Roman KONARSKI
Instytut Psychologii
Uniwersytet Gdański

MODEL CECHY LATENTNEJ W ANALIZIE PSYCHOMETRYCZNEJ TESTÓW I POZYCJI TESTOWYCH

Model wyniku zadania testowego (item response theory, IRT) oraz model konfirmacyjnej analizy czynnikowej (confirmatory factor analysis, CFA) dla zmiennych dyskretnych są specjalnymi przypadkami bardziej ogólnego modelu cechy latentnej (latent trait model, LTM) (Takane, de Leeuw, 1987; Bartholomew, Knott, 1999; McDonald, 1999). Na wstępie przedstawię podstawowe założenia modelu IRT i modelu CFA dla zmiennych dyskretnych oraz omówię relacje pomiędzy parametrami obu modeli psychometrycznych. Następnie na przykładach konkretnych danych empirycznych zaprezentuję możliwości zastosowania modelu cechy latentnej w typowej analizie psychometrycznej pozycji testowych.

ZMIENNE LATENTNE I ZMIENNE OBSERWOWALNE

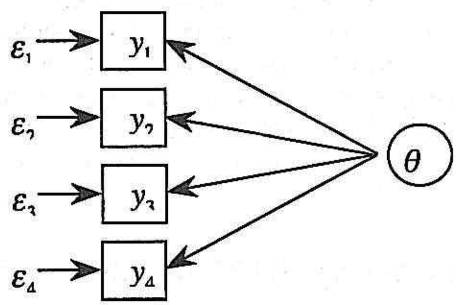
W diagnostyce edukacyjnej problem pomiaru jest szczególnie złożony, gdyż zmienne nas interesujące nie są bezpośrednio obserwowalne. Takie konstrukty psychologiczne jak inteligencja, kompetencja lub wiedza są zmiennymi latentnymi, które posiadają znacznie empiryczne jedynie poprzez obserwowalne konsekwencje tych zmiennych (Hornowska, 1989). Obserwowalne konsekwencje zmiennych latentnych nazywamy wskaźnikami, pozycjami testowymi, lub zadaniami testowymi. Zauważmy, że takie ujęcie relacji pomiędzy zmiennymi latentnymi i ich obserwowalnymi wskaźnikami implikuje określoną relację przyczynową, w której konstrukty teoretyczne przyjmują rolę zmiennych wyjaśniających, a wskaźniki przyjmują rolę zmiennych wyjaśnianych[1] (Borsboom, Mellenbergh, van Heerden, 2003).

Relacja pomiędzy zmienną latentną θ i jej wskaźnikami y_1, y_2, y_3, i y_4 jest przedstawiona za pomocą wykresu ścieżkowego w rysunku 1. Zgodnie z ogólno przyjętą konwencją[2], komponenty wykresu ścieżkowego są przedstawiane w następujący sposób:

[1] Modele postulujące odwrotny kierunek relacji pomiędzy zmienną latentną i jej wskaźnikami były także formułowane (np., Bollen, Lennox, 1991). Tak zwane „modele formatywne" to modele dla wskaźników definicyjnych (Nowak, 1965), które są prawdopodobnie bardziej odpowiednie dla takich konceptów socjobgicznych jak status socjoekonomiczny. Ontologiczny charakter zmiennych latentnych i ich relacji ze zmiennymi obserwowalnymi jest dyskutowany w Borsboom i in. (2003).
[2] Istnieje także alternatywna konwencja RAM (McArdle, McDonald, 1984).

- zmienne obserwowalne (jakościowe lub ilościowe) są wskazane jako kwadrat i są nazwane literami rzymskimi (y);
- zmienna latentna jest wskazana jako koło i jest nazwane literą grecką (θ);
- jednokierunkowa strzałka pomiędzy dwoma zmiennymi postuluje bezpośredni wpływ jednej zmiennej na drugą zmienną;
- nieznane wartości parametrów modelu są oznaczone literami greckimi (λ, ε).

Przykład przedstawiony na rysunku 1 ilustruje model jednoczynnikowy, w którym cecha wspólna (θ) ma wskaźniki y_1, y_2, y_3, i y_4. Ponadto błąd pomiarowy każdej zmiennej obserwowalnej jest oznaczone jako ε_k. Relacja pomiędzy czynnikiem wspólnym i wskaźnikami jest oznaczona jako λ_k.



Rysunek 1. Wykres ścieżkowy modelu pomiarowego

Wykres modelu pomiarowego na rysunku 1 skutecznie pokazuje związek pomiędzy zmienną latentną i jej obserwowalnymi wskaźnikami oraz demonstruje, że reakcja na daną pozycję testową jest wynikiem dwóch latentnych komponentów: pozycji danej osoby na skali cechy latentnej oraz (latentnego) błędu pomiarowego. Zależność pomiarowa przedstawiona na rysunku 1 ujawnia dalej, że informacja zawarta w pojedynczej pozycji testowej nie jest wystarczająca aby oszacować dwa latentne komponenty odpowiedzi na daną pozycję testową. Z tego względu wymagane jest posiadanie wielokrotnego pomiaru danej cechy latentnej. Informacja o danej cesze latentnej jest zawarta w strukturze relacyjnej zmiennych obserwowalnych.

