prof. dr hab. Barbara Ciżkowicz
Uniwersytet Kazimierza Wielkiego

Zastosowanie modelowania równań strukturalnych w badaniu związków przyczynowych na przykładzie danych PISA 2012

W poniższym opracowaniu przedstawiono zastosowanie modelowania równań strukturalnych w analizie związków przyczynowych między wybranymi uwarunkowaniami procesu kształcenia a umiejętnościami matematycznymi polskich 15-latków w badaniach PISA Głównym celem opracowania jest wskazanie zalet tej strategii analizy danych oraz zwrócenie uwagi na ograniczenia w zakresie formułowania wniosków dotyczących zależności przyczynowych.

Wstęp

W badaniach edukacyjnych pytanie o przyczyny zachodzenia określonych zjawisk jest pytaniem ważnym z punktu widzenia organizacji procesu kształcenia. Interesuje nas, co jest przyczyną niskich osiągnięć z matematyki tegorocznych maturzystów, wzrastającej agresji wśród młodzieży gimnazjalnej, nasilania się stanów depresyjnych wśród młodzieży itd. Znajomość przyczyn pozwala projektować proces kształcenia w taki sposób, by minimalizować niepożądane skutki. Eksperyment jest jedyną metodą pozwalającą na pełną kontrolę zmiennej niezależnej i pewne wnioskowanie o zależnościach przyczynowych Jednak w badaniach edukacyjnych realizacja schematu eksperymentalnego jest trudna i mało efektywna. Stąd interesujące wydają się analizy statystyczne umożliwiające testowanie modeli przyczynowo-skutkowych na podstawie danych nieeksperymentalnych. Reprezentantem tego rodzaju strategii analiz statystycznych jest modelowanie równań strukturalnych (StatSoft, 2011; Pearl, 2000).

Badania PISA

Program Międzynarodowej Oceny Umiejętności Uczniów (*Programme for International Student Assessment* PISA) jest programem sprawdzającym kluczowe kompetencje uczniów 15-letnich. Są to umiejętności przydatne na współczesnym rynku pracy, wymagającym elastyczności i zdolności samodzielnego nabywania nowych kwalifikacji. Międzynarodowe diagnozy tych umiejętności są szczególnie ważne ze względu na nasilającą się migrację ludności. Badania realizowane są w odstępach trzyletnich, począwszy od 2000 roku. Daje się odnotować wzrost liczby krajów uczestniczących. W roku 2000 w badaniach udział wzięło łącznie 40 krajów, z czego 25 było członkami OECD. W ostatnio przeprowadzanych badaniach, w roku 2012 uczestniczyło 69 krajów, w tym 35 członków OECD. PISA jest największym porównawczym badaniem umiejętności uczniów przeprowadzanym na reprezentatywnych próbach. Celem tego badania jest diagnoza poziomu umiejętności uczniów w zakresie matematyki, czytania oraz nauk przyrodniczych. Stosowana

metodologia umożliwia zarówno porównania podłużne, pozwalające ocenić zmiany umiejętności uczniów zachodzące w czasie, jak i poprzeczne, określające pozycję danego kraju na tle innych krajów uczestniczących.

W badaniach PISA w 2012 roku matematyka po raz drugi była przedmiotem wiodącym (poprzednio w 2003 roku). Coraz częściej znajomość matematyki postrzegana jest jako ważny element przygotowania młodych ludzi do życia w nowoczesnym społeczeństwie. Wzrastający odsetek problemów i sytuacji spotykanych w życiu codziennym, których rozwiązywanie wymaga pewnego poziomu umiejętności i rozumowania matematycznego, nasila zainteresowanie miejscem matematyki w procesie kształcenia (PISA, 2012, s. 27). Szczególnie w tym kontekście osiągnięcia polskich 15-latków w badaniach PISA w 2012 roku dostarczają wiele powodów do zadowolenia. Wynik testu umiejętności matematycznych przeciętnego polskiego gimnazjalisty wzrósł o 23 punkty w stosunku do roku 2009 (PISA, 2012).

PISA, oprócz diagnozy kluczowych kompetencji uczniów z matematyki, czytania i nauk przyrodniczych, diagnozuje szeroki zakres zmiennych związanych z uwarunkowaniami procesu kształcenia. Dane te pozyskiwane są za pomocą badań kwestionariuszowych przeprowadzanych wśród uczniów i ich rodziców oraz dyrektorów szkół. Dotyczą one grupy zmiennych związanych ze szkołą, z uczniem oraz z jego środowiskiem rodzinnym (PISA, 2012b).

