

Test podłużnej niezmienności modelu podwójnego czynnika na przykładzie Kwestionariusza poczucia integracji rówieśniczej

PAWEŁ GRYGIEL

Instytut Badań Edukacyjnych*

W badaniach podłużnych zmiany badanego konstruktów mogą być interpretowane jedynie pod warunkiem spełnienia założenia o podłużnej niezmienności pomiarowej. Celem artykułu jest systematyczne testowanie podłużnej niezmienności pomiarowej modelu podwójnego czynnika z użyciem równań strukturalnych. Badany konstrukt zmierzono Kwestionariuszem poczucia integracji rówieśniczej (PIR) w trzech falach ogólnopolskiego badania szkolnych uwarunkowań efektywności kształcenia ($N = 4349$). Wyniki wskazują, że PIR jest narzędziem rzetelnym, substancjalnie jednowymiarowym, o strukturze podwójnego czynnika, niezmiennym konfiguralnie, metrycznie i skalarnie, ale nie ściśle. Można więc przyjąć, że w kolejnych falach badania układu czynników, wielkość ładunków czynnikowych i progów nie różnią się znacząco od siebie, ale rzetelność pomiaru jest różna: niższa w klasie 3 niż w klasach 5–6. Artykuł ukazuje także konsekwencje niespełnienia założeń związanych z niezmiennością pomiarową dla wyników analiz statystycznych.

SŁOWA KLUCZOWE: socjologia, podłużna niezmiennność pomiarowa, model podwójnego czynnika, postrzeganie relacji rówieśniczych, rzetelność, zmiany rozwojowe.

W szeroko pojętych naukach społecznych badacze często są zainteresowani konstruktami bezpośrednio nieobserwowalnymi, określanymi mianem zmiennych ukrytych lub cech latentnych (Byrne, 2012). Konstrukty są dostępne tylko pośrednio. Każdy z nich jest reprezentowany przez zestaw obserwowalnych zachowań, których wzajemne podobieństwo opiera się na wspólnym ukrytym źródle – postulowanej cesze latentnej (Raykov i Marcoulides, 2011). Możliwość wnioskowania o cesze latentnej wymaga zbadania wewnętrznej

struktury narzędzia, tj. określenia jego: (a) wymiarowości, (b) rzetelności oraz (c) pomiarowej niezmienności.

Jednowymiarowość i model podwójnego czynnika (*bifactor*)

Jedną z najstarszych i najlepiej rozwiniętych metod statystycznych służących do badania wewnętrznej struktury narzędzia jest analiza czynnikowa. Oddziela ona efekt oddziaływania cechy ukrytej, czyli wariację wspólną wskaźników, od (a) wariacji swojej, której źródłem może być specyficzna

* Adres: ul. Górczewska 8, 01-180 Warszawa.
E-mail: p.grygiel@ibe.edu.pl

© Instytut Badań Edukacyjnych

cecha ukryta (inna niż leżąca u podstaw wariacji wspólnej), oraz (b) od losowego błędu wariacji wskaźnika związanego z błędem pomiaru lub brakiem rzetelności wskaźnika (Brown, 2006).

Klasyczne modele czynnikowe opierały się na założeniu lokalnej niezależności, tj. wzajemnej niezależności wariacji swoistej w stosunku do czynnika latentnego. Jeżeli źródłem leżącym u podstaw relacji występujących między zmiennymi obserwowalnymi (wskaźnikami) jest jeden czynnik latentny, to kontrolowanie jego poziomu powinno eliminować wszelkie dodatkowe związki między wskaźnikami (Reise, Morizot i Hays, 2007). Tak rozumiana jednowymiarowość, określana jako ścisła (Stout, 1987), okazała się jednak zbyt restrykcyjna w praktyce badawczej (Bonifay, Reise, Scheines i Meijer, 2015; McDonald, 1981). W zamian zaproponowano pojęcie jednowymiarowości wystarczającej (zasadniczej, substancjalnej; Stout, 1990), która dopuszcza oprócz głównego wymiaru także wymiary poboczne (czynniki specyficzne, lokalne), pod warunkiem że to, co wspólne dla wszystkich wskaźników, będzie dominowało nad tym, co łączy tylko ich część (Reise, Moore i Haviland, 2010).

W ostatnich latach rozwinięto procedury testowania jednowymiarowości wystarczającej w ramach równań strukturalnych (*structural equation modeling*, SEM). Do najważniejszych można zaliczyć modele podwójnego czynnika (*bifactor*) oraz struktur wyższego rzędu (*higher-order factor models*; Humenny i Grygiel, 2015b; Reise, Waller, i Comrey, 2000). Oba zakładają, że zmienność wskaźników można podzielić na dwie grupy: (a) wspólną dla wszystkich wskaźników i (b) specyficzną jedynie dla ich części. W modelu podwójnego czynnika czynnik główny jest definiowany przez ładunki czynnikowe wszystkich pozycji skali, a czynniki poboczne – przez mniejsze wiązki pozycji. Wariancja pozycji jest więc

rozdzielana na trzy części: (a) wspólną dla wszystkich; (b) wspólną dla części (reprezentującą tę część wariacji pozycji kwestionariusza, której nie daje się wyjaśnić czynnikiem głównym, a których jednocześnie nie można zredukować do losowego błędu pojedynczego wskaźnika); (c) charakterystyczną tylko dla pojedynczego wskaźnika (Chen, West i Sousa, 2006; Chen, Hayes, Carver, Laurenceau, Zhang, 2012).

W modelach wyższego rzędu to, co wspólne dla wszystkich wskaźników, nie jest estymowane – jak ma to miejsce w modelach podwójnego czynnika – jako część wspólna pozycji skali, lecz jako część wspólna wariacji czynników niższego rzędu¹. W modelach wyższego rzędu relacje między wskaźnikami a czynnikiem wyższego rzędu są zapośredniczone przez czynniki pierwszego rzędu, natomiast w modelach podwójnego czynnika są one bezpośrednie². W konsekwencji modele podwójnego czynnika pozwalają zidentyfikować sytuacje, w których czynnik poboczny został niepoprawnie zdefiniowany, np. przez włączenie pozycji niemających de facto wspólnej wariacji z innymi pozycjami tworzącymi ten czynnik poza wariacją

¹ Modele struktur wyższego rzędu stanowią proste rozwinięcie idei leżącej u podstaw analizy czynnikowej – że wspólną wariację zmiennych obserwowalnych można tłumaczyć leżącą u jej podstaw zmienną lub zmiennymi latentnymi. Rozwinięcie polega na zastąpieniu analizy korelacji lub kowariancji występujących między zmiennymi obserwowalnymi korelacjami między zmiennymi latentnymi. W pewnym uproszczeniu można powiedzieć, że w modelach struktur wyższego rzędu „zmienia się” (poszerza) jednostka analizy. O ile w klasycznej analizie czynnikowej jest nią macierz korelacji między wskaźnikami obserwowalnymi, o tyle w modelach wyższego rzędu jest nią dodatkowo także macierz korelacji między czynnikami. Korelacje zachodzące między zmiennymi – same w sobie oparte na korelacjach między zmiennymi obserwowalnymi – stają się macierzą wejściową dla analiz wyższego rzędu. Per analogiam, jeśli czynniki drugiego rzędu są ze sobą skorelowane, to istnieje również możliwość wyłonienia czynników trzeciego czy kolejnego rzędu (Thompson, 2004).

² Dlatego modele podwójnego czynnika nazywane są także modelami bezpośrednio hierarchicznymi (*direct hierarchical*; Gignac, 2008).

związaną z czynnikiem głównym³. Modele podwójnego czynnika są więc bardziej użyteczne na etapie testowania struktury czynnikowej, szczególnie gdy obejmuje ona także ocenę trafności wykorzystanych zmiennych obserwowalnych (Chen i in., 2006; 2012; Reise i in., 2010).

Rzetelność czynnika głównego i czynników pobocznych

Jeśli model podwójnego czynnika okazuje się lepiej dopasowany do danych niż model zakładający jeden wymiar lub większą ich liczbę, to fakt ten przemawia na rzecz jednowymiarowości skali, ale nie stanowi przesądzającego dowodu. Wystąpić może bowiem sytuacja, w której zarówno czynnik główny, jak i czynniki poboczne będą wyjaśniać znaczną część zmienności wskaźników. W takich wypadkach uzasadnione będzie przetwarzanie informacji niesionej przez wszystkie pozycje skali łącznie i przez grupy pozycji (podskale). Nie jest to sytuacja rzadka. Mamy z nią do czynienia np. w pomiarze jakości życia, w ramach której wyróżnia się ogólną satysfakcję z życia oraz satysfakcję z określonych jego aspektów, np. relacji międzyludzkich czy zabezpieczenia materialnego (Gibbons i in., 2007).

W modelach podwójnego czynnika porównaniu proporcji wariancji wspólnej (związanej z czynnikiem głównym) ze specyficzną (związaną z czynnikami pobocznymi) służą takie miary, jak omega hierarchiczna (ω_h), omega specyficzna (ω_s) lub współczynnik wyjaśnionej wspólnej wariancji (*explained common variance*, ECV). Pierwsza wskazuje, jaką częścią wariancji wyników wszystkich pozycji jest wariancja czynnika głównego. Proponuje się, aby wartość ω_h powyżej 0,7 uznać za

świadczącą o substancjalnej jednowymiarowości skali (Reise, Bonifay i Haviland, 2013; Reise, Scheines, Widaman i Haviland, 2013). Druga, omega specyficzna, informuje o rzetelności pomiaru dokonywanego tylko na podstawie pozycji tworzących określony czynnik poboczny, a więc o tym, o ile informacja z nim związana wykracza poza to, co przekazuje czynnik główny. Wartość 0,5 uznaje się za dolną granicę, od której można uznać, że dany czynnik poboczny ma znaczenie merytoryczne niezależne od czynnika głównego (Reise, Bonifay i in., 2013). Trzecia, ECV, jest uznawana za najlepszą miarę jednowymiarowości, informuje bowiem o proporcji wspólnej wariancji wyjaśnianej przez czynnik główny (Ten Berge i Sočan, 2004). Wartość ECV powyżej 0,6 uznaje się za świadczącą o przynajmniej substancjalnie jednowymiarowym charakterze narzędzia (Reise, Scheines i in., 2013), zaś wartość powyżej 0,9 – o jego ścisłej jednowymiarowości (Quinn, 2014).