MODELE POMIAROWE

Wczesny rozwój modelu cechy latentnej (LTM) możemy przypisać pracy Spearmana (1904) nad współczynnikiem korelacji oraz wysiłkom Lawleya (1943), który postulował, że odpowiedzi na pozycje testowe są wynikiem nieobserwowalnej (latentnej) cechy lub kompetencji danej osoby. Lord (1953) później sformalizował postulat Lawleya (1943) poprzez różnorodne modele matematyczne pozwalające na predykcję poprawnej odpowiedzi na zadanie testowe bazując na właściwościach danej pozycji testowej oraz poziomu cechy latentnej (kompetencji/wiedzy).

Obecnie wyróżniamy cztery ogólne kategorie modeli zmiennej latentnej różniące się założeniami odnośnie skali zmiennej latentnej oraz skali pozycji testowych (Bartholomew, 1987; Borsboom i in., 2003). Bartholomew (1987) zaproponował klasyfikację modeli zmiennych latentnych, która jest przedstawiona w tabeli 1. Modele dla ciągłej zmiennej latentnej są nazywane modelami cechy latentnej (LTM) (Mellenbergh, 1994; Bartholomew, Knott, 1999). Analiza czynnikowa jest modelem pomiarowym dla relacji pomiędzy ciągłymi zmiennymi latentnymi i ciągłymi zmiennymi obserwowalnymi. Analiza czynnikowa została sformułowana jako formalny model statystyczny przez Lawleya i Maxwella (1963) oraz jako model konfirmacyjnej analizy czynnikowej (confirmatory factor analysis, CFA) przez Jöreskoga (1971). W późniejszych latach model CFA został rozszerzony do dyskretnych pozycji testowych przez Christoffersona (1975), Muthéna (1984) oraz Jöreskoga (1990). Równolegle do rozwoju analizy czynnikowej, dzięki pracy Rascha (1960), Birnbauma (1968) oraz Lorda (1980), w tradycji edukacyjnej powstał model wyniku zadania testowego (item response theory, IRT). Początkowo, IRT określał relację pomiędzy ciągłą zmienną latentną i dychotomicznymi pozycjami testowymi. W późniejszych latach model IRT został rozszerzony, aby uwzględnić politomiczne pozycje testowe dzięki publikacjom Samejima (1969) oraz Bocka (1972). W równoległym programie badawczym, powstały modele statystyczne dla dyskretnych zmiennych latentnych (Lazarsfeld, 1950; Green, 1951; Lazarsfeld, Henry, 1968; Goodman, 1974). Wysiłki Lazarsfelda (1950), Lazarsfelda i Henrya (1968), oraz Goodmana (1974) zaowocowały analizą latentnych struktur (latent structure analysis, LSA). Gdy w modelu LSA zmienne obserwowalne są dyskretne, model taki jest nazywany modelem latentnych klas (latent class analysis, LCA) (Green, 1951). Gdy zaś w modelu LSA zmienne obserwowalne są ciagłe, model jest nazywany modelem latentnych profili (latent profile analysis, LPA) (Lazarsfeld, Henry, 1968; Bartholomew, 1987).

Tabela 1. Klasyfikacja modeli zmiennych latentnych (Bartholomew, 1987)

		Zmienne obserwowalne	
		Ciągłe	Dyskretne
Zmienna latentna	Ciągła	analiza czynnikowa	analiza czynnikowa dla zmiennych dyskretnych teoria wyniku zadania testowego
	Dyskretna	analiza latentnych profili	analiza latentnych klas

Początkowo modele pomiarowe przedstawione w tabeli 1 powstawały w oddzielnych tradycjach statystycznych, co sprzyjało podkreślaniu różnic pomiędzy tymi modelami. Modele te dzielą jednak wspólne założenia odnośnie cechy latentnej oraz zmiennych obserwowalnych. Dwa fundamentalne założenia modeli zmiennej latentnej to jednowymiarowość testu (unidimensionality) oraz lokalna (warunkowa) niezależność (local independence) (Bartholomew, 1987). Wymóg jednowymiarowości testu oznacza, że statystyczna zależność pomiędzy pozycjami testowymi może być wyjaśniona przez jedną zmienną latentną, co oznacza, że pozycje testowe są statystyczne zależne w całej populacji osób testowanych i jednocześnie są statystycznie niezależne w każdej podpopulacji osób testowanych, których członkowie są homogeniczni pod względem wartości cech latentnej. Ponieważ ta niezależność jest zdefiniowana dla podpopulacji osób testowanych ulokowanych w danym punkcie na skali cechy latentnej, warunek ten nazywamy lokalną niezależnością (Crocker, Algina, 1986). Celem analizy zmiennej latentnej jest znalezienie (lub test a priori określonej) takiej zmiennej latentnej, która spełnia założenie lokalnej niezależności.