W poniższych rozważaniach przedstawiono zastosowanie modelowania równań strukturalnych do testowania związków przyczynowych między wybranymi zmiennymi kontekstowymi a poziomem umiejętności matematycznych polskich 15-latków w badaniach PISA w 2012 roku. W grupie zmiennych kontekstowych uwzględniono *motywację do uczenia się matematyki*, *samoocenę zdolności matematycznych*, *lęk matematyczny* oraz *kapitał kulturowy ucznia*. Są to zmienne ukryte (latentne) będące teoretycznymi konstruktami, których pomiar możliwy jest tylko w sposób pośredni poprzez właściwie dobrane wskaźniki empiryczne. Wybór tych zmiennych spośród wielu innych zawartych w badaniach PISA i mających uzasadnienie teoretyczne podyktowany był zainteresowaniami badawczymi autorki.

Celem poniższego opracowania jest ocena stopnia dopasowania przyjętych modeli strukturalnych do danych empirycznych oraz wskazanie możliwości i ograniczeń wniosków dotyczących przyczynowości formułowanych na podstawie analizy danych nieeksperymentalnych. W artykule świadomie pominięto teoretyczne uzasadnienie przyjętych modeli.

Metoda

Pomiar

Umiejętności matematyczne są zmienną zależną w modelu. Poziom tej zmiennej latentnej jest badany za pomocą pięciu wskaźników empirycznych, które obejmują różne grupy umiejętności, począwszy od prostych, odtwórczych, wymagających jedynie zastosowania gotowego algorytmu po myślenie strategiczne oraz rozumowanie i argumentację (współczynnik α-Cronbacha

= 0,98). Poziom umiejętności uczniów, w tym umiejętności matematycznych, wyrażany jest na skali akademickiej (M = 500; SD = 100).

Zmienne objaśniające uwzględnione w modelu zostały zaczerpnięte z kwestionariusza ucznia (PISA, 2012c). Przy każdej zmiennej w nawiasach podano jej kod identyfikacji w zbiorze danych. Wszystkie zmienne zostały tak zakodowane, że im wyższa wartość punktowa, tym wyższy poziom zmiennej.

Kapitał kulturowy (ST28) – jego wskaźnikiem empirycznym jest wielkość domowego księgozbioru. Zmienna ta przyjmuje wartości od 1 do 6, dla których górne granice przedziałów są odpowiednio do**:** 10; 25; 100; 200; 500; więcej niż 500 książek.

Motywacja do uczenia się matematyki (ST29) jest skalą zbudowaną z ośmiu pozycji. W skali tej występują dwie podskale, po cztery pozycje każda. Jedna podskala dotyczy **motywacji wewnętrznej** (np. *nie mogę doczekać się lekcji matematyki*), druga – **zewnętrznej** (np. *będę uczył się wielu rzeczy z matematyki, które pomogą mi dostać pracę*). Badani określali, w jakim stopniu zgadzają się z danym stwierdzeniem, używając czterostopniowej skali szacunkowej. W efekcie zakres możliwych wartości każdej ze zmiennych mieści się w przedziale od 4 do 16 punktów. Rzetelność pomiarów szacowano metodą zgodności wewnętrznej. Współczynnik α-Cronbacha jest równy 0,88 dla motywacji wewnętrznej i 0,90 dla motywacji zewnętrznej.

Lęk matematyczny (ST42) – wskaźnikami empirycznymi poziomu lęku jest pięć z dziesięciu stwierdzeń (np. *jestem bardzo podenerwowany, rozwiązując problemy matematyczne*) dotyczących uczenia się matematyki. Badani określali zgodność z tymi stwierdzeniami na czterostopniowej skali szacunkowej. Rozstęp lęku jest równy 15 punktów (od 5 do 20). Pomiar lęku matematycznego jest wysoce rzetelny (α-Cronbacha = 0,87).

Samoocena zdolności matematycznych. Poziom samooceny jest sumą punktów uzyskanych za pozostałych pięć pozycji (np. *uczę się matematyki szybko)* zawartych w (ST42). Poziom samooceny zdolności matematycznych może przybierać wartości z przedziału od 5 do 20 punktów. Również w tym przypadku rzetelność pomiaru jest wysoka α-Cronbacha = 0,91).

Osoby badane

Dane do analizy zostały zaczerpnięte z baz danych PISA 2012 (PISA, 2012b). W PISA 2012 uczestniczyło ok. 510 tys. uczniów. Z populacji polskich 15-latków pobrano reprezentatywną próbę losową liczącą 4607 uczniów z 184 szkół, w tym 4594 uczniów gimnazjum (PISA, 2012). W szkołach publicznych naukę pobiera 97% uczniów objętych badaniem. W próbie znalazło się 51% dziewcząt.

