

Badanie trafności egzaminu z przedmiotu „Diagnoza psychologiczna”¹

Władysław Jacek Paluchowski*

Uniwersytet im. Adama Mickiewicza, Poznań

Piotr Haładziński

Uniwersytet im. Adama Mickiewicza, Poznań

INVESTIGATING THE VALIDITY OF PSYCHOLOGICAL ASSESSMENT ACADEMIC COURSE EXAM

In this article the validity of a psychological assessment knowledge test is discussed. Validity is an absolutely crucial characteristic of the educational test. It provides information on the adequacy of operationalization and the level to which the test meets established demands. An efficient method to define construct validity might be the analysis of non random changes in test scores. It applies comparing test scores – before and after inputting experimental manipulation. The results ought to reflect the causal effect of the manipulation. This method can be successfully implemented in determining validity of an educational test-by investigating the impact of class attendance on test scores. In this study we compare scores for psychological assessment knowledge test. Two comparison groups were formed on the basis of completing the academic course of psychological assessment. The analysis revealed a significant interaction between the completion of the course and the test scores, confirming that the test scores varied significantly depending on the completion of the psychological assessment course. In the latter section we conducted an analysis of the questions which differentiated comparison groups.

WPROWADZENIE

Wymóg trafności stosowanych narzędzi jest oczywisty w odniesieniu do testów psychologicznych. Mniej oczywisty jest odnośnie do testów dydaktycznych. W dalszej części przedstawiamy dane dotyczące trafności testu wiadomości² z przedmiotu „Diagnoza psychologiczna”.

TESTY DYDAKTYCZNE I ICH CECHY

W dydaktyce rozróżnia się – biorąc za kryterium funkcje, jakie spełniają w procesie nauczania – trzy rodzaje ocen: diagnostyczne (informujące o zdolnościach uczniów), formacyjne (sprawdzian kształtujący, ukierunkowujący; pozwalający na kontrolę bieżącą i sprzyjający doskonaleniu procesu dydaktycznego w sensie „jak się dalej uczyć” oraz „jak najlepiej nauczać”) i sumaryczne (sprawdzian sumujący, sumatywny, poświadczający, atestujący; umożliwiający kontrolę końcową i nastawiony na oszacowanie wyników procesu kształcenia)³. Sprawdzanie osiągnięć studentów jest ważne zarówno dla prowadzących zajęcia, studentów i ich przyszłych pracodawców. Na studiach psychologicznych do sprawdzania wiedzy (głównie teoretycznej) służą sprawdziany sumujące.

W ostatnich latach głównym sposobem sprawdzania wiedzy studentów są – zamiast egzaminów ustnych – testy wiadomości. Jest to forma akceptowana przez studentów ze względu na ten sam stopień trudności i ten sam zakres wiedzy dla wszystkich zdających, możliwość spokojnego zastanowienia się nad odpowiedzią, poprawienia błędnej odpowiedzi oraz obiektywne kryteria oceniania⁴. Także z punktu widzenia egzaminatorów pisemne testy wiedzy mają przewagę nad ustnymi egzaminami – pozwalają szybciej sprawdzić i ogłosić wyniki (szczególnie gdy osób zdających jest dużo) oraz łatwiej udokumentować brak stroniczości przy ocenianiu. Niewątpliwie testy pisemne mają też wiele wad: nie dają wglądu w proces myślenia osoby egzaminowanej i wiedzy o jej sprawności komunikacyjnej, nie dają jej informacji zwrotnej o rodzaju błędów popełnionych podczas uczenia się, mogą dawać większą szansę na dobry wynik w efekcie niesamodzielnej pracy.

¹ Wykorzystano m. in. wyniki pracy magisterskiej Agaty Król „Elementy analizy ilościowej i jakościowej testu wiadomości z przedmiotu »Diagnoza psychologiczna«”, za których udostępnienie dziękujemy.

² W tekście będziemy zamiennie używać terminu „test wiadomości” i „test wiedzy”.

³ Używając analogii: jeżeli kucharz smakuje zupę to jest to ocena kształtująca, jeżeli robi to klient, to jest to ocena podsumowująca.

⁴ Na różnych studenckich forach podaje się też argument łatwiejszego ściągania.