Obecnie obserwujemy tendencję do rozważania tych modeli pomiarowych w tabeli 1 jako specjalnych przypadków bardziej ogólnego modelu zmiennych latentnych (Bartholomew, 1987; Bartholomew, Knott, 1999; McDonald, 1999; Borsboom, i in., 2003). W obecnym artykule skupimy się na relacji pomiędzy tak zwanym dwuparametrycznym modelem IRT i modelem CFA dla zmiennych dyskretnych. W ostatnich latach kilku autorów zademonstrowało formalną (np., Takane, de Leeuw, 1987; Bock, Gibbons, Muraki, 1988; Bartholomew, 1987) oraz empiryczną (np., Knol, Berger, 1991; Hoijtink, Rooks, Wilmink, 1991; Raju, Laffitte, Byrne, 2002) ekwiwalencję pomiędzy tymi modelami statystycznymi. Na najbardziej ogólnym poziomie modele IRT i CFA różnią się zakresem wykorzystania informacji zawartej w danych. Model IRT jest tak zwanym modelem pełnej informacji (full-information model), ponieważ wykorzystuje całą informację zawartą we wzorcach odpowiedzi uzyskanych z p dychotomicznych pozycji testowych. Natomiast model CFA dla zmiennych dyskretnych jest nazywany modelem ograniczonej informacji (limited-information model), ponieważ wykorzystuje jedynie informacje zawarte w korelacjach tetrachorycznych pomiędzy parami p pozycji testowych.

MODEL IRT

W literaturze tematu wyróżnia się kilka modeli IRT (patrz Lord, 1980; Hambelton, Swaminathan, 1985). Modele te różnią się szczególną funkcją matematyczną wyjaśniającą prawdopodobieństwo poprawnej odpowiedzi na daną pozycję testową. My rozważymy jedynie tak zwany dwuparametryczny model logistyczny (two-parameter logistic model, 2PLM).

Z formalnego punktu widzenia problem analizy cechy latentnej może być sformułowany następująco: analizowane dane są zawarte w macierzy danych zawierającej obserwacje dla próby n osób na p zmiennych obserwowalnych (pozycji testowych)

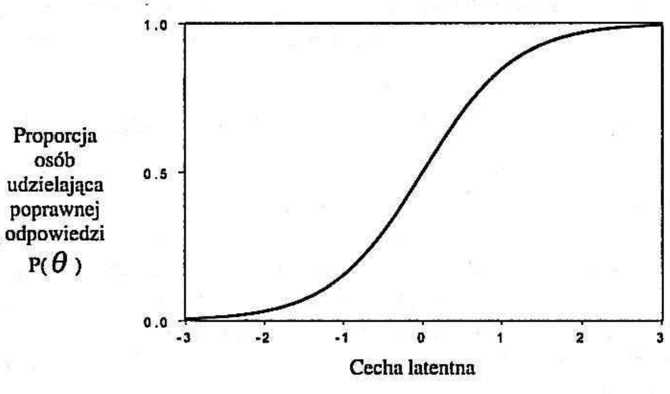
$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix}$$

Element macierzy danych y_{ik} są zaobserwowaną wartością zmiennej losowej y_k ($k = 1, 2, ..., p$) dla osoby i ($i = 1, 2, ..., n$). W konsekwencji możemy zdefiniować $\mathbf{y}^T = \left[y_1, y_2, \ldots, y_p \right]$ jako wektor p zmiennych obserwowalnych. W naszych rozważaniach zakładamy, że zmienne y_k są dychotomiczne i kodowane w następujący sposób:

$$y_k = \begin{cases} 1, & \text{dla poprawnej odpowiedzi} \\ 0, & \text{w przeciwnym razie} \end{cases}$$

W modelu IRT zakładamy, że odpowiedzi na pozycję testową mogą być wyjaśnione przez model matematyczny postulujący jedną zmienną latentną. Ta latentna cecha (zdolność) θ posiada skalę standardową normalną (ze średnią wartością zero i odchyleniem standardowym jeden) w zakresie plus minus nieskończoności (-∞, +∞). Prawdopodobieństwo podania poprawnej odpowiedzi[3] $P(\theta)$ na dychotomiczne zadanie testowe jest monotonicznie rosnącą i nieliniową funkcją cechy latentnej θ.

Centralnym konceptem modelu IRT jest krzywa charakterystyczna pozycji testowej (item charakteristic curve, ICC). Typowa krzywa ICC posiada kształt litery S[4] tak jak krzywe ICC przedstawione na rysunku 2. Wysokość krzywej ICC powyżej danego poziomu latentnej cechy θ reprezentuje proporcję osób testowanych na tym poziome uzdolnień, którzy mogą odpowiedzieć poprawnie na daną pozycję testową. Jak możemy zauważyć na rysunku 2, krzywa ICC określa prawdopodobieństwo poprawnej odpowiedzi $P(\theta)$ na pozycję testową jako funkcję cechy latentnej (θ), leżącej u podłoża wyniku testowego, oraz parametrów pozycji testowej.