Analiza wyników badań

Analizy przeprowadzono przy użyciu pakietu statystycznego Statistica wersja 10. W badaniu związków przyczynowych między *motywacją do uczenia się matematyki*, *samooceną zdolności matematycznych*, *lękiem matematycznym* oraz *kapitałem kulturowym uczniów* a *umiejętnościami matematycznymi*

polskich 15-latków zastosowano modelowanie równań strukturalnych. Przed przystąpieniem do oszacowania poprawności modelu sprawdzono rozkłady zmiennych oraz zastosowano analizę wielowymiarowej regresji.

Modelowanie równań strukturalnych (*Structural Equation Modeling*; *SEM*) jest zbiorem procedur wielowymiarowych analiz statystycznych, które w klasycznym ujęciu bazują na ogólnym modelu liniowym. Modele strukturalne powstały z dwóch głównych technik: konfirmacyjnej analizy czynnikowej (Bollen, 1989) oraz wielowymiarowej regresji i analizy ścieżek (Hollander i Wolfe, 1999). SEM jest techniką testowania i oceny relacji przyczynowych przy użyciu danych empirycznych i jakościowych założeń przyczynowych (Pearl, 2000). Technika ta zyskuje coraz większą popularność wśród badaczy. Umożliwia ona testowanie stopnia dopasowania teoretycznych modeli do danych. Modele te pozwalają uwzględniać, obok związków między zmiennymi ukrytymi i obserwowalnymi (modele pomiarowe), również struktury powiązań między zmiennymi ukrytymi. Modele strukturalne umożliwiają badanie jednoczesnego wpływu wielu źródeł na zmienną zależną (Kline, 2004). Niewątpliwą zaletą SEM jest możliwość analizowania zarówno bezpośrednich, jak i pośrednich zależności. Ponadto zmienne zawarte w modelach mogą być mierzone na różnych skalach pomiarowych. Dodatkową, użyteczną własnością SEM jest dostępność wielu metod estymacji, które w określonych sytuacjach pozwalają rozluźnić wymagania odnośnie rozkładów zmiennych (Pleśniak, 2009). Jednak pociąga to za sobą określone wymagania dotyczące wielkości próby. Najczęściej stosowanymi estymatorami są metoda największej wiarygodności (maximum likelihood, ML), uogólniona metoda najmniejszych kwadratów (generalized least squares, GLS) oraz metoda asymptotycznie niewrażliwa na rozkład (asymptotically distribution free, ADF). Estymatory ML i GLS mogą być stosowane dla wielowymiarowych rozkładów normalnych. Przy niespełnieniu warunku normalności można stosować GLS lub ADF. Obie metody w tym zastosowaniu wymagają bardzo dużych prób, GLS daje dobre oszacowanie parametrów przy próbach N>2500 (Byrne, 2010; Bedyńska i Książek, 2012), zaś ADF wymaga minimum 500 obserwacji (Curran, West i Finch, 1996) dla najprostszych modeli, a w przypadku złożonych modeli N>5000 (Hu, Bentler i Kano, 1992). Z przeglądu publikacji wynika, że w zastosowaniach SEM wymagania dotyczące wielkości próby często pozostają niespełnione (Westland, 2010).

W ostatnich latach prace teoretyczne doprowadziły do opracowania metod testowania także efektów nieliniowych (Lee, Song i Poon, 2004).

Wyniki

Statystyki opisowe zmiennych uwzględnionych w modelu zamieszczono w tabeli 1. Rozstęp dla każdej zmiennej objaśniającej pokrywa cały zakres skali. Prawie wszystkie rozkłady są lekko prawoskośne i platokurtyczne. Jedynie *motywacja zewnętrzna* ma nieco silniejsze i lewoskośne odchylenie rozkładu od symetrii i, w przeciwieństwie do rozkładów pozostałych zmiennych, jest nieznacznie leptokurtyczna. Uwzględniając wartości skośności, kurtozy oraz wielkość próby, można przyjąć spełnienie założenia normalności rozkładów.