* Korespondencję dotyczącą artykułu można kierować na adres: Władysław J. Paluchowski, Instytut Psychologii UAM, ul. Szamarzewskiego 89, 60-568 Poznań.

e-mail: Wladyslaw.Paluchowski@amu.edu.pl

Pomiar sprawdzający, zwany też pomiarem odnoszonym do kryteriów (*criterion-referenced measurement*) to pomiar dydaktyczny, w którym wynikowi każdego badanego ucznia jest przyporządkowana pewna wartość skali pomiarowej, pozwalająca na wskazanie poziomu opanowanej wiedzy z zakresu programu⁵ (Niemierko, 1990; Hornowska, 2009). W dydaktyce stosowany jest również pomiar różnicujący, zwany pomiarem odnoszonym do norm (*norm-referenced measurement*), w którym wynik badania wiedzy ucznia porównujemy z przeciętnym poziomem wyników innych osób z określonej grupy odniesienia (Niemierko, 2009).

Testy stosowane w dydaktyce dzielą się na wielostopniowe i jednostopniowe. W teście jednostopniowym wszystkie zadania traktuje się jako równorzędne (dla danego zakresu treści kształcenia), różnice między uczniami odzwierciedla liczba poprawnie rozwiązanych przez nich zadań. Istotą testu sprawdzającego wielostopniowego jest hierarchizacja wymagań (Niemierko, 1999, s. 54). W klasycznym teście wielostopniowym przygotowuje się odrębne grupy zadań (o różnym udziale w całym teście) dla każdego poziomu wiedzy (treści konieczne, podstawowe, rozszerzone, dopełniające i wykraczające poza program kształcenia; na każdym poziomie równorzędne). Formą testu wielostopniowego jest test hierarchiczno-kryterialny, polegający na tworzeniu zadań otwartych, by można je było rozwiązać na wielu poziomach – od intuicyjno-zdroworozsądkowego do uzasadnionych literaturą. Korzystanie z podręczników i notatek podczas takiego testu jest dozwolone, a nawet zalecane. Tutaj każde z tych rozwiązań jest akceptowane, ale punktacja silnie różnicuje poziomy (Niemierko, 1991).

Zadania testowe mogą być zamknięte lub otwarte, jednokrotnego lub wielokrotnego wyboru (kiedy więcej niż jedna odpowiedź jest prawidłowa). Zadania otwarte nie mają gotowej propozycji odpowiedzi, student samodzielnie formułuje i wpisuje rozwiązanie; w zadaniach zamkniętych student wybiera odpowiedź spośród gotowych. W zadaniach otwartych można wymagać krótkiej (1–2 słowa czy liczby) lub długiej wypowiedzi osoby egzaminowanej albo uzupełnienia zwrotu, zdania, fragmentu tekstu itp. Typowe pisemne zadanie zamknięte składa się z trzonu (w którym zawarty jest problem bądź pytanie), odpowiedzi prawidłowej oraz dystraktorów. Zestaw gotowych odpowiedzi do wyboru nazywany jest kafeterią. Odpowiedź prawidłowa lub najlepsza z gotowych odpowiedzi podanych w zadaniu zamkniętym nazywana jest werstraktorem. Pozostałe proponowane przez autora testu odpowiedzi stanowią dystraktory, czyli odpowiedzi mogące stwarzać pozory prawidłowych lub stanowić alternatywę dla właściwego rozwiązania. Generalnie ich

zadaniem jest zakłócanie procesu wyboru odpowiedzi, odciąganie uwagi od odpowiedzi poprawnych dzięki powierzchniowemu podobieństwu do odpowiedzi poprawnych lub odwołaniu się do „ekonomii poznawczej” (łatwości) albo do błędnej wiedzy. Dystraktory – jeśli nie mają być tylko pozorne – nie powinny odbiegać znacznie pod względem treściowym i formalnym od odpowiedzi prawidłowej, nie powinny być banalne i w sposób oczywisty niedorzeczne. Dystraktory, podobnie jak werstraktory, można analizować określając ich atrakcyjność oraz moc różnicującą (związek z niskim wynikiem w teście wiedzy) (Niemierko, 2009). Trzon zadania zamkniętego w formie zdania prostego lub złożonego może wymagać potwierdzenia prawdziwości jakiejś tezy, zdefiniowania czegoś, określenia celu, ustalenia przyczyny, przewidywania skutku, ustalenia związku, znajdowania (wskazywania) błędu, udzielenia odpowiedzi na pytania o to, jaką zasadę naruszono albo jak skorygować błędne działanie lub jego skutki, ustalenia różnicy czy podobieństwa, porządkowania czy też uzasadnienia opinii (Niemierko, 1999, s. 121).