Niezmiennność pomiarowa

Jeśli narzędzie badawcze ma być zastosowane do osób o dynamicznie zmieniających się cechach poznawczych, emocjonalnych itp., to określenie struktury czynnikowej oraz rzetelności w jednym okresie może okazać się niewystarczające (Widaman, Ferrer i Conger, 2010). Wskutek zmian rozwojowych stanowiących immanentną część dzieciństwa i młodości (Kail, 2010) przekształceniom mogą podlegać nie tylko możliwości zrozumienia treści pytań kwestionariusza, lecz także sam konstrukt, który za jego pomocą chcemy badać. W wypadku relacji rówieśniczych pojawiają się nowe formy interakcji (np. kliki lub diadyczne relacje z płcią przeciwną), które zyskują lub tracą na znaczeniu w różnych okresach. Badania wskazują, że choć już w wieku 6 lat dzieci potrafią oceniać swoje relacje społeczne, to poziom adekwatności tej oceny, czyli zgodność z zewnętrznymi kryteriami,

³ Model podwójnego czynnika wykaże w takim przypadku zerowe ładunki wskaźnika na czynniku specyficznym i znaczące ładunki na czynniku głównym. Sytuacja taka będzie trudna do wykrycia w wypadku modeli wyższego rzędu.

ulega zmianom – stopniowo rośnie od środowkowego okresu dzieciństwa (wiek 6–12 lat) aż po okres dorastania (wiek 13–18 lat; Malloy, Sugarman, Montvilo i Ben-Zeev, 1995; Malloy, Yaras, Montvilo i Sugarman, 1996). Wymaga to wykazania, że narzędzie pomiaru na różnych etapach rozwoju mierzy ten sam konstrukt, czyli że jest czasowo niezmiennie (Millsap, 2011).

Warunek ten nabiera szczególnego znaczenia, gdy zamierzamy wnioskować o stabilności określonej cechy w różnych okresach – czyli o tym, czy wysoki poziom cechy w jednym okresie wiąże się z wysokim poziomem w innych (jak ma to miejsce w modelach autokorelacyjnych lub autoregresyjnych) – lub o różnicach w średnich natężenia interesującej nas zmiennej latentnej w następujących po sobie falach badania (np. czy jej poziom generalnie rośnie, spada, czy też nie ulega zasadniczym zmianom, co pokazują różne postacie latentnych krzywych rozwojowych).

Celem przyświecającym testowaniu pomiarowej niezmienności jest weryfikacja założenia, że badani z różnych grup lub w różnych okresach, którzy mają taki sam wynik prawdziwy (w sensie klasycznej teorii testu), będą mieć taki sam wynik obserwowalny. Innymi słowy, że wynik obserwowalny nie zależy od przynależności do grupy lub od okresu pomiaru, lecz jedynie od poziomu badanego konstrukt (Meredith, 1964; 1993). Wśród badaczy istnieje powszechna zgoda, że o ile np. analizy stabilności wymagają jedynie metrycznego poziomu niezmienności (równości ładunków), o tyle sensowność porównań różnic między średnimi musi opierać się na występowaniu przynajmniej częściowej niezmienności typu skalarnego (Byrne, Shavelson i Muthén, 1989; Little, 1997; Little, Preacher, Selig i Card, 2007), która zakłada, że zarówno średnie pozycji (zaś w wypadku zmiennych jakościowych – progii), jak i ładunki czynnikowe, nie różnią się znacząco między

grupami lub falami badania (Meredith, 1993). Jeżeli model czynnikowy przedstawimy w postaci równania regresji (Lubke, Dolan, Kelderman i Mellenbergh, 2003), to ładunki czynnikowe będą odpowiadać za nachylenie linii regresji wyniku obserwowalnego do latentnego, zaś średnie pozycji (progii) – za to, czy centrum zmiennej latentnej w różnych grupach (okresach pomiaru) znajdzie się w tym samym miejscu. Brak równości ładunków lub średnich (progów) oznacza brak równoległości między liniami regresji, a więc konieczność odrzucenia hipotezy, że w analizowanych grupach (okresach) taki sam wynik obserwowalny wiąże się z takim samym poziomem zmiennej latentnej.

Poziomy niezmienności pomiarowej i ich testowanie

Najczęściej stosowana procedura testowania niezmienności pomiarowej opiera się na sekwencyjnej estymacji serii hierarchicznie zagnieżdżonych modeli ze wzrastającą liczbą ograniczeń (Vandenberg i Lance, 2000). Pierwszym testowanym modelem jest model konfiguracyjny, w którym nie zakłada się jakiegokolwiek niezmienności narzędzia pomiędzy kolejnymi falami badań – wszystkie parametry modelu w każdym z okresów są testowane jako potencjalnie niezależne od siebie. W kroku drugim jest estymowany model metryczny, z ograniczeniem na wzdłużną wielkość ładunków czynnikowych (tzw. słaba niezmienność). W trzecim estymowany jest model skalarny, w którym ograniczenie równości nakłada się na wielkość ładunków czynnikowych i na odpowiadające sobie w kolejnych turach pomiarowych progii twierdzeń.

Ostatnim możliwym do testowania typem niezmienności jest tzw. niezmienność ścisła, zakładająca przybliżoną równość nie tylko ładunków i progów, lecz także błędów związanych z poszczególnymi pozycjami. Ten poziom niezmienności

zakłada, że różnice w wariancji zmiennych obserwowalnych wynikają jedynie z różnic w wariancji zmiennych latentnych, wariancje błędów zaś nie różnią się zasadniczo od siebie między grupami lub okresami. Jest to więc wskaźnik międzygrupowej lub czasowej rzetelności pomiaru. Uważa się, że brak niezmienności tego typu nie stanowi bariery dla dokonywania porównań między średnimi cechy latentnej w różnych punktach pomiaru (Little, 1997; 2013), uniemożliwia jednak eksplorację danych przez uśrednienie lub zsumowanie pozycji. Innymi słowy: jej brak oznacza, że nieadekwatne są metody analizy, które nie pozwalają na uwzględnienie zróżnicowania związanego z błędami resztowymi, a więc m.in. opierające się na estymacji kontraktu latentnego przez proste uśrednienie lub zsumowanie wartości liczbowych przypisanych do poszczególnych obserwowalnych wskaźników (odpowiedzi na pytania).

Przyjęcie lub odrzucenie hipotezy o niezmienności pomiarowej (na danym poziomie) odbywa się przez sprawdzenie, czy nałożone ograniczenia w znaczącym stopniu pogarszają dopasowanie modelu do danych w stosunku do modelu bez tych ograniczeń. Jeśli model zostanie odrzucony (tj. stwierdzony zostanie brak niezmienności), to analizie poddaje się indeksy modyfikacyjne (Saris, Satorra i Veld, 2009) w celu znalezienia parametru w największym stopniu odpowiadającego za brak niezmienności. Model jest estymowany ponownie bez ograniczenia nałożonego na zmienny (nie niezmienny) parametr oraz ponownie porównywany z modelem bazowym. Proces powtarza się aż do momentu, w którym miary dopasowania nie wskażą na brak znaczącego pogorszenia testowanego modelu. Przyjmuje się, że pomiar można uznać za niezmienny, jeśli większość analizowanych wskaźników (ładunków, progów itp.) wykazuje cechę niezmienności (Reise, Widaman i Pugh, 1993).

Cel analiz

Zasadniczym celem tego artykułu jest ukazanie użyteczności modeli podwójnego czynnika (*bifactor*) w analizie struktury latentnej narzędzi pomiaru stosowanych w badaniach podłużnych. Zostanie on zrealizowany na przykładzie kompleksowej analizy struktury czynnikowej Kwestionariusza poczucia integracji rówieśniczej (PIR) użytego w trzech falach ogólnopolskiego badania uczniów szkół podstawowych zrealizowanych w ramach projektu *Badania szkolnych uwarunkowań efektywności kształcenia* (SUEK). Według wiedzy autora, jest to pierwszy w literaturze przedmiotu artykuł, w którym podejmuje się wątek podłużnej niezmienności pomiarowej modelu podwójnego czynnika, przyjmując jakościowy (porządkowy) charakter wskaźników.

Kwestionariusz poczucia integracji rówieśniczej

Do testowania podłużnej niezmienności pomiarowej modelu podwójnego czynnika wykorzystano Kwestionariusz poczucia integracji rówieśniczej. Powstał on jako część większego narzędzia – Fragebogen zur Erfassung von Dimensionen der Integration von Schülern (FDI 4–6; Haeblerlin, Moser, Bless i Klaghofer, 1989) – służącego do pomiaru samooceny integracji ucznia ze zbiorowością uczniów w szkole. Za pomocą PIR uzyskujemy informację o opinii danego ucznia na temat postrzegania zarówno pozytywnych, jak i negatywnych relacji z kolegami lub koleżankami z oddziału szkolnego. Narzędzie składa się z ośmiu pozycji, których akceptacja oznacza pozytywne relacje rówieśnicze, oraz siedmiu, z którymi zgoda oznacza negatywne ich postrzeganie. Zadaniem badanych jest ocena tego, jak dobrze każde z piętnastu twierdzeń skali odzwierciedla sytuację respondenta. Zgodność tę uczniowie opisują na czteropunktowej skali: od

tak, przez raczej tak, raczej nie, aż do nie. Polska wersja narzędzia – przygotowana przez Grzegorza Szumskiego (2010) – była z powodzeniem wykorzystywana w wielu badaniach (Grygiel, Humenny, Rębisz, Bajcar i Świtaj, 2014; Grygiel, Humenny, Świtaj, Rębisz i Anczewska, 2014; Szumski i Karwowski, 2012; 2014).

