Edukacja 2013, 4(124), 20–41
ISSN 0239-6858

Modele psychometryczne w pomiarze diagnostycznym

Artur Pokropek

Zespół Analiz Osiągnięć Uczniów, Instytut Badań Edukacyjnych*

W artykule przedstawione zostały podstawowe kategorie modeli psychometrycznych, które mogą zostać zastosowane w diagnostycznym pomiarze edukacyjnym. Szczególną uwagę poświęcono nowej dla polskiego pomiaru edukacyjnego kategorii modeli określanej mianem „statystycznych modeli diagnostycznych" lub „kognitywnych modeli diagnostycznych". W artykule szczegółowo opisany został jeden z modeli diagnostycznych: DINA, a następnie pokazano jego zastosowanie na polskich danych uzyskanych na egzaminie gimnazjalnym w części matematyczno-przyrodniczej. Artykuł pokazuje korzyści płynące z nowego podejścia, jak również problemy związane z jego implementacją dla diagnozy edukacyjnej.

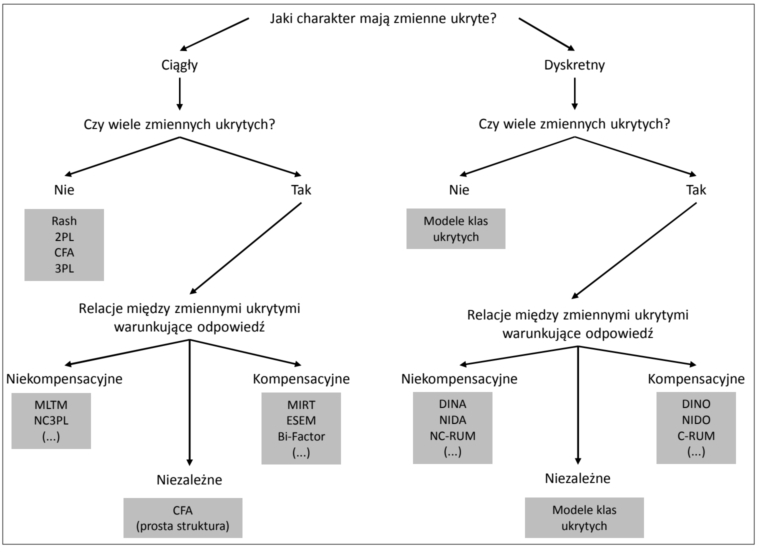
Słowa kluczowe: pomiar edukacyjny, egzaminy, IRT, matematyka, model DINA, modele cech ukrytych, modele diagnostyczne, pomiar diagnostyczny.

Modele zmiennych ukrytych

W światowej literaturze najpopularniejszymi metodami statystycznymi wykorzystywanymi w pomiarze edukacyjnym są analizy oparte na klasycznej teorii testów (KTT) oraz na modelach teorii odpowiedzi na zadania testowe w skrócie nazywanych IRT (od *Item Response Theory*). W ostatnich latach do głosu dochodzi nowa kategoria modeli określanych mianem „statystycznych modeli diagnostycznych" lub „kognitywnych modeli diagnostycznych" (por. Rupp, Templin i Henson, 2010). Modele diagnostyczne (będziemy się tutaj posługiwać akronimem CDM od *Cognitive Diagniostic Models*) często próbuje się przeciwstawiać analizom opartym na IRT czy klasycznej teorii testów. Takie podejście nie jest podejściem trafionym. To cele badawcze pomiaru edukacyjnego mogą mieć charakter diagnostyczny – sam model statystyczny jest tylko zbiorem procedur wykonywanych przez komputer. Odpowiednio przeprowadzone analizy w paradygmacie KTT czy IRT w określonych przypadkach mogą okazać się równie dobre lub nawet bardziej użyteczne z punktu widzenia diagnozy, niż analizy oparte na modelach CDM. Zarówno modele diagnostyczne, jak i modele IRT należą do szerszej kategorii modeli zmiennych ukrytych. Modele CDM to modele zmiennych ukrytych o określonych właściwościach, które wydają się wychodzić naprzeciw oczekiwaniom badawczym związanym z diagnozą edukacyjną.

Na Rysunku 1 przedstawiono schemat ułatwiający orientację wśród modeli zmiennych ukrytych. Schemat został skonstruowany w taki sposób, by w przybliżeniu odpowiadał ścieżce pytań i odpowiedzi, jaką musi

*Adres do korespondencji: ul. Górczewska 8, 01-180 Warszawa. E-mail: a.pokropek@ibe.edu.pl



Rysunek 1. Modele zmiennych ukrytych.

przejść badacz chcący dokonać trafnego pomiaru edukacyjnego. Jedno z pierwszych pytań, jakie trzeba sobie zadać na takiej ścieżce, brzmi: Jaki charakter mają zmienne latentne: ciągły czy porządkowy? Innymi słowy, czy mierzoną cechę można przedstawić na liczbowym kontinuum jak wzrost czy wagę, czy ma ona raczej „ziarnisty" charakter. Odpowiedź na to pytanie zależy od wielu czynników i nie jest nigdy oczywista. Weźmy na przykład umiejętność czytania. Czy jest to cecha stopniowalna w sposób ciągły? Czy badanych respondentów da się uszeregować wedle ich umiejętności tak, aby każdy respondent sąsiadował z jednym badanym, który będzie czytał lepiej i z jednym, który czytał będzie gorzej? Czy może respondenci stanowią kilka kategorii, np. „nie umie czytać", „czyta słabo", „czyta biegle", a porównywanie umiejętności respondentów wewnątrz tych kategorii nie ma sensu? Lub weźmy inną umiejętność: dodawania i odejmowania liczb naturalnych w zakresie od 0 do 100. Czy można dodawać lepiej lub gorzej, czy może umiejętność ta ma dychotomiczny charakter? Uczeń potrafi dodawać w zakresie od 0 do 100 lub nie potrafi. Odpowiedź na to pytanie musi być przemyślana, nie zależy ona tylko od typu sprawdzanej umiejętności, ale też od grupy badawczej, celów badawczych i podejścia teoretycznego, jakie ma zostać wykorzystane.

Drugie pytanie, jakie musi sobie zadać badacz brzmi: Czy badamy wiele zmiennych latentnych (umiejętności), czy nie? To znaczy, czy umiejętność, którą chcemy zmierzyć, ma wiele wymiarów (atrybutów) i należy ją opisywać kilkoma wartościami, czy można bez szkody dla wnioskowania zredukować ją do jednego wymiaru. I znowu, choć istnieją statystyczne narzędzia eksploracyjne, odpowiedź na to pytanie uwikłana jest w wiele uwarunkowań i również zależy od typu umiejętności, grupy badawczej, celów badawczych i podjęcia teoretycznego.

Jeżeli badacz odpowie sobie na pierwsze dwa pytania i stwierdzi, że charakter badanej zmiennej jest ciągły oraz ma do czynienia z jedną zmienną ukrytą, statystyczne modele, jakimi może się posługiwać, należą do klasy jednowymiarowych modeli IRT (zob. Kondratek i Pokropek, 2013) lub w wielu wypadkach z powodzeniem może używać klasycznej teorii testów. Jeżeli zaś charakter zmiennych jest dyskretny i badacz ma do czynienia z cechą ukrytą o jednowymiarowym charakterze, typ modeli statystycznych, do których powinien się odwołać, nazywany jest modelem klas ukrytych (zob. Hagenaars i McCutcheon, 2002).

Sytuacja staje się bardziej skomplikowana, gdy badanych jest kilka cech latentnych (zarówno w przypadku zmiennych o charakterze ciągłym, jak i dyskretnym). W takiej sytuacji musimy rozpatrzyć, jaki charakter mają relacje między zmiennymi ukrytymi warunkującymi odpowiedzi na dane zadanie. Czy relacje są kompensacyjne, niekompensacyjne, czy może niezależne?

Warto zaznaczyć, że nie chodzi tutaj o korelację między zmiennymi ukrytymi, ale o relacje, które bezpośrednio wpływają na poprawne bądź niepoprawne odpowiadanie na zadanie testowe. W tym kontekście, gdy mówimy, że relacje między umiejętnościami są „niezależne", mamy na myśli to, że prawdopodobieństwo poprawnej odpowiedzi na jedno zadanie jest warunkowane tylko przez jedną zmienną ukrytą. Tak jak w wielowymiarowej konfirmacyjnej analizie czynnikowej, w której nie ma krzyżujących się ładunków czynnikowych (jedna zmienna ukryta mierzona jest przez jeden unikalny zestaw wskaźników).

