ten może dotyczyć dwóch przedmiotów *medycyny katastrof* oraz *medycyny ratunkowej – ratownictwa specjalistycznego*, dla których to zaobserwowano negatywny wpływ oceny na ogólną zgodność pomiaru. Wyraźnie niższą rzetelność zaobserwowano natomiast w przypadku oceniania z przedmiotów ogólnouniwersyteckich ($\alpha = 0,42$; SEM = 1,54) oraz przedklinikcznych ($\alpha = 0,36$; SEM = 1,36). Jednakże w tym przypadku należy wziąć pod uwagę bardzo duży stopień zróżnicowania ocenianych efektów, które w większości przypadków mogą dotyczyć, szczególnie w kształceniu ogólnym, bardzo odległych cech i właściwości zdających. Inną przyczyną niskiej wartości współczynnika α ($< 0,7$) może być duży udział błędów losowych w wynikach danego pomiaru (SEM dla obszaru klinicznego wyniósł 2,35). Przypadkowe wahania wyników oceniania opisane w klasycznej teorii testu egzaminacyjnego, mogą obniżać wartość współczynnika α [25]. Przy wartości $\alpha < 0,7$ błędy losowe stanowią więcej niż 30% zmienności uzyskanych wyników, a pomiar w takich warunkach według Guilforda może być stosowany właściwie jedynie przy porównywaniu międzygrupowym, a nie do różnicowania indywidualnego [26]. Należy także zauważyć, że dążenie jedynie do uzyskania wysokich wartości współczynnika α nie rozwiązuje problemu rzetelności, ponieważ duża wartość α oznacza jedynie minimalizację wpływu błędów przypadkowych na uzyskane wyniki, nie daje jednak pewności w zakresie istnienia, czasami poważnych, błędów systematycznych związanych ze stronniczością pomiaru [26].

Oprócz określenia stopnia rzetelności pomiaru, która odnosi 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 roznieć, jako stopień zgodności, z jaką narzędzie pomiarowe mierzy to, do mierzenia czego zostało zaprojektowane. Innymi słowy pojęcie trafności dotyczy poprawności wniosków

wyprowadzanych na podstawie wyników badania z użyciem danego narzędzia pomiarowego [27]. Możemy więc mówić o użyteczności danej metody w ocenie konkretnego zestawu cech i właściwości zdającego [28]. Jeśli wybrana metoda sprawdza *de facto* umiejętność dopasowania się studenta do stosowanego narzędzia pomiarowego (dosłownie „*What Do They Want Me To Say?*”), to ocena nie jest ukierunkowana na te cechy, które chcemy mierzyć [29]. 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 (treściowa), wewnętrzna, diagnostyczna, prognostyczna i teoretyczna [9, 16, 25]. W przypadku egzaminów przewidzianych programem kształcenia na kierunku ratownictwo medyczne, istotna w zakresie trafności programowej jest zgodność treści zadań egzaminacyjnych z celami kształcenia dla poszczególnych przedmiotu. 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 interkorelację [30].

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 α -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, a w niektórych przypadkach współczynnik r-Spearmana był ujemny. W zakresie analizy korelacji dla zależności: wyniki kształcenia przedklinikcznego vs. wyniki dla zakresu klinicznego, szczególnie korzystne relacje zaobserwowano w przypadku *anatomii* i w mniejszym stopniu *mikrobiologii*. Dla przykładu, można przytoczyć wartość współczynnika korelacji r-Spearmana dla pary przedmiotów *anatomia – traumatologia narządów ruchu*, który osiągnął poziom 0,45. Wynik ten w doskonały sposób prezentuje pozytywną zależność między oceną z pierwszego przedmiotu (zajęcia w pierwszym roku nauki), a wynikami studenta osiągniętymi w ramach kształcenia z przedmiotu klinicznego (ostatni rok nauki). Korelacja taka dowodzi silnego związku między osiągniętymi efektami kształcenia dla obu przedmiotów. Z drugiej strony zaobserwowano, że dla takich przedmiotów przedklinikcznych jak *fizjologia* i *patofizjologia* brak jest istotnych statystycznie korelacji z wynikami kształcenia z zakresu *chirurgii*. W oparciu o powyższe obserwacje, należy dołożyć wszelkich starań w kierunku szczegółowej analizy jakościowej ocenianych efektów kształcenia w zakresie adekwatności używanych metod pomiaru edukacyjnego.