Rysunek 2. Krzywa charakterystyczna (ICC) pozycji testowej

Obecnie dominującą parametryzacją krzywej ICC jest funkcja logistyczna (Lord, 1980). W logistycznym modelu dwuparametrycznym (2PLM) zakładamy, że prawdopodobieństwo udzielenia poprawnej odpowiedzi na daną pozycję testową *k* zależy od dwóch parametrów krzywej ICC

[3] Znaczenie terminu „poprawnej odpowiedzi" jest arbitralne i ma szczególne zastosowanie w sprawdzających testach edukacyjnych. W psychologicznych testach diagnostycznych będzie się to odnosiło się do szczególnej kategorii odpowiedzi.

[4] Inne funkcje są także możliwe, na przykład, funkcja krokowa (step function) taka jak we wzorcach odpowiedzi charakteryzowanych jako simplex lub perfect Guttman scale (Guttman, 1944).

$$P_k(\theta) = \frac{1}{1 + e^{-D\alpha_k(\theta - \beta_k)}},$$

gdzie stała D = 1,7 jest wartością maksymalizującą dopasowanie krzywej logistycznej do ogivy rozkładu normalnego (Lord, Novick, 1969). Parametr α jest określany mianem parametru mocy dyskryminacyjnej pozycji testowej, gdyż określa różnicę w prawdopodobieństwa poprawnej odpowiedzi $P(\theta)$ pomiędzy dwoma grupami osób ulokowanych dwóch na różnych poziomach cechy latentnej θ. Chociaż parametr mocy dyskryminacyjnej α jest teoretycznie zdefiniowany w przedziale $(-\infty, +\infty)$, to pozycje z negatywną lub zerową mocą dyskryminacyjną są niepożądane, a pozycje z mocą dyskryminacyjną większą niż 2,0 są bardzo rzadko obserwowane (Hambelton, Swaminathan, 1985). Pozycje testowe, które nie są efektywne w dyskryminacji pomiędzy różnymi poziomami uzdolnień, posiadają relatywnie płaskie krzywe ICC i niskie wartości parametru α. Pozycje testowe, które są efektywne w dyskryminacji pomiędzy różnymi poziomami uzdolnień posiadają relatywnie strome krzywe ICC i wysokie wartości parametru α. Parametr β określa trudność pozycji testowej i zwykle znajduje się w zakresie ±2,0 (Hambelton, Swaminathan, 1985). Wartość parametru β oznacza poziom cechy latentnej, na którym prawdopodobieństwo poprawnej odpowiedzi na pozycję testową jest równe 50 procent. Wartości β blisko — 2,0 charakteryzują łatwe pozycje testowe, natomiast wartości β blisko 2,0 charakteryzują relatywnie trudne pozycje testowe. Pozycje testowe z wartością β około 0,0 posiadają średnią trudność.

MODEL CFA DLA ZMIENNYCH DYSKRETNYCH

Możliwość i zalety stosowania konfirmacyjnej analizy czynnikowej (CFA) w analizie psychometrycznej testów i pozycji testowych były od dawna rozpoznane (patrz Jöreskog, 1971). Jednak tradycyjne metody analizy czynnikowej zakładają normalność wielozmiennową (multivariate normality) rozkładu analizowanych pozycji testowych. Wiąże się to z faktem, iż tradycyjne metody wykorzystują jedynie informację zawartą w macierzy kowariancji (lub korelacji Pearsona) analizowanych zmiennych. W przypadku analizy dychotomicznych pozycji testowych założenie normalności jest w oczywisty sposób pogwałcone. W konsekwencji tradycyjna analiza zmiennych dychotomicznych (oraz politomicznych) prowadziła do wielu trudności numerycznych oraz artefaktów metodologicznych. Ograniczenia tradycyjnych metod analizy czynnikowej w przypadku analizy zmiennych dyskretnych zostały przezwyciężone poprzez wykorzystanie korelacji tetrachorycznych (lub polychorycznych w przypadku zmiennych porządkowych) oraz specjalistycznych metod estymacji (Christofferson, 1975; Muthén, 1984, Jöreskog, 1990; Bartholomew, 1987).