W modelach kompensacyjnych pozwala się na to, żeby między zmiennymi latentnymi dochodziło do interakcji warunkujących poprawną lub niepoprawną odpowiedź na zadanie. Zakłada się tutaj, że braki w jednej umiejętności mogą być rekompensowane przez inne umiejętności. W modelach niekompensacyjnych posiadanie odpowiedniego poziomu umiejętności (lub posiadanie danej umiejętności w ogóle) jest niezbędne do poprawnej odpowiedzi na zadanie – jedna umiejętność nie może być całkowicie zastąpiona drugą.

Dwa proste deterministyczne modele odpowiedzi na zadania testowe mogą wytłumaczyć różnicę między relacjami kompensacyjnymi i niekompensacyjnymi. Załóżmy, że mamy trzy zmienne ukryte (umiejętności): U_1, U_2, U_3, załóżmy dalej, że umiejętności te mają charakter dychotomiczny. U_i przybiera wartość 1, jeżeli dana umiejętność jest opanowana, lub 0, jeżeli umiejętność nie jest opanowana. Niech X oznacza odpowiedź na zadanie tak, że przybiera wartość 1 dla poprawnej odpowiedzi i 0 dla odpowiedzi niepoprawnej. Prosty niekompensacyjny model może być wyrażony następująco:

$$X = U_1 \cdot U_2 \cdot U_3$$

Zauważmy, że w tym modelu posiadanie wszystkich umiejętności jest niezbędne do uzyskania poprawnej odpowiedzi. Jeżeli przynajmniej jedna umiejętność nie została opanowana, iloczyn po prawej stronie równania przybierze wartość 0. Taki typ interakcji nazywany jest regułą koniunktywną (*conjuctive*). Przeciwną tej regule jest tzw. reguła dysjunktywna (*disjunctive*) i charakteryzuje ona model kompensacyjny (por. Rupp, Templin i Henson, 2010):

$$X = 1 - [(1 - U_1) \cdot (1 - U_2) \cdot (1 - U_3)]$$

W tym równaniu widać, że do poprawnego rozwiązania zadania potrzebna jest jedynie jedna umiejętność. Gdy przynajmniej jedna wartość U_i równa jest 1, iloczyn po prawej stronie równania przybiera wartość 1.

Modele kompensacyjne wydają się najodpowiedniejsze dla zadań mających wiele możliwych strategii odpowiedzi (Reckase, 1997). Modele niekompensacyjne wydają się odpowiednie dla takich zadań, w których niezbędne są pewne umiejętności. Na przykład do zadania matematycznego niezbędne są dwie umiejętności: umiejętność czytania, aby przeczytać zadanie, oraz matematyczne umiejętności niezbędne do rozwiązania takiego zadania (Bolt i Lall, 2003).

W przypadku ciągłych zmiennych ukrytych dla każdego typu relacji między tymi zmiennymi, wpływającymi na prawdopodobieństwo poprawnej odpowiedzi, dysponujemy całkiem pokaźną liczbą modeli statystycznych mogących zostać wykorzystanych w analizach. Jeżeli zakładamy, że odpowiedzi na pytania są zależne tylko od jednej zmiennej latentnej, mamy do dyspozycji konfirmacyjną analizę czynnikową. Jeżeli relacje między zmiennymi latentnymi są kompensacyjne, można wykorzystać model MLTM (*Multicomponent Latent Triat Model*; Whitely, 1980) lub wielowymiarowy model IRT (np. NC3PL, Sympson, 1978). Gdy zaś relacje uznane zostaną za kompensacyjne, badacz ma do dyspozycji np. *Multinominal 2 Parameter Logistic Model* (Reckase, 1985) lub inne modele IRT zaliczane do tak zwanej klasy modeli MIRT. Również inne modele statystyczne, takie jak *Bi-Factor* (Gibbons i Hedeker, 1992) czy ESEM (Asparouhov i Muthén, 2009) mogą odzwierciedlać kompensacyjne relacje między zmiennymi ukrytymi.

Jeżeli analizowane zmienne ukryte mają charakter dyskretny i mamy do czynienia z niezależnością zmiennych latentnych, przechodzimy do modeli klas ukrytych. Natomiast gdy relacje warunkujące odpowiedzi między zmiennymi latentnymi są typu kompensacyjnego lub niekompensacyjnego, wtedy badacz przechodzi do gałęzi modeli nazywanych diagnostycznymi (CDM). Sześć najbardziej popularnych modeli diagnostycznych zostało przedstawionych w Tabeli 1, w której podano typ, skróconą nazwę oraz referencje do pełnego opisu modelu.

Jak zostało pokazane, mnogość modeli statystycznych mogących opisywać umiejętności uczniów jest zawrotna i nie ogranicza badacza do silnie zakorzenionych założeń mówiących o jednowymiarowości i ciągłym charakterze badanej cechy. O ile w pomiarze mającym charakter różnicujący założenia takie mogą być pożądane, o tyle w pomiarze mającym charakter diagnostyczny przedstawiona na Rysunku 1 ścieżka pytań i odpowiedzi powinna być za każdym razem konstruowana, a model najbliższy teoretycznym założeniom powinien zostać empirycznie zweryfikowany. W dalszej części artykułu na przykładzie danych z części matematyczno-przyrodniczej egzaminu gimnazjalnego zaprezentowane zostaną analizy odnoszące się do dwóch typów pomiaru: różnicującego i diagnostycznego. W pierwszym typie analiz zastosowany zostanie dwuparametryczny model IRT, w drugim typie jeden z modeli diagnostycznych.

Matematyka na egzaminie gimnazjalnym 2012

Test mierzący umiejętności matematyczne będący częścią egzaminu gimnazjalnego w 2012 r. posłuży nam za przykładowy zestaw danych, co do których można zastosować różne podejścia pomiarowe i typy modeli statystycznych. W przedstawionych

Tabela 1
Popularne modele diagnostyczne

Typ modelu	Akronim	Pełna nazwa	Źródło
Niekompensacyjny	DINA	Deterministic Inputs, Noisy AND Gate	Haertel (1989)
	NIDA	Noisy Inputs, Deterministic AND Gate	Maris (1995)
	RUM	Reparameterized Unified Model	Rupp, Templin i Henson (2010)
Kompensacyjny	DINO	Deterministic Inputs, Noisy ‘OR’ Gate	Templin i Henson (2006)
	NIDO	Noisy Inputs, Deterministic ‘OR’ Gate	Rupp, Templin i Henson (2010)
	C-RUM	Compensatory Reparameterized Unified Model	Xu (2012)

niżej analizach posłużono się losowo dobraną próbą 30 000 uczniów. Losowanie miało charakter prosty. Ograniczenie się do tak dobranej próby uczniów było praktyczne i miało za zadanie przyśpieszyć procedury estymacyjne wykorzystywane w zastosowanych analizach. Dla uproszczenia analiz zadania wielopunktowe (w arkuszu egzaminacyjnym: 21, 22, 23) zostały zdychotomizowane w taki sposób, by maksymalizować wariancje dla poszczególnych zadań.

Należy również podkreślić, że analizy przeprowadzane niżej mają przede wszystkim charakter demonstracyjny. Głównym ich zadaniem jest pokazanie możliwości, jakie otwierają się przed badaczem, który zechce wykorzystać modele diagnostyczne. Inne cele poznawcze schodzą w tym artykule na drugi plan.