Zadania testu dydaktycznego (testu wiedzy) mogą być dla danej grupy łatwiejsze lub trudniejsze. Łatwość zadania (*item easiness*) to stosunek liczby punktów uzyskanych przez testowanych studentów za rozwiązanie danej pozycji skalowej, do maksymalnej możliwej do osiągnięcia przez nich liczby punktów. Jej odwrotnością jest trudność zadań testowych. Jeśli zadania oceniane są jako prawidłowo i nieprawidłowo rozwiązane (1 i 0) to proporcja osób udzielających poprawnej odpowiedzi jest wskaźnikiem łatwości, a udzielających nieprawidłowej odpowiedzi wskaźnikiem trudności zadania testowego [może być to proporcja wśród wszystkich zdających – wtedy w grę wchodzi i osoby, które opuściły zadanie lub do niego nie dotarły i osoby odpowiadające nieprawidłowo – lub tylko wśród osób rozwiązujących dane zadanie].

Trafność testu wiadomości jest jego najważniejszą charakterystyką. Wg Hornowskiej (2009) trafność jest to stopień adekwatności operacjonalizacji danej wielkości oraz stopień, w jakim test jest w stanie osiągnąć stawiane mu cele. Jej wartość liczbową wskazuje na wielkość błędu spowodowanego rozbieżnością między zamierzeniami autora testu a tym, co test faktycznie mierzy. W pedagogice większy nacisk kładzie się na trafność szacowania aktualnych i/lub przyszłych osiągnięć uczniów (Niemierko, 1990). W testowaniu wyróżnia się⁶ trafność wewnętrzną testu osiągnięć szkolnych (zgodność treści testu z treściami programu nauczania; trafność treściowa), trafność diagnostyczną (zgodność wyników testu z wynikami innych sposobów badania wiedzy; trafność kryterialna) oraz trafność prognostyczną (zgodność wyników testu z wynikami późniejszych osiągnięć; inna forma trafności kryterialnej). Dla testów dydaktycznych największe znaczenie ma trafność treściowa – odnoszona do wymagań programowych.

⁵ Częste definiowanie w pedagogice pomiaru odnoszonego do kryteriów jako pomiaru, w którym układem odniesienia wyniku każdego ucznia są wymagania programowe może prowadzić do nieporozumień – oczekiwane osiągnięcia uczniów nie muszą oznaczać opanowania treści podręcznika w 100%. Lepsze jest określenie oczekiwanego poziomu kompetencji.

⁶ Jak pisze Hornowska (2007) współcześnie nie rozróżnia się rodzajów trafności, ale różne źródła danych o trafności.

Trafność kryterialną testu wiedzy ocenić można określając korelację danego testu i innymi miarami, co do których można założyć, że wiedza na dany temat jest z nimi powiązana. Można też sprawdzać trafność testu wiedzy przez porównanie wyników testu i rzeczywistego zachowania. Hornowska pisze (2009), że jedną z form badania teoretycznej trafności testu może być analiza zmian nieprzypadkowych wyników testu – porównywanie wyników dwukrotnego badania tym samym testem gdzie w przerwie między badaniami wprowadza się manipulację eksperymentalną. Wynik porównywania powinien być zgodny z założonymi efektami manipulacji. Wydaje się, że podobnie można postępować w odniesieniu do trafności testu wiadomości, badając wpływ udziału w zajęciach na wyniki w takim teście.

