

# Pytania testowe i zachowania studentów: analiza testu wiadomości z „Diagnozy psychologicznej”

Władysław Jacek Paluchowski\*

Uniwersytet im. Adama Mickiewicza, Poznań

Piotr Haładziński

Uniwersytet im. Adama Mickiewicza, Poznań

## TEST ITEMS AND STUDENTS' BEHAVIOR ANALYSIS PERTAINING TO PSYCHOLOGICAL ASSESSMENT ACADEMIC COURSE EXAM

Multiple-choice tests are now the most common and readily used assessment tools arranged to measure learning outcomes at most universities and academies. A proper and well constructed test, ingrained in an academic course learning programme, ought to estimate not only the student's knowledge on a certain subject, but also serve as a tool to conduct inner evaluation of the didactic system. Measuring is always burdened with random and systematic biases related to the test itself, or to people tested. The first step to control biases is to identify their sources, frequency. It can be obtained by conducting a statistical analysis of test results. The aim of this paper is to present the results of analyses conducted on psychological assessment academic course exam for such item features as locus of the answer, item difficulty, and distractor's validity. The analysis also includes behaviours manifested during the test, like cheating or carelessness.

### WPROWADZENIE

Testy wiadomości są bodaj najczęściej wykorzystywanym narzędziem do pomiaru osiągnięć studentów psychologii na wyższych uczelniach. Ze względu na rosnącą ilość studentów oraz oferowaną przez testy ekonomię czasową (związaną z przeprowadzeniem egzaminu oraz oceną wyniku każdego studenta), wypierają one coraz częściej egzaminy ustne. Jednak stosunkowo krótki czas potrzebny na przeprowadzenie pomiaru i ogłoszenie wyników nie jest jedynym czynnikiem rozstrzygającym na korzyść stosowania testów wiadomości. Są nimi także większa, niż w przypadku egzaminów ustnych, obiektywność oceniania<sup>1</sup> oraz (zazwyczaj) częściej przestrzegana standaryzacja procedury przeprowadzania egzaminu.

### KRYTERIA DOBROCI TESTÓW WIADOMOŚCI

Dobry test wiadomości powinien spełniać kryteria, które wywodzą się z teorii pomiaru dydaktycznego (Niemierko, 1990, 1999, 2004). Są to:

- bezstronność testu
- dokładność punktowania
- rzetelność testu
- trafność testu
- obiektywizm testu

Bezstronność testu (nazywana także niezależnością sytuacji pomiarowej) oznacza, że zadania nie powinny dyskryminować osób egzaminowanych ze względu na indywidualne cechy, niezwiązane bezpośrednio z tym, co test mierzy. Zdobyte punkty nie powinny wynikać z formalnych cech zadań, zgadywania, czy sprytu testowego (*test wiseness*). Można zatem powiedzieć, że wymóg bezstronności oznacza, że zróżnicowane funkcjonowanie pozycji testowych (termin za: Hornowska, 1999; *differential item functioning DIF*) nie powinno wynikać z właściwości pytań, tylko z poziomu wiedzy studentów. Dokładność punktowania (inaczej: obiektywizm punktowania, rzetelność punktującego) to wymóg, aby zadania były punktowane zgodnie z opracowanym wcześniej schematem (kluczem), niezależnie od osoby sprawdzającej wynik w teście. Rzetelność testu rozumiana jest tu tak samo, jak w przypadku testów psychologicznych i oznacza najkrócej dokładność pomiaru (Hornowska, 2009, s. 28). Trafność testu wiadomości, a więc obszar jego zastosowania, określana jest za pomocą jego trafności wewnętrznej (*content validity*), kryterialnej (*criterion-related validity*) i teoretycznej (*construct validity*). Pierwsza, zwana także trafnością programową (*curricu-*

\* 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

<sup>1</sup> Wynik egzaminu testowego nie jest obciążony błędami związanymi z próbami kierowania wrażeniem, efektem halo, czy efektem kontrastu; nie oznacza to jednak, że treść pytań jest zawsze odpowiednio zrozumiała wśród studentów, a pytania zawsze konstruowane są poprawnie.