Analizy interkorelacji między wynikami uzyskanymi przez studentów z poszczególnych przedmiotów klinicznych wykazała, że jedynie w przypadku *farmakologii* osiągnięcia studentów z tego obszaru są istotnie skorelowane ze wszystkimi przedmiotami kształcenia klinicznego. Podobne, ale nieco gorsze wyniki zaobserwowano dla pięciu innych przedmiotów: *medyczne czynniki*

ratunkowe – techniki zabiegów medycznych, medyczne czynności ratunkowe – zaawansowane czynności ratunkowe ALS, neurologia, pediatria oraz położnictwo i ginekologia. Z drugiej strony w przypadku medycyny katastrof i medycyny ratunkowej – ratownictwa specjalistycznego, analiza interkorelacji wskazuje na duży stopień niezgodności w ocenianiu z innymi przedmiotami klinicznymi. Potwierdza to wcześniejsze obserwacje z analizy rzetelności. Brak korelatów międzyprzedmiotowych pozwala postawić tezę, że uzyskane przez studentów wyniki z tych dwóch przedmiotów, 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 szczegółową ocenę efektów kształcenia dla przedmiotów, dla których brak jest pozytywnych korelatów. Bardziej prawdopodobne jest że ocena osiągnięć studentów dla medycyny katastrof i medycyny ratunkowej – ratownictwa specjalistycznego została uzyskana w wyniku źle zaplanowanego egzaminu lub świadczy o niepoprawnym przeprowadzeniu procesu kształcenia. W innym przypadku należałoby podważyć trafność oceny dla pozostałych zakresów kształcenia klinicznego, dla których nie wykazano pozytywnych interkorelacji z tymi przedmiotami.

Przedstawiona powyżej analiza trafności pomiaru edukacyjnego ma za zadanie zapobiegać nadużyciom w interpretacji wyników pomiaru [31]. 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 jakości pomiaru jest szacowanie trafności prognostycznej, która odnosi się do oceny w jakim stopniu dane wyniki kształcenia mogą służyć 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. Podobnie cenne wskazówki co do trafności systemu oceniania można uzyskać w analizie właściwości diagnostycznych pomiaru edukacyjnego. W tym celu konieczne jest sprawdzenie osiągniętych przez studenta/absolwenta kompetencji w zewnątrznie przeprowadzonym sprawdzianie. Jeżeli w przyszłości dla wszystkich absolwentów kierunku w celu uzyskania prawa wykonywania zawodu zostanie wprowadzony obowiązek zdania Państwowego Egzaminu Ratownictwa Medycznego (PERM), to będzie możliwe także zweryfikowanie jakości procesu dydaktycznego oraz trafności diagnostycznej systemu oceniania studentów na danej uczelni.

Ze względu na fakt, że przyjmowanie w badaniach trafności, jako wskaźnika oceny, współczynnika korelacji rang Spearmana jest uproszczonym jednowymiarowym modelem, znacznie bardziej wiarygodne wyniki można uzyskać z wielowymiarowej analizy z użyciem metody regresji wielorakiej. Zaproponowany model regresji uwzględniał czternaście zmiennych niezależnych (objaśniających): wiek i płeć studenta oraz oceny dla dwunastu przedmiotów kończących się egzaminem. Wyniki analizy regresji prezentowane w niniejszej pracy, potwierdzają obserwacje uzyskane z jednowymiarowej analizy korelacji. Wyznaczone

standaryzowane współczynniki β w przedziale od 0,04 do 0,11 dla dwunastu przedmiotów, wskazują na ich zróżnicowany stopień oddziaływania na wyniki końcowe mierzone średnią ocen z toku studiów. W nawiązaniu do wyników uzyskanych z analizy interkorelacji, także prezentowany model regresji potwierdza kluczową rolę oceny uzyskanej przez studenta z farmakologii. Także dobrym prognostykiem są wyniki uzyskane z zakresu neurologii i pediatrii ($\beta = 0,09$). Wyniki kształcenia z tych trzech przedmiotów są najsilniejszym prognostykiem sukcesu studenta w postaci odpowiednio wysokiej średniej ocen na koniec studiów.