Próba

Analizy przeprowadzono na danych pochodzących z trzech fal ogólnopolskiego badania podłużnego przeprowadzonego w Instytucie Badań Edukacyjnych w ramach projektu SUEK. Pierwsza fala została przeprowadzona w drugim semestrze roku szkolnego 2010/2011 wśród ponad 5 000 uczniów z losowo dobranych 181 oddziałów klasy 3 szkoły podstawowej. Kolejne dwie fale przeprowadzono wśród tych samych uczniów w pierwszym semestrze klasy 5 (rok

szkolny 2012/2013) oraz w drugim semestrze klasy 6 (rok szkolny 2014/2015). W analizach wykorzystano dane od uczniów, którzy w każdej z trzech fal wypełnili kwestionariusz PIR. Analizy uwzględniają więc łącznie odpowiedzi 4349 uczniów (49,7% dziewcząt). Średni wiek badanych mierzony w trakcie pierwszej fali (tj. w klasie 3) wyniósł 9,6 roku (z wariancją 0,1). Szczegółowy opis metodologii badania można znaleźć w publikacjach książkowych (Dolata, 2014; Dolata i in., 2015)⁴, a rozkłady uzyskanych odpowiedzi przedstawionow Tabeli 1.

Plan analiz

Analiza podłużnej struktury czynnikowej kwestionariusza PIR zostanie przeprowadzona w pięciu etapach. Pierwszy polega

⁴ Obie można pobrać ze strony Instytutu Badań Edukacyjnych (<http://eduentuzjasci.pl/publikacje-suek.html>).

Tabela 1

Rozkłady odpowiedzi akceptujących (suma „tak” oraz „raczej tak”) w PIR w trzech falach badania (w %)

Nr poz.	Treść twierdzenia	Klasa		
		3	5	6
2	Bardzo chętnie przebywam z koleżankami i kolegami ze szkoły. (R)	94,7	95,6	95,2
5	W klasie mam bardzo wielu przyjaciół. (R)	90,6	87,0	85,0
8	Moje koleżanki i koledzy z klasy zawsze chętnie mi pomagają. (R)	84,1	85,4	84,6
11	Niewiele koleżanek i kolegów z klasy lubię.	36,1	29,6	24,0
14	Chętnie spędzam wolny czas z koleżankami i kolegami z klasy. (R)	90,4	93,2	91,2
17	Często się złościę na moje koleżanki i kolegów z klasy.	14,0	14,3	13,7
20	Żyję w całkowitej zgodzie z moimi koleżankami i kolegami z klasy. (R)	84,7	83,0	82,2
23	W klasie jest wielu uczniów, którzy mi dokuczają.	20,6	16,4	11,1
26	W naszej klasie wszyscy jesteśmy dobrymi przyjaciółmi. (R)	77,2	67,8	67,5
29	W klasie czuję się samotny(-a).	8,0	8,7	6,9
32	Dobrze się rozumiem z moimi koleżankami i kolegami z klasy. (R)	91,1	90,4	90,2
35	Z moimi koleżankami i kolegami z klasy chętnie spotykam się też po szkole. (R)	82,9	88,3	88,7
38	W naszej klasie jest wielu uczniów, których nie lubię.	24,7	26,5	21,7
41	Nie zawsze dobrze się czuję w naszej klasie.	29,1	33,5	28,5
44	Nudzę się z moimi koleżankami i kolegami z klasy.	10,0	10,2	10,0

Numery pozycji zgodnie z kolejnością w Kwestionariuszu poczucia integracji rówieśniczej. (R) = twierdzenie, które należy przekodować tak, aby wyższy wynik oznaczał wyższy poziom zintegrowania z rówieśnikami.

na określeniu optymalnej liczby czynników niezbędnych do wyjaśnienia zmienności występującej między wskaźnikami. Liczba czynników zostanie wyznaczona na podstawie dwóch najczęściej wykorzystywanych w praktyce kryteriów (Schmitt, 2011): (a) Kaisera wartości własnych większych niż 1 oraz (b) wykresu osypiska. Poszczególne rozwiązania będą oceniane także ze względu na: (a) ich interpretowalność i (b) warunek posiadania przez czynnik przynajmniej trzech pozycji z ładunkami powyżej wartości 0,3, które to pozycje jednocześnie nie będą posiadały ładunków wyższych w ramach żadnego innego czynnika (Brown, 2006). Innymi słowy, jako warunek istnienia czynnika przyjmujemy nie tylko możliwość jego sensownej interpretacji, lecz także „swoistość” tworzących go pozycji.

Dodatkowo każde rozwiązanie zostanie ocenione ze względu na miary dopasowania. Wykorzystane zostaną: (a) pierwiastek ze średniego kwadratu błędu przybliżenia (*root mean square error of approximation*, RMSEA; Steiger, 1990), (b) nienormowany indeks dopasowania (*Tucker-Lewis index*, TLI; Tucker i Lewis, 1973) oraz (c) porównawczy indeks dopasowania (*comparative fit index*, CFI; Bentler, 1990). Zgodnie z powszechnie przyjętymi regułami (Marsh, Hau i Grayson, 2005; Yu, 2002) model wskazujący na adekwatne dopasowanie do danych powinien cechować się wartościami RMSEA równymi lub mniejszymi niż 0,06 oraz wartościami CFI i TLI większymi niż 0,90. Jednocześnie o lepszym dopasowaniu modelu do danych, bardziej adekwatnym odzwierciedleniu „faktycznej” struktury czynnikowej, będą świadczyć niższe wartości współczynników RMSEA oraz wyższe wartości współczynników CFI i TLI.

W drugim etapie analizy, po określeniu liczby czynników oraz wstępnym przyporządkowaniu do nich pozycji – na podstawie wybranego rozwiązania opartego na

eksploracyjnej analizie czynnikowej (*exploratory factor analysis*, EFA) – będzie estymowany confirmacyjny model podwójnego czynnika z jednym czynnikiem głównym oraz czynnikami pobocznymi w liczbie określonej w ramach EFA. Podobnie jak w wypadku analiz eksploracyjnych, także na tym etapie do oceny jakości dopasowania modelu do danych zostaną wykorzystane trzy miary: RMSEA, CFI oraz TLI oraz ta sama reguła (model dobrze odzwierciedlający strukturę danych, to model o wartościach RMSEA równych lub mniejszych niż 0,06, oraz wartościach CFI i TLI większych niż 0,90).

W trzecim etapie zostaną oszacowane miary rzetelności czynnika głównego oraz czynników pobocznych z rozwiązania podwójnego czynnika. Wykorzystane zostaną trzy współczynniki: ω_h , ω_s oraz ECV. Zakładamy, że hipoteza o substancjalnej jednowymiarowości zostanie przyjęta, gdy ω_h przekroczy wartość 0,7, zaś współczynnik ECV będzie wyższy niż 0,6. O możliwości wykorzystywania czynników pobocznych jako źródła informacji niezależnego od czynnika głównego będą świadczyć wartości współczynnika ω_s przekraczające 0,5.

Etap czwarty zostanie poświęcony weryfikacji niezmienności pomiaru dokonywanego kwestionariuszem PIR pomiędzy klasami 3 a 6. Sekwencyjnie będzie estymowana seria zagnieżdżonych modeli ze wzrastającą liczbą ograniczeń. Pierwszym testowanym będzie model konfiguracyjny (M1). Jest to model, w którym nie zakłada się jakiegokolwiek niezmienności narzędzia pomiędzy trzema falami badań (wszystkie parametry modelu w każdym z okresów będą testowane jako potencjalnie niezależne od siebie). W kroku drugim będzie estymowany model metryczny (M2) z ograniczeniem na wzdłużną wielkość ładunków czynnikowych (tzw. słaba niezmienność). W ramach kroku trzeciego będzie estymowany model skalarny (M3), w którym

– oprócz wielkości ładunków czynnikowych
 – ograniczenie równości zostanie nałożone także na progi twierdzeń odpowiadające sobie w kolejnych falach pomiarowych. W ostatnim kroku ograniczenia zostaną nałożone także na wariancję reszt poszczególnych wskaźników (M4). Testowana więc będzie niezmienność ścisła.

W celu sprawdzenia, czy nałożone ograniczenia w znaczącym stopniu pogarszają dopasowanie modelu do danych w stosunku do modelu bez ograniczeń, uwzględniona zostanie – zgodnie z propozycją zespołu Adama W. Meade’a (Meade, Johnson i Braddy, 2008) – zmiana dwóch miar: CFI i RMSEA. Zakładamy dosyć liberalną (łatwo pozwalającą odrzucić hipotezę zerową) regułę, że hipotezę o niezmienności pomiarowej odrzucimy, gdy różnica (Δ) między modelem z większą liczbą ograniczeń a modelem z mniejszą ich liczbą w wypadku CFI okaże się niższa od -0,002, zaś w przypadku RMSEA – wyższa od 0,007.

Uzyskanie przynajmniej częściowej niezmienności skalarnej jest punktem wyjścia do ostatniego (piątego) etapu analiz, tj. przeprowadzenia porównań między średnimi latentnymi opisującymi poziom satysfakcji z relacji rówieśniczych w klasach 3, 5 i 6, a także określenia poziomu stabilności pomiaru dokonywanego kwestionariuszem PIR w badanym okresie (autoregresji wyników między 3, 5 oraz 6 klasą).