Podejście różnicujące

W podejściu różnicującym badaczom chodzi o uszeregowanie uczniów ze względu na (zazwyczaj) jedną cechę. Ponadto zakłada się, że cecha ta ma charakter ciągły. W takiej sytuacji odpowiednimi narzędziami badawczymi mogą być modele IRT lub analizy oparte na KTT. Przykład takich analiz przedstawiony został w Tabeli 2. Znalazły się w niej informacje o mocy różnicującej zadania (liczonej na dwa sposoby: korelacją Pearsona i korelacją polichoryczną), łatwości zadań oraz o parametrach modelu IRT: dyskryminacji i trudności. Analizę tę można by rozszerzyć o analizę dystraktorów, graficzną analizę dopasowania modelu IRT do danych, lecz w tym wypadku nie wniosłoby to wiele. Moc różnicująca zadań jest zadowalająca, ich rozkład łatwości wydaje się dobrze dobrany, choć liczba zadań jest zdecydowanie za krótka dla zapewnienia rzetelnego pomiaru indywidualnego (α Cronbacha = 0,85, błąd standardowy pomiaru wynosi 2 pkt., czyli precyzja pomiaru przy 95-procentowym przedziale ufności wynosi +/- 4 pkt.). Słowem, zadania mające zbudować ciągłą skalę umiejętności mają przyzwoite właściwości psychometryczne. Skala skonstruowana na ich podstawie, niestety, nie będzie skalą dobrze różnicującą z powodu niewielkiej liczby zadań.

Nawet jeżeli założymy, że precyzja, z jaką skala może różnicować uczniów jest do jakichś celów wystarczająca, analizy statystyczne i wnioskowanie na ich podstawie kończą się w tym momencie. Oczywiście można prowadzić dalsze dociekania w odniesieniu do różnych grup uczniów (np. chłopcy i dziewczęta), przeprowadzić analizę stronniczości zadań, ale wszystkie te analizy niewiele wniosą do wiedzy na temat szczegółowych właściwości umiejętności uczniów. Nie jest to zarzut w kontekście postawionego sobie zadania, które polega na pomiarze różnicującym. Pomiar różnicujący został dokonany. Jego właściwości (rzetelność i błąd pomiaru) przedstawione. Skonstruowane skale dalej mogą być używane do różnicowania uczniów ze względu na ogólne umiejętności matematyczne.

Podejście diagnostyczne

Podejście diagnostyczne polega na ocenie stanu wiedzy, a w szczególności specyficznych jej braków, odkrywa ich przyczyny i kieruje ku sposobom ich kompensacji. Narzędzie pomiarowe w tej sytuacji musi być zatem inne niż w przypadku pomiaru różnicującego. Jak łatwo wywnioskować, proste jednowymiarowe skalowanie zdaje się być nieadekwatne do celów diagnostycznych.

Pomiar diagnostyczny wymaga wnikliwego przejścia ścieżki pytań i odpowiedzi przedstawionej na Rysunku 1. Nie chcemy w tym artykule wnikać w teoretyczne rozważania, prezentowane analizy traktowane są tu instruktażowo. Załóżmy zatem, że teoretyczna analiza została przeprowadzona i umiejętności mierzone na teście matematycznym uznane zostały za dyskretne (precyzyjniej, dwuwartościowe: uczeń posiada lub nie posiada danej umiejętności) oraz że

Tabela 2
Analiza zadań za pomocą klasycznej teorii testów (KTT) i dwuparametrycznego modelu IRT

Zadanie	Parametry KTT			Parametry IRT			
	Moc	Moc (polichoryczna)	Łatwość	Dyskryminacja		Trudność	
				Parametr	*se*	Parametr	*se*
mat. 1	0,38	0,53	0,71	0,52	0,01	-1,17	0,03
mat. 2	0,50	0,59	0,45	0,65	0,01	0,21	0,02
mat. 3	0,47	0,59	0,57	0,62	0,01	-0,37	0,02
mat. 4	0,49	0,58	0,29	0,66	0,01	0,95	0,02
mat. 5	0,36	0,47	0,66	0,42	0,01	-1,02	0,03
mat. 6	0,33	0,65	0,89	0,85	0,02	-1,92	0,03
mat. 7	0,51	0,63	0,56	0,74	0,01	-0,29	0,01
mat. 8	0,48	0,69	0,73	0,89	0,02	-0,96	0,02
mat. 9	0,49	0,63	0,61	0,72	0,01	-0,51	0,01
mat. 10	0,50	0,63	0,60	0,74	0,01	-0,47	0,01
mat. 11	0,40	0,71	0,84	0,98	0,02	-1,44	0,02
mat. 12	0,51	0,61	0,47	0,70	0,01	0,08	0,01
mat. 13	0,53	0,62	0,41	0,72	0,01	0,35	0,01
mat. 14	0,56	0,67	0,26	0,91	0,01	0,93	0,02
mat. 15	0,41	0,5	0,49	0,47	0,01	0,04	0,02
mat. 16	0,32	0,4	0,47	0,31	0,01	0,25	0,03
mat. 17	0,46	0,68	0,76	0,90	0,02	-1,09	0,02
mat. 18	0,47	0,58	0,56	0,61	0,01	-0,34	0,02
mat. 19	0,49	0,59	0,52	0,64	0,01	-0,15	0,02
mat. 20	0,49	0,61	0,54	0,67	0,01	-0,21	0,01
mat. 21	0,67	0,77	0,31	1,37	0,02	0,57	0,01
mat. 22	0,57	0,70	0,23	0,99	0,02	1,04	0,02
mat. 23	0,68	0,80	0,26	1,54	0,02	0,74	0,01

badanych jest kilka umiejętności jednocześnie, lub – inaczej mówiąc – kilka wymiarów wiedzy matematycznej mierzonych jest na podstawie zadań tego testu matematycznego. Założenia wydają się zdroworozsądkowe i do przyjęcia.

W tym punkcie analiz niezbędne jest wyspecyfikowanie badanych umiejętności oraz relacji, jakimi zadania połączone są z umiejętnościami. Zacznijmy od wyspecyfikowania umiejętności. Istnieją do tego dwie główne drogi. Po pierwsze, można odwołać się do planu testu i sprawdzić, wedle jakiego schematu zadania zostały skonstruowane. Jeżeli plan testu jest poprawny, a zadania zostały stworzone wedle jego wytycznych, zyskujemy wysokiej jakości informacje. Druga metoda polega na zatrudnieniu ekspertów, którzy post hoc skategoryzują umiejętności i przypiszą odpowiednie zadania do odpowiednich umiejętności.

W tym artykule zostały wykorzystane obydwa podejścia. Pierwsza specyfikacja umiejętności odnosi się do oficjalnej dokumentacji testu przedstawionej przez Centralną Komisję Egzaminacyjną (CKE). W dokumentacji tej możemy znaleźć informacje, że zadania korespondują z wymaganiami ogólnymi zapisanymi w podstawie programowej (*Rozporządzenie*…, 2012). W dokumentacji CKE każdemu zadaniu przyporządkowane zostały wymagania ogólne i szczegółowe (CKE, 2012). Wymagania szczegółowe są zbyt rozdrobnione, by móc je wykorzystać do analizy. Wymagania ogólne wydają się za to możliwym do przyjęcia opisem struktury wiedzy ucznia. Składają się one z pięciu umiejętności: wykorzystywania i tworzenia informacji, wykorzystywania i interpretowania reprezentacji, modelowania matematycznego, użycia i tworzenia strategii, rozumowania i argumentacji. Opis tych umiejętności przedstawiony został w Tabeli 3.

Druga klasyfikacji powstała na podstawie wiedzy eksperckiej (zaangażowanych w jej stworzenie było dwóch matematyków i dwóch pedagogów). Eksperci nie byli zgodni co do klasyfikacji, jednak zostało wybrane kompromisowe rozwiązanie obejmujące siedem umiejętności. Pewne uproszczenia musiały być również uwzględnione z powodu tego, że jedna umiejętność była mierzona jedynie przez jedno zadanie (rachunek prawdopodobieństwa i elementy statystyki). Z praktycznego punktu widzenia uwzględnianie takiej umiejętności w analizach nie ma sensu, gdyż jedno zadanie i tak nie pozwoli na precyzyjny jego pomiar. Umiejętności wyszczególnione do analiz to: sprawność rachunkowa, posługiwanie się procentami, posługiwanie się wyrażeniami algebraicznymi i równaniami, odczytywanie wykresów, znajomość figur płaskich i wykorzystanie ich właściwości, znajomość brył i wykorzystanie ich właściwości, rozumowanie i argumentacja. Opis umiejętności wyszczególnionych przez ekspertów przedstawiony został w Tabeli 4.