W CFA dla zmiennych dyskretnych zakładamy, że dychotomiczna pozycja testowa y_k odzwierciedla jedynie nieobserwowalną zmienną ciągłą y_k^*, która posiada rozkład normalny oraz próg τ_k:

$$y_k = \begin{cases} 1, & \text{jeżeli } y_k^* \geq \tau_k \\ 0, & \text{jeżeli } y_k^* < \tau_k \end{cases}$$

Gdy wykorzystujemy macierz korelacji tetrachorycznych pomiędzy zmiennymi latentnymi y_k^* możemy zastosować analizę czynnikową, ponieważ zmienne y_k^* spełniają założenia odnośnie normalności rozkładu przyjmowane w modelu analizy czynnikowej

$$y_k^* = \tau_k + \lambda_k \theta + \varepsilon_k,$$

gdzie y_k^* jest latentną odpowiedzią na pozycje testową y_k, τ_k jest stałą regresji pozycji testowej na latentną cechę (czynnik) θ, λ_k jest ładunkiem czynnikowym danej pozycji testowej na θ, a ε_k reprezentuje błąd pomiarowy.

Stosując standardowe założenia modelu analizy czynnikowej odnośnie rozkładu cechy latentnej wraz z założeniami odnośnie rozkładu błędu pomiarowego (patrz Everitt, 1984), otrzymujemy macierz kowariancji Σ pomiędzy zmiennymi y^* jako

$$\Sigma_{y^*} = \Lambda\Lambda^T + \Psi^2,$$

gdzie Λ jest wektorem ładunków czynnikowych a Ψ^2 jest diagonalną macierzą wariancji błędu pomiarowe. Ponieważ obserwujemy jedynie zmienne dyskretne $\mathbf{y}$, nie jest możliwe zidentyfikowanie diagonalnych elementów implikowanej macierzy Σ_{y^*}. W konsekwencji diagonalne elementy macierzy Σ_{y^*} są ustalone do wartości jeden, a diagonalne elementy macierzy Ψ^2 są zdeterminowane jako

$$\Psi^2 = \mathrm{I} - \mathrm{diag}\left(\Lambda\Lambda^T\right).$$

Parametry modelu CFA dla zmiennych dyskretnych są typowo szacowane za pomocą estymatora ważonych najmniejszych kwadratów (weighted least squares, WLS)[5].

RELACJA POMIĘDZY PARAMETRAMI MODELI IRT i CFA

Takane i de Leeuw (1987) oraz Bartholomew (1987) niezależnie zademonstrowali, że model 2PLM oraz model CFA dla zmiennych dyskretnych są formalne ekwiwalentne. Parametry modelu 2PLM α_k i β_k ($k = 1, 2, ..., p$) mogą być wyrażone w postaci parametrów modelu CFA τ_k, λ_k, i ψ_k jako

$$\alpha_k = \lambda_k / \psi_k$$

oraz

$$\beta_k = -\tau_k / \psi_k,$$

gdzie λ_k jest ładunkiem czynnikowym pozycji testowej k, ψ_k jest pierwiastkiem wariancji błędu pomiarowego pozycji testowej k, a τ_k jest stałą regresji pozycji testo-

[5] Jöreskog i Moustaki (2001) oraz Oranje (2003) ostatnio porównali empiryczne zachowanie szeregu alternatywnych estymatorów.

wej k. Możemy także wyrazić parametry modelu CFA w postaci parametrów modelu 2PLM

$$\lambda_k = \alpha_k \left(1+\alpha_k^2\right)^{-1/2},$$

$$\tau_k = -\beta_k \left(1+\alpha_k^2\right)^{-1/2}$$

oraz

$$\psi_k = \left(1+\alpha_k^2\right)^{-1/2}.$$

PRZYKŁAD ANALIZY POZYCJI TESTOWYCH

Analiza pozycji testowych jest istotnym elementem procesu konstruowania testów. Ograniczenia klasycznych indeksów jakości pozycji testowych są od dawna dobrze udokumentowane w literaturze anglojęzycznej (patrz Allen i Yen, 1979 lub Crocker i Algina, 1986). Ostatnio, ograniczenia klasycznej teorii zostały także wskazane w literaturze polskojęzycznej (patrz Hornowska, 2001 lub Kryniewski, 2003). Zastosowanie IRT w analizie pozycji testowych pozwala przezwyciężyć ograniczenia klasycznych wskaźników jakości zadań testowych. Analiza pozycji testowych oparta o RT polega na szacowaniu krzywych ICC analizowanych pozycji testowych (Lord, 1980; Hambelton, 1983).

W naszym przykładzie analizy pozycji testowych posłużyliśmy się danymi pochodzącymi z podłużnego badania dzieci i młodzieży kanadyjskiej (National Longitudinal Survey of Children and Youth, NLSCY). Dane, które wykorzystaliśmy pochodzą z drugiej fali tego badania, przeprowadzonej na przełomie lat 1996 i 1997. Przeanalizowaliśmy sześć pozycji testowych mierzących cechę agresji fizycznej dla $n = 3284$ chłopców i dziewczynek w wieku od 8 do 9 lat. Respondent, którym była osoba najbardziej poinformowana na temat wyselekcjonowanego dziecka (w większości przypadków jedno z rodziców) odpowiadał czy każde z sześciu analizowanych zachowań agresywnych było zaobserwowane u danego dziecka. Pozycje testowe wraz z proporcją zaobserwowanych zachowań są pokazane w tabeli 2. Respondenci odpowiadali na 3-stopniwej skali Likerta: 1 = nigdy; 2 = czasami; 3 = często. Na potrzeby zaprezentowanych analiz psychometrycznych odpowiedzi respondentów zostały przekodowane w następujący sposób: 0 = nigdy; 1 = czasami lub często.