W zaleceniach dotyczących analizy regresji wielorakiej, podaje się jako ważne predyktory, zmienne demograficzne, takie jak: wiek oraz płeć studenta [32]. Opublikowane badania dotyczące oceniania na studiach medycznych wskazują, że generalnie to kobiety lepiej radzą sobie w trakcie studiów niż mężczyźni [33, 34]. Jak podkreślają Ferguson i wsp. na podstawie systematycznego przeglądu piśmiennictwa, zaleca się uwzględnić zmienną płeć, jako jednego z ważnych czynników (zmienna niezależna) w badaniach predykcyjnych [32]. Biorąc pod uwagę powyższe obserwacje w modelu regresji uwzględniono zmienne płeć i wiek. Jednakże wyniki analizy nie wskazują na znaczącą rolę tych zmiennych jako istotnych predyktorów ($p > 0,05$). Niemniej jednak, biorąc pod uwagę brak równoważności poszczególnych roczników pod względem struktury wiekowej oraz płci, należy powyższe wyniki traktować z ograniczonym zaufaniem. Uzyskanie równolicznych grup studentów i zbliżonej struktury wiekowej wymaga wyodrębnienia próby w losowaniu warstwowym. Ponadto analiza musiałaby dotyczyć tylko tych roczników, dla których w cyklu kształcenia przewidziano identyczne przedmioty o tym samym zakresie mierzonych efektów kształcenia. Powyższe uwarunkowania są trudne do spełnienia przy obecnej dostępności danych.

Rzetelność i trafność to dwie hierarchicznie powiązane ze sobą cechy, dlatego też zawsze dla danego pomiaru należy obie właściwości rozpatrywać łącznie. Stosunkowa wysoka rzetelność pomiaru w zakresie przedmiotów klinicznych przyczynia się do dość dużej dokładności (precyzji) oceny wiedzy i umiejętności zdających, a co za tym idzie dobrego różnicowania indywidualnego. Jednakże spore zastrzeżenia budzą wyniki analizy trafności. W niektórych przypadkach można uznać, że wyniki uzyskiwane przez studentów nie odzwierciedlają rzeczywistych właściwości i cechy zdających. Centralną rolę w badanym systemie oceniania spełnia przedmiot farmakologia, której kluczową rolę potwierdza zarówno analiza korelacji jak i regresji. Należy jednak w pogłębionej analizie programu kształcenia dokonać krytycznego osądu, na ile przedmiot ten rzeczywiście reprezentuje zakres kompetencji istotnych dla wszystkich pozostałych obszarów klinicznych. Natomiast przeciwne relacje zaobserwowano w przypadku medycyny ratunkowej - ratownictwa specjalistycznego, co także wymaga jakościowej oceny reprezentowanych przez ten przedmiot końcowych efektów kształcenia oraz adekwatności proponowanych metod pomiaru edukacyjnego. Sytuacja w której dość dobra rzetelność idzie w parze z niską trafnością skutkuje obniżeniem wiarygodności całego systemu oceniania kompetencji. Ma to szczególne znaczenie dla oceny 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 ratownictwa medycznego.

WNIOSKI

Każda z trzech strategii analitycznych – szacowanie współczynnika korelacji r-Spearmana, Tau Kendalla oraz α -Cronbacha, mimo że ma różne podstawy logiczne i formuły obliczeniowe, prowadzi do podobnych wniosków. Średni stopień zgodności oceniania osiągnięć studentów w kolejnych latach świadczy o dostatecznej spójności systemu pomiaru efektów kształcenia na kierunku ratownictwo medyczne. Mimo ograniczonego zakresu analizy trafności, można wskazać na niski stopień adekwatności oceny osiągniętych efektów kształcenia dla niektórych przedmiotów z grupy nauk klinicznych. Krytyczna ocena obecnie stosowanych rozwiązań pozwala na zaplanowanie szczegółowej ewaluacji stosowanych kryteriów oceniania w oparciu o zidentyfikowane słabe punkty w obszarze nauczania. Konieczne jest także w przyszłości oszacowanie trafności prognostycznej i diagnostycznej stosowanych metod oceny kompetencji studentów ratownictwa medycznego w WUM.