lar validity), określa zgodność treści testu z programem nauczania. Druga pozwala na ustalenie, czy wyniki testu korelują z zewnętrznym kryterium, np. wynikami innego testu badającego ten sam obszar wiedzy. Trafność teoretyczna z kolei odnosi się do tego, na ile wyniki pomiaru odzwierciedlają założenia teorii psychologicznej i pedagogicznej, w oparciu o którą test był konstruowany. Niemierko opisuje również (1999, s. 175–176) inne rodzaje trafności, odnoszące się *stricte* do pomiaru sprawdzającego, będące rozszerzeniem lub uproszczeniem wspomnianych wyżej rodzajów<sup>2</sup>. Obiektywizm testu w ujęciu Niemierko (*op. cit.*, s. 181) zbliżony jest do trafności wewnętrznej, ponieważ wyznacza on dokładność z jaką wyniki testu odzwierciedlają założenia programowe. Wyniki są obiektywne wówczas, gdy „odpowiadają na pytanie o spełnienie założonych wymagań” (*ibidem*). Aby test wiadomości spełniał powyższe kryteria, koncepcja testu musi bezpośrednio wynikać z programu kształcenia. Staranny dobór pytań pod kątem zarówno treściowym, jak i formalnym oraz opracowany system punktowania (oparty na modelu kumulatywnym) zagwarantują z kolei bezstronność, dokładność punktowania oraz rzetelność testu. Niezrozumienie treści trzonu pytania lub opcji odpowiedzi stanowiących kafeterię jest jednym ze źródeł zmienności odpowiedzi o charakterze systematycznym. Może ona wynikać z indywidualnych cech osoby odpowiadającej na pytania lub z wadliwej konstrukcji pytań (Niemierko, 1999, s. 121–132).

Poprawnie skonstruowany test wiadomości, zakorzeniony w opracowanym programie nauczania przedmiotu, służy nie tylko jako narzędzie do oceny wiedzy studentów, ale jest także pomocny w wewnętrznej ewaluacji systemu dydaktycznego (Niemierko, 1999, s. 287). Wyniki testu mogą dostarczać bowiem danych empirycznych do oceny trudności przedmiotu, czy jasności i klarowności sposobu przekazywania wiedzy przez osobę prowadzącą. Jednak wymaga to założenia, że studenci uczęszczają na zajęcia, aktywnie słuchają, zadają pytania i czytają literaturę<sup>3</sup>. Ponadto wariancja wyników kształtowana jest nie tylko przez zróżnicowany poziom wiedzy studentów. Czynniki wpływającymi na ostateczny wynik mogą być m.in.: wspomniana wcześniej niejasność zadań, spryt testowy, nieuczciwość przejawiająca się w pracy niesamodzielnej – ściąganie, konsultowanie odpowiedzi, kopiowanie testów z lat ubiegłych oraz dowiadywanie się od koleżanek i kolegów z innych grup, jakie pytania pojawiły się na egzaminie. Ten ostatni przypadek skutkuje uczeniem się odpowiedzi i bezrefleksyjnym podejściem do zadań.

Sposobem na identyfikację źródeł zmienności wyników testu wiadomości może być jego analiza ilościowa i jakościowa. Dokonanie takiej analizy pozwoli ustalić, czy konstrukcja testu jest wolna od wad, na ile cechy treściowe i formalne zadań decydują o wyniku oraz w jakim stopniu ów wynik kształtowany jest przez nieuczciwe

zachowania studentów. Ocena takich cech egzaminu, jak łatwość pozycji testowych (*item easiness*), charakterystyka pytań trudnych, trafność dystraktorów, mogą stanowić punkt wyjścia dla korygowania testu, bieżącego kontrolowania jego rzetelności oraz trafności oraz ewentualnych korekt w programie nauczania. W dalszej części przedstawiona zostanie krótka analiza wyników egzaminu z przedmiotu „Diagnoza psychologiczna”.

