

BARBARA CIŻKOWICZ

Instytut Pedagogiki, Akademia Bydgoska, Bydgoszcz

KLASYCZNA I PROBABILISTYCZNA TEORIA TESTU – ANALIZA PORÓWNAWCZA

Egzaminy zewnętrzne, które od kilku lat stały się trwałym elementem polskiego systemu kształcenia, spowodowały znaczny wzrost zainteresowania testami – zarówno wśród badanych, jak i osób badających osiągnięcia uczniów. Nauczycielowi przygotowującemu testy, szczególnie te wykorzystywane przez Centralną Komisję Egzaminacyjną, potrzebna jest wiedza teoretyczna na temat ich konstrukcji i analizy. Powszechnie stosowane procedury szacowania rzetelności oraz oceny właściwości pozycji testowych są wyprowadzane głównie z klasycznej teorii testu¹.

Innym podejściem teoretycznym jest teoria odpowiadania na pozycje testu (ang. *Item Response Theory* – IRT). Wskutek popularyzacji komputerów wzrasta zainteresowanie tą teorią, jednak liczba jej zastosowań nadal nie jest imponująca. Niewątpliwym utrudnieniem w jej upowszechnieniu jest zarówno duża liczebność próby konieczna do jej stosowania, jak potrzeba wykorzystania profesjonalnych aplikacji do przeprowadzania analiz².

W poniższym opracowaniu przedstawię porównanie wyników analizy testu uzyskanych w efekcie zastosowania klasycznej i probabilistycznej teorii. Obliczenia wykonano wykorzystując odpowiednio moduł (*Analiza rzetelności i pozycji*) pakietu Statistica 7 oraz program RUMM2010³.

¹ Niemierko B., *Testy osiągnięć szkolnych*, WSiP, Warszawa 1975; Magnuson D., *Wprowadzenie do teorii testów*, PWN, Warszawa 1981; Brzeziński J., *Elementy metodologii badań psychologicznych*, PWN, Warszawa 1980.

² Baker F.B., *The basic of item response theory*, NH: Heinemann. Portsmouth 1985; Hambleton R., Swaminathan H., Rogers H., *Fundamental of Item Response Theory*, SAGE Publications 1991.

³ www.rummlab.com.au

1. Podstawy teoretyczne

Klasyczna teoria testu (KTT) przyjmuje, że na wynik otrzymany testu składa się wynik prawdziwy i błąd pomiaru. Ponadto zakłada ona, że między wynikiem prawdziwym i otrzymanym jest zależność prostoliniowa, obciążona tej samej wielkości błędem dla całego zakresu wyników. Rzetelność pomiaru, definiowana jako stosunek wariancji wyników prawdziwych do wariancji wyników otrzymanych, może być szacowana na wiele sposobów. Najczęściej stosowaną metodą jej szacowania jest badanie zgodności wewnętrznej. Używany w niej współczynnik α Cronbacha jest powszechnie wykorzystywanym wskaźnikiem rzetelności. Wyliczane są również parametry poszczególnych pozycji (trudność i moc różnicująca), które wpływają na rzetelność oraz błąd standardowy pomiaru⁴.

Teoria odpowiadania na pozycje testu (IRT) pozwala przyporządkować każdemu badanemu wynik umożliwiający umieszczenie go na skali cechy nieobserwowalnej (latentnej θ), stanowiącej kontinuum poziomu umiejętności (przedział od $-\infty$ do $+\infty$). Możliwe jest wyznaczenie prawdopodobieństwa prawidłowej odpowiedzi na konkretne pytanie $[P(\theta)]$ dla każdej wartości umiejętności θ . Związek między poziomem umiejętności a prawdopodobieństwem udzielenia poprawnej odpowiedzi na pozycję nazywany jest w IRT krzywą charakterystyczną pozycji testowej (ang. *item characteristic curve* – ICC). Określone funkcje matematyczne stosowane do opisu kształtu ICC implikują występowanie różnych modeli w IRT (rys. 1.). W praktyce często stosowany jest model logistyczny występujący w trzech wariantach: jako jedno - , dwu - i trójparametryczny⁵.