Tabela 2. Pozycje testowe oraz proporcja (p) zaobserwowanych zachowań dla próby (n = 3284) dzieci w wieku od 8 do 9 lat

Pozycja (k)	Treść pozycji testowej	p_k
Q6G	bije się z rówieśnikami	0,348
Q6X	reaguje ze złością, gdy inne dziecko nienaumyślnie zrani go/ją (na przykład, gdy wpadnie na niego/nią)	0,368
Q6AA	fizycznie atakuje ludzi	0,151
Q6NN	kopie, gryzie, uderza inne dzieci	0,149
Q6FF	grozi ludziom	0,109
Q6JJ	jest okrutny lub surowy w stosunku do innych	0,108

Analizę pozycji testowych rozpoczęliśmy od testu założenia jednowymiarowości[6]. Hattie (1985) argumentuje, że model CFA dla zmiennych dyskretnych może być zastosowany do określona wymiarowości pozycji testowych. Test modelu jednoczynnikowego wykazał zadawalające dopasowanie modelu do danych[7]: $\chi^2(9) = 31{,}601$, $p <$ 0,001; RMSEA = 0,028. W konsekwencji, możemy przyjąć, że założenie jednowymiarowości testu jest spełnione. Szacunki parametrów modelu CFA oraz modelu IRT są przedstawione w tabeli 3.

Tabela 3. Szacunki parametrów pozycji testowych dla modeli CFA i IRT

	Model CFA			Model IRT	
Pozycja (k)	λ_k	ψ_k^2	τ_k	α_k	β_k
Q6G	0,719	0,484	-0,390	1,033	0,561
Q6X	0,689	0,526	-0,338	0,950	0,466
Q6AA	0,918	0,157	-1,033	2,317	2,607
Q6NN	0,867	0,248	-1,041	1,741	2,090
Q6FF	0,832	0,308	-1,232	1,499	2,220
Q6JJ	0,803	0,355	-1,238	1,348	2,078

[6] Wszystkie zaprezentowane analizy przeprowadziliśmy za pomocą pakietu statystycznego *Mplus* 1.04 (Muthén, Muthén, 2000).

[7] W typowych zastosowaniach CFA, obok testu chi-kwadrat często stosuje się dodatkowe statystyki dobroci dopasowania (*goodness-of-fit statistics*). Jedną taką statystyką jest RMSEA (Steiger, 1990). Zastosowanie dodatkowych statystyk dobroci dopasowania modelu jest uzasadnione przez wrażliwość testu chi-kwadrat na nieistotne błędy w specyfikacji modelu w dużych próbach badawczych.

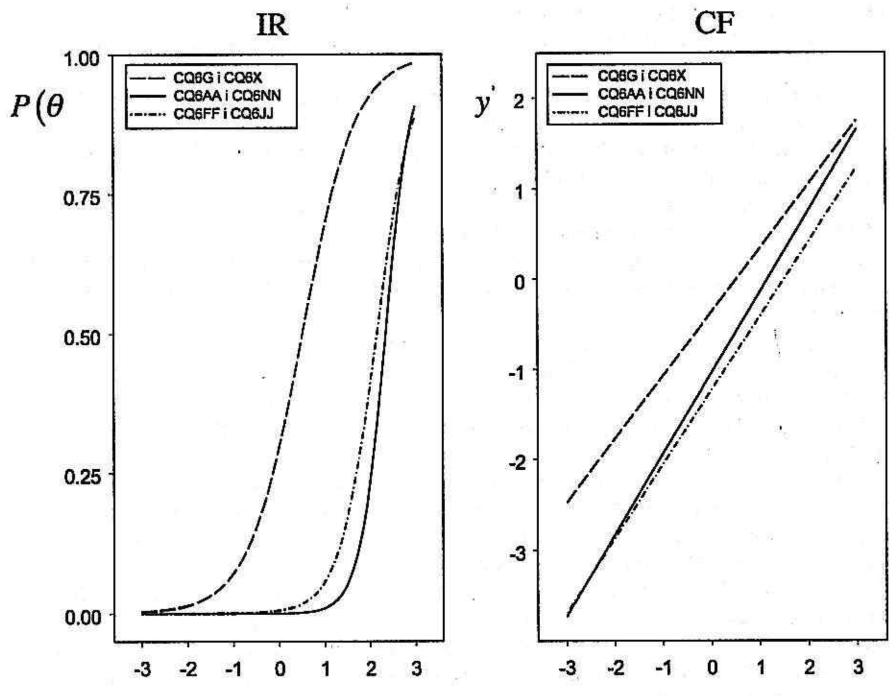
W następnym etapie analizy przetestowaliśmy hipotezę o równości krzywych charakterystycznych dla następujących par pozycji testowych: CQ6G i CQ6X, CQ6AA i CQ6NN, oraz CQ6FF i CQ6JJ. Jeżeli dwie pozycje testowe posiadają takie same ICC to z punktu widzenia parametrów trudności i mocy dyskryminacyjnej mogą być stosowane zamiennie. Test równości krzywych charakterystycznych jest przeprowadzony poprzez test modelu jednoczynnikowego, w którym ładunki czynnikowe oraz stałe regresji są ograniczone do równości dla każdej pary pozycji testowych: $\lambda_{CQ6G} = \lambda_{CQ6X}$, $\lambda_{CQ6AA} = \lambda_{CQ6NN}$, $\lambda_{CQ6FF} = \lambda_{CQ6JJ}$, oraz $\tau_{CQ6G} = \tau_{CQ6X}$, $\tau_{CQ6AA} = \tau_{CQ6NN}$, $\tau_{CQ6FF} = \tau_{CQ6JJ}$. Test tych ograniczeń może być przeprowadzony poprzez porównanie statystyki testowej χ^2 otrzymanej z testu modelu, w którym określone ograniczenia są nałożone ze statystyką testową χ^2 otrzymaną z testu modelu bez takich ograniczeń. Test modelu z nałożonymi ograniczeniami równości krzywych ICC dla par pozycji testowych dał wielkości statystyki dopasowania $\chi^2(15) = 42,768$, $p < 0,001$. Ta wielkość statystyki dopasowania modelu stanowi statystycznie nieistotny spadek w dopasowaniu modelu do danych, $\Delta\chi^2(\Delta df = 6) = 11,167$, $p > 0,05$. To potwierdza, że założenie o równych krzywych ICC dla par pozycji testowych jest uzasadnione.