## METODA BADAWCZA

Egzamin z przedmiotu „Diagnoza psychologiczna” jest testem sumującym (umożliwiającym końcową ocenę wiedzy studentów), opartym na pomiarze sprawdzającym (Niemierko, 1990, 1999, s. 53), tj. takim, w którym każdy pojedynczy wynik odnoszony jest do wymagań programowych, nie zaś do wyników pozostałych osób. Test składa się z 40 pytań zamkniętych, na rozwiązanie których przeznaczonych jest 35 minut. Kafeteria składa się z 2–5 odpowiedzi, z których najczęściej poprawna jest jedna. Istnieje także wariant egzaminu, w którym do każdego zadania przyporządkowane są zawsze 4 możliwe odpowiedzi<sup>4</sup>. Możliwość wystąpienia dwóch, lub więcej, werstraktorów wyraźnie wskazana jest w trzonie zadania: „wskaż wszystkie”. Każde z 40 pytań losowane jest z puli 550, za pomocą programu *Question Mark teacher software v.1.1*. Podobnie losowo przydzielana jest kolejność ułożenia dystraktorów i werstraktora w każdym pytaniu. Pula 550 zadań powstała na podstawie programu dydaktycznego przedmiotu „Diagnoza psychologiczna” i odpowiada zakresowo wszystkim treściom poruszanym na wykładach oraz zawartych w podanej literaturze obowiązkowej przedmiotu (Paluchowski, 2006, 2007; Stemplewska-Żakowicz i Paluchowski, 2008). W trakcie jednego kursu wykorzystywane są równoległe wersje testu. Zadania testowe przyjmują różną formę i wymagają od studentów potwierdzania prawdziwości tezy (pytania typu prawda/fałsz), definiowania, określania celu lub funkcji, ustalania przyczyny lub skutku, czy wreszcie znajdowania błędów. Obok pytań wymagających jedynie rozpoznania pojawiają się także pytania problemowe, w których udzielenie poprawnej odpowiedzi uzależnione jest od stopnia przetworzenia wiedzy. Procedura przeprowadzania egzaminu jest wystandaryzowana. Za każdym razem studenci informowani są o: czasie przeznaczonym na rozwiązanie testu, sposobie nanoszenia odpowiedzi na arkusz, ilości możliwych poprawnych odpowiedzi, progu zaliczeniowym oraz ilości punktów odpowiadających konkretnym ocenom.

Do przedstawionych poniżej analiz wykorzystano wyniki egzaminu studentów Instytutu Psychologii UAM ( $N=260$ ), Instytutu Psychologii Uwr ( $N=196$ ) oraz wyniki studentów SWPS ( $N=499$ ) we Wrocławiu.

<sup>2</sup> Są to: trafność wyboru dziedziny, trafność opisowa oraz trafność funkcjonalna.

<sup>3</sup> Niestety to założenie wydaje się daleko idącą idealizacją.

<sup>4</sup> Będziemy się do niego odwoływać analizując zjawisko zmiany trudności pytań powtarzających się w różnych terminach egzaminu.

## REZULTATY I OMÓWIENIE WYNIKÓW

### OCENA ROZKŁADU WERSTRAKTORÓW W PYTANIACH

Korzystanie z oprogramowania – baz pytań – w procesie układania testu wiadomości ma wiele zalet. Są nimi przede wszystkim: możliwość gromadzenia dużej ilości pytań egzaminacyjnych z opcją ich bieżącej edycji, szybkie losowanie zestawów i tworzenie wersji równoległych testu, losowy rozkład kolejności odpowiedzi w kafeterii. O ile dwie pierwsze wymienione zalety odnoszą się głównie do ekonomii związanej z czasem administrowania testem, o tyle ta ostatnia pozwala na wyeliminowanie wpływu tendencyjności osoby układającej test na miejsce poprawnej odpowiedzi wśród dystraktorów. Owa tendencyjność, wiążąca się z faworyzowaniem (świadomym lub nie) konkretnej opcji odpowiedzi przez konstruktora testu, może istotnie wpływać na dokonanie ostatecznego wyboru wariantu przez studenta. Takiego wpływu można by najczęściej oczekiwać w sytuacji, gdy osoba odpowiadająca nie jest pewna swojego wyboru lub po prostu nie zna odpowiedzi. Działając w sposób racjonalny, a więc maksymalizując możliwe zyski przy minimalizowaniu kosztów, może podjąć próbę zgadywania odpowiedzi lub kierowania się innymi względami nie związanymi z treścią. Jednym z takich względów jest pamięć wyboru odpowiedzi na pytania (poprzedzające), na które osoba ta знаła poprawne odpowiedzi. Zauważenie wyłaniającej się tendencji w umiejscowieniu werstraktora w kafeterii może prowadzić do wykształcenia się stylu odpowiadania uruchamianego w podobnych sytuacjach.

Egzamin z przedmiotu „Diagnoza psychologiczna” konstruowany jest w 4 wersjach równoległych z 2–5 możliwościami odpowiedzi na każde pytanie. Określenie, czy rozkład werstraktorów w pytaniach egzaminacyjnych istotnie odchyła się od losowego pozwoli zidentyfikować ewentualne źródło wariancji odpowiedzi. Może stanowić także ocenę użyteczności programu *Question Mark*.