Tabela 1. Statystyki opisowe analizowanych zmiennych latentnych

Zmienne	statystyki opisowe					
	M	min	max	SD	skośność	kurtoza
umiejętności matematyczne	521	245	818	88,4	0,12	-0,36
kapitał kulturowy	3,2	1	6	1,4	0,30	-0,58
motywacja wewnętrzna	8,6	4	16	2,9	0,26	-0,37
motywacja zewnętrzna	11,2	4	16	2,9	-0,51	0,19
lęk matematyczny	11,8	5	20	3,6	0,13	-0,33
samoocena	12,0	5	20	3,9	0,22	-0,67

M – średnia arytmetyczna; SD – odchylenie standardowe

W celu sprawdzenia, jaką część wariancji *umiejętności matematycznych uczniów* wyjaśniają przyjęte zmienne objaśniające, zastosowano wielowymiarową analizę regresji liniowej. Współczynnik korelacji wielokrotnej R jest równy 0,70. Wariancja wyjaśniona przez zmienne objaśniające przyjęte w modelu stanowi prawie połowę (49%) całkowitej zmienności *umiejętności matematycznych uczniów* ($R^2 = 0,49$). Model jest bardzo dobrze dopasowany do danych ($F = 257,5$; $df = 5$; $p < 0,001$).

W tabeli 2. zamieszczono standaryzowane współczynniki regresji (b) oraz współczynniki b wraz z błędami oszacowania każdego z nich. Porównanie wartości b wskazuje, że najsilniejszym predyktorem *umiejętności matematycznych uczniów* jest *samoocena zdolności matematycznych* (b = 0,47). *Motywacja zewnętrzna,* choć znacząca statystycznie, w niewielkim stopniu poprawia właściwości predykcyjne modelu (D $R^2 = 0,002$).

Tabela 2. Predyktory poziomu umiejętności matematycznych polskich 15-latków

Zmienne objaśniające	poziomu umiejętności matematycznych				
	β	bł. std. β	b	bł. std. b	t (1444)
W. wolny			438,89	15,37	28,56***
kapitał kulturowy	0,28	0,02	18,48	1,24	14,87***
motywacja wewnętrzna	0,21	0,03	6,44	0,91	7,07***
motywacja zewnętrzna	0,06	0,03	1,86	0,78	2,37**
lęk matematyczny	-0,23	0,03	-6,07	0,68	-8,92***
samoocena	0,47	0,03	10,96	0,73	14,95***

**** p< 0,01; ***** p< 0,001**

Modele strukturalne

Modele strukturalne przedstawiające hipotetyczne zależności pomiędzy motywacją do *uczenia się matematyki wewnętrzną* (*mot_w*) i *zewnętrzną* (*mot_z*), *lękiem matematycznym* (*lęk*), *samooceną zdolności matematycznych* (*samooc*) i *kapitałem kulturowym* (*kap_kul*) a *umiejętnościami matematycznymi uczniów* (*mat*) zamieszczono na rycinie 1. i 2. W modelach tych można wyróżnić wewnętrzną strukturę ścieżkową, opisującą związki przyczynowo-skutkowe

między zmiennymi nieobserwowalnymi oraz pięć modeli pomiarowych charakteryzujących relacje między zmiennymi latentnymi a ich obserwowalnymi wskaźnikami (na ryc. 2 pominięto modele pomiarowe).

Należy podkreślić, że hierarchia ważności wpływów poszczególnych zmiennych objaśniających na zmienną objaśnianą może okazać się inna przy uwzględnieniu wpływów łącznych efektów niż przy uwzględnieniu tylko efektów bezpośrednich analizowanych w wielowymiarowej regresji.

Wybór metody największej wiarygodności jako funkcji rozbieżności poprzedzono sprawdzeniem liniowości związków między zmiennymi oraz normalności rozkładów. Wartości współczynników asymetrii mieszczą się w przedziale [–0,65; +0,66], a kurtozy – [-0,92; +0,23], co pozwala na zastosowanie tej metody jako estymatora modelu.

Stopień dopasowania modelu teoretycznego do danych oceniany jest przez szereg miar, wśród których wyróżnić można miary absolutne i miary względne (porównywanie z alternatywnymi modelami). Najczęściej spotykanymi absolutnymi wskaźnikami dopasowania są wartość statystyki χ^2 oraz RMSEA (Root Mean Square Error of Approximation). Przy użyciu statystyki χ^2 testowana jest hipoteza zerowa mówiąca o idealnym dopasowaniu modelu empirycznego i teoretycznego. Tak więc wartość testu powinna być nieistotna statystycznie ($p>0,05$) dla dobrze dopasowanego modelu. Jednak liczne analizy wskazują, że nie powinno się przywiązywać zbyt dużej wagi do tej statystyki (Konarski, 2009).

RMSEA jest pierwiastkiem średniego kwadratu błędu aproksymacji. Poprzez uwzględnienie stopni swobody wskaźnik ten premiuje proste modele. Wartości RMSEA < 0,05 świadczą o *dobrym dopasowaniu* modelu (Hu i Bentler, 1999), przedział [0,06; 0,08] – *zadowalające dopasowanie*; [0,08; 0,10] – *mierne dopasowanie*. Wartości RMSEA >0,10 wskazują na złe dopasowanie modelu (Browne i Cudeck, 1992).