Szacunki parametrów modelu z nałożonymi ograniczeniami na parametry pozycji testowych są przedstawione w tabeli 4. Jak możemy zauważyć w tabeli 4, pozycje testowe Q6G i Q6X posiadają umiarkowane wielkości parametrów trudność oraz mocy dyskryminacyjnej. Natomiast pozycje Q6AA i Q6NN posiadają szacunki parametru mocy dyskryminacyjnej oraz parametru trudności przekraczające wielkość 2,0, którą Hambelton i Swaminathan (1985) uważają za ekstremalną dla typowych pozycji testowych.

Tabela 4. Szacunki parametrów pozycji testowych dla modeli CFA i IRT z nałożonymi ograniczeniami na parametry modelu

	Model CFA			Model IRT	
Pozycja (k)	λ_k	ψ_k^2	τ_k	α_k	β_k
Q6G	0,704	0,505	-0,365	0,991	0,514
Q6X	0,704	0,505	-0,365	0,991	0,514
Q6AA	0,896	0,197	-1,039	2,019	2,341
Q6NN	0,896	0,197	-1,039	2,019	2,341
Q6FF	0,819	0,329	-1,233	1,428	2,150
Q6JJ	0,819	0,329	-1,233	1,428	2,150

Wykresy krzywych ICC dla modelu IRT oraz funkcji regresji dla modelu CFA są pokazane na rysunku 3. Jak możemy zauważyć oba modele psychometryczne dają porównywalny obraz funkcjonowania analizowanych pozycji testowych. Pozycje Q6G i Q6X najlepiej różnicują dzieci charakteryzujące się umiarkowanym poziomem agresji, natomiast pozostałe pozycje testowe (Q6AA, Q6NN, Q6FF i Q6JJ) jedynie różnicują dzieci charakteryzujące się relatywnie wysokim poziomem agresji.



Rysunek 3. Krzywe ICC dla modelu IRT oraz funkcje regresji dla modelu CFA

LITERATURA

Allen M.J., Yen W.M., (1979), *Introduction to measurement theory.* Monterey, CA: Brooks/Cole.

Bartholomew D.J., (1987), *Latent variable models and factor analysis.* New York: Oxford University Press.

Bartholomew D.J., Knott, M., (1999), *Latent variable models and factor analysis* (2nd ed.). London: Arnold Kendall's Library of Statistics.

Bock R.D., (1972), *Estimating item parameters and latent ability when responses are scored in two or more nominal categories.* Psychometrika, 37, 29-51.

Bock R.D., Gibbons R., Muraki E., (1988), *Full-information item factor analysis.* Applied Psychological Measurement, 12, 261-280.

Bollen K., Lennox, R., (1991), *Conventional wisdom on measurement: A structural equation perspective.* Psychological Bulletin, 110, 305-314.

Borsboom D., Mellenbergh, van Heerden J., (2003), *The theoretical status of latent variables*. Psychological Review, 110, 203-219.

Christofferson A., (1975), *Factor analysis of dichotomized variables*. Psychometrika, 40, 5-32.