Metody estymacji

Biorąc pod uwagę, że odpowiedzi na poszczególne pytania kwestionariusza PIR były mierzone na czteropunktowej skali porządkowej, analizy czynnikowe zostaną oparte na macierzy korelacji polichorycznej z wykorzystaniem estymatora ważonych najmniejszych kwadratów ze skorygowaną średnią i wariancją (*weighted least squares means and variance adjusted*, WLSMV). Analizy statystyczne zostaną przeprowadzone w pakiecie Mplus 7.3 (Muthén

i Muthén, 2012). Ze względu na hierarchiczny charakter danych (uczniowie zagrzeźdzeni w oddziałach) w analizach zastosowano odpowiednią poprawkę (opcja *complex* w Mplus), redukująca obciążenie błędów standardowych i testów statystycznych. Przed przeprowadzeniem analiz stwierdzenia „pozytywne” (tj. pozycje o numerach: 2, 5, 8, 14, 20, 26, 32 oraz 35) zostały przekodowane w taki sposób, aby wyższy wynik całej skali wskazywał na pozytywną ocenę klasowych relacji rówieśniczych.

Wyniki

Modele eksploracyjne

Zgodnie z zarysowanym wcześniej planem analizy rozpoczęto od określenia optymalnej liczby czynników niezbędnych do wyjaśnienia zmienności między wskaźnikami kwestionariusza PIR. Metoda wartości własnych większych niż jeden sugeruje wyodrębnienie trzech czynników w klasie 3 oraz dwóch w klasie 5 oraz 6 (Tabela 2).

Tabela 2

Wartości pierwszych pięciu składowych macierzy korelacji zmiennych kwestionariusza PIR

Klasa	Numer składowej				
	1	2	3	4	5
3	6,41	1,43	1,07	0,78	0,71
5	7,40	1,40	0,96	0,74	0,66
6	8,14	1,22	0,90	0,70	0,59

Analiza miar dopasowania wskazuje jednak, że w każdej fali badań modelem lepiej odzwierciedlającym strukturę danych jest model trójczynnikiowy niż dwu- czy jedno-czynnikowy. W każdym z badanych okresów w jego wypadku odnotowano najniższe wartości RMSEA oraz najwyższe wartości współczynników CFI oraz TLI (Tabela 3).

Istotne, że w wypadku rozwiązania trójczynnikiowego układ zmiennych „ładujących” poszczególne czynniki jest podobny w każdym z badanych okresów (Tabela 4).

Tabela 3

Miary dopasowania eksploracyjnych modeli czynnikowych w trzech turach badań (klasy 3, 5 oraz 6)

Klasa	Liczba czynników	χ^2 (df)	RMSEA	RMSEA 90% CI	CFI	TLI
3	1	2 475,09 (90)	0,078	[0,075–0,080]	0,634	0,573
	2	1 168,56 (76)	0,057	[0,054–0,060]	0,832	0,768
	3	395,91 (63)	0,035	[0,031–0,038]	0,949	0,915
5	1	3 767,68 (90)	0,096	[0,094–0,099]	0,902	0,886
	2	1 902,74 (76)	0,074	[0,071–0,077]	0,952	0,933
	3	759,68 (63)	0,050	[0,047–0,053]	0,982	0,969
6	1	3 858,13 (90)	0,098	[0,095–0,100]	0,931	0,920
	2	1 845,85 (76)	0,073	[0,070–0,076]	0,968	0,955
	3	779,42 (63)	0,051	[0,048–0,054]	0,987	0,978

Wykorzystano rotację Geomin, estymator WLSMV oraz macierz korelacji polichorycznej.

Tabela 4

Wartości ładunków czynnikowych eksploracyjnej analizy czynnikowej

Nr poz.	Klasa 3			Klasa 5			Klasa 6		
	1	2	3	1	2	3	1	2	3
2	0,13*	0,64*	0,05	0,00	0,72*	0,13*	0,02	0,65*	0,21*
5	-0,01	0,28*	0,54*	-0,02	0,30*	0,58*	-0,02	0,34*	0,57*
8	-0,01	0,30*	0,43*	-0,01	0,28*	0,52*	0,04	0,30*	0,47*
11	0,55*	0,14*	-0,07	0,63*	0,11*	-0,06	0,64*	0,11*	-0,03
14	0,02	0,87*	-0,01	0,00	0,91*	0,00	-0,03	0,93*	0,03
17	0,47*	0,03	0,21*	0,52*	-0,00	0,17	0,59*	-0,01	0,11*
20	-0,02	0,01	0,81*	0,06*	0,01	0,76*	0,05	0,00	0,80*
23	0,70*	-0,06	0,08	0,60*	-0,01	0,22*	0,65*	-0,01	0,16*
26	0,05	-0,03	0,70*	0,01	-0,05	0,79*	0,01	0,00	0,77*
29	0,52*	0,25*	0,01	0,42*	0,23*	0,27*	0,48*	0,27*	0,16*
32	0,02	0,24*	0,64*	0,04	0,18*	0,72*	0,00	0,25*	0,68*
35	-0,04	0,64*	0,07*	0,03	0,73*	0,06	0,01	0,86*	0,01
38	0,69*	-0,03	0,06	0,72*	-0,06	0,07	0,79*	-0,07	0,03
41	0,54*	0,05	0,01	0,54*	0,05	0,13	0,56*	0,09*	0,12*
44	0,46*	0,35*	-0,04	0,42*	0,35*	-0,02	0,46*	0,43*	-0,13*

Odpowiedzi zostały przekodowane w ten sposób, aby wyższy wynik oznaczał lepszą ocenę relacji rówieśniczych. Pogrubione pola oznaczają pozycje z najwyższymi ładunkami w danym czynniku.

* $p < 0,05$.

Ważne także, że rozwiązanie trójczynnikowe poddaje się sensownej interpretacji. W każdym z okresów czynnik pierwszy tworzą zmienne negatywne (twierdzenia o numerach: 11, 17, 23, 29, 38, 41 i 44). Czynnik drugi składa się z trzech twierdzeń pozytywnych odnoszących się do chęci utrzymywania relacji z rówieśnikami z klasy poza środowiskiem szkolnym (pozycje o numerach: 2, 14, 35). Czynnik trzeci reprezentują pozostałe twierdzenia pozytywne (o numerach: 5, 8, 20, 26, 32). Wstępne analizy eksploracyjne wskazują więc, że zarówno w klasie 3, jak 5 czy 6 trójczynnikowa struktura skali na pierwszy rzut oka wydaje się podobna.

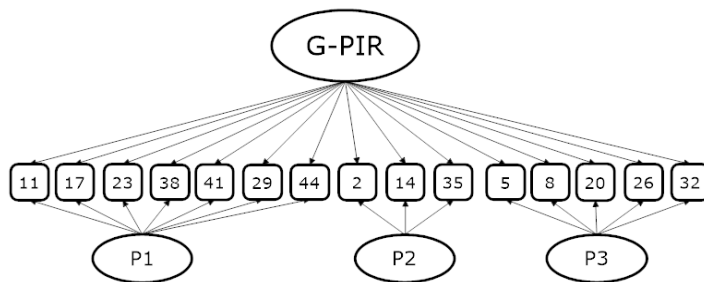
Konfirmacyjny model podwójnego czynnika

Na podstawie rozwiązania eksploracyjnego określono konfirmacyjny model podwójnego czynnika z jednym czynnikiem głównym oraz trzema pobocznymi (Rysunek 1). Miary dopasowania wskazują

na dobre dopasowanie modelu do danych z każdego okresu (Tabela 5)⁵.

Wszystkie pozycje w każdym okresie mocniej (por. Tabela 6) ładują czynnik główny niż czynniki specyficzne (z wyjątkiem pozycji 38 w klasie 5). Ładunki czynnika głównego są nie tylko statystycznie różne od zera, lecz także wyraźnie przekraczają wartość 0,3. Średnia wielkość ładunków czynnika głównego w klasie 3 wyniosła 0,60, w klasie 5: 0,66, a w klasie 6 – 0,70. Ładunki czynników pobocznych okazały się znacznie niższe. W klasie 3 średnia ładunków pierwszego czynnika pobocznego wyniosła 0,39, drugiego 0,47, trzeciego 0,26; w klasie 5, odpowiednio: 0,38, 0,50, 0,21; w klasie 6: 0,34, 0,48, 0,21.

⁵ Dodajmy, że testowaniu poddano również model konfirmacyjny zakładający występowanie trzech skorelowanych czynników, lecz bez czynnika głównego (klasyczny konfirmacyjny model trójczynnikowy). W każdym z okresów okazał się on nieco gorzej dopasowany do danych niż model podwójnego czynnika. Współczynnik RMSEA w klasie 3 wyniósł 0,041; w klasie 5: 0,057; w klasie 6 zaś 0,060. Odpowiednie dane dla współczynnika CFI wyniosły: 0,899; 0,967, 0,975, zaś dla TLI: 0,879; 0,960; 0,969.



Rysunek 1. Schemat testowanego konfirmacyjnego modelu podwójnego czynnika (BI-CFA).

Tabela 5

Miary dopasowania dla alternatywnych modeli pomiarowych kwestionariusza PIR

Klasa	Test χ^2 (df)	CFI	TLI	RMSEA	RMSEA CI 90%
3	655,20 (75)	0,911	0,901	0,040	[0,039–0,043]
5	962,29 (75)	0,976	0,967	0,052	[0,049–0,055]
6	1 141,23 (75)	0,981	0,973	0,057	[0,054–0,060]

Tabela 6

Wielkości standaryzowanych ładunków czynnikowych modelu podwójnego czynnika

Nr poz.	Klasa 3				Klasa 5				Klasa 6			
	G	P1	P2	P3	G	P1	P2	P3	G	P1	P2	P3
2	0,64		0,35		0,70		0,41		0,74		0,32	
5	0,64			0,18	0,67			0,08	0,73			0,08
8	0,70			0,10	0,78			0,04	0,80			0,02
11	0,64	0,38			0,72	0,45			0,73	0,40		
14	0,64		0,61		0,70		0,66		0,74		0,63	
17	0,69	0,32			0,74	0,36			0,76	0,35		
20	0,50			0,40	0,61			0,31	0,66			0,38
23	0,62	0,51			0,66	0,43			0,70	0,39		
26	0,58			0,37	0,74			0,43	0,77			0,34
29	0,79	0,33			0,84	0,25			0,83	0,22		
32	0,50			0,24	0,63			0,21	0,70			0,22
35	0,56		0,45		0,51		0,44		0,55		0,48	
38	0,51	0,50			0,50	0,55			0,59	0,53		
41	0,42	0,38			0,54	0,37			0,64	0,31		
44	0,57	0,28			0,59	0,22			0,63	0,18		

G: czynnik główny; P1, P2, P3: czynniki poboczne. Wszystkie ładunki istotne na poziomie $p < 0,05$.