PROBLEM, HIPOTEZY I ICH UZASADNIENIE

Celem podjętych badań było sprawdzenie, czy studenci, którzy zdają kursowy egzamin z przedmiotu „Diagnoza psychologiczna” uzyskują lepsze wyniki w teście wiadomości z tego przedmiotu, niż studenci z innych lat. Inaczej mówiąc sprawdzono, czy przygotowywanie się do egzaminu (chodzenie na wykłady, czytanie zalecanej literatury przedmiotu) wpływa na wyniki testu wiedzy z zakresu diagnozy psychologicznej. Wyniki tych osób zestawiono z osobami z pierwszego roku studiów stacjonarnych, którzy nigdy nie mieli tego przedmiotu oraz czwartego roku takich studiów, czyli tymi, którzy zdawali taki egzamin wcześniej. Pierwsze pytanie badawcze brzmiało: Czy odbycie kursu z przedmiotu „Diagnoza psychologiczna” ma dodatni wpływ na wyniki uzyskane przez studentów w teście wiadomości z tego przedmiotu? Było to pośrednie badanie trafności wykorzystanego testu wiadomości. Gdyby się okazało, że faktycznie taki wpływ daje się zaobserwować, można postawić następane pytanie: Czym różnią się formalnie i treściowo zadania które różnicują między grupami?

METODA BADAWCZA

Test wiadomości z przedmiotu „Diagnoza psychologiczna” składa się z 40 zadań zamkniętych wielokrotnego wyboru. Każda pozycja zawierała kafeterię od trzech do pięciu propozycji rozwiązania. Pytanie oparte były na wskazanej studentom literaturze (Paluchowski, 2006, 2007; Stemplewska-Żakowicz, Paluchowski, 2008) oraz treści wykładów.

Dla celów egzaminacyjnych przygotowano 4 wersje (zestawy) testu. Zestawy zadań zostały wylosowane przez program komputerowy Quest v.1.1. firmy Question Mark Computing, z puli 550 zadań. Program przydzielił losowo zadania do zestawów, losowo ustalił również kolejność odpowiedzi do wyboru. Z gotowych odpowiedzi student musi wybrać zazwyczaj jedną prawidłową. Czas na wykonanie całości wynosił 35 minut.

OSOBY BADANE I PRZEBIEG BADAŃ

Łącznie zbadano 320 studentów psychologii UAM: 30 studentów pierwszego roku trybu stacjonarnego, 64 drugiego roku trybu niestacjonarnego, 196 trzeciego roku trybu stacjonarnego oraz 30 studentów czwartego roku. Studenci są obcy z tą formą egzaminowania. Wykonywanie testów podczas egzaminów (rok drugi i trzeci) oraz badań studentów pierwszego i czwartego roku przebiegało w typowy sposób: studenci otrzymali cztery różne warianty testu, dodatkowe osoby nadzorowały przebieg, a ograniczony czas na rozwiązanie zdecydowanie utrudniał próby kontaktu pomiędzy testowanymi. Na miejsce badania dostarczono odpowiednią liczbę arkuszy testowych oraz kart odpowiedzi, a pomieszczenia były przygotowane na przewidywaną liczbę osób. Każda z grup podchodzących do egzaminu (badania) otrzymała instrukcje ustne od osób kierujących przebiegiem testowania, dodatkowo na kartach odpowiedzi zawarte były informacje o konieczności podania swoich danych osobowych wraz z numerem indeksu, oznaczenia karty literą symbolizującą wersję rozwiązywanego kwestionariusza oraz polecenie sposobu zakreślania wybranych rozwiązań.

REZULTATY

Średnie wyniki w poszczególnych grupach przedstawiono w Tabeli 1 (wynik progowy to 25 punktów – wynik poniżej oznaczał ocenę niedostateczną).

Tabela 1
Średnie wyniki badanych grup

	średnia	odch. statyst.
rok 1	16.33	3.68
rok 2	24.16	4.27
rok 3	27.04	4.47
rok 4	19.43	4.36
rok 5	24.75	5.61

Trudność zadań wahała się od 0 do 1. Jej krańcowe wartości w poszczególnych wersjach przedstawiono w Tabeli 2.

Aby mieć podstawę do traktowania wersji testu wiadomości jako równoległych, zastosowano test Levene’a oraz 1-czynnikową analizę wariancji. Okazało się, że wariancje i odchylenia standardowe we wszystkich czterech grupach są równe (F -Levene’a_{3, 316}=1.73; p_a =.16), a różnica między średnimi jest nieistotna statystycznie ($F_{3, 316}$ =.9166; p_a =.4331). Na tej podstawie przyjęto, że do dalszej analizy wykorzystane będą – na równych prawach – wszystkie cztery wersje testu wiedzy.