Po wyodrębnieniu umiejętności mierzonych przez test, następnym krokiem niezbędnym do poprawnej analizy jest przeprowa-

Tabela 3
Umiejętności uczniów według wymagań ogólnych podstawy programowej

Wymagania ogólne	Opis nabytych umiejętności
Wykorzystanie i tworzenie informacji	Uczeń interpretuje i tworzy teksty o charakterze matematycznym, używa języka matematycznego do opisu rozumowania i uzyskanych wyników.
Wykorzystywanie i interpretowanie reprezentacji	Uczeń używa prostych, dobrze znanych obiektów matematycznych, interpretuje pojęcia matematyczne i operuje obiektami matematycznymi.
Modelowanie matematyczne	Uczeń dobiera model matematyczny do prostej sytuacji, buduje model matematyczny danej sytuacji.
Użycie i tworzenie strategii	Uczeń stosuje strategię jasno wynikającą z treści zadania, tworzy strategię rozwiązania problemu.
Rozumowanie i argumentacja	Uczeń prowadzi proste rozumowania, podaje argumenty uzasadniające poprawność rozumowania.

Źródło: Podstawa programowa kształcenia ogólnego dla gimnazjów i szkół ponadgimnazjalnych, s. 172–173.

Tabela 4
Umiejętności uczniów wyszczególnione przez ekspertów

Umiejętność	Opis nabytych umiejętności
Sprawność rachunkowa	Uczeń potrafi dodawać, odejmować, mnożyć i dzielić w zakresie liczb wymiernych.
Posługiwanie się procentami	Uczeń zna pojęcie procentu i potrafi się nim posługiwać.
Posługiwanie się wyrażeniami algebraicznymi i równaniami	Uczeń potrafi posługiwać się wyrażeniami algebraicznymi oraz potrafi rozwiązywać równania i wykorzystywać je w sytuacjach problemowych.
Odczytywanie wykresów	Uczeń potrafi odczytywać i interpretować informacje zawarte na wykresie.
Znajomość figur płaskich i wykorzystanie ich właściwości	Uczeń zna figury płaskie i ich właściwości, umie obliczyć ich powierzchnie.
Znajomość brył i wykorzystanie ich właściwości	Uczeń zna podstawowe bryły i ich właściwości, umie obliczyć ich pole i powierzchnie.
Rozumowanie i argumentacja	Uczeń prowadzi proste rozumowania, podaje argumenty uzasadniające jego poprawność.

dzenie mapowania zadań do umiejętności – innymi słowy – przypisanie każdemu zadaniu umiejętności, jakie są za jego pomocą mierzone. Wyniki takiego mapowania zapisywane są zwyczajowo w macierzach nazywanych macierzami *Q* (*Q-matrices*). Macierze korespondujące z Tabelami 3 i 4 zostały przedstawione w Tabeli 5.

W macierzach *Q* w wierszach mamy do czynienia z zadaniami, w kolumnach – z umiejętnościami. Macierz wypełniona jest zerami i jedynkami. Jedynka oznacza, że dana umiejętność jest mierzona przez konkretne zadanie, zero – że umiejętność ta jest niepotrzebna do rozwiązania danego zadania. I tak, np. do rozwiązania zadania pierwszego wedle macierzy przedstawionej w Tabeli 5 potrzebne są dwie umiejętności: wykorzystywanie i tworzenie informacji oraz wykorzystywanie i interpretowanie reprezentacji. Wedle macierzy do rozwiązania tego zadania potrzebne są również dwie umiejętności, tyle że reprezentujące inny poziom abstrakcji: sprawność rachunkowa i odczytywanie wykresów.

Po specyfikacji macierzy *Q* pozostała jeszcze jedna decyzja, od której zależeć będzie wybór modelu statystycznego wykorzystanego w diagnozie. Pytanie dotyczy relacji między zmiennymi ukrytymi a procesem odpowiedzi na pytanie: Czy relacje są kompensacyjne, niekompensacyjne, czy niezależne? Ostatnia odpowiedź została wykluczona przez samą konstrukcję macierzy *Q*. Pozostaje zatem rozstrzygnąć, czy relacje są niekompensacyjne, czy kompensacyjne, innymi słowy, czy do rozwiązania jednego zadania niezbędne są wszystkie umiejętności, czy wystarczy jedna z nich. W naszym wypadku konstrukcja zadań oraz macierzy *Q* wskazuje raczej na pierwsze rozwiązanie, czyli niekompensacyjny charakter relacji. Zakładamy tutaj, że do rozwiązania zadania niezbędne jest posiadanie wszystkich umiejętności wyszczególnionych w macierzy *Q*.

Taka droga rozumowania prowadzi do kilku modeli statystycznych, które mogą być wykorzystane w diagnozie. Posłużymy się tutaj najbardziej znanym, a zarazem najprostszym modelem diagnostycznym.

Model DINA
Specyfikacja modelu DINA podawana jest za José de la Torre (2004, s. 116–117). Niech

Tabela 5

Macierz Q oparta na umiejętnościach wyszczególnionych przez podstawę programową i ekspertów

Zadanie	Umiejętności wyszczególnione w podstawie programowej					Umiejętności wyszczególnione przez ekspertów						
	(1) Wykorzystanie i tworzenie informacji	(2) Wykorzystanie i interpretowanie reprezentacji	(3) Modelowanie matematyczne	(4) Użycie i tworzenie strategii	(5) Rozumowanie i argumentacja	(1) Sprawność rachunkowa	(2) Posługiwanie się procentami	(3) Posługiwanie się wyrażeniami algebraicznymi i równaniami	(4) Odczytywanie wykresów	(5) Znajomość figur płaskich i wykorzystanie ich właściwości	(6) Znajomość brył i wykorzystanie ich właściwości	(7) Rozumowanie i argumentacja
1	1	1	0	0	0	1	0	0	1	0	0	0
2	0	1	0	0	0	1	0	0	0	0	0	0
3	0	0	1	0	0	1	0	0	0	0	0	1
4	0	1	0	0	0	1	0	0	0	0	0	0
5	0	1	0	0	0	1	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0	1	0	0	0
7	0	1	0	0	0	1	1	0	0	0	0	0
8	0	1	1	0	0	1	1	0	0	0	0	0
9	0	1	1	0	0	1	0	0	0	0	0	1
10	0	1	0	0	0	1	0	0	0	0	0	1
11	1	1	0	0	0	0	0	0	1	1	0	0
12	1	0	1	0	1	0	0	1	1	0	0	0
13	1	0	1	0	0	0	0	1	1	0	0	0
14	0	0	0	1	0	1	0	0	0	0	0	1
15	0	1	0	0	0	0	0	0	0	1	0	1
16	0	0	1	0	0	0	0	0	0	1	0	1
17	0	0	0	0	1	0	0	0	0	1	0	1
18	0	1	1	0	0	1	0	1	0	1	0	0
19	0	0	0	1	0	1	0	0	0	1	1	1
20	0	1	0	0	0	0	0	1	0	0	1	0
21	1	0	0	1	1	1	0	0	0	0	0	1
22	0	0	0	0	1	0	0	0	0	1	0	1
23	0	0	0	1	0	1	0	1	0	1	0	1

X_{ij} oznacza odpowiedź ucznia i na zadanie j; $i = 1, ..., I$; $j = 1, ..., J$ oraz $\boldsymbol{\alpha}_i = \{\alpha_{ik}\}$ oznacza wektor dychotomicznych atrybutów (umiejętności) ucznia, $k = 1, ..., K$, gdzie 1 dla t-tego elementu oznacza obecność atrybutu lub umiejętności k, a 0 brak tego atrybutu lub umiejętności. Jak większość modeli ukierunkowanych na diagnozę, także model DINA wymaga wyspecyfikowania macierzy Q zbudowanej z J wierszy i K kolumn zawierających zera i jedynki. Komórka tej macierzy oznaczona jest przez wyrażenie q_{jk}, wskazujące na k-tą umiejętność wymaganą do poprawnej odpowiedzi na zadanie j.

W modelu „wejścia" DINA wektor umiejętności uczniów i macierz Q tworzą latentny wektor odpowiedzi $\eta_i = \{\eta_{ij}\}$, gdzie:

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}} \qquad (1)$$

Latentna odpowiedź przedstawiona w równaniu (1) zakłada wartość 1, jeżeli uczeń i posiada wszystkie umiejętności niezbędne do rozwiązania zadania j i wartości 0, jeżeli uczeń nie posiada przynajmniej jednej z potrzebnych umiejętności.