Przedstawione dane pozwalają na wyciągnięcie dwóch (wstępnych) wniosków. Po pierwsze, w każdym okresie narzędzie nadaje się do pomiaru ogólnego postrzegania integracji z rówieśnikami w klasie. Wprawdzie nie cechuje go ściśła jednowymiarowość – o czym świadczy choćby słabe dopasowanie do danych – ale nie jest ściśle wielowymiarowy. To, że czynnik główny jest zdecydowanie lepiej określony niż czynniki poboczne, wskazuje na fakt, że strukturę narzędzia charakteryzuje jednowymiarowość typu substancjalnego. Po drugie, rosnące wraz z czasem wartości ładunków czynnikowych czynnika głównego mogą wskazywać na postępujący wzrost rzetelności pomiaru między klasami 3 a 6.

Rzetelność czynnika głównego i czynników pobocznych

Jak wspomniano na wstępie, problem substancjalnej jednowymiarowości narzędzia można rozwiązać dzięki estymacji współczynników rzetelności specyficznych

dla modeli podwójnego czynnika, tj. różnych wersji współczynnika omega oraz współczynnika ECV⁶. Analiza współczynników omega (Tabela 7) wskazuje, że gdy

⁶ Wszystkie obliczenia zostały przeprowadzone na standaryzowanych ładunkach czynnikowych (por. Reise, Bonifay i in., 2013).

Tabela 7

Miary rzetelności modelu podwójnego czynnika w trzech falach badania

Klasa	Miara	G	P1	P2	P3
3	ω_n/ECV	0,81/0,82	–	–	–
	ω_s	–	0,25	0,31	0,14
5	ω_n/ECV	0,85/0,84	–	–	–
	ω_s	–	0,22	0,36	0,09
6	ω_n/ECV	0,88/0,87	–	–	–
	ω_s	–	0,17	0,31	0,07

G: czynnik główny; P1, P2, P3: czynniki poboczne.

wyłączymy wariancję wszystkich pozostałych czynników, tj. innych czynników specyficznych i czynnika głównego, to rzetelność specyficzna (ω_s) dla każdego czynnika pobocznego będzie stosunkowo niska (między 0,07 a 0,36) i nie przekroczy granicznej wartości 0,5. Jednocześnie omega przyjmie wysokie wartości dla czynnika głównego – w każdym z okresów $\omega_h > 0,8$. Oznacza to, że czynniki poboczne charakteryzują się niskim poziomem swoistej rzetelności. Poszczególne pozycje narzędzia przenoszą więc informację raczej o jednym konstrukcie niż o trzech. To, co łączy zbiór wszystkich wskaźników, jest znacznie mocniejsze od tego, co łączy ich podzbiory. Optymalnym rozwiązaniem jest więc uznanie kwestionariusza PIR za substancjalnie jednowymiarowy. Potwierdzają to wartości współczynnika ECV, przekraczające graniczną wartość 0,6 w każdym z okresów badania.

Przedstawione dane przynoszą jeszcze jedną interesującą informację. Rzetelność pomiaru czynnika głównego rośnie stopniowo w kolejnych falach badania. Wzrost ten uwidacznia się w obu analizowanych współczynnikach: ω_h i ECV. Dzieci w młodszym wieku (w klasie 3) mają większe problemy z rzetelnym wypełnieniem kwestionariusza PIR niż w klasie 5 czy 6. Do tego zagadnienia powrócimy w dalszej części tekstu.

Podłużna niezmienność pomiarowa

Zgodnie z planem analizę podłużnej niezmienności rozpoczęto od testu

niezmienności typu konfiguracyjnego, w której zarówno wszystkie ładunki czynnikowe, jak i progi mogą się różnić między trzema falami badania (model M1). Dane zawarte w Tabeli 8 wskazują, że ten model dobrze oddaje relacje występujące między zmiennymi (RMSEA < 0,06; CFI > 0,90; TLI > 0,90). Wyjątkiem jest test chi kwadrat ($p < 0,05$), lecz wskaźnika tego – jak już powiedziano – nie bierzemy pod uwagę ze względu na jego wrażliwość na wielkość próby.

W następnym kroku przetestowano niezmienność typu metrycznego (M2), zakładającą, że w trzech okresach ładunki czynnikowe nie różnią się od siebie. Estymowany model (Tabela 8) nie okazał się znacząco gorzej dopasowany do danych niż model konfiguracyjny (Δ RMSEA < 0,007; Δ CFI > -0,002), co nie pozwala na odrzucenie hipotezy. Również model skalarny, w którym oprócz ładunków czynnikowych założono międzyokresową równość progów, był dopasowany nie gorzej niż model testowany wcześniej (Tabela 9). Nie ma więc podstaw do odrzucenia hipotezy o zbliżonej we wszystkich trzech okresach „trudności” pozycji tworzących skalę PIR.

W analizach możemy pójść krok dalej i nałożyć na model dodatkowe ograniczenie: że reszty we wszystkich trzech falach nie różnią się między sobą. Model taki odpowiada niezmienności nazywanej ścisłą (M4). Niestety, okazuje się on gorzej dopasowany do danych niż skalarny (M3). Różnica między wielkością współczynnika

Tabela 8

Miary dopasowania modeli przy różnych poziomach podłużnej niezmienności pomiarowej^(a)

Model	χ^2 (df)	$\Delta\chi^2$ (Δ df)	RMSEA	RMSEA 90% CI	Δ RMSEA	CFI	Δ CFI	TLI
M1	2 008,74 (849)	–	0,018	[0,017–0,019]	–	0,985	–	0,982
M2	2 081,16 (899)	174,60 (50)	0,017	[0,016–0,018]	-0,001	0,984	-0,001	0,983
M3	2 318,05 (953)	526,31 (54)	0,018	[0,017–0,019]	0,001	0,982	-0,002	0,981
M4	2 799,30 (983)	890,98 (84)	0,021	[0,020–0,021]	0,003	0,976	-0,006	0,976

^(a) $\Delta\chi^2$ i Δ df estymowano procedurą DIFFTEST z pakietu Mplus.

Tabela 9

Wartości niestandardyzowanych i standaryzowanych ładunków czynnikowych modelu podwójnego czynnika (skalarna podłużna niezmiennosc pomiarowa)

Nr poz.	Czynnik główny				Podczynniki			
	Niestandaryzowane	Standaryzowane			Niestandaryzowane	Standaryzowane		
		3	5	6		3	5	6
2	0,75	0,62	0,69	0,73	1	0,38	0,38	0,36
5	1	0,77	0,78	0,80	1	0,15	0,13	0,13
8	0,72	0,66	0,71	0,74	0,40	0,07	0,07	0,07
11	0,40	0,39	0,47	0,55	1	0,43	0,45	0,44
14	1,14	0,63	0,70	0,73	2,59	0,64	0,64	0,60
17	0,53	0,50	0,52	0,55	0,97	0,40	0,36	0,31
20	0,85	0,68	0,74	0,75	2,20	0,34	0,32	0,32
23	0,59	0,52	0,60	0,67	1,21	0,46	0,47	0,43
26	0,77	0,60	0,65	0,67	3,27	0,49	0,47	0,47
29	0,76	0,66	0,75	0,78	0,67	0,25	0,25	0,22
32	1,04	0,76	0,82	0,83	1,62	0,23	0,22	0,21
35	0,55	0,50	0,63	0,71	1,03	0,43	0,48	0,49
38	0,59	0,48	0,51	0,58	1,57	0,56	0,52	0,49
41	0,46	0,46	0,56	0,63	0,76	0,33	0,35	0,32
44	0,60	0,58	0,59	0,62	0,58	0,24	0,21	0,18

Ładunki niestandardyzowane są jednakowe w klasach 3, 5 i 6.

CFI dla modelu skalarnego i modelu ściślego ($\Delta CFI = -0,006$) jest wyraźnie niższa od granicznej wartości $-0,002$. Oznacza to, że reszt związanych z twierdzeniami nie można uznać za niezmiennie w czasie – a zatem, że poziom rzetelności pomiaru jest różny w trzech okresach.

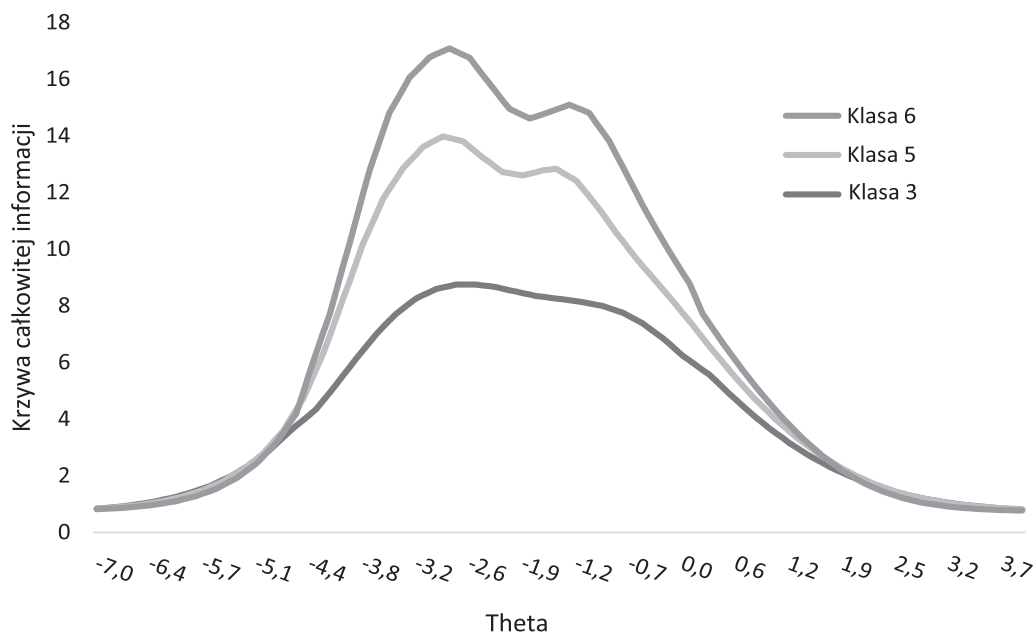
Poziom rzetelności narzędzia można określić za pomocą krzywej całkowitej informacji (Mellenbergh, 1994). W klasie 6 maksymalna wartość krzywej wyniosła 17,10, w klasie 5 mniej: 13,98, a w klasie 3 tylko 8,76 (Rysunek 2). Rzetelność skali PIR – zgodnie z przewidywaniami – wyraźnie wzrasta wraz z wiekiem uczniów.