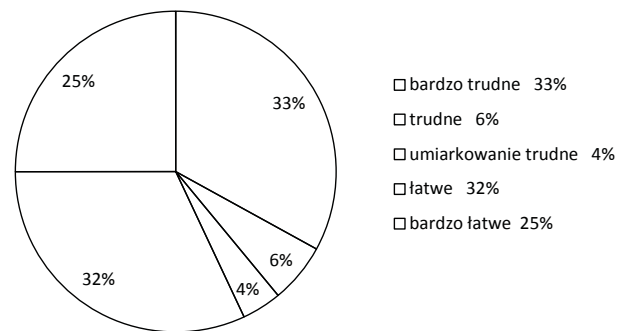
W pierwszym etapie badania rozkładu werstraktorów, podzielono 160 wylosowanych pytań ze względu na objętość kafeterii. Pytania z 4 wariantami odpowiedzi stanowiły ponad 53% wszystkich pytań. Drugie co do liczności były pytania z 3 opcjami odpowiedzi – ponad 41%. Pozostałe niecałe 6% to pytania z 2 lub 5 możliwościami odpowiedzi. Ze względu na małą liczebność pominięto je w dalszej analizie. Następnie dla dwóch wcześniejszych grup zastosowano osobno test *chi* kwadrat dla jednej zmiennej. W obu przypadkach wyniki okazały się nieistotne statystycznie, odpowiednio:  $\chi^2(3, N=85)=3.24$ ,  $p_a=.36$  dla pytań o 4 opcjach odpowiedzi oraz  $\chi^2(2, N=65)=1.81$ ,  $p_a=.40$  dla pytań z 3 wariantami. Należy więc przyjąć, że rozkład werstraktorów w pytaniach testu ułożonego za pomocą programu *Question Mark* jest losowy i nieobarczony błędem tendencyjności.

### OCENA TRUDNOŚCI PYTAŃ

Trudność w odniesieniu do zadań punktowanych 0–1 (tzn. punkt przyznawany za wskazanie poprawnej lub

wszystkich poprawnych odpowiedzi – „wszystko albo nic”) określa się poprzez stosunek ilości osób, które błędnie rozwiązały dane zadanie, do liczby osób poddanych testowaniu (Niemierko, 1999, s. 152–153, 2009, s. 108–109). Analogicznie określa się także łatwość pytań poprzez stosunek liczby osób, które udzieliły poprawnej odpowiedzi na konkretne pytanie, do ilości osób poddanych testowaniu. Wskaźnik trudności określane jest symbolem *q*, łatwości – symbolem *p*. Oba przyjmują wartości {0;1}.

Ocena trudności poszczególnych zadań oraz występowanie pytań bardzo łatwych i bardzo trudnych w teście może dostarczyć informacji o poprawności dydaktycznej egzaminu. Ponadto analiza pytań bardzo trudnych i trudnych mogłaby pomóc w identyfikacji obszarów wymagających interwencji dydaktycznej. Na Rycinie 1 zaprezentowano udział procentowy pytań o różnym stopniu trudności na egzaminie z „Diagnozy psychologicznej” ( $N=955$ ).



Ryc. 1. Rozkład procentowy pytań ze względu na ich trudność

Niski udział procentowy pytań bardzo łatwych oraz bardzo trudnych oznacza, że egzamin z „Diagnozy psychologicznej” jest skonstruowany poprawnie pod kątem dydaktycznym. Jego średni stopień trudności wynosi  $q=.6$ .

### CECHY PYTAŃ TRUDNYCH I BARDZO TRUDNYCH

Analiza cech pytań trudnych i bardzo trudnych pozwoli ustalić poprawność dydaktyczną samych zadań, wyjaśnić przyczyny niskiej frakcji odpowiedzi poprawnych oraz wskazać ewentualne obszary wymagające interwencji dydaktycznej.