Komponenty modelu, które odnoszą się do sformułowania «*„and" gate*» w nazwie modelu odsyłają do koniunkcyjnego procesu określania η_{ij}, w którym poprawna odpowiedź na zadanie wymaga posiadania wszystkich atrybutów (umiejętności) określonych jako niezbędnych do prawidłowego wykonania zadania. DINA jest podobny w tym aspekcie do standardowego niekompensacyjnego wielowymiarowego modelu IRT. Jeżeli proces odpowiedzi byłby całkowicie deterministyczny, latentny wektor odpowiedzi byłby tożsamy z obserwowalnym wektorem odpowiedzi. Zakładamy jednak, że proces odpowiadania na zadanie jest stochastyczny i wektory te nie są tożsame.

Szum wprowadzony przez proces stochastyczny parametryzowany jest w modelu DINA poprzez parametry „pomyłki" (*slip*) oraz „zgadywania" (*guessing*). Tak, że uczeń posiadający wszystkie wymagane do rozwiązania zadania parametry może pomylić się i odpowiedzieć niepoprawnie, a uczeń nieposiadający wszystkich parametrów ma szanse poprawnie odpowiedzieć na zadanie. W modelu DINA parametry pomyłki (s_j) oraz zgadywania (g_j) zdefiniowane są następująco:

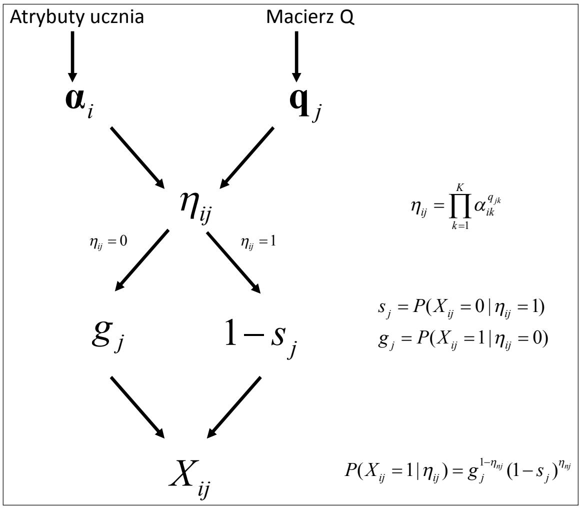
$$\begin{aligned} s_j &= P(X_{ij} = 0 \mid \eta_{ij} = 1) \\ g_j &= P(X_{ij} = 1 \mid \eta_{ij} = 0) \end{aligned} \qquad (3)$$

Prawdopodobieństwo poprawnej odpowiedzi na zadanie j ucznia i posiadającego wektor atrybutów $\boldsymbol{\alpha}_i$ wyrażane jest jako:

$$P_j(\boldsymbol{\alpha}_i) = P(X_{ij} = 1 \mid \boldsymbol{\alpha}_i) = g_j^{1-\eta_{nj}} (1 - s_j)^{\eta_{nj}} \qquad (4)$$

Jeżeli nie ma zgadywania i pomyłek, prawdopodobieństwo poprawnej odpowiedzi na zadanie j dla ucznia, który posiadł wszystkie umiejętności wyspecyfikowane do poprawnej odpowiedzi na zadanie wynosi 1, dla ucznia, który nie ma przynajmniej jednej umiejętności, wynosi 0. Odpowiedź zależy od interakcji między $\boldsymbol{\alpha}_i$ a macierzą Q.

Rysunek 2 przedstawia graficzną reprezentację modelu DINA. Latentny wektor odpowiedzi na zadanie j wynika bezpośrednio z latentnego wektora atrybutów ucznia i wyspecyfikowanej macierzy Q. Prawdopodobieństwo poprawnej odpowiedzi na zadanie dla ucznia, którego latentna odpowiedź jest poprawna, zależy od parametru pomyłki, natomiast prawdopodobieństwo poprawnej odpowiedzi dla ucznia, którego latentna odpowiedź wynosi 1, zależy bezpośrednio od parametru zgadywania. Przy czym warto zwrócić uwagę (za: de la Torre i Douglas, 2004), że parametr zgadywania oznacza coś więcej, niż mogłoby to wynikać z jego nazwy. Nie chodzi tutaj tylko o zgadywanie, które możemy nazwać „strzelaniem" na testach jednokrotnego wyboru,



Rysunek 2. Graficzna reprezentacja modelu DINA.

ale zawiera też w sobie sytuację, w której uczniowie rozwiązują poprawnie zadanie za pomocą alternatywnych strategii niewyspecyfikowanych w macierzy *Q*.

Estymacja modelu DINA opiera się na procedurze *Marginal Maximum Likelihood*, analogicznej jak w procedurze stosowanej w IRT. W modelach IRT maksymalizowana funkcja wiarygodności przybiera następującą postać:

$$L(\mathbf{X}) = \prod_{i=1}^{I} L(\mathbf{X}_i) = \prod_{i=1}^{I} \int L(\mathbf{X}_i \mid \theta) g(\theta) d\theta \quad (5)$$

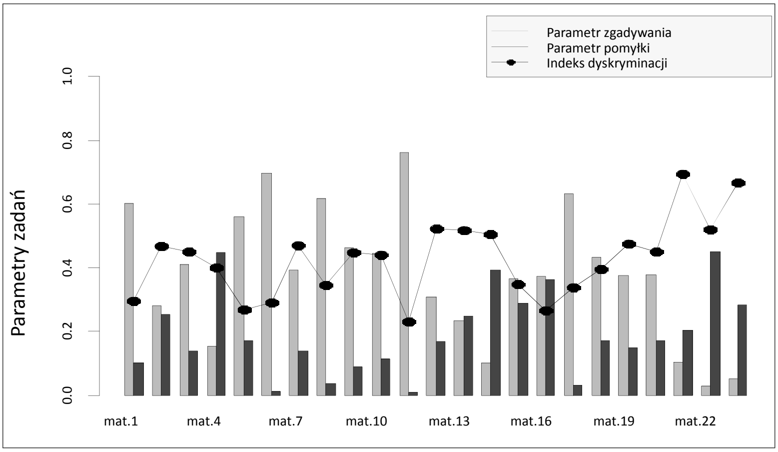
Podczas gdy w modelu DINA

$$L(\mathbf{X}) = \prod_{i=1}^{I} L(\mathbf{X}_i) = \prod_{i=1}^{I} \sum_{l=1}^{L} L(\mathbf{X}_i \mid \boldsymbol{\alpha}_l) p(\boldsymbol{\alpha}_l) \quad (6)$$

$L(\mathbf{X}_i)$ jest funkcją wiarygodności dla ucznia *i*, $p(\boldsymbol{\alpha}_l)$ to prawdopodobieństwo a priori dla wektora atrybutów $\boldsymbol{\alpha}_l$, natomiast $l = 2^K$, gdzie *K* jest liczbą umiejętności (atrybutów). Procedury estymacyjne są analogiczne jak w jednowymiarowych modelach IRT. Całkowanie przez ciągłą umiejętność, która zostaje zdyskretyzowana w procencie estymacyjnym modeli IRT, w modelu DINA zostaje zastąpiona przez ważoną sumę warunkowych wiarygodności dla 2^K możliwych wzorów atrybutów (umiejętności). Tak jak w modelach IRT do estymacji parametrów wykorzystane mogą zostać zarówno algorytm EM, jak i MCMC (de la Torre, 2009 s. 119). Modele przedstawione w tym artykule szacowane były przy użyciu algorytmu EM za pomocą pakietu CDM (Robitzsch, Kiefer, George, Uenlue i Robitzsch, 2013).

Zastosowanie modelu DINA do wymagań ogólnych określonych w podstawie programowej

Na Rysunku 3 przedstawione zostały parametry zadań dla modelu DINA, w którym wykorzystana została macierz *Q* reprezentująca klasyfikację umiejętności według wymagań ogólnych podstawy programowej. Dokładne wartości parametrów przedstawiono w aneksie. Rysunek wyraża trzy parametry: zgadywania, pomyłki i dyskryminacji. Pierwsze dwa były omawiane wcześniej, lecz ostatni wymaga dopowiedzenia. Dyskryminacja jest tutaj zoperacjonalizowana jako: (1 - s_i) - g_i. Mimo odmiennego wzoru dyskryminacje w modelu DINA można interpretować podobnie do mocy różnicującej w KTT. Niska dyskryminacja



Rysunek 3. Parametry modelu DINA dla macierzy *Q* reprezentującej klasyfikację umiejętności według wymagań ogólnych podstawy programowej.