WERYFIKACJA HIPOTEZ

Chcąc sprawdzić, czy wykonanie testu wiadomości jest różne w badanych grupach najpierw sprawdzono jedno-

rodność wariancji – ze względu na duże nierówności liczebności grup – testem Browna-Forsythe’a ($F_{3,316}=.4669$; $p_a=.7056$). W związku z tym, że wariancje okazały się jednorodne, zastosowano następnie 1-czynnikową analizę wariancji i okazało się, że różnica między średnimi jest istotna statystycznie ($F_{3,316}=70.6716$; $p_a=.0001$). Aby sprawdzić, które różnice są istotne ($p_a \leq .05$), wykorzystano test Scheffé’go. Nieistotna statystycznie okazała się jedynie różnica między studentami pierwszego i czwartego roku trybu stacjonarnego ($p_a \leq .05714$)⁷. W efekcie stworzono dwie grupy: „bez wykładu”, obejmująca studentów pierwszego roku oraz czwartego roku trybu stacjonarnego oraz „po wykładzie”, w której znaleźli się studenci drugiego roku trybu niestacjonarnego i trzeciego roku trybu stacjonarnego.

Tabela 2
Wartości wskaźnika trudności zadań w poszczególnych grupach i wersjach testu

	wersja testu	wartość min. (trudne)	wartość maks. (łatwe)
grupa „bez wykładu”	A	.071	.857
	B	.000	.875
	C	.067	.933
	D	.067	.933
grupa „bez wykładu”	A	.081	.959
	B	.174	1.000
	C	.148	.963
	D	.190	.936

Tabela 3
Relacja między trudnością zadań grupy „bez wykładu” i grupy „po wykładzie”

		grupa „bez wykładu”				
		0 – .19 (b. trudne)	.20 – .49	.50 – .69	.70 – .89	.90 – 1 (b. łatwe)
grupa „po wykładzie”	0 – .19 (b. trudne)	0	1	0	0	0
	.20 – .49	1	0	1	0	0
	.50 – .69	4	12	0	0	0
	.70 – .89	4	13	5	0	0
	.90 – 1 (b. łatwe)	1	6	11	4	0

Ze względu na charakter zmiennych i wielkość grup, analizując różnice między sposobem odpowiadania na poszczególne zadania w zestawach wykorzystano test dokładnego prawdopodobieństwa Fishera; do dalszej analizy brano te wyniki, dla których $p_a \leq .05$ (test dwustronny).

Relację między grupami wśród zadań różnicujących między trudnością zadań grupy „bez wykładu” i grupy „po wykładzie” pokazuje Tabela 3 (niezależnie od wersji testu). Generalnie więc różnice obserwowano tam, gdzie w grupie „po wykładzie” były zadania umiarkowanie trudne i łatwe, a w grupie „bez wykładu” relatywnie trudne i umiarkowanie trudne (por. Tabela 3).

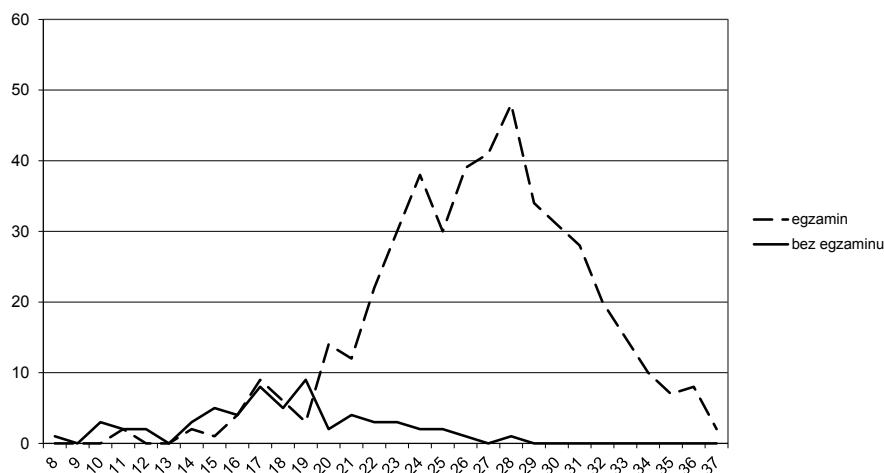
OMÓWIENIE WYNIKÓW

Potwierdziła się hipoteza, że odbycie kursu z przedmiotu „Diagnoza psychologiczna” ma dodatni wpływ na wyniki uzyskane przez studentów w teście wiadomości z tego przedmiotu. Zgodnie z oczekiwaniem najniższe wyniki osiągnęli studenci pierwszego roku trybu stacjonarnego, następnie czwartego roku oraz kolejno studenci drugiego roku trybu niestacjonarnego i trzeciego roku trybu stacjonarnego. Wyniki te potwierdzają trafność egzaminu z przedmiotu.