Do opisu ICC w trójparametrycznym modelu stosowana jest funkcja postaci:

$$P(\theta) = c_i + (1 - c_i) \cdot \frac{e^{D \cdot a_i \cdot (\theta - b_i)}}{1 + e^{D \cdot a_i \cdot (\theta - b_i)}}$$

gdzie:

a_i – moc różnicująca (dyskryminacyjna) pozycji testu;

b_i – trudność pozycji testu;

c_i – współczynnik zgadywania;

D – stała maksymalizująca dopasowanie krzywej logistycznej do ogiwy rozkładu normalnego ($D = 1,7$).

Model dwuparametryczny otrzymujemy, przyjmując w powyższym równaniu założenie o zerowej wartości c_i . Jeśli ponadto przyjąć, że a_i (moc różnicująca) wszystkich pozycji testowych jest równa i ma wartość 1, wówczas otrzymamy model jednoparametryczny.

Trudność pozycji testowej (b_i) jest wartością cechy latentnej (θ), dla której prawdopodobieństwo prawidłowej odpowiedzi na daną pozycję $[P(\theta)]$ jest równe

⁴ Niemierko B., op. cit.; tenże, *Pomiar wyników kształcenia*, WSiP, Warszawa 1999.

⁵ Hambleton R., Swaminathan H., Rogers H., op. cit.

0,5. Pozycja jest tym trudniejsza, im wyższa wartość b_i ⁶. Trudność, czyli mediana poziomu cechy latentnej dla pozycji, informuje, w którym miejscu na skali θ dana pozycja najlepiej różnicuje badanych. Łatwe pozycje dobrze różnicują osoby o niskich umiejętnościach, trudne – o wysokich. Teoretyczny zakres zmienności tego parametru obejmuje przedział $(-\infty; +\infty)$, jednak wartości typowe ograniczają się do przedziału $(-3; +3)$.

Moc różnicująca pozycji testowej (a_i) jest proporcjonalna do kąta nachylenia ICC w punkcie przegięcia. Informuje ona, w jakim stopniu pozycja pozwala różnić badanych o poziomie umiejętności niższym od trudności pozycji (b_i) od tych, u których poziom ten jest wyższy. Moc dyskryminacyjna jest tym większa, im wyższą wartość przyjmuje parametr a_i . Wartości teoretyczne a_i mogą zawierać się w przedziale od $-\infty$ do $+\infty$. Praktyczny zakres mieści się między $-2,8$ a $+2,8$ ⁷. Ujemna wartość tego parametru, związana z malejącym przebiegiem ICC, oznacza wysokie prawdopodobieństwo wyniku poprawnego u osób o niskim poziomie θ . Oczywiście, pozycje o ujemnym współczynniku a_i nie powinny występować w prawidłowo funkcjonującym teście.

Współczynnik zgadywania c_i informuje, jakie jest prawdopodobieństwo uzyskania prawidłowej odpowiedzi tylko przez zgadywanie. Parametr c_i nie zmienia się wraz z funkcją umiejętności. Ma on stałą wartość, co oznacza, że badani posiadają tę samą szansę uzyskania niezerowego wyniku, niezależnie od reprezentowanego przez nich poziomu θ . Teoretycznie c_i może przyjmować wartości z przedziału $(0,0-1,0)$, jednak najczęściej mieści się w zakresie $(0,0-0,4)$ ⁸. Pożądaną są jak najniższe wartości tego współczynnika.

2. Ilościowa analiza testu

Poniżej przedstawiono porównanie wyników analizy testu, przeprowadzonej zgodnie z klasyczną teorią i IRT. W IRT zastosowano dwuparametryczny model logistyczny. Opracowanie nie zawiera pełnej, systematycznej analizy wyników, a jedynie pewne jej elementy możliwe do porównania w obu teoriach. Obliczenia przeprowadzono na przykładzie wyników testu do badania wyuczonej bezradności⁹. Testem składającym się z 20 pozycji punktowanych 0–2 przebadano próbę 100 osób.