W grupie miar absolutnych znajdują się też wskaźnik dobroci dopasowania (goodness-of-fit index, GFI) i wskaźnik dobroci dopasowania skorygowany o stopnie swobody (adjusted goodness-of-fit index, AGFI). Dla modeli dobrze dopasowanych wartości GFI i AGFI powinny być większe od 0,90. Badania symulacyjne wskazują nawet wartość 0,95 jako dolną granicę dobrego dopasowania modelu (Sivo i in., 2006; Hu i Bentler, 1999). Miary te nie są w pełni niezależne od wielkości próby.

Każdy z pięciu modeli pomiarowych zmiennych latentnych zawartych w modelu strukturalnym (ryc. 1) poddano odrębnie ocenie, stosując jednoczynnikową CFA. Miary dopasowania modeli do danych zestawiono w tabeli 3. Zastosowana konfirmacyjna analiza czynnikowa pozwala przyjąć strukturę czynników w każdym z pięciu ocenianych modeli. Najlepsze dopasowanie wystąpiło dla *umiejętności matematycznych* (CFA_mat) oraz dla *motywacji zewnętrznej do uczenia się matematyki* (CFA_mat_z). W przypadku zmiennych latentnych *lęk matematyczny* oraz *samoocena zdolności matematycznych* błąd aproksymacji (RMSEA) jest na tyle duży, że modele te możemy uznać za dopasowane do danych, ale tylko miernie. Analiza ładunków czynnikowych wskazuje, że w każdym modelu są one istotne statystycznie ($p<0,001$).

W tabeli 3. zamieszczono też miary dopasowania dwóch modeli strukturalnych (por. ryc. 1 i 2).

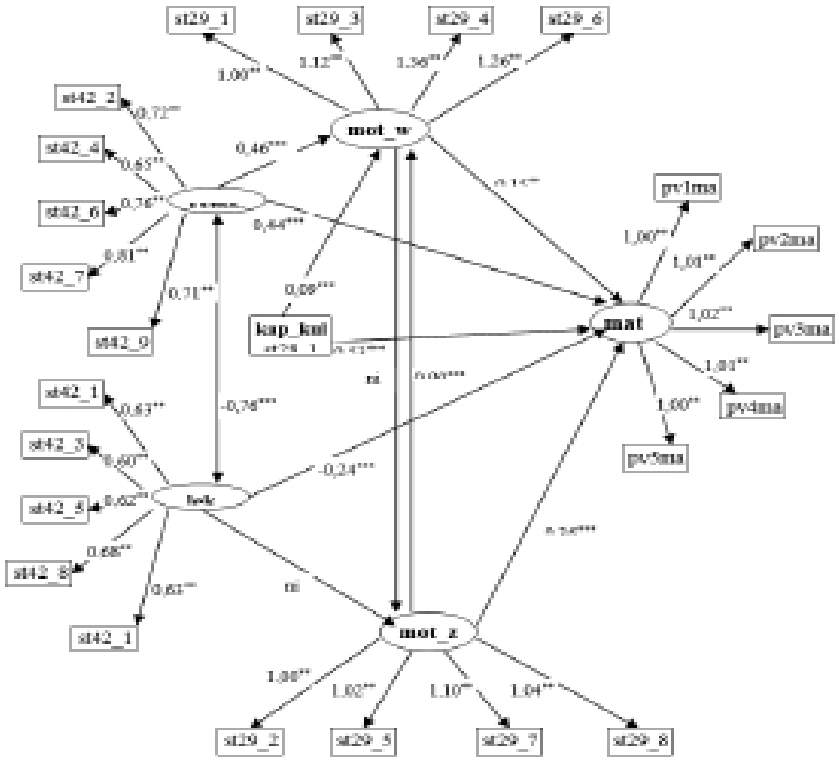
Tabela 3. Miary dopasowania modeli pomiarowych i ogólnych modeli strukturalnych

Model	χ^2	df[a]	RMSEA[b]	CI[c] 90%RMSEA	GFI[d]	AGFI[e]
CFA_mot_z	5,11	2	0,023	0,000; 0,480	0,99	0,99
CFA_mot_w	20,57***	2	0,055	0,035; 0,078	0,99	0,98
CFA_lek	177,0***	5	0,087	0,076; 0,098	0,99	0,95
CFA_samooc	132,3***	5	0,093	0,079; 0,107	0,98	0,95
CFA_mat	3,09	5	0,00	0,000; 0,016	1	1
Model 1	1332,2***	242	0,058	0,055; 0,061	0,924	0,906
Model 2	1343,5***	244	0,058	0,055; 0,061	0,924	0,906