Choć generalnie globalne wyniki testu wiadomości różnicują grupy, to średnio tylko 39.4% zadań różnicuje grupy studentów „po wykładzie” i „bez wykładu”. Inaczej mówiąc, bez wysłuchania wykładu, a jedynie na podstawie wiedzy z innych zajęć (a nawet spoza studiów) można udzielić średnio 16 poprawnych odpowiedzi. Można z tego wyciągnąć wniosek, że test był trafny, ale stosunkowo łatwy: w grupie osób „po wykładzie” średni wynik

to 67.6% wyniku maksymalnego (dla studentów niestacjonarnych 60.4%). Jeszcze silniej widać to przy analizie wyników w grupie osób „bez wykładu”: średni wynik to 40.8% wyniku maksymalnego dla studentów pierwsze-

⁷ Sprawdzono także tę różnicę testem *t* Studenta – i w tej analizie różnica okazała się istotna statystycznie ($t = -2.976$; $p_a = .0042$).



Ryc. 1. Wyniki testu wiadomości w obu grupach

go roku i 48.6% dla studentów czwartego roku. Jednak porównanie rozkładu wyników w dwóch podstawowych grupach – po i bez wykładu – wyraźnie pokazuje, że w grupie studentów „bez wykładu” wyniki w większości lokują się po stronie niskich (zdecydowanie poniżej wyniku progowego, czyli 25 punktów) (Rycina 1).

Biorąc pod uwagę treść pytań, różnicujące pytania dotyczyły: rodzaju danych wg Cattella, standaryzacji, kwestionariuszowych skal kontrolnych, hipotezy dewiacyjnej Berga, charakterystyki i okoliczności powstania systemów klasyfikacji DSM i ICD, istoty wyjaśnień psychologicznych. Inaczej mówiąc istotną trudność sprawiały grupie „bez wykładu” odpowiedzi na pytania dotyczące specyficznych treści zawartych w zalecanej literaturze (i przekazywanych podczas wykładu).

Szukając wspólnych cech formalnych zadań różnicujących analizowano długość trzonu oraz całości, liczbę alternatyw do wyboru oraz pozycje (miejsce) tak odpowiedzi prawidłowych, jak i dystraktorów. Nie zauważono żadnych cech charakterystycznych dla tych pozycji. Także w odniesieniu do miernika łatwości czytania Flescha (*Flesch Reading Ease Level*) jak i miernika szkolnego Flescha-Kincaida (*Flesch-Kincaid Grade Level*) oraz wskaźnika Mglistości Gunninga (*Gunning Fog Index*) nie zauważono poważnych różnic, choć zadania nie różnicujące w poszczególnych wersjach wydają się trochę trudniejsze do zrozumienia (por. Tabela 4⁸).

Można też (na podstawie Tabeli 2) zaryzykować wniosek, iż studenci z grupy „bez egzaminu” lepiej odpowiadają na zadania, których treść jest zbliżona do języka potocznego oraz przywołuje codzienne okoliczności, niż

na polecenia napisane w ściśle specjalistycznym języku. Rozwiązując te pozycje studenci mogli posłużyć się również wiedzą z innych obszarów psychologii (psychometria, statystyka, etyka).

Jak pisze Król (2001, str. 51) „...zaobserwowano wiele spontanicznych (...) reakcji studentów: ci z I roku, mimo wyrażenia zgody na badanie i entuzjazmu, byli zupełnie zdezorientowani zadaniami testu i ... oceniali egzamin jako ‘abstrakcyjny’ i ‘bardzo trudny’. Odwrotnie niż studenci IV roku, którzy ... pytania uznawali za ‘całkiem oczywiste’ (jeden z wielu komentarzy), a niepewność co do prawidłowych odpowiedzi przypisywali swojej niepamięci lub nie dość uważnej i skutecznej nauce (wielokrotnie powtarzany komentarz ‘trzeba się było lepiej tego nauczyć’)”.