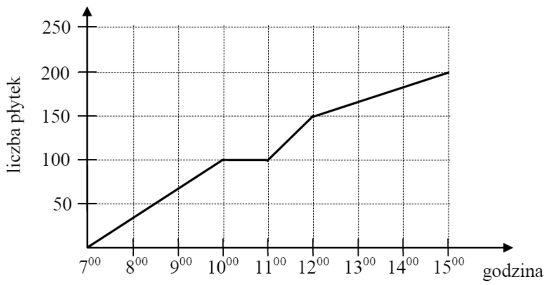
oznacza, że odpowiedź na zadanie słabo różnicuje uczniów posiadających umiejętności niezbędne do rozwiązania zadania i uczniów nieposiadających takich umiejętności. Wysoka dyskryminacje wskazuje na to, że poprawna bądź niepoprawna odpowiedź na zadanie jest dobrym wskaźnikiem odnoszącym się do tego, czy uczeń posiada wyszczególnione w macierzy *Q* umiejętności, czy nie (Rupp, Templin i Henson, 2010).

Jak widać na Rysunku 3, zadania w tym wypadku odznaczają się bardzo przeciętnymi charakterystykami psychometrycznymi. Parametry zgadywania są wysokie, ale należy pamiętać, że test ten dla większości zadań miał charakter pytań zamkniętych. Dlatego parametry zgadywania w okolicy wartości 0,25 dla pytań zamkniętych są do zaakceptowania. Warto zwrócić uwagę, że trzy ostatnie zadania, które były zadaniami otwartymi (brak możliwości zgadywania) charakteryzują się bardzo niską dyskryminacją. Jednak dla części zadań zamkniętych parametry zgadywania są bardzo wysokie. Aż dla 10 zadań parametr zgadywania wynosi powyżej 0,4. Najwyższymi parametrami zgadywania charakteryzują się zadania 6 i 11. Treść ich i graficzna forma przedstawione zostały na Rysunku 4.

Według twórców testu, aby rozwiązać zadanie 6, uczeń musi posiadać umiejętność „wykorzystywania i tworzenia informacji". Inne umiejętności nie są mierzone w tym zadaniu. Aby rozwiązać zadanie 11, uczeń musi posiadać dodatkową umiejętność „wykorzystania i interpretowania reprezentacji". W przypadku obydwu zadań wydaje się, że zakres wyspecyfikowanych umiejętności jest zbyt wąski, dlatego parametry zgadywania są tak wysokie. Parametry pomyłki są zdecydowanie niższe niż parametry zgadywania. Tylko w dwóch przypadkach przekraczają wartość 0,4, a w czterech 0,3 (dla zadań 4, 14, 16, 22). Wysokie parametry pomyłki mogą oznaczać, że umiejętności przypisane zadaniom faktycznie nie są potrzebne do poprawnej odpowiedzi lub charakter zadania (rozkład dystraktorów, sformułowanie pytania) są nieadekwatne. Rozstrzygnięcie pozostawiamy czytelnikowi. Parametr dyskryminacji zależy bezpośrednio od parametru zgadywania i pomyłki, nie dziwi zatem niska dyskryminacja zadań: 11, 5, 16 i 6.

Charakterystyki zadań sugerują, że model odznacza się średnim dopasowaniem do danych (warto zaznaczyć, że statystyki RMSEA dla poszczególnych zdań nie

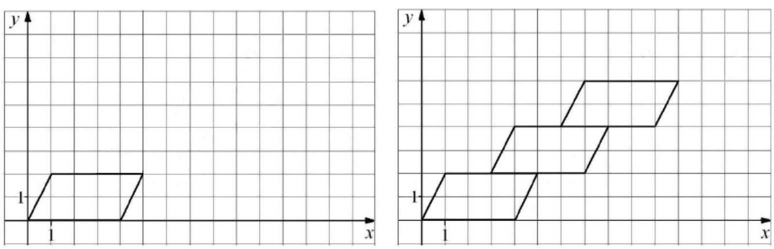
Zadanie 6.
Glazurnik układał płytki. Wykres przedstawia liczbę ułożonych płytek w zależności od czasu w trakcie ośmiogodzinnego dnia pracy.



Na podstawie wykresu wybierz zdanie fałszywe.

A. O godzinie 10^{00} glazurnik rozpoczął godzinną przerwę.
B. Od 7^{00} do 8^{00} glazurnik ułożył mniej płytek niż od 11^{00} do 12^{00}.
C. W ciągu każdej godziny glazurnik układał taką samą liczbę płytek.
D. Przez ostatnie trzy godziny pracy glazurnik ułożył 50 płytek.

Informacje do zadań 11.–13.
Małgosia narysowała równoległobok położony w układzie współrzędnych tak jak na pierwszym rysunku. Kolejne przystające do niego równoległoboki rysowała w taki sposób, że dolny lewy wierzchołek rysowanego równoległoboku był środkiem górnego boku poprzedniego równoległoboku (rysunek 2.).



Rysunek 1. Rysunek 2.

Zadanie 11.
Dokończ zdanie tak, aby otrzymać zdanie prawdziwe.

Małgosia narysowała w opisany sposób czwarty równoległobok. Współrzędna y prawego górnego wierzchołka tego równoległoboku jest równa

A. 8 **B.** 9 **C.** 10 **D.** 11

Rysunek 4. Zadanie 6 i 11 z testu matematycznego egzaminu gimnazjalnego przeprowadzonego w 2012 r. Źródło: arkusz egzaminacyjny (CKE, 2012).

są bardzo wysokie i sugerują przynajmniej dostateczne dopasowanie modelu do danych; patrz aneks), dlatego pozostałe konkluzje płynące z modelu należy przyjmować ostrożnie.

Do tej pory omawialiśmy parametry analogicznie (do pewnego stopnia) z parametrami jednowymiarowych modeli IRT i KTT. Innymi słowy, uzyskaliśmy informacje na temat psychometrycznych właściwości wykorzystanych w pomiarze zadań. Za pomocą modelu DINA można jednak uzyskać znacznie więcej.

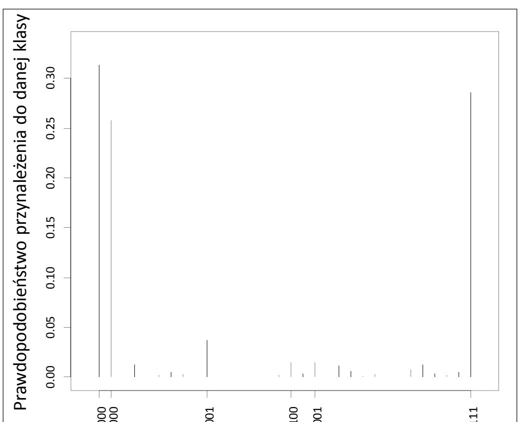
Tabela 6
Szacowany procent uczniów posiadający umiejętności według wymagań ogólnych podstawy programowej (w %)

Umiejętności	Odsetek uczniów posiadających daną umiejętność
Wykorzystanie i tworzenie informacji	66,3
Wykorzystanie i interpretowanie reprezentacji	35,5
Modelowanie matematyczne	35,7
Użycie i tworzenie strategii	31,5
Rozumowanie i argumentacja	38,3

Pierwszym przykładem niezwykle ważnej informacji z punktu widzenia diagnozy jest estymowany w populacji procent uczniów, którzy posiedli wyszczególnione w podstawie programowej umiejętności. W Tabeli 6 przedstawione zostały szacunki. Około 2/3 uczniów potrafi wykorzystywać i tworzyć informacje. Jeżeli chodzi o pozostałe umiejętności, to każdą z nich opanowało około tylko 1/3 uczniów.