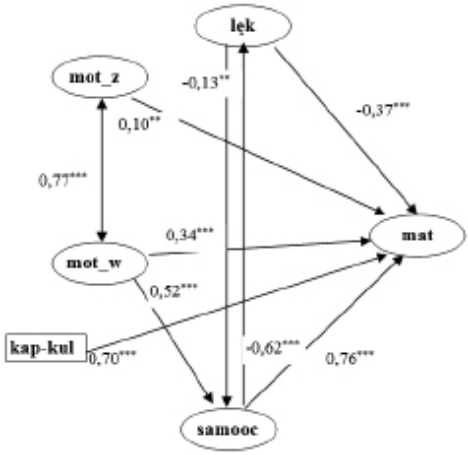
[a] – liczba stopni swobody
[b] – średniokwadratowy błąd aproksymacji
[c] – przedział ufności
[d] – wskaźnik dopasowania modelu Joreskga i Sorboma
[e] – wskaźnik dopasowania modelu skorygowany o df

Obydwa modele mają tę samą liczbę zmiennych latentnych i tę samą liczbę wskaźników empirycznych, a różnią się strukturą powiązań między zmiennymi latentnymi. W obu modelach zmienną objaśniającą jest *kapitał kulturowy ucznia* a zmienną objaśnianą *umiejętności matematyczne*. Modele 1. i 2. różnią się rolą zmiennych *lęk*, *samoocena* i *motywacja*. W modelu 1. *lęk* i *samoocena* pełnią rolę zmiennych objaśniających, natomiast w modelu 2. zmiennymi objaśniającymi są *motywacje*. Miary dopasowania uprawniają do uznania, że każdy z tych modeli dobrze odtwarza dane. Wartości RMSEA wskazują na dobre dopasowanie modeli. GFI i AGFI (> 0,90) pozwalają uznać modele za dopasowane w zadowalającym stopniu. Na podkreślenie zasługuje fakt prawie identycznych wartości miar dopasowania obu modeli. Jedyna nieznaczna różnica dotyczy χ^2.

W modelu strukturalnym zamieszczonym na rycinie 1. nieistotny okazał się wpływ *motywacji wewnętrznej* na *motywację zewnętrzną* oraz *lęku* na *motywację zewnętrzną*. Wpływ *motywacji wewnętrznej* na *umiejętności matematyczne uczniów* jest istotny na poziomie p<0,01. Pozostałe zależności są istotne na poziomie p<0,001. W modelu uwzględnione są związki bezpośrednie i pośrednie. I tak, całkowity wpływ *samooceny* na *umiejętności matematyczne uczniów* jest sumą wpływu bezpośredniego i pośredniego – przez *motywację wewnętrzną*, jako zmienną pośredniczącą.



Rycina 1. Diagram ogólnego modelu strukturalnego (model 1) z wartościami standaryzowanymi współczynników ścieżkowych i ładunków czynnikowych
**** p <0,01; *** p<0,001; ni p>0,05**



Rycina 2. Diagram modelu strukturalnego (model 2) z wartościami standaryzowanymi współczynników ścieżkowych
**** p <0,01; *** p<0,001**

Na rycinie 2. pominięto modele pomiarowe, które są identyczne jak w modelu 1. (por. ryc. 1). Przedstawione na rycinie 2. współczynniki ścieżek dają inną strukturę wpływu poszczególnych zmiennych na umiejętności matematyczne uczniów. I tak, na łączny efekt wpływu *motywacji wewnętrznej* na *umiejętności matematyczne* oprócz efektu bezpośredniego składa się efekt pośredni, będący rezultatem wpływu mediowanego przez *samoocenę zdolności matematycznych*. Uwzględnienie wszystkich łącznych efektów zmiennych objaśniających pozwala ustalić hierarchę ich wpływu na zmienną objaśnianą w modelach.

Dokonanie wyboru między modelem 1. i 2. jest niemożliwe na podstawie miar statystycznych. Niezbędne jest w tym celu mocne osadzenie modelu w teorii na etapie bez konstrukcji.