Goodman L.A., (1974), *Exploratory latent structure analysis using both identifiable and unidentifiable parameters*. Biometrika, 61, 215-231.

Green B.F. (1951). *A general solution for the latent class model of latent structure analysis*. Psychometrika, 16, 151-166.

Crocker L., Algina J., (1986), *Introduction to classical and modern test theory*. San Francisco, CA: Holt, Rinehart and Winston.

Guttman L.A., (1944), *A basis for scaling qualitative data*. American Sociological Review, 9, 139-150.

Hambelton R.K. (Ed.)., (1983), *Applications of item response theory*. Vancouver, BC: Educational Research Institute of British Columbia.

Hambelton R.K., Swaminathan H., (1985), *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff Publishing.

Hattie J., (1985), *Methodology review: Assessing unidimensionality of test and items*. Applied Psychological Measurement, 9, 139-164.

Hoijtink H., Rooks G., Wilmink F.W, (1999), *Confirmatory factor analysis of items with dichotomous response format using the multidimensional Rasch model*. Psychological Methods, 4, 300-314.

Hornowska E., (1989), *Operacjonalizacja wielkości psychologicznych. Założenia – struktura – konsekwencje*. Wrocław: Ossolineum.

Hornowska E., (2001), *Testy psychologiczne. Teoria i praktyka*. Warszawa: Scholar.

Jöreskog K.G., (1971), *Statistical analysis of sets of congeneric tests*. Psychometrika, 36, 109-133.

Jöreskog K.G., (1990), *New developments in LISREL: analysis of ordinal variables using polychoric correlations and weighted least squares*. Quality & Quantity, 24, 387-404.

Jöreskog K.G., Mustaki I., (2001), *Factor analysis of ordinal variables: A comparison of three approaches*. Multivariate Behavioral Research, 36, 347-387.

Knol D.L., Berger M.P., (1991), *Empirical comparison between factor analysis and multidimensional item response models. Multivariate Behavioral Research. 26*, 457-477.

Kryniewski M., (2003), *Klasyczne i probabilistyczne miary jakości zadania testowego – nowe możliwości*, [w:] B. Niemierko (red.), *Trafność pomiaru jako podstawa obiektywizacji egzaminów szkolnych*. Łódź: Druk.

Lawley D.N., (1943), *On problems connected with item selection and test construction*. Proceedings of the Royal Society of Edinburgh, 61, 273-287.

Lawley D.N., Maxwell A.E., (1963), *Factor analysis as a statistical method*. London: Butterworth.

Lazarsfeld P.F., Henry N.W., (1968), *Latent structure analysis*. New York: Houghton-Miffli

Lazarsfeld P.F., (1950), *The logical and mathematical foundations of latent structure analysis*, [w:] S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, J.A. Clausen. (Eds.). *Measurement and prediction*. Princeton, NJ: Princeton University Press.

Lord F.M., (1953), *The relation of test score to the trait underlying the test*. Educational and Psychological Measurement, 13, 517-548.

Lord F.M., (1980), *Applications of item response theory to practical testing problems*. Hillsdale, NJ: LEA Publishers.

Lord F.M., Novick, M.R., (1968), *Statistical theories of mental test scores*. New York: Addison Wesley.

McDonald R.P, (1999), *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

McArdle J.J., McDonald R.P., (1984), *Some Algebraic Properties of the Reticular Action Model.* British Journal of Mathematical and Statistical Psychology, 37, 234 -251.
Mellenbergh G.J., (1994), *A unidimensional latent trait model for continuous item responses.* Multivariate Behavioral Research, 29, 223-236.
Muthén B.O., (1984), *A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators.* Psychometrika, 49, 115-132.
Muthén L.K., Muthén B.O., (1998-2000), *Mplus: Statistical analysis with latent variables.* Los Angeles, CA: Muthén & Muthén.
Nowak S., (1965), *Pojęcia i wskaźniki,* [w:] S. Nowak (red.), *Studia z metodologii nauk* społecznych. Warszawa: PWN.
Oranje A., (2003), *Comparison of estimation methods in factor analysis with categorized variables: Applications to NAEP data.* Educational Testing Service.
Raju N.S., Laffitte L.J., Byrne B.M., (2002), *Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory.* Journal of Applied Psychology, 87, 517-529.
Rasch G., (1960), *Probabilistic models for some intelligence and attainment tests.* Copenhagen: The Danish Institute for Educational research.
Samejima F., (1969), *Estimation of latent ability using a response pattern of graded scores.* Psychometrika, Monograph, 17.
Spearman C., (1904), *General intelligence objectively determined and measured.* American Journal of Psychology, 15, 201-293.
Staiger J.H., (1999), *Structural model evaluation and modification. An interval estimation approach.* .Multivariate Behavioral Research, 25, 173-180.
Takane Y., de Leeuw J., (1987), *On the relationship between item response theory and factor analysis of discredited variables.* Psychometrika, 52, 393-408.