PIŚMIENNICTWO

1. Ustawa z dnia 25 lipca 2001 r. o Państwowym Ratownictwie Medycznym (Dz.U. 2001 nr 113 poz. 1207).
2. Ustawa z dnia 8 września 2006 r. o Państwowym Ratownictwie Medycznym (Dz.U. 2006 nr 191 poz. 1410).
3. Poździoch S. Ustawa o Państwowym Ratownictwie Medycznym - komentarz. 2. ed. Warszawa: Wolters Kluwer; 2013.
4. Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 12 lipca 2007 r. w sprawie standardów kształcenia dla poszczególnych kierunków oraz poziomów kształcenia, a także trybu tworzenia i warunków, jakie musi spełniać uczelnia, by prowadzić studia międzykierunkowe oraz makrokierunki (Dz.U. 2007 nr 164 poz. 1166).
5. Jakubaszko J, Boznański A. Ratownik medyczny: podręcznik. Górnicki Wydawnictwo Medyczne, Wrocław, 2010.
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;1:137–144.
8. Zawadzki A. Medycyna ratunkowa i katastrof. Wydawnictwo Lekarskie PZWL, Warszawa, 2007.
9. Niemierko B. Diagnostyka edukacyjna. Wydawnictwo Naukowe PWN, Warszawa, 2009.
10. Rydzewska-Włodarczyk M. Efekty kształcenia i kompetencje w programach studiów. Folia Pomeranae Universitatis Technologiae Stetinensis Oeconomica. 2009;273(56):163–170.
11. Wyrozębski P. Podejście do tworzenia programów nauczania oparte na efektach kształcenia. e-Mentor. 2009;3(30):18–25.
12. Schuwirth LW, van der Vleuten CP. General overview of the theories used in assessment: AMEE Guide No. 57. Med Teach. 2011;33(10):783–797.
13. Harden RM, Crosby JR, Davis MH. AMEE Guide No. 14: Outcome-based education: Part 1 - An introduction to outcome-based education. Medical Teacher. 1999;21(1):7–14.
14. Norman GR, Vleuten C, Newble DI. International handbook of research in medical education: Springer, 2002.
15. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R et al. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 conference. Medical Teacher. 2011;33(3):206–214.
16. Niemierko B. Pomiar wyników kształcenia. Wydawnictwo Szkolne i Pedagogiczne, Warszawa, 1999.
17. Rowley J. Measuring quality in higher education. Quality in Higher Education. 1996;2(3):237–255.
18. Tam M. Measuring Quality and Performance in Higher Education. Quality in Higher Education. 2001;7(1):47–54.
19. Sood R, Singh T. Assessment in medical education: Evolving perspectives and contemporary trends. National Medical Journal of India. 2012;25(6):357–364.
20. Kendall MG. A new measure of rank correlation. Biometrika. 1938;8:1–93.
21. Babbie E. The practice of social research. 13th ed. Cengage Learning, Belmont, 2013.
22. Guttman L. A basis for analyzing test-retest reliability. Psychometrika. 1945;10(4):255–282.
23. Feldt LS. A test of hypothesis that Cronbachs alpha or Kuder-Richardson coefficient 20 is same for 2 tests. Psychometrika. 1969;34(3):363–367.
24. Nunnally JC, Bernstein IH. Psychometric theory. 3 edn. New York: McGraw-Hill New York; 1967.
25. Niemierko B. Testy osiągnięć szkolnych. Podstawowe pojęcia i techniki obliczeniowe. 1st ed. Warszawa: Wydawnictwo Szkolne i Pedagogiczne; 1975.
26. Guilford JP. Psychometric methods. 2nd ed. New York: McGraw-Hill; 1954.
27. Moss PA. Shifting conceptions of validity in educational measurement: Implications for performance assessment. Review of Educational Research. 1992;62(3):229–58.
28. Goodwin LD. Changing conceptions of measurement validity: an update on the new standards. The Journal of nursing education. 2002;41(3):100–106.
29. White J, Brownell K, Lemay JF, Lockyer JM. „What do they want me to say?” The hidden curriculum at work in the medical school selection process: a qualitative study. BMC Medical Education. 2012;12:17.
30. Meagher DG, Pan T, Wegner R, Olson AT, Overgaard SL, Mehle JJ. PCAT Reliability and Validity. 3rd ed. Pearson Executive Office, San Antonio, 2012.
31. Kubielski W. Podstawy pomiaru, konstruowania i ewaluacji testu dydaktycznego: Wydawnictwo Wyższej Szkoły Pedagogicznej TWP; 2006.
32. Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. BMJ. 2002;324(7343):952–957.
33. Pai MRS, Sanji N, Pai PG, Kotian S. Comparative assessment in pharmacology multiple choice questions versus essay with focus on gender differences. Journal of Clinical and Diagnostic Research. 2010;4(3):2515–2520.
34. McDonough CM, Horgan A, Codd MB, Casey PR. Gender differences in the results of the final medical examination at University College Dublin. Med Educ. 2000;34(1):30–34.

ADRES 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. (22)520 490
e-mail: mariusz.panczyk@wum.edu.pl

Nadesłano: 20.02.2015

Zaakceptowano: 21.03.2016