Tabela 4
Porównanie wskaźników czytelności dla zadań

	zadania różnicujące	zadania nie różnicujące
wskaźnik łatwości czytania Flescha	23.09	28.29
wskaźnik szkolny Flescha-Kincaida	15.55	15.24
średnia liczba sylab w słowie	1.91	1.82
wskaźnik Gunning Fog	23.39	27.02

Powodem do refleksji może być stosunkowo niski wynik studentów czwartego roku trybu stacjonarnego (por. Tabela 1). Potwierdziły się oczekiwania, że wypadną oni w teście wiadomości lepiej niż studenci pierwszego roku (różnica była – jak wspomniano – istotna statystycznie), ale jednak uzyskali oni średni wynik zdecydowanie poniżej progów kwalifikacyjnych. Gdyby założyć, że ich motywacją do uczenia się nie jest jedynie chęć zdania

⁸ Obliczenia wykonano za pomocą programu Flesh 2.0 (<http://flesh.sourceforge.net>) oraz strony Simona Bonda (<http://gunning-fog-index.com/fog.cgi>). Znaczenia tych wskaźników opisuje Anna Seretny (2006).

egzaminu, ale też (głównie?) zdobywanie zawodowych kompetencji (Stemplewska-Żakowicz i Paluchowski, 2008), to ich wynik musi niepokoić. Zbyt skutecznie bowiem „pozbywają się” oni wiadomości nabytych podczas zajęć z przedmiotu „Diagnoza psychologiczna” w efekcie udziału w innych zajęciach i praktykach studenckich.

Przedstawione powyżej wyniki badania dowodzą, że egzamin z przedmiotu „Diagnoza Psychologiczna” jest trafnym narzędziem pomiarowym. Biorąc pod uwagę fakt, że testy wiadomości są najchętniej wykorzystywaną metodą ewaluacji osiągnięć studentów na wyższych uczelniach, badania trafnościowe dotyczące egzaminów powinny być standardowym krokiem w projektowaniu programu dydaktycznego, bowiem tylko na podstawie wyników trafnego testu można adekwatnie ocenić stopień realizacji założonych celów dydaktycznych.

LITERATURA

- Hornowska, E. (2007). Testy psychologiczne w świetle Standardów dla procedur stosowania testów w psychologii i edukacji. W: W.J. Paluchowski (red.), *Diagnoza psychologiczna. Proces, narzędzia, standardy* (s. 198–210). Warszawa: Wydawnictwa Akademickie i Profesjonalne.
- Hornowska, E. (2009). *Testy psychologiczne. Teoria i praktyka*. Warszawa: Wydawnictwo Naukowe Scholar.
- Król, A. (2011). *Elementy analizy ilościowej i jakościowej testu wiadomości z przedmiotu „Diagnoza psychologiczna”*. Nie opublikowana praca magisterska. Instytut Psychologii UAM, Poznań.
- Niemierko, B. (1990). *Pomiar sprawdzający w praktyce*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Niemierko, B. (1991). *Pomiar sprawdzający w dydaktyce. Teoria i zastosowania*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Niemierko, B. (1999). *Pomiar wyników kształcenia*. Warszawa: Wydawnictwa Szkolne i Pedagogiczne.
- Niemierko, B. (2009). *Diagnostyka edukacyjna*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Paluchowski, W. J. (2006). *Diagnoza psychologiczna. Podejście ilościowe i jakościowe* (wyd. 2). Warszawa: Wydawnictwo Naukowe Scholar.
- Paluchowski, W.J. (2007). *Diagnoza psychologiczna. Proces, narzędzia, standardy*. Warszawa: Wydawnictwa Akademickie i Profesjonalne.
- Seretny, A. (2006). Wskaźnik czytelności tekstu jako pomoc w określaniu stopnia jego trudności. *LingVaria*, 2, 87–98.
- Stemplewska-Żakowicz, K., Paluchowski, W.J. (2008). Podstawy diagnozy psychologicznej. W: J. Strelau, D. Doliński (red.), *Psychologia. Podręcznik akademicki* (t. 2, s. 23–94). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.