Na pytania trudne i bardzo trudne poprawnych odpowiedzi udzielało od 10% do 48% studentów. Analizując polecenia zawarte w trzonie pytań, nie znaleziono żadnych cech, które wyróżniałyby je na tle innych zadań. Polecenia te dotyczyły:

- zastosowania zasady
- wskazania typowych cech, definiowania
- uzupełnienia luki
- porządkowania

- uzasadnienia opinii
- ustalenia podobieństw i różnic
- ustalenia związku

Analizując pytania trudne (43) pod kątem obszarów wiedzy, których dotyczą, można stwierdzić, że prawie połowa z nich (42%) odnosi się do tematyki związanej z problematyką pomiaru, testowania, metod kwestionariuszowych i wskaźników. Może to wskazywać na problemy studentów w przygotowywaniu się z zakresu tych obszarów do egzaminu. Co wydaje się ciekawe, pytania bardzo trudne (6), czyli takie, na które odpowiedzi poprawnych udzielało od 10% do 19% studentów, odnoszą się do problematyki związanej z testowaniem i metodami kwestionariuszowymi, ale także do zagadnień takich jak model wyboru narzędzi diagnostycznych, czy efekt Barnuma. Pytania bardzo łatwe (8) z kolei odnoszą się do podstawowych funkcji diagnozowania oraz pojęć łatwo zapamiętywanych przez studentów, jak zasada Rumpelstilzchena, czy wspomniany efekt Barnuma. O ile jednak najczęściej są to zadania wymagające od studentów zdefiniowania lub rozpoznania zjawiska, o tyle bardzo trudnym okazuje się zadanie dotyczące wskazania konsekwencji wystąpienia tego efektu oraz jego związku z komunikowaniem diagnozy.

Pod kątem formalnym pytania trudne i bardzo trudne nie różnią się od pytań łatwych i bardzo łatwych. Długości trzonu i opcji odpowiedzi w kafeterii jest względnie taka sama. Nie zauważono także różnic w wielkości miernika łatwości czytania Flescha (*Flesch Reading Ease Level*) oraz miernika szkolnego Flescha–Kincaida (*Flesch–Kincaid Grade Level*), przedstawionych w Tabeli 1<sup>5</sup>. Interpretując powyższe wskaźniki można powiedzieć, że pytania z testu wiadomości z „Diagnozy psychologicznej” są trudne, ale powinny być zrozumiałe („czytelne”) dla studentów (DuBay, 2004, s. 22).

**Tabela 1**  
Wskaźniki czytelności dla pytań trudnych i bardzo trudnych oraz łatwych i bardzo łatwych

	Pytania trudne i bardzo trudne	Pytania łatwe i bardzo łatwe
Wskaźnik czytelności Flescha	32.49	30.48
Wskaźnik Flescha–Kincaida	13.67	13.27

#### TRAFNOŚĆ DYSTRAKTORÓW

Związek frakcji wyboru odpowiedzi błędnych z wynikiem ogólnym w teście określa moc dyskryminacyjną dane go dystraktora oraz jego trafność. Trafny dystraktor to zatem taki, który dobrze różnicuje osoby ze względu na ogólny wynik testu. Jeżeli więc osoba, która uzyskuje

wysoki wynik ogólny częściej wybiera odpowiedź prawidłową, natomiast osoba o niskim wyniku częściej wybiera odpowiedź błędną, możemy mówić o wysokiej trafności dystraktora. Będzie to bowiem oznaczało, że spełnia on swoje zadanie w odciąganiu uwagi od odpowiedzi poprawnej. Z drugiej strony jednak jest rozłączny z werstraktorem i dzięki posiadanej wiedzy osoba rozwiązująca zadanie jest w stanie trafnie go rozpoznać i odrzucić.

Oceny trafności dystraktorów w teście wiadomości z „Diagnozy psychologicznej” dokonano na podstawie frakcji wyborów oraz korelacji, za pomocą współczynnika *V–Cramera*, pomiędzy wynikiem w teście (wysoki vs niski), a wyborem odpowiedzi. Przeanalizowano łącznie 136 pytań (4 wersje testu po usunięciu pytań powtórzonych) o 2–5 opcjach odpowiedzi. W przypadku 82% pytań dystraktory były częściej wybierane przez osoby z ogólnym niskim wynikiem, czasem były to jednak różnice nieznaczne. W przypadku 44% pytań (60) zanotowano istotny statystycznie związek pomiędzy wyborem odpowiedzi, a wynikiem ogólnym testu. Współczynnik *V–Cramera* przyjmował wartości z zakresu  $V=.25$  do  $V=.72^6$ . W Tabeli 2 zaprezentowano rozkład procentowy odpowiedzi ze względu na wynik ogólny w teście.