Prezentowany model diagnostyczny pozwala dodatkowo szacować, jakie profile umiejętności uwidaczniają się w badanej sytuacji. Wyniki te przedstawione zostały na Rysunku 5. Jak widać, około 30% uczniów nie posiadło żadnej z wyszczególnionych umiejętności. Niespełna 30% posiada wszystkie umiejętności wyszczególnione na podstawie wymagań ogólnych podstawy programowej. Około 25% uczniów posiada jedynie umiejętność wykorzystywania i tworzenia informacji. Pozostałe profile pojawiają się dużo rzadziej.

Kolejnym ważnym zestawem informacji jest macierz korelacji między umiejętnościami. Wynika z niej wyraźnie, że umiejętności są silnie powiązane. Słabiej powiązana z innymi jest umiejętność wykorzystywania i tworzenia informacji. Może stanowić



Rysunek 5. Model DINA: estymowane profile umiejętności według macierzy *Q* skonstruowanej na podstawie kategoryzacji opartej na wymaganiach ogólnych podstawy programowej.

Tabela 7
Korelacje między umiejętnościami według macierzy Q skonstruowane na podstawie kategoryzacji opartej na wymaganiach ogólnych podstawy programowej

Wymaganie ogólne podstawy programowej	(1)	(2)	(3)	(4)	(5)
(1)	1	0,85	0,74	0,82	0,85
(2)	0,85	1	0,98	0,99	0,96
(3)	0,74	0,98	1	0,98	0,95
(4)	0,82	0,99	0,98	1	0,97
(5)	0,85	0,96	0,95	0,97	1

to wskazówkę dydaktyczną, że temat ten można traktować w szczególny sposób. Ciekawa jest również najniższa korelacja, którą znajdujemy w macierzy, wynosząca 0,74. Jest to korelacja między wykorzystaniem i tworzeniem informacji a modelowaniem matematycznym. Wydaje się, że taki wynik ma uzasadnienie teoretyczne.

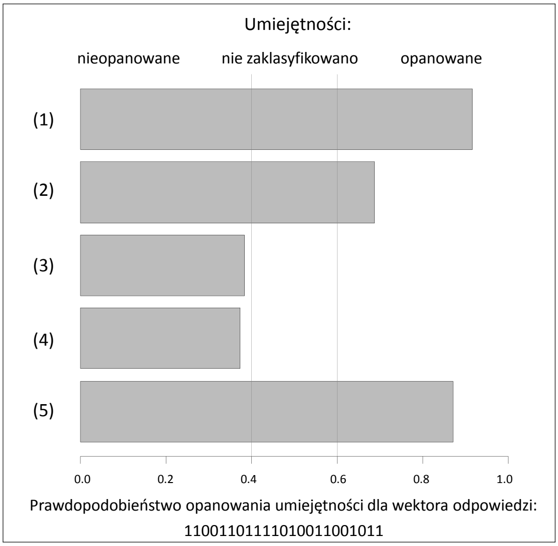
Najciekawsze informacje z modelu DINA otrzymujemy dla poszczególnych uczniów. Nie jest to jedna liczba, tak jak w przypadku pomiaru różnicującego, ale zestaw informacji, który może okazać się ważny z dydaktycznego punktu widzenia. Dwa przykłady takich informacji przedstawione na Rysunku 6.

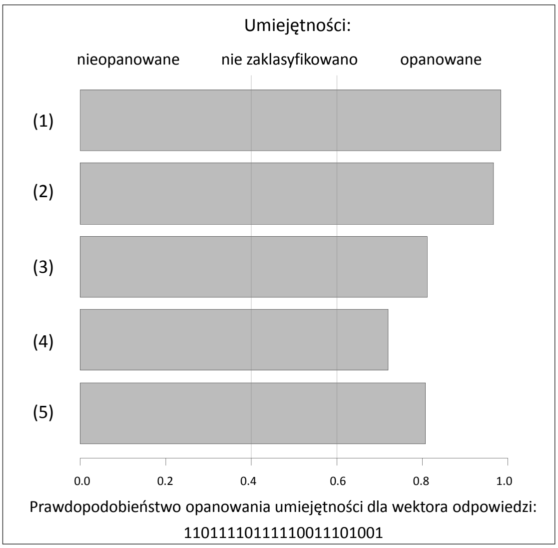
Rysunek 6 przedstawia wykresy określające prawdopodobieństwo opanowania poszczególnych umiejętności. Zakłada się tutaj, że prawdopodobieństwo posiadania umiejętności 0,4 i niższe oznacza, że uczeń jej nie opanował. Prawdopodobieństwo powyżej 0,6 oznacza, że uczeń opanował daną umiejętność, zaś obszar między 0,4 a 0,6 pozostaje obszarem niepewności. W przykładzie z Rysunku 6 (panel górny) widać, że uczeń nie posiadł umiejętności (3) „modelowania matematycznego" oraz (4) „użycia i tworzenia strategii". Natomiast uczeń, którego wyniki przedstawiono w panelu dolnym, posiadł wszystkie umiejętności definiowane przez podstawę programową. Warto zauważyć, że niesie to stosunkowo dużo informacji, które mogą zostać wykorzystane w diagnozie umiejętności ucznia, a także w dalszej indywidualizacji toku kształcenia.

Podobny zestaw informacji można uzyskać dzięki modelowaniu IRT (wielowymiarowemu lub jednowymiarowemu), które ukierunkowane będzie diagnostycznie. W takiej sytuacji prawdopodobieństwo może być zastąpione poziomem umiejętności mierzonym na abstrakcyjnej skali zmiennej ukrytej, a próg informujący o tym, czy uczeń posiadł daną umiejętność, czy nie, wyznaczony może zostać dzięki procedurze *standard setting.* Istnieje jednak kilka ważnych zalet prezentowanego modelu (abstrahując, które założenia co do umiejętności, stojące za modelami, są prawdziwe). Okazuje się, że model DINA uzyskuje znacząco wyższą rzetelność od modeli IRT (procent dobrze zakwalifikowanych uczniów) nawet przy mniejszej liczbie zadań (Rupp, Templin i Henson, 2010, s. 232–247). Problem szacowania progu umiejętności w zasadzie nie istnieje w modelu DINA, gdyż jest „wmontowany" w parametryzacje modelu, podczas gdy dla modeli IRT musi być każdorazowo szacowany przez ekspertów. Dodatkowo interpretacja wyników modelu przedstawiona głównie za pomocą procentów jest dużo łatwiejsza od interpretowania wyników na skali logitowej.

Zastosowanie modelu DINA do sprawdzania umiejętności wyszczególnionych przez ekspertów

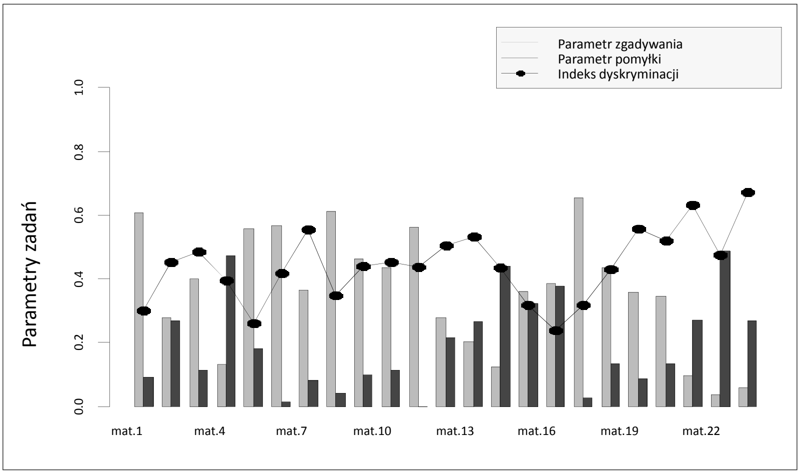
Na Rysunku 7 przedstawiono parametry zadań modelu DINA estymowanego przy użyciu macierzy *Q* wyspecyfikowanej za pomocą wiedzy eksperckiej (patrz Tabela 5). Wyniki estymacji parametrów są do pewnego stopnia zbliżone z estymacją opartą na macierzy *Q*, reprezentującą klasyfikację opartą na podstawie programowej. Parametry





Rysunek 6. Prawdopodobieństwo opanowania poszczególnych umiejętności dla dwóch przykładowych wzorów odpowiedzi (identyfikujących jednoznacznie uczniów). Model DINA według macierzy *Q* skonstruowanej na podstawie kategoryzacji opartej na wymaganiach ogólnych podstawy programowej.

zgadywania w tym modelu nie przekraczają wartości 0,7, lecz lista zadań o wysokich parametrach zgadywania (zadania: 1, 5, 6, 8, 11, 17) w znacznym stopniu pokrywa się. Macierze *Q* nie są całkowicie rozłączne, taka sytuacja nie może zatem dziwić. Należy zaznaczyć, że dopasowanie modeli jest bardzo podobne. Statystyki AIC i BIC (patrz aneks) sugerują, że model oparty na macierzy *Q*, reprezentującej kategoryzację umiejętności ze względu na podstawę programową, jest odrobinę lepiej dopasowany od modelu opartego na wiedzy eksperckiej. Różnice te nie są jednak duże i nie należy ich przeceniać.