Podsumowanie

Modelowanie równań strukturalnych **(SEM)** jest strategią analizy danych pozwalającą testować rozbudowane modele teoretyczne uwzględniające różne formy relacji przyczynowych między zmiennymi. W literaturze przedmiotu podkreślane są liczne walory stosowania tych wielowymiarowych analiz statystycznych. SEM pozwala na weryfikację hipotez o występowaniu określonej struktury zależności między zmiennymi (Korol, 2005). Modele równań strukturalnych są podobne do modeli regresji wielowymiarowej, są jednak bardziej elastyczne w przypadku opisu interakcji między zmiennymi, zwłaszcza jakościowymi, dla których nie są spełnione klasyczne założenia regresji (Sagan, 2003). Ponadto umożliwiają testowanie, oprócz efektów bezpośrednich, efektów pośrednich, dostarczając informacji na temat roli mediatorów i moderatorów. Wprowadzenie do modelu strukturalnego modeli pomiarowych (zmiennych latentnych wraz z ich wskaźnikami empirycznymi) pozwala uwzględnić w oszacowaniach siły związków między zmiennymi latentnymi także błąd związany z pomiarem.

Stosowanie SEM w wielu dziedzinach nauki ma długą tradycję. Jednak nadal trwają dyskusje na temat przydatności tej metody (Breckler, 1990; McKim i Turner, 1997). Część badaczy SEM uważa, że jest to metoda pozwalająca testować związki przyczynowe na podstawie danych nieeksperymentalnych (Meehl i Waller, 2002; Kozak, 2011), inni postrzegają równania strukturalne jako metodę bezwartościową i podważają zasadność jej stosowania (Freedman, 1987, 1997; Rogosa, 1987).

W dyskusji nad możliwością stosowania SEM do testowania związków przyczynowo-skutkowych podkreśla się niespełnienie warunku o czasowym poprzedzaniu skutku przez przyczynę. Ponadto w wielu przypadkach zmiana kierunku zależności między dwoma konstruktami nie zmienia dopasowania modelu i oszacowania parametrów (Mueller, 1997). Kolejnych argumentów podważających możliwości stosowania SEM do testowania związków przyczynowych dostarcza analiza modeli ekwiwalentnych, czyli modeli o tej samej liczbie parametrów, stopni swobody i różnych relacjach między zmiennymi. Wykazują one podobne dopasowanie do tych samych danych empirycznych (MacCallum i in., 1993).

Decyzje dotyczące kierunku zależność muszą być formułowane na podstawie teorii i projektu badań. Przy użyciu SEM zaproponowane sekwencje przyczynowo-skutkowe mogą nie zostać sfalsyfikowane, co nie oznacza ich potwierdzenia. Silniejszych dowodów na czasową sekwencję konstruktów w modelu dostarczają liczne badania podłużne. Podłużne procedury SEM oferują lepsze, choć wciąż niewystarczające możliwości wstępnego dowodzenia kierunku przyczynowości (Bullock, Harlow i Mulaik, 1994; Hoyle i Smith, 1994).

Ta utrzymująca się różnica zdań nie ma podłoża matematycznego, lecz raczej metodologiczne lub filozoficzne (Meehl i Waller, 2002).

Przedstawione w części badawczej dwa modele strukturalne miały być praktycznym przykładem trudności w rozstrzyganiu związków przyczynowo-skutkowych na podstawie danych nieeksperymentalnych przy użyciu SEM. Pokazano istnienie dwóch modeli równoważnych co do puli zmiennych, które zawierają odmienne struktury ścieżkowe pasujące równie dobrze do tych samych danych empirycznych. Brak we wstępnej części artykułu uzasadnienia teoretycznego dla jednego z przyjętych modeli pozbawił możliwości podjęcia racjonalnej decyzji. Wskazuje to wyraźnie, że o wyborze jednego z alternatywnych modeli decydują nie wskaźniki statystyczne, lecz spójność testowanego modelu z teorią (Konarski, 2009). Dlatego też konieczna jest świadomość ograniczeń w rozległości wniosków dotyczących przyczynowości, do których uprawnia stosowanie tej strategii badawczej.

Bibliografia

1. Bedyńska, S., Książek, M. (2012). *Statystyczny drogowskaz 3*, Warszawa: Wydawnictwo Akademickie Sedno.
2. Breckler, S. J. (1990). *Applications of covariance structure modeling in psychology: Cause for concern?* "Psychological Bulletin"*, 107,* 260-273.
3. Bollen, K. (1989).*With new incremental structural index for general equation models made.* "Sociol Methods Res" ,17, 303–316.
4. Browne, M., Cudeck, R. (1992). *Alternative ways of assessing model fit*, "Sociological Methods and Research". 21, 2, 230-259.
5. Bullock, H., Harlow, L., & Mulaik, S. (1994). *Causation issues in structural equation modeling research.* "Structural Equation Modeling", *1,* 253-267.
6. Byrne B.*, Structural Equation Modeling with AMOS. Basic Concepts, Applications, and Programming*, Routledge, New York, 2010.
7. Curran, P., West, S., Finch, J. (1996). *The robustnees of test statistics to nonnormality and specification error in confirmatory factor analisis*, "Psychological Methods", 1,16-29.
8. Freedman, D. A. (1987). *As others see us: A case study in path analysis. "*Journal of Educational Statistics"*, 12,* 101-128.
9. Freedman, D. A. (1997). *From association to causation via regression*. In V. R. McKim & S. P. Turner (Eds.), *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences* (pp. 113-161, 177-182). Notre Dame, IN: University of Notre Dame Press.