**Tabela 2**  
Przykładowy silny związek między ogólnym wynikiem w teście a wyborem odpowiedzi

	Dystraktor 1	Dystraktor 2	Dystraktor 3	Werstraktor
Wynik niski	24% (6)	12% (3)	20% (5)	44% (11)
Wynik wysoki	0% (0)	4.3% (1)	0% (0)	95.7% (22)
$V=.57, p<.001$				

W odniesieniu do 76 pytań (56%) nie zanotowano istotnej korelacji.

#### OSZUSTWO EGZAMINACYJNE

Zmiany trudności pytań w równoległych wersjach testu mogą wynikać ze zróżnicowanego poziomu wiedzy, ale także z prób oszukiwania, czyli, jak pisze Niemierko (2006): „ukrytego posługiwania się informacją uzyskaną z innych źródeł niż własna wiedza i praca, dokonywanego w celu uzyskania pożądanego wyniku egzaminu”. W sytuacji, kiedy egzamin odbywa się w kilku terminach, może dochodzić do wymiany informacji pomiędzy studentami na temat pytań i (rzekomo poprawnych) odpowiedzi. W czterech równoległych wersjach testu wiadomości z „Diagnozy psychologicznej” 15% pytań powtarza się w zmienionej formie: treść trzonu i kafeterii pozostaje ta sama, zmienia się jedynie kolejność następowania po sobie dystraktorów i werstraktora. Zauważono, że w wariancie egzaminu, w którym kafeteria składała się

<sup>5</sup> Porównania dokonano za pomocą programu *Flesh 2.0*, dostępnego na stronie internetowej: <http://flesh.sourceforge.net/>

<sup>6</sup> Czyli od słabego po bardzo silny związek.

z 4 odpowiedzi, dla prawie połowy z powtórzonych pytań (10) zmalał wskaźnik trudności w sytuacji, kiedy to samo zadanie pojawiało się w późniejszym terminie egzaminu. Można więc pokusić się o hipotezę, że studenci, korzystając z nadarzającej się okazji, są skłonni oszukiwać.

Obok zmian trudności pytań w czasie kolejnym źródłem hipotez dotyczących oszustwa egzaminacyjnego może być analiza rozkładu wyników egzaminu<sup>7</sup>. Odchylenia od rozkładu normalnego, a w szczególności wystąpienie rozkładu dwumodalnego, może oznaczać istnienie dodatkowego źródła zmienności, np. kiedy część osób zna wcześniej pytania egzaminacyjne. Taka sytuacja może ujawniać się w rozkładzie w postaci dwóch wartości dominujących: jedna odnosić się będzie do osób uczciwie rozwiązujących test, druga jest wynikiem znajomości pytań i związanego z nią podobnego profilu odpowiadania. W związku z podejrzeniami o oszustwo w wariancie egzaminu z czterema odpowiedziami, zdecydowano się wykonać test Kołmogorowa-Smirnowa, aby zbadać zgodność obserwowanego rozkładu z normalnym. W związku z tym, że wynik testu okazał się nieistotny statystycznie ( $Z=1.18$ ;  $p_a=.12$ ), przyjęto, że wyniki egzaminu posiadają rozkład normalny, gdzie:  $M=21.7$ ,  $SD=4.73$ . Na podstawie niniejszej analizy przypuszczenia dotyczące oszukiwania nie znajdują więc potwierdzenia.

#### ODPOWIADANIE PRZYPADKOWE<sup>8</sup>

Niedbałość (*carelessness*) często wiązana jest z odpowiadaniem na pytania kwestionariusza. Stanowi jedno ze źródeł zakłóceń związanych z osobą badaną (Paluchowski, 1983). Za przyczynę niedbałości można uznać brak motywacji, zmęczenie, problemy z koncentracją uwagi, które przejawiają się w udzielaniu odpowiedzi przypadkowych, niespójnych. Sposobem na identyfikację niedbałości jest wbudowywanie w kwestionariusz skal kontrolnych składających się z powtórzonych pytań, w postaci identycznej lub odwróconej.