Stabilność i zmiana

Przybliżona równość ładunków czynnikowych (niezmiennosc na poziomie metrycznym) pozwala na sprawdzenie poziomu stabilności pomiaru, czyli stopnia, w jakim

wynik na skali latentnej w jednym okresie wiąże się z takim wynikiem w następnych okresach. Niestandardyzowany współczynnik autoregresji między czynnikiem głównym pochodzącym z pomiaru pod koniec klasy 3 a analogiczną zmienną z początku klasy 5 wyniósł 0,47 ($se = 0,04$; $p < 0,01$), zaś między klasą 5 a klasą 6: 0,61 ($se = 0,02$; $p < 0,01$). Różnica wielkości obu współczynników – estymowana testem t dla zmiennych latentnych (opcja *Model Constraint* w Mplus) – okazała się statystycznie istotna ($\Delta = 0,14$; $se = 0,05$; $p < 0,01$). Łatwiej zatem przewidzieć wynik badanych na skali PIR w klasie 6 na podstawie wyników PIR z klasy 5 niż w klasie 5 na podstawie wyników z klasy 3. Stabilność poczucia integracji rówieśniczej rośnie z wiekiem badanych.

Brak podstaw do odrzucenia hipotezy o przybliżonej równości ładunków czynnikowych oraz progów (niezmiennosc na



Rysunek 2. Krzywa całkowitej informacji czynnika głównego w trzech turach pomiaru.

poziomie skalarnym) umożliwia natomiast zweryfikowanie hipotezy o braku statystycznie istotnych różnic pomiędzy średnim poziomem cechy latentnej w kolejnych falach badania. Przyjęty sposób estymacji modelu skalarnego zakłada, że w okresie odniesienia (w naszym wypadku jest nim klasa 3) wartość średniej jest równa zero, zaś w kolejnych falach średnia zmiennych latentnych jest estymowana jako parametr wolny (może przyjmować dowolne wartości). Odnotowanie poziomu istotności średniej konstruktum latentnego niższego niż graniczne 0,05 oznacza, że wartość tego parametru w znaczący sposób odchyła się od zera, a więc różni od wyniku uzyskanego w klasie 3.

Zastosowany do danych model zakładający podłużną równość ładunków czynnikowych i progów wykazuje spadek postrzeganej integracji z rówieśnikami między klasą 3 a klasą 5 ($M_{KL5} = -0,30$; $se = 0,05$; $p < 0,01$, przy $M_{KL3} = 0$) oraz między klasą 3 a kla-

są 6 ($M_{KL6} = -0,46$; $se = 0,05$; $p < 0,01$, przy $M_{KL3} = 0$). Dzieci uczące się w klasie 3 cechuje bardziej optymistyczna ocena integracji niż w okresach późniejszych. Test t dla zmiennych latentnych pozwala odrzucić hipotezę o ich równości między klasami 5 a 6 ($\Delta M_{KL5-KL6} = -0,16$; $se = 0,029$; $p < 0,01$). Między obu późniejszymi falami badania następuje więc statystycznie istotne pogorszenie postrzegania integracji. Ogólnie jednak dynamika spadku postrzegania integracji nie jest taka sama w całym okresie związanym z II etapem nauki w szkole podstawowej. Głębsze, bardziej negatywne zmiany zachodzą na początku tego okresu, mniejsze – pod koniec.

Interesujący i pouczający jest fakt, że gdybyśmy przeprowadzili analizy na uśrednionych wynikach obserwowalnych tych samych respondentów – np. za pomocą analizy wariancji dla pomiarów powtarzanych – to doszlibyśmy do innych wniosków.

Okazałoby się, że między klasą 3 a klasą 6 poczucie integracji z rówieśnikami pogarsza się ($\Delta M_{KL5-KL3} = -0,24$; $se = 0,01$; $p < 0,01$), lecz między klasami 5 a 6 ulega poprawie ($\Delta M_{KL6-KL5} = 0,08$; $se = 0,01$; $p < 0,01$). Dla modelu liniowego η^2 wyniosłaby 0,08, dla kwadratowego zaś byłaby dwa razy większa (0,16). Obserwowany trend nie miałby więc charakteru liniowego, lecz krzywoliniowy (najpierw spadek, potem wzrost). Dodajmy, że satysfakcja z relacji rówieśniczych nie wróciłaby jednak do poziomu odnotowanego w klasie 3 ($\Delta M_{KL6-KL3} = -0,17$; $se = 0,01$; $p < 0,01$).

Ciekawe konsekwencje przyniosłoby także zbagatelizowanie informacji o braku równości błędów wskaźników pomiędzy różnymi fazami badania. Gdybyśmy przyjęli założenie o podłużnej niezmienności ścisłej, to różnice w poziomach średniej latentnej byłyby znacznie mniejsze niż przy uwzględnieniu jedynie niezmienności typu skalarnego ($M_{KL5} = -0,06$; $se = 0,04$; $p = 0,13$, przy $M_{KL3} = 0$; natomiast $M_{KL6} = -0,11$; $se = 0,05$; $p = 0,02$, przy $M_{KL3} = 0$). Dodajmy, że w takiej sytuacji nie odnotowano by również statystycznie istotnych różnic między średnią latentną w klasach 5 i 6 ($\Delta M_{KL6-KL5} = -0,05$; $se = 0,04$; $p = 0,20$). Zatem zignorowanie różnego poziomu rzetelności cechującego poszczególne okresy badania prowadziłoby do wniosku o stabilności poczucia integracji z rówieśnikami między klasami 3 a 5 oraz o niewielkim spadku zachodzącym między tymi klasami (najpierw stabilność, potem spadek).

Wnioski i dyskusja wyników

Na podstawie wyników trzech fal ogólnopolskiego badania podłużnego uczniów szkół podstawowych zrealizowanych w ramach projektu SUEK wykazano, że Kwestionariusz postrzeganej integracji rówieśniczej jest stosunkowo krótkim, a jednocześnie rzetelnym i podłużnie

niezmiennym narzędziem do pomiaru poczucia integracji z rówieśnikami klasowymi, i że powodzeniem można go stosować już do uczniów klasy 3. Wskaźniki stabilności pomiaru (współczynniki regresji pomiędzy kolejnymi falami badania) świadczą o relatywnie wysokiej rzetelności pomiaru dokonywanego za jego pomocą. Zbliżone współczynniki (rok do roku) uzyskali np. Lisa Jobe-Shields, Robert Cohen i Gilbert R. Parra (2011) w badaniu uczniów w podobnym wieku (9, 10 i 11 lat) – wahały się one między 0,36 a 0,56. Również wyższy poziom stabilności między klasami 5 a 6 niż 3 a 5 został potwierdzony w innych badaniach. Przykładowo analizy przeprowadzone przez Davida A. Cole'a, Farrę M. Jacques i Tracy L. Maschman (2001) wskazują, że o ile półroczna stabilność wśród uczniów klas 3 wyniosła 0,49, o tyle w klasie 6 już 0,64. Także uchwycenie przez skalę PIR stopniowego spadku poczucia integracji z rówieśnikami między klasami 3 a 6 pozostaje w zgodzie z dotychczasowym dorobkiem badawczym i dobrze świadczy o właściwościach psychometrycznych narzędzia. Wyniki badań wskazują, że o ile poczucie integracji z rówieśnikami w okresie przedszkolnym jest stosunkowo wysokie i nie ulega większym czasowym wahaniom, o tyle ulega ono wyraźnemu obniżeniu między pierwszym a drugim rokiem nauki w szkole powszechnej (Ladd i Burgess, 1999), i że ten negatywny trend utrzymuje się w kolejnych latach nauki (Galanaki i Kalantzi-Azizi, 1999; Quay, 1992).

Zasadniczy cel, jaki postawiono w ramach niniejszego artykułu, wykraczał jednak poza uchwycenie struktury czynnikowej kwestionariusza PIR oraz poziomu jego rzetelności. Było nim ukazanie użyteczności zastosowania modeli podwójnego czynnika do analizy struktury latentnej do badania właściwości skali psychologicznej takiej jak PIR.

Spośród testowanych modeli zdecydowanie najgorzej dopasowany okazał się model

jednoczynnikowy. Gdybyśmy więc ograniczyli się do porównania jedynie modelu jedno- i trójczynnиковego kwestionariusza PIR, to miary dopasowania zmusiłyby nas do odrzucenia hipotezy o jego jednowymiarowości i przyjęcia tezy o wielowymiarowości. W konsekwencji poczucie integracji z rówieśnikami byłibyśmy skłonni rozbić na trzy różne aspekty, ignorując jego ogólny poziom. Użycie modelu podwójnego czynnika wskazuje, że takie rozbieżności – ignorujące istnienie czynnika głównego, leżącego u podstawy odpowiedzi na wszystkie pozycje skali (Reise, Bonifay i in., 2013) – byłoby przedwczesne. Co ważne, zastosowane miary rzetelności wykazały, że w porównaniu z czynnikiem głównym trzy czynniki poboczne cechuje niewielka wariancja swoista. Należy je więc potraktować raczej jako „szum” informacyjny, wymiary „uciążliwe”, utrudniające pomiar interesującej nas cechy – poczucia integracji z klasowymi rówieśnikami (Reise i in., 2010).

W kwestionariuszu PIR ich wystąpienie ma dwa nakładające się na siebie źródła. Po pierwsze, wynika z nieintencjonalnego pomiaru specyficznej zmiennej latentnej (reprezentowanej przez czynnik drugi), wiążącej się z pozytywnymi opiniami na temat spędzania z rówieśnikami z klasy czasu wolnego po zajęciach szkolnych. Przypomnijmy w tym miejscu, że analizowane wyniki pochodzą od uczniów szkół podstawowych mających ograniczoną możliwość decydowania, gdzie i z kim spędzają czas wolny. Mogą więc oni lubić swoich rówieśników z klasy, ale mieszkać od nich na tyle daleko, że akceptacja twierdzeń tworzących czynnik (przede wszystkim 14 i 35) staje się niemożliwa. Innymi słowy, pozycje te – w porównaniu z twierdzeniami tworzącymi czynnik trzeci (także odnoszący się do pozytywnych ocen dotyczących integracji, lecz bez odniesienia do pozaszkolnego czasu wolnego) – zawierają dodatkowy kontekst znaczeniowy. Ich akceptacja oznacza nie

tylko, że uczeń lubi swoich rówieśników, lecz także, że może spędzać z nimi czas wolny i chętnie to czyni.

Po drugie, pojawienie się czynników ubocznych wynika ze zrównoważonego charakteru analizowanej skali. Do jej konstrukcji wykorzystano zarówno twierdzenia pozytywne, jak i negatywne. Badacze wykorzystujący skale zrównoważone często nie radzą sobie dostatecznie dobrze z dokładnym treściowo i logicznie odwróceniem pozycji pozytywnych. Brak tej dokładności skutkuje pojawieniem się nadprogramowych czynników o charakterze metodologicznym (Billiet i McClendon, 2000; Schweizer i Schreiner, 2010). Zaznaczmy, że taka metodologiczna niejednorodność jest w wypadku skal zrównoważonych – w tym narzędzi przeznaczonych do badania poczucia integracji z rówieśnikami wśród młodzieży szkolnej (Ebesutani i in., 2012a; 2012b; Humenny i Grygiel, 2015a; Maes, Van den Noortgate, Vanhalst, Beyers i Goossens, 2015) – raczej normą niż wyjątkiem.

Przedstawione wyniki nie tylko pokazują, jak ważne jest zastosowanie odpowiedniego modelu do analizy struktury latentnej, lecz także, że ciągle częsta (Borsboom, 2006) praktyka analizowania wielopozycyjnych kwestionariuszy przez uśrednienie lub zsumowanie informacji niesionych przez poszczególne pozycje (DiStefano, Zhu i Míndrilá, 2009), a następnie wykorzystanie takich metod jak test *t*, ANOVA czy MANOVA, może prowadzić do błędnych wniosków (Cole, Maxwell, Arvey i Salas, 1993; Yuan i Bentler, 2006). Proste porównanie różnic między średnimi PIR w trzech badanych okresach estymowanych dwoma sposobami – z jednej strony za pomocą konfirmacyjnego modelu podwójnego czynnika (z przetestowaną niezmiennością skalarną), z drugiej zaś przez uśrednienie wartości wskaźników – doprowadziło do odmiennych wniosków. Analizy czynnikowe wskazały na stopniowy spadek poczucia

integracji między klasami 3 a 6, natomiast analizy oparte na wartościach średnich – na jego początkowy spadek, następnie zaś na wzrost. Czym zasadniczo różnią się obie te metody? Pierwsze umożliwiają uwzględnienie w odkrywaniu interesującej nas struktury latentnej (a) różnej wagi wskaźników, (b) istnienia oprócz czynnika głównego także czynników ubocznych oraz (c) błędów losowych związanych pojedynczymi wskaźnikami, drugie zaś nie.

Przeprowadzona analiza ukazuje znaczenie weryfikacji założeń związanych z niezmiennością pomiarową, w naszym wypadku – z podłużnym charakterem danych. Jej wyniki dowodzą, że zignorowanie informacji o braku niezmienności ścisłej (przypomnijmy: zakładającej niezmiennosc nie tylko ładunków czy progów, lecz także błędów) nieuchronnie doprowadziłoby do błędnego oszacowania wielkości średniej cechy latentnej w poszczególnych falach badania i w konsekwencji do zaniżenia występujących między nimi różnic oraz stwierdzenia braku zmian w natężeniu poczucia integracji z rówieśnikami między klasami 3 a 5 i niewielkiego spadku między klasami 3 a 6.

Należy zauważyć, że niezmiennosc pomiarową należy testować nie tylko w wypadku danych podłużnych, lecz także przekrojowych (jako niezmiennosc międzygrupową). W gruncie rzeczy przedstawiony sposób analizy danych (począwszy od analiz eksploracyjnych po różne poziomy niezmiennosci modeli confirmacyjnych) ma charakter uniwersalny – może być stosowany do każdego typu wielopozycyjnych narzędzi zakładających występowanie cechy latentnej (psychologicznych, edukacyjnych, socjologicznych) i do danych pochodzących zarówno z badań podłużnych, jak i przekrojowych.

Na zakończenie dodajmy, że wykorzystane w zasadniczej części artykułu modele opierały się na rozwiązaniu

confirmacyjnym, które nie uwzględnia możliwości występowania ładunków krzyżowych. Jest to wyraźne ograniczenie przeprowadzonych analiz. Badania symulacyjne (Reise, Moore i Maydeu-Olivares, 2011) wskazały, że w wypadku błędnej specyfikacji confirmacyjnego modelu podwójnego czynnika, polegającej na nieuwzględnieniu istotnych i znaczących ($\lambda > 0,3$) ładunków jednej zmiennej na więcej niż jednym czynniku, może prowadzić do problemów z estymacją prawdziwych wartości ładunków na czynniku głównym i czynnikach pobocznych. Ryzyko takie jest jednym z powodów, dla których w ostatnich latach rozwija się modele łączące elementy analiz eksploracyjnych (EFA) z confirmacyjnymi (CFA; Marsh, Morin, Parker i Kaur, 2014; Morin, Marsh i Nagengast, 2013), w tym modele z rotacją celową (*target rotation*; Browne, 2001).

Niestety, zastosowanie rotacji celowej wymusza symultaniczne powiązanie (lub niepowiązanie) wszystkich estymowanych w jej ramach czynników z innymi zmiennymi w modelu (Marsh i in., 2014; Morin i in., 2013) i dlatego jej zastosowanie do podłużnych modeli typu podwójnego czynnika (zakładających wzajemną ortogonalność czynników pobocznych i czynnika głównego) nie jest w chwili pisania tego artykułu możliwe. Nie ma możliwości estymowania modelu, w którym odpowiadające sobie czynniki z różnych fal badania byłyby skorelowane, przy jednoczesnym zachowaniu wzajemnej ortogonalności pozostałych czynników, zarówno w ramach każdej z fal, jak i między falami. Innymi słowy: nie istnieje możliwość estymacji modelu, w którym między poszczególnymi falami badania skorelowane byłyby ze sobą czynniki główne, pozostając jednocześnie ortogonalnymi w stosunku do wszystkich czynników pobocznych. To ograniczenie dowodzi, że projektu naukowego pod nazwą „modele czynnikowe” nie można uznać zakończony i że czeka nas z jego strony jeszcze wiele niespodzianek.

Literatura

- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246.
- Billiet, J. B. i McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(4), 608–628.
- Bonifay, W. E., Reise, S. P., Scheines, R. i Meijer, R. R. (2015). When are multidimensional data unidimensional enough for structural equation modeling? An evaluation of the DETECT multidimensionality index. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(4), 504–516.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71(3), 425–440.
- Briggs, S. R. i Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54(1), 106–148.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36(1), 111–150.
- Byrne, B. M. (2012). *Structural equation modeling with Mplus: basic concepts, applications, and programming*. New York: Routledge Academic.
- Byrne, B. M., Shavelson, R. J. i Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456–466.
- Chen, F. F., West, S. i Sousa, K. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41(2), 189–225.
- Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P. i Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: comparison of the bifactor model to other approaches. *Journal of Personality*, 80(1), 219–251.
- Cole, D. A., Jacquez, F. M. i Maschman, T. L. (2001). Social origins of depressive cognitions: a longitudinal study of self-perceived competence in children. *Cognitive Therapy and Research*, 25(4), 377–395.
- Cole, D. A., Maxwell, S. E., Arvey, R. i Salas, E. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. *Psychological Bulletin*, 114(1), 174–184.
- DiStefano, C., Zhu, M. i Mindrila, D. (2009). Understanding and using factor scores: considerations for the applied researcher. *Practical Assessment, Research i Evaluation*, 20(14), 1–11.
- Dolata, R. (red.). (2014). *Czy szkoła ma znaczenie? Analiza zróżnicowania efektywności nauczania na pierwszym etapie edukacyjnym* (t. 1). Warszawa: Instytut Badań Edukacyjnych.
- Dolata, R., Grygiel, P., Jankowska, D. M., Jarnutowska, E., Jasińska-Maciążek, A., Karwowski, M., Modzelewski, M. i Pisarek, J. (2015). *Szkolne pytania. Wyniki badań nad efektywnością nauczania w klasach IV–VI*. Warszawa: Instytut Badań Edukacyjnych.
- Ebesutani, C., Drescher, C. F., Reise, S. P., Heiden, L., Hight, T. L., Damon, J. D. i Young, J. (2012a). The importance of modeling method effects: resolving the (uni)dimensionality of the loneliness questionnaire. *Journal of Personality Assessment*, 94(2), 186–195.
- Ebesutani, C., Drescher, C. F., Reise, S. P., Heiden, L., Hight, T. L., Damon, J. D. i Young, J. (2012b). The Loneliness Questionnaire–Short Version: an evaluation of reverse-worded and non-reverse-worded items via item response theory. *Journal of Personality Assessment*, 94(4), 427–437.
- Galanaki, E. P. i Kalantzi-Azizi, A. (1999). Loneliness and social dissatisfaction: its relation with children's self-efficacy for peer interaction. *Child Study Journal*. Pobrano z <http://www.highbeam.com/doc/1G1-57603783.html>
- Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., ... Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, 31(1), 4–19.
- Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: *g* as superordinate or breadth factor? *Psychology Science*, 50(1), 21–43.
- Grygiel, P., Humenny, G., Rębisz, S., Bajcar, E. i Świtaj, P. (2014). Peer rejection and perceived quality of relations with schoolmates among children with ADHD. *Journal of Attention Disorders*. doi: 10.1177/1087054714563791
- Grygiel, P., Humenny, G., Świtaj, P., Rębisz, S. i Anczewska, M. (2014). Between isolation and loneliness: social networks and perceived integration with peers of children diagnosed with ADHD in regular classrooms. W: N. Popov, C. Wolhuter, K. Skubic Ermenc, G. Hilton, J. Ogunleye i O. Chi-

- gisheva (red.), *Education's role in preparing globally competent citizens* (s. 303–310). Sofia: Bulgarian Comparative Education Society.
- Haerberlin, U., Moser, U., Bless, G. i Klaghofer, R. (1989). *Integration in die schulklasse: fragebogen zur erfassung von dimensionen der integration von schülern: FDI 4-6*. Bern–Stuttgart: Verlag Paul Haupt.
- Humenny, G. i Grygiel, P. (2015a). Poza ścisłą jedno- i wielowymiarowość. Struktura czynnikowa skali samotności de Jong Gierveld wśród dzieci. W: A. Pokropek (red.), *Modele cech ukrytych w badaniach edukacyjnych, psychologii i socjologii. Teoria i zastosowania* (s. 400–424). Warszawa: Instytut Badań Edukacyjnych.
- Humenny, G. i Grygiel, P. (2015b). Wielowymiarowa struktura latentna w perspektywie analizy czynnikowej. W: A. Pokropek (red.), *Modele cech ukrytych w badaniach edukacyjnych, psychologii i socjologii. Teoria i zastosowania* (s. 130–165). Warszawa: Instytut Badań Edukacyjnych.
- Jobe-Shields, L., Cohen, R. i Parra, G. R. (2011). Patterns of change in children's loneliness: trajectories from third through fifth grades. *Merrill-Palmer Quarterly*, 57(1), 25–47.
- Kail, R. V. (2010). *Human development: a life-span view* (wyd. 5). Australia–Belmont: Wadsworth Cengage Learning.
- Ladd, G. W. i Burgess, K. B. (1999). Charting the relationship trajectories of aggressive, withdrawn, and aggressive/withdrawn children during early grade school. *Child Development*, 70(4), 910–929.
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: practical and theoretical issues. *Multivariate Behavioral Research*, 32(1), 53–76.
- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: Guilford Press.
- Little, T. D., Preacher, K. J., Selig, J. P. i Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, 31(4), 357–365.
- Lubke, G. H., Dolan, C. V., Kelderman, H. i Mellenbergh, G. J. (2003). Weak measurement invariance with respect to unmeasured variables: an implication of strict factorial invariance. *British Journal of Mathematical and Statistical Psychology*, 56(2), 231–248.
- Maes, M., Van den Noortgate, W., Vanhalst, J., Beyers, W. i Goossens, L. (2015). The Children's Loneliness Scale: factor structure and construct validity in Belgian children. *Assessment*. doi: 10.1177/1073191115605177
- Malloy, T. E., Sugarman, D. B., Montvilo, R. K. i Ben-Zeev, T. (1995). Children's interpersonal perceptions: a social relations analysis of perceiver and target effects. *Journal of Personality and Social Psychology*, 68(3), 418–426.
- Malloy, T. E., Yarlas, A., Montvilo, R. K. i Sugarman, D. B. (1996). Agreement and accuracy in children's interpersonal perception: a social relations analysis. *Journal of Personality and Social Psychology*, 71(4), 692–702.
- Marsh, H. W., Hau, K.-T. i Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. W: A. Maydeu-Olivares, J. J. McArdle i R. P. McDonald (red.), *Contemporary psychometrics: a festschrift for Roderick P. McDonald* (s. 275–340). Mahwah: Erlbaum.
- Marsh, H. W., Morin, A. J. S., Parker, P. D. i Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10(1), 85–110.
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34(1), 100–117.
- Meade, A. W., Johnson, E. C. i Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, 93(3), 568–592.
- Mellenbergh, G. J. (1994). Generalized linear item response theory. *Psychological Bulletin*, 115(2), 300–307.
- Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, 29(2), 177–185.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Psychology Press.
- Morin, A. J. S., Marsh, H. W. i Nagengast, B. (2013). Exploratory structural equation modeling. W: G. R. Hancock i R. O. Mueller (red.), *Structural equation modeling: a second course* (wyd. 2, s. 395–436). Charlotte: Information Age Publishing.
- Muthén, L. K. i Muthén, B. O. (2012). *Mplus user's guide* (wyd. 7). Los Angeles: Muthén&Muthén.
- Quay, L. C. (1992). Personal and family effects on loneliness. *Journal of Applied Developmental Psychology*, 13(1), 97–110.

- Quinn, H. O. (2014). *Bifactor models, explained common variance (ECV), and the usefulness of scores from unidimensional item response theory analyses*. Chapel Hill: University of North Carolina.
- Raykov, T. i Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York: Routledge.
- Reise, S. P., Bonifay, W. E. i Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95(2), 129–140.
- Reise, S. P., Moore, T. M. i Haviland, M. G. (2010). Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544–559.
- Reise, S. P., Moore, T. i Maydeu-Olivares, A. (2011). Target rotations and assessing the impact of model violations on the parameters of unidimensional item response theory models. *Educational and Psychological Measurement*, 71(4), 684–711.
- Reise, S. P., Morizot, J. i Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16(S1), 19–31.
- Reise, S. P., Scheines, R., Widaman, K. F. i Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: a bifactor perspective. *Educational and Psychological Measurement*, 73(1), 5–26.
- Reise, S. P., Waller, N. G. i Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, 12(3), 287–297.
- Reise, S. P., Widaman, K. F. i Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychological Bulletin*, 114(3), 552–566.
- Saris, W. E., Satorra, A. i Veld, W. M. van der (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, 16(4), 561–582.
- Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, 29(4), 304–321.
- Schweizer, K. i Schreiner, M. (2010). Avoiding the effect of item wording by means of bipolar instead of unipolar items: an application to social optimism. *European Journal of Personality*, 24(2), 137–150.
- Steiger, J. H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research*, 25(2), 173–180.
- Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52(4), 589–617.
- Stout, W. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55(2), 293–325.
- Szumski, G. (2010). *Wokół edukacji włączającej: efekty kształcenia uczniów z niepełnosprawnością intelektualną w stopniu lekkim w klasach specjalnych, integracyjnych i ogólnodostępnych*. Warszawa: Wydawnictwo Akademii Pedagogiki Specjalnej.
- Szumski, G. i Karwowski, M. (2012). Funkcjonowanie psychospołeczne uczniów z lekką niepełnosprawnością intelektualną na tle uczniów sprawnych intelektualnie – znaczenie formy kształcenia i płci. *Ruch Pedagogiczny*, LXXXIII(3), 33–55.
- Szumski, G. i Karwowski, M. (2014). Psychosocial functioning and school achievement of children with mild intellectual disability in Polish special, integrative, and mainstream schools: social functioning and school achievement. *Journal of Policy and Practice in Intellectual Disabilities*, 11(2), 99–108.
- Ten Berge, J. M. F. i Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, 69(4), 613–625.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: understanding concepts and applications* (wyd. 1). Washington: American Psychological Association.
- Tucker, L. R. i Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1–10.
- Vandenberg, R. J. i Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70.
- Widaman, K. F., Ferrer, E. i Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: measuring the same construct across time. *Child Development Perspectives*, 4(1), 10–18.
- Yu, C. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Los Angeles: University of California.
- Yuan, K.-H. i Bentler, P. M. (2006). Mean comparison: manifest variable versus latent variable. *Psychometrika*, 71(1), 139–159.

Artykuł powstał na podstawie danych pochodzących z *Badania szkolnych uwarunkowań efektywności kształcenia*, przeprowadzonego w Instytucie Badań Edukacyjnych w ramach projektu systemowego „Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego” współfinansowanego ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Kapitał Ludzki 2007–2013. Priorytet III: Wysoka jakość systemu oświaty.

Tekst złożony 11 grudnia 2015 r., recenzowany 30 marca 2016 r., przyjęty do druku 21 kwietnia 2016 r.

**Testing the longitudinal invariance of the bifactor model
using the example of The Perceived Peer Integration Questionnaire**

In longitudinal data, changes in constructs over time can only be sensibly interpreted if the measured variables are assumed to be invariant across time. This article uses the empirical example of The Perceived Peer Integration Questionnaire (PPI) and three rounds of the nationwide study on *School conditions of education effectiveness* ($N = 4349$) to illustrate the use of structural equation modeling to systematically test the measurement invariance of the bifactor model across time. The results prove that the PPI questionnaire is a reliable tool; it is substantially one-dimensional, with a bifactor structure, a longitudinally invariant measurement: configural, metric and scalar, but not strict. We can therefore assume that even though the factor patterns, size of factor loadings and thresholds do not differ significantly in successive editions of the study, the level of reliability of the measurement cannot be considered invariant over time. A lower level of measurement reliability was recorded for grade 3 than for grades 5 or 6. The article also shows the consequences of ignoring the assumptions relating to longitudinal invariance on the results of the statistical analysis.

KEYWORDS: sociology, longitudinal invariance, bifactor model, peer relationships, reliability, developmental changes.