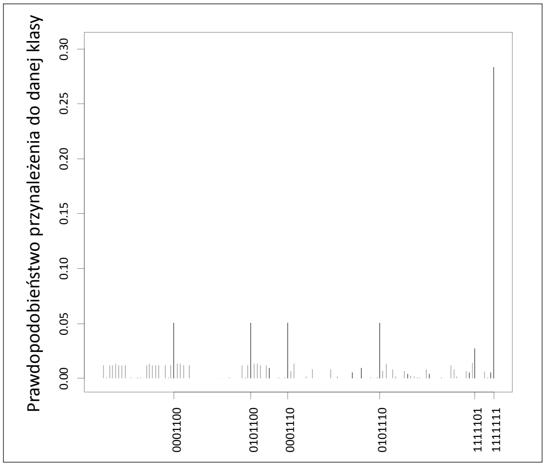


Rysunek 7. Parametry modelu DINA dla macierzy *Q* reprezentującej klasyfikację umiejętności według wiedzy eksperckiej.

Znacząco inna jest interpretacja wyników opartych na obydwu macierzach *Q*. W Tabeli 8 przedstawiono szacowany procent uczniów posiadających umiejętności według *kategoryzacji stworzonej na podstawie wiedzy eksperckiej*. Wedle szacunków opartych na modelu DINA około 38% uczniów posiadło umiejętność nazwaną przez ekspertów sprawnością rachunkową. Ponad 66% uczniów potrafi posługiwać się procentami. 47% uczniów potrafi posługiwać się wyrażeniami algebraicznymi i równaniami. Umiejętnością najczęściej opanowaną przez uczniów jest odczytywanie wykresów. Znajomość figur płaskich estymowana jest na poziomie 65%, a brył na 51%. Warto zwrócić uwagę na ostatnią 7. kategorię. Rozumowanie i argumentacja pojawiają się w obydwu modelach, lecz charakteryzowane są przez inny zestaw zadań. W modelu opartym na podstawie programowej są to 4 zadania, w modelu opartym o wiedzę ekspercką to aż 11 zadań. Stąd też duże rozbieżności. W pierwszym modelu szacujemy, że 38% posiada umiejętność rozumowania i argumentacji, w drugim modelu procent ten wynosi 53%.

Tabela 8
Szacowany procent uczniów posiadających umiejętności według kategoryzacji stworzonej na podstawie wiedzy eksperckiej (w %)

Umiejętności wyszczególnione przez ekspertów	Odsetek uczniów posiadających daną umiejętność
(1) Sprawność rachunkowa	38,4
(2) Posługiwanie się procentami	66,3
(3) Posługiwanie się wyrażeniami algebraicznymi i równaniami	47,2
(4) Odczytywanie wykresów	72,7
(5) Znajomość figur płaskich i wykorzystanie ich właściwości	65,1
(6) Znajomość brył i wykorzystanie ich właściwości	61,1
(7) Rozumowanie i argumentacja	53,4



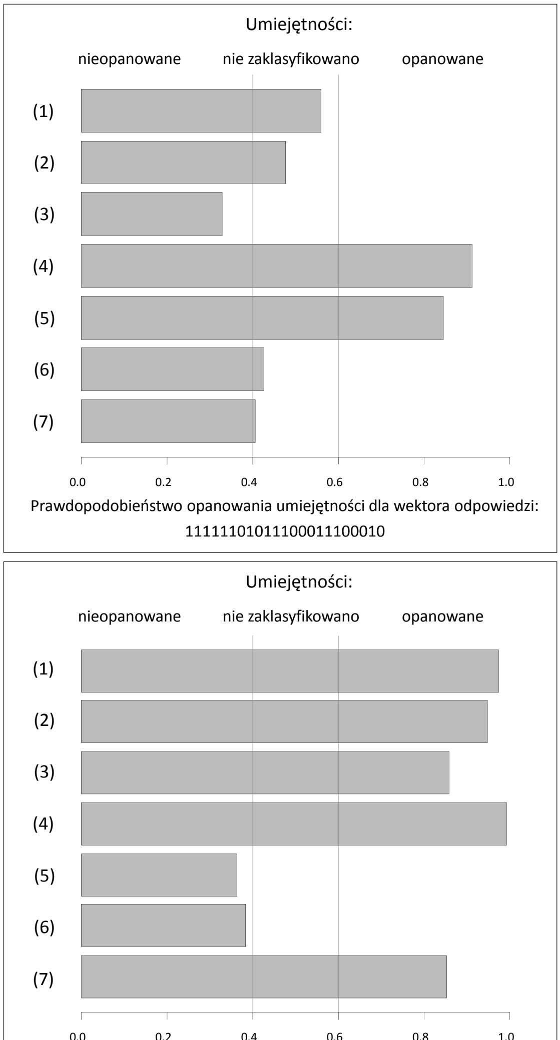
Rysunek 8. Model DINA: estymowane profile umiejętności według macierzy *Q* skonstruowanej na podstawie kategoryzacji opartej na wiedzy eksperckiej.

Na Rysunku 8 przedstawiono estymowane profile umiejętności. Najczęściej zdarza się, że uczniowie posiedli wszystkie umiejętności (około 28%). Inne wzory sytuacji są znacznie rzadziej obserwowalne. Wśród wzorów umiejętności, które pojawiają się przynajmniej u 5% uczniów możemy znaleźć następujące: uczniowie potrafiący odczytywać wykresy i znający figury płaskie; uczniowie potrafiący odczytywać wykresy i znający figury płaskie oraz potrafiący posługiwać się pojęciem procentu; uczniowie potrafiący odczytywać wykresy i znający figury płaskie oraz bryły; uczniowie potrafiący odczytywać wykresy i znający figury płaskie, bryły oraz potrafiący posługiwać się pojęciem procentu.

W Tabeli 9 przedstawiona została macierz korelacji między umiejętnościami wyszczególnionymi przez ekspertów. Na szczególną uwagę zasługują tutaj niskie korelacje między umiejętnościami powiązanymi z geometrią a innymi umiejętnościami mierzonymi przez test matematyczny. Wynik korelacji umiejętności (5) z (7), czyli

Tabela 9
Korelacje między umiejętnościami według macierzy Q skonstruowane na podstawie kategoryzacji opartej o wiedzę ekspercką

Umiejętności wyszczególnione przez ekspertów	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)	1	0,72	0,84	0,81	0,79	0,66	0,74
(2)	0,72	1	0,50	0,35	0,38	0,36	0,46
(3)	0,84	0,50	1	0,17	0,56	0,37	0,64
(4)	0,81	0,35	0,17	1	0,60	0,45	0,39
(5)	0,79	0,38	0,56	0,60	1	0,42	0,06
(6)	0,66	0,36	0,37	0,45	0,42	1	0,45
(7)	0,74	0,46	0,64	0,39	0,06	0,45	1



Rysunek 9. Prawdopodobieństwo opanowania poszczególnych umiejętności dla dwóch przykładowych wzorów odpowiedzi (identyfikujących jednoznacznie uczniów). Model DINA według macierzy *Q* skonstruowanej na podstawie kategoryzacji opartej na wiedzy eksperckiej.

znajomość figur płaskich oraz rozumowanie na poziomie bliskim 0, wskazuje na to, że zadania mierzące znajomość figur płaskich w tym tekście nie były skomplikowane i nie wymagały przeprowadzania trudniejszego rozumowania i argumentacji, niekoniecznie zaś, że umiejętności te nie są skorelowane ze sobą w ogóle. Trzeba pamiętać, że we wnioskowaniu na temat umiejętności musimy się odnosić