Presja czasowa związana z egzaminem, dodatkowe napięcie, fluktuacje uwagi i wynikające z tego niedokładne czytanie poleceń mogą wpływać na jakość udzielanych odpowiedzi w teście wiadomości. Aby sprawdzić na ile odpowiadanie niespójne i stojąca za nim niedbałość przejawiają się w teście wiadomości z „Diagnozy psychologicznej”, zanalizowano odpowiedzi na powtórzone przypadkowo w jednej wersji testu pytanie. Pojawiło się ono na początku i w drugiej połowie arkusza z pytaniami ze zmienioną kolejnością odpowiedzi. Rozkład odpowiedzi przedstawia się następująco: ze 120 osób, tylko 7

(6%) udzieliło niespójnych odpowiedzi na dwa pytania. Pozostali odpowiadali spójnie. Warto jednak dodać, że 8 osób (7%) poprawiło swoje odpowiedzi z niespójnych na spójne, dzięki czemu można przypuszczać, że chociaż marginalnie, ale jednak zjawisko niedbałości występuje w sytuacji rozwiązywania testu wiadomości.

#### WNIOSKI

Na podstawie przedstawionych powyżej wyników można sformułować następujące wnioski:

- 1) Test wiadomości z „Diagnozy psychologicznej” spełnia dobrze funkcję sprawdzianu sumującego
  - a) układ werstraktorów w teście jest losowy i nie indukuje atreściowych stylów odpowiadania
  - b) zadania są poprawne dydaktycznie; nie dominują pytania o bardzo wysokim lub bardzo niskim poziomie trudności;
  - c) dystraktory stosowane w pytaniach są częściej wybierane przez osoby o niskim ogólnym wyniku w teście
- 2) Analiza pytań trudnych pozwala na identyfikację obszarów wymagających interwencji dydaktycznej
  - a) prawie połowa pytań trudnych związana jest z problematyką testowania, kwestionariuszy, pomiaru i wskaźników
- 3) Analiza wyników pozwala wysuwać hipotezy na temat źródeł wariancji związanych z osobami rozwiązującymi test, niewynikających jednak ze różnicowania poziomów wiedzy
  - a) oszukiwanie przez uczenie się pytań,
  - b) niedbałość i odpowiadanie przypadkowe.

#### LITERATURA

- DuBay, W.H. (2004). *The principles of readability*. Costa Mesa: Impact Information.
- Hornowska, E. (1999). *Stronniczość testów psychologicznych. Problemy – kierunki – kontrowersje*. Poznań: Wydawnictwo Fundacji Humaniora.
- Hornowska, E. (2009). *Testy psychologiczne. Teoria i praktyka*. Warszawa: Wydawnictwo Naukowe Scholar.
- Niemierko, B. (1990). *Pomiar sprawdzający w dydaktyce. Teoria i zastosowania*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Niemierko, B. (1999). *Pomiar wyników kształcenia*. Warszawa: Wydawnictwa Szkolne i Pedagogiczne.
- Niemierko, B. (2004). *Perspektywy badań psychologicznych, dydaktycznych, socjologicznych i politologicznych wspomagających system egzaminowania zewnętrznego*. W: Z. Lisiecka (red.), *Biuletyn Badawczy nr 1 Centralnej Komisji Egzaminacyjnej* (s. 4–10). Warszawa: Pracownia Ewaluacji CKE.
- Niemierko, B. (2006, listopad). *Oszustwo egzaminacyjne*. Referat wygłoszony na XII Konferencji Diagnostyki Edukacyjnej w Lublinie.
- Niemierko, B. (2009). *Diagnostyka edukacyjna*. Warszawa: Państwowe Wydawnictwo Naukowe.

<sup>7</sup> Ciekawym przykładem wykrycia oszustwa egzaminacyjnego m.in. poprzez analizę rozkładu wyników, jest *casus* profesora Richarda Quinna, demaskatora ponad 200 osób, uczestników egzaminu, którzy wcześniej znali pytania: <http://www.telegraph.co.uk/news/newsvideo/weirdnewsvideo/8140456/200-students-admit-cheating-after-professors-online-rant.html>

<sup>8</sup> W analizie wykorzystano wyniki egzaminu z Diagnozy psychologicznej przeprowadzonego w 2010 roku ( $N=120$ ).

- Paluchowski, W.J. (1983). Źródła zakłóceń w kwestionariuszowym badaniu osobowości i ich kontrola. W: W.J. Paluchowski (red.), *Z zagadnień diagnostyki osobowości* (s. 249–272). Wrocław: Wydawnictwo Polskiej Akademii Nauk.
- 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.
- 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.