Trudność i moc różnicująca, jako wskaźniki charakteryzujące pozycję, występują zarówno w klasycznej, jak probabilistycznej teorii testu. Jednak informacje, jakie można uzyskać o właściwościach pozycji, w obu teoriach są istotnie

⁶ Baker F.B., *Methodology review: Item parameter estimation under the one-, two- and three-parameter logistic models*, „Applied Psychological Measurement” 1987, 11, s. 111–142.

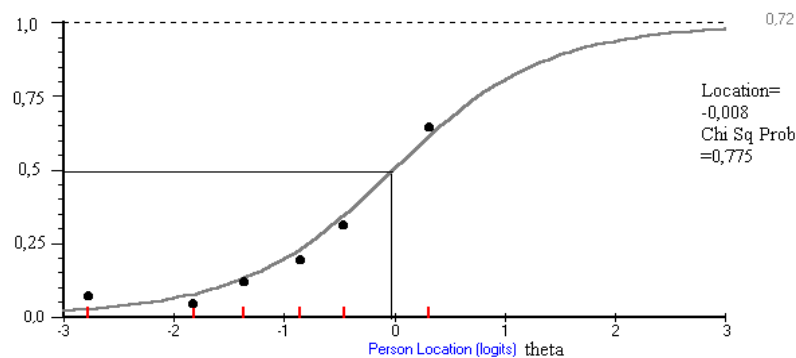
⁷ Baker F.B., *The basic of item response theory*, op. cit., s. 2.

⁸ Hornowska E., *Testy psychologiczne. Teoria i praktyka*, Wydawnictwo Naukowe SCHOLAR, Warszawa 2001, s. 199.

⁹ Ciżkowicz B., *Oceny szkolne a bezradność intelektualna*, [w:] B. Niemierko, J. Brzdąk (red.), *Dwa rodzaje oceniania szkolnego*, Katowice 2002.

różne. Przewaga krzywych ICC nad klasycznymi wskaźnikami dobroci pozycji testowych polega na tym, że na ich podstawie można określić zależność między prawdopodobieństwem poprawnej odpowiedzi na konkretną pozycję testową a różnymi wartościami cechy latentnej¹⁰. Mimo iż trudność pozycji testowej w IRT została zdefiniowana jako jedna liczba (mediana θ), to krzywa charakterystyczna pozwala ustalić, jakie jest prawdopodobieństwo prawidłowej odpowiedzi (jak trudna jest dana pozycja) dla badanego o konkretnym poziomie umiejętności θ .

Program RUMM2010 wyznacza kształt oraz parametry krzywej charakterystycznej dla każdej pozycji testu. Rys. 1. przedstawia ICC dla pozycji nr 3 z analizowanego testu bezradności. Test statystyczny χ^2 stosowany jest jako miara dobroci dopasowania krzywej teoretycznej do punktów empirycznych. Jeśli poziom istotności χ^2 jest większy od 0,05, dopasowanie krzywej można uznać za zadowalające¹¹. Wyliczana trudność pozycji oznacza poziom umiejętności θ , dla którego prawdopodobieństwo udzielenia poprawnej odpowiedzi na daną pozycję jest równe 0,5. Na wykresie podana jest też moc różnicująca pozycji.



Rys. 1. Krzywa charakterystyczna pozycji nr 3 testu (ICC)

gdzie: *Location* – trudność pozycji,

Slope – moc różnicująca,

Chi Sq Prob – poziom istotności testu statystycznego χ^2 .

W celu porównania wyników analizy pozycji uzyskanych w dwóch różnych podejściach teoretycznych, zestawiono w tab.1. parametry wyznaczone w pakiecie Statistica (klasyczna teoria) i w programie RUMM2010 (IRT). W KTT trudność pozycji (q) jest dopełnieniem do jedności stosunku sumy punktów uzyskanych przez wszystkich badanych za tę pozycję do maksymalnej liczby punktów, jaką mogli za nią uzyskać. Miarą mocy różnicującej pozycji wielopunktowych w klasycznej teorii testu jest współczynnik korelacji r Pearsona między wynikiem rozpatrywanej pozycji a wynikiem testu¹².

¹⁰ Hornowska E., op. cit., s.199.

¹¹ Baker F.B., Kim S., *Item Response Theory. Parameter Estimation Techniques*, Marcel Dekker, Inc., N. Y. 2004.

¹² Magnuson D., op. cit.

Tab.1. Parametry pozycji testu w klasycznej i probabilistycznej teorii

Nr poz.	Klasyczna teoria		IRT		Nr poz.	Klasyczna teoria		IRT	
	q	r	Locat	Slope		q	r	Locat	Slope
P1	0,47	0,38	-0,09	0,70	P9	0,53	0,39	0,29	0,77
P2	0,34	0,48	-0,86	0,62	P10	0,47	0,57	-0,05	0,66
P3	0,54	0,40	0,29	0,82	P11	0,46	0,60	-0,03	0,53
P4	0,53	0,48	0,34	0,69	P12	0,54	0,30	0,54	0,60
P5	0,50	0,50	0,08	0,72	P13	0,55	0,42	0,34	0,88
P6	0,52	0,48	0,26	0,66	P14	0,53	0,32	0,32	0,73
P7	0,68	0,51	0,02	0,65	P15	0,54	0,47	0,39	0,75
P8	0,46	0,42	-0,19	0,79	P16	0,21	0,39	-1,64	0,64

q, *Locat* – trudność pozycji testu odpowiednio w klasycznej teorii i w IRT

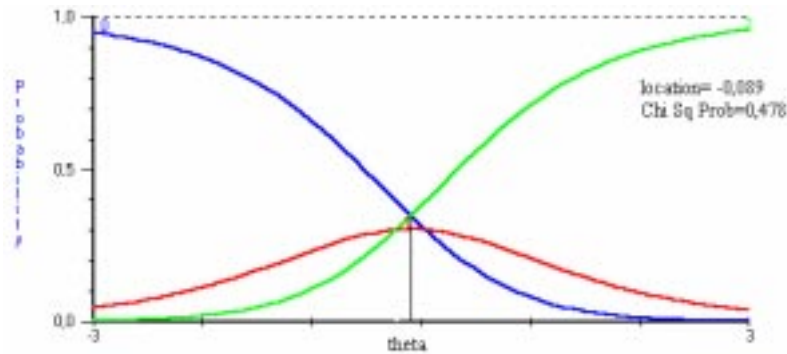
r, *Slope* – moc różnicująca pozycji testu odpowiednio w klasycznej teorii i w IRT

Ponieważ wartości parametrów w obu podejściach teoretycznych są niewspółmierne, obliczono zależność między trudnościami pozycji oraz mocami różnicującymi wyznaczonymi w KTT i w IRT. Użyty do tego celu współczynnik r Pearsona przyjął odpowiednio wartości $r_{xy} = 0,993$ i $r_{xy} = -0,354$. Niemal pełna zależność między trudnościami wskazuje, że zadania trudne w IRT okazały się relatywnie trudne również w podejściu klasycznym. W przypadku mocy różnicującej zależność jest ujemna, przeciętna. Wiąże się to z innym znaczeniem tych parametrów w obu podejściach. W klasycznej teorii wskaźnik mocy różnicującej informuje, czy pozycja testu dobrze czy źle różnicuje badanych o wysokich i niskich wynikach w teście, natomiast w IRT pokazuje on, jak dobrze rozdziela badanych o poziomie umiejętności (θ) równym trudności analizowanej pozycji.

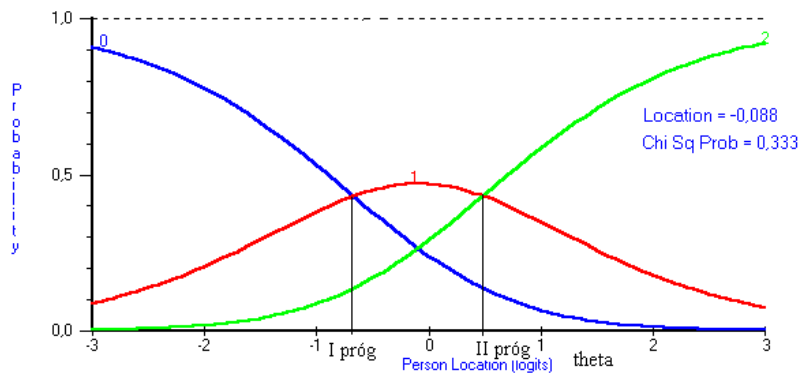
IRT umożliwia także analizę rozkładu prawdopodobieństwa każdego wyniku możliwego do uzyskania za daną pozycję. W rozpatrywanym przykładzie są to wartości 0, 1 i 2. Układ progów, czyli miejsc, w których następuje przecięcie się krzywych prawdopodobieństwa, pozwala ocenić, czy sposób punktowania pozycji jest logicznie spójny. Na rys. 2. przedstawiono rozkłady prawdopodobieństwa możliwych do uzyskania wyników za pozycję w zależności od poziomu umiejętności θ . Dla niskich wartości zmiennej θ najbardziej prawdopodobne jest uzyskanie 0 punktów (krzywa malejąca), aż do miejsca, w którym θ osiągnie wartość -0,09, czyli do przecięcia się tej krzywej z krzywą przedstawiającą rozkład prawdopodobieństwa uzyskania 2 punktów (krzywa rosnąca). Miejsca przecięcia się krzywych wyznaczają progi dla pozycji nr 1. Progi są niewłaściwie uporządkowane, a prawdopodobieństwo pojawienia się wyniku 1 (krzywa symetryczna) nigdy nie osiąga wartości najwyższej.

Na rys. 3. przedstawiono progi dla pozycji 7, których ułożenie jest prawidłowe. Osoby o niskim poziomie umiejętności ($\theta < -0,67$) uzyskują z największym prawdopodobieństwem 0 punktów za pozycję. Dla wyższych wartości ($-0,67 < \theta < 0,50$) maksymalnie prawdopodobne jest uzyskanie 1 punktu, a dla

$\theta > 0,50$ najbardziej prawdopodobny staje się wynik 2. Trudność pozycji nr 7 jest nieznacznie niższa od średniego poziomu umiejętności badanych testem (Location = -0,088). Krzywa charakterystyczna jest dobrze dopasowana (różnice między krzywą a punktami empirycznymi są nieistotne: Chi Sq Prob = 0,333).



Rys. 2. Progi dla pozycji nr 1 (rozkład prawdopodobieństw wyników 0; 1; 2)



Rys. 3. Progi dla pozycji nr 7 (rozkład prawdopodobieństw wyników 0; 1; 2)

Poniżej poddano analizie niezmienniczość parametrów pozycji testowych. Zasadniczą zaletą IRT jest fakt, że parametry pozycji testowych są własnością samej pozycji i nie zależą ani od próby badanych, ani od próby pozycji, z którymi współtworzą test. Oczywiście, nie oznacza to równości liczbowych wartości estymowanych parametrów¹³.

W celu sprawdzenia niezmienniczości estymatorów parametrów konkretnej pozycji względem zbioru pozycji, z którymi tworzy ona test, przeprowadzono obliczenia dla trzech podtestów. Otrzymano je poprzez wycięcie z tej samej macierzy wyników odpowiednio: 16 pierwszych, 16 ostatnich oraz 10 środkowych pozycji z 20, z których składa się test bezradności. Wyniki dotyczące tylko wspólnej części zbiorów pozycji zestawiono w tab.2.

¹³ Baker, F., *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD 2001.

Tab.2. Parametry pozycji w wydzielonych podtestach

Nr poz.	Trudność pozycji w podtestach			Moc różnicująca pozycji w podtestach		
	Poz. 1–16	Poz. 5–20	Poz. 6–15	Poz. 1–16	Poz. 5–20	Poz. 6–15
5	0,08	-0,01	-	0,72	0,72	-
6	0,26	0,20	0,08	0,66	0,66	0,66
7	0,02	-0,06	-0,17	0,65	0,66	0,66
8	-0,19	-0,28	-0,38	0,79	0,80	0,79
9	0,29	0,24	0,12	0,77	0,76	0,77
10	-0,05	-0,13	-0,24	0,66	0,68	0,67
11	-0,03	-0,09	-0,22	0,53	0,53	0,54
12	0,54	0,48	0,36	0,60	0,61	0,61
13	0,34	0,26	0,13	0,88	0,88	0,88
14	0,32	0,25	0,13	0,73	0,72	0,72
15	0,39	0,33	0,21	0,75	0,75	0,75
16	-1,64	-1,77	-	0,64	0,63	-

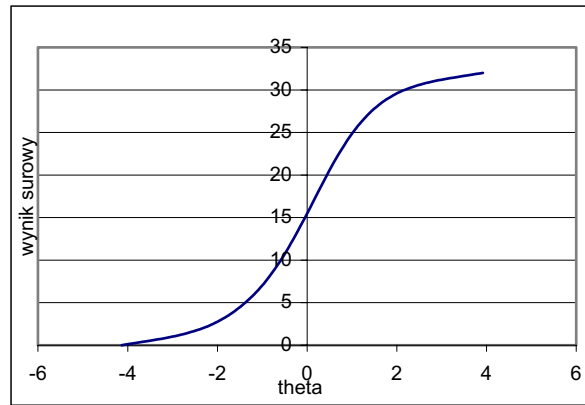
Estymatory trudności tej samej pozycji zmieniają się nieco w zależności od podzbioru pozycji, z którymi współpracuje. Analizując wielkości różnic między estymatorami trudności dla tej samej pozycji w trzech podtestach (od 0,09 do 0,19) oraz biorąc pod uwagę wartość błędu standardowego popełnianego przy ich szacowaniu, który mieści się w przedziale 0,15–0,18, można uznać, że różnice są nieistotne. Niezmienniczość mocy różnicującej konkretnej pozycji względem różnych grup pozycji, z którymi współtworzy test, jest wyraźna. Jeśli występują jakiegokolwiek różnice, to osiągają maksymalną wartość 0,02 (por. tab.2.).

Inaczej przedstawia się sytuacja w klasycznej teorii testu. Z definicji parametrów wynika ich zależność od próby osób, na podstawie której są wyznaczane. Na trudność pozycji nie mogą mieć również wpływu pozostałe pozycje testu, jednak mogą one zmieniać moc różnicującą.

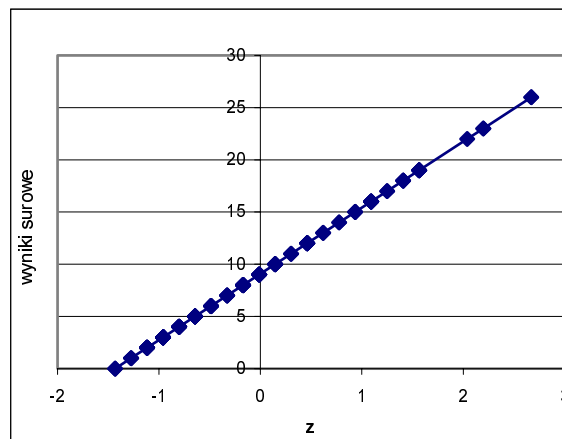
Po przeprowadzeniu analizy charakterystyk poszczególnych pozycji testu, w dalszej części przedstawione zostaną wyniki testu jako całości.

Na rys. 4. pokazano wyznaczoną przez program RUMM2010 zależność między wynikami surowymi testu a poziomem umiejętności (θ). Wykres ten umożliwia określenie poziomu cechy latentnej osoby, która w teście uzyskała określoną liczbę punktów. Można również odpowiedzieć na pytanie: ile punktów powinna uzyskać osoba, u której cecha latentna jest na określonym poziomie?

Zależność między wynikami surowymi i poziomem zmiennej θ w IRT ma kształt krzywoliniowy i, bez względu na wyniki w próbie, jest ekstrapolowana na cały zakres zmienności narzędzia. W klasycznej teorii testu zależność wyników surowych i wartości standardowej (z) jest prostoliniowa i ograniczona wartościami minimalnego i maksymalnego wyniku w próbie (rys. 5.).



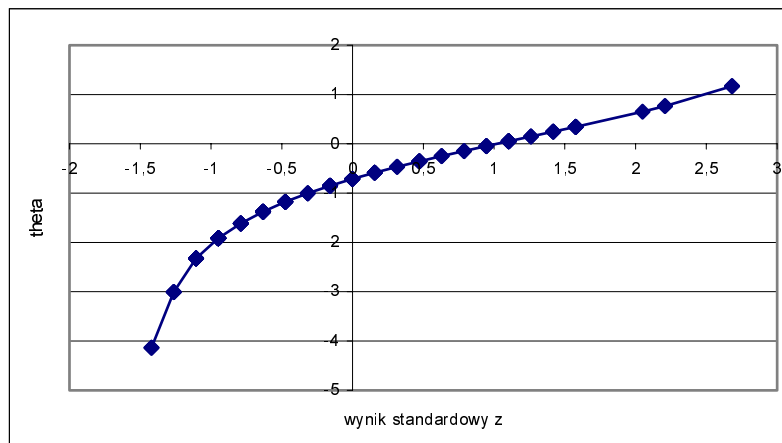
Rys. 4. Wynik surowy testu a poziom umiejętności

Rys. 5. Wyniki surowe a wartość standardowa z

Należy zwrócić uwagę na różnice w relacjach między wynikiem otrzymanym w teście a wartościami odpowiednio z i θ . W ujęciu klasycznej teorii, osobie, która uzyskała określoną liczbę punktów w teście, będzie przyporządkowana zawsze ta sama wartość zmiennej standardowej (z). W IRT taka jednoznaczność przyporządkowania obowiązuje tylko w przypadku jednoparametrycznego modelu Rascha. Jest to konsekwencja przyjętego w tym modelu założenia o stałej (równej 1) wartości mocy różnicującej wszystkich pozycji. W modelu dwu- i trójparametrycznym poziom umiejętności badanego zależy nie tylko od liczby punktów uzyskanych w teście, ale i od specyfiki rozkładu odpowiedzi na pozycje. Oznacza to, że dwóm osobom, które uzyskały tę samą ilość punktów w teście, może zostać przypisany inny poziom cechy latentnej¹⁴.

¹⁴ Baker F.B., Kim S., op. cit.

W celu sprawdzenia, jakie są relacje między usytuowaniem badanych na skali zmiennej standardowej z (klasyczna teoria) oraz na skali cechy latentnej θ (IRT), przedstawiono wyniki na wykresie (rys. 6.). Wszystkie punkty ułożyły się na gładkiej krzywej, co uprawnia do stwierdzenia, że zależność jest funkcją krzywoliniową. Uporządkowanie badanych na obu skalach jest takie samo, jednak odległości nie są zachowane. Jest to szczególnie widoczne przy wynikach najniższych i najwyższych.

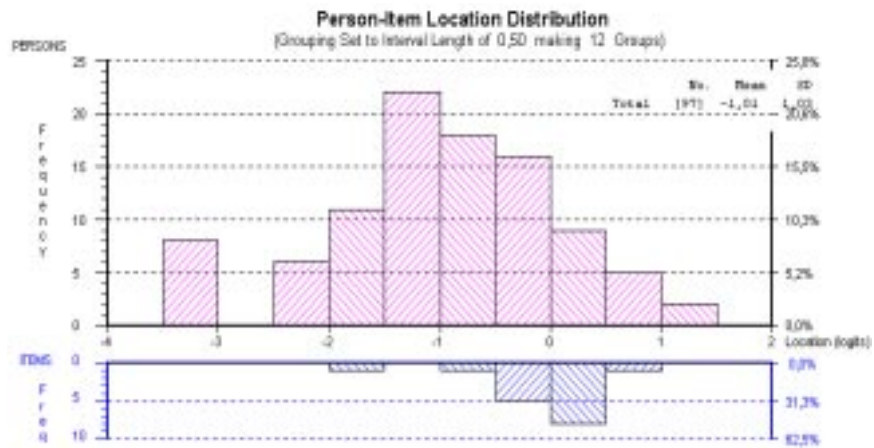


Rys. 6. Zależność thety od z dla pozycji 1–16

Na rys. 7. przedstawiono rozkłady wyników testu (część górna wykresu) oraz trudności zadań (część dolna wykresu), wyrażone na standaryzowanej skali θ (oś odciętych X). Utworzono 12 przedziałów o szerokości 0,5. Oś rzędnych (Y) z lewej strony dotyczy liczności – w górnej części jest to liczba osób, w dolnej – liczba zadań w teście. Po prawej stronie oś Y została wyskalowana w procentach. Podana jest też ogólna liczba badanych umieszczonych na rozkładzie (No.), średni wynik testu w próbie (Mean) oraz dyspersja wyników testu mierzona odchyleniem standardowym (SD).

Interesujący dla konstruktora testu jest rozkład zadań. W analizowanym teście składającym się z 16 pozycji trudność 8 z nich mieści się w przedziale $0 < \theta < 0,5$. W przedziale $-0,5 < \theta < 0$ znajduje się 5 następujących zadań. Biorąc pod uwagę definicję trudności i mocy różnicującej, można powiedzieć, że w tym obszarze test dobrze różnicuje badanych. W obszarze $\theta < -0,5$ występują tylko trudności 2 zadań. Zważywszy na liczbę osób, których cecha latentna $\theta < -0,5$, można stwierdzić, że test ten nie różnicuje w sposób zadowalający osób o niskim poziomie bezradności.

Informacje o trudności i mocy różnicującej pozycji testu w IRT pozwalają tak dobierać je do konstruowanego narzędzia, by w sposób równomierny pokryć cały obszar zmienności badanej cechy latentnej. Takich informacji nie dostarcza analiza prowadzona w ramach klasycznej teorii. Umożliwia ona sporządzenie rozkładu wyników testu dla badanych, jednak nie daje rozkładów pozycji.



Rys. 7. Rozkłady wyników testu i trudności zadań w ujęciu IRT

Wnioski

Wydaje się, że w pełni kompetentna ocena zastosowania teorii powinna obejmować trzy wymiary. Pierwszy z nich to stosowany w każdej teorii aparat matematyczny. Jest on stosunkowo prosty w klasycznej, zaś skomplikowany w probabilistycznej teorii. Szczególnie ważne są metody estymacji i świadomość konsekwencji ich stosowania. Drugi wymiar dotyczy oceny jakości aplikacji programowych, których instrukcje dostarczane użytkownikowi wcale nie są czytelne. Trzeci wymiar to praktyczne zastosowania i konsekwencje wyboru teorii dla jakości otrzymywanych narzędzi badawczych. W tab. 3. zestawiono dostrzeżone wady i zalety dwóch możliwych do stosowania teorii testu.

Tab. 3. Zalety i wady KTT i IRT

Klasyczna teoria testu	Probabilistyczna teoria testu
Zalety	
1. Próba niewielka 2. Prosty aparat matematyczny do oszacowania parametrów 3. Nieskomplikowane obliczenia możliwe do przeprowadzenia nawet w Excelu 4. Intuicyjnie zrozumiała interpretacja wyników	1. Parametry pozycji testu i błędu standardowego pomiaru są niezależne od próby. 2. Trudność pozycji i poziom cechy latentnej wyrażone są w tych samych jednostkach. 3. Krzywa charakterystyczna pozwala oszacować poziom cechy latentnej (θ) na podstawie każdej pozycji testowej.
Wady	
1. Związek między wynikiem otrzymanym a wynikiem prawdziwym jest prostoliniowy. 2. Parametry zadań (trudność i moc różnicująca) zależą od próby, na podstawie której były szacowane. 3. Trudność pozycji i testu odnosi się do <i>przeciętnego</i> ucznia.	1. Charakterystyki pozycji powinny być otrzymywane na podstawie dużej próby (powyżej 1000 jednostek). 2. Iteracyjne dobieranie parametrów ICC do danych empirycznych wymaga zastosowania profesjonalnych aplikacji. 3. Pojęcia nie zawsze intuicyjnie zrozumiałe