10. Hollander M, Wolfe D. (1999). *Nonparametric statistical methods*. Wiley series in probability and statistics: applied probability and statistic. New York: Wiley.
11. Holmbeck, Grayson N. (1997) . *Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples child-clinical and pediatric psychology literatures*. "Journal of Consulting and Clinical Psychology", 65*,* 599-610.
12. Hoyle, R., Smith, G. (1994). *Formulating clinical research hypotheses as structural equation models: Aconceptual overview.* "Journal of Consulting and Clinical Psychology", 62*,* 429-440.
13. Hu, L., Bentler, P., Kano, Y., (1992). *Can test statistics in covariance structure analisis be trusted?* "Psychological Bulletin", 112, 2, 351-362.
14. Jöreskog K. G. (2002), *Structural Equation Modeling with Ordinal Variables Using LISREL*, http//www.ssicentral.com/lisrel/techdocs/ordinal.pdf
15. Kline, R. (2011). *Principles and practice of structural equation modeling (methodology in the social sciences)*, NY: The Guilford Press.
16. Konarski, R. (2009). *Modele równań strukturalnych*, Warszawa: Wydawnictwo Naukowe PWN.
17. Korol J. (2005): *Modele równań strukturalnych i sieci neuronowe w modelowaniu rozwoju zrównoważonego*. Gorzów Wielkopolski: Akademickie Wydawnictwo Ekonomiczne.
18. Kozak, M. (2011). *Analiza zwi**ązków przyczynowo-skutkowych w agronomii i hodowli roślin*, Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin, 259, 1-19.
19. Lee, S., Song, X., Poon, W. (2004). Comparison of approaches in estimating interaction and quadratic effects of latent variables. *Multivariate Behavioral Research*, 39(1), 37-67.
20. MacCallum, R., Wegener, D., Uchino, B., Fabrigar, L. (1993). *The problemof equivalent models in applications of covariance structure analysis. "*Psychological Bulletin", 114, 185–199.
21. McKim, V., Turner, S. (1997). *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences.* Notre Dame, IN: University of Notre Dame Press.
22. Meehl, P., Waller, N. (2002). *The Path Analysis Controversy: A New Statistical Approach to Strong Appraisal of Verisimilitude,* "Psychological Methods", 7, 3, 283–300.
23. Mueller, R. (1997). *Structural equation modeling: Back to basics*. "Structural Equation Modeling"*, 4,* 353-369.
24. Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.
25. *PISA (2012). Wyniki badania 2012 w Polsce*. http://ifispan.waw.pl/pliki/wyniki_pisa.pdf
26. *PISA (2012a) Assessment and Analytical Framework Mathematics, Reading, Science, Problem Solving and Financial Literacy*, http://www.oecd.org/pisa/pisaproducts/PISA%202012%20framework%20e-book_final.pdf
27. *PISA (2012b). Database – PISA 2012,* http://pisa2012.acer.edu.au/downloads.php
28. *PISA (2012c). Student questionnaire Form B*, http://pisa2012.acer.edu.au/downloads/MS12_StQ_FORM_B_ENG.pdf
29. Pleśniak, A. (2009). *Wybór metody estymacji w budowie skali czynnikowej*, "Wiadomości statystyczne", 11, 1-16.
30. Rogosa, D. (1987). *Causal models do not support scientific conclusions: A comment in support of Freedman.* "Journal of Educational Statistics"*, 12,* 185–195.
31. Sagan A. (2003): *Model pomiarowy satysfakcji i lojalności*. StatSoft Polska.
32. Sivo**,** S., Xitao, F., Witta, E., Willse, J (2006). *The search for "optimal" cutoff properties: fit index criteria in structural equation modeling, "*Journal of Experimental Education". 2006, 74, 3, 267-288.

33. StatSoft (2011). *Modelowanie równań strukturalnych* (SEPATH) http://www.statsoft.pl/textbook/stathome_stat.html?http%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstsepath.html
34. von Oertzen, T. (2010). *Power equivalence in structural equation modelling*, "British Journal of Mathematical and Statistical Psychology", 62, 2, 257-272.
35. Westland, J. (2010). *Lower bounds on sample size in structural equation modeling,* Electron. Comm. Res. Appl., 9, 6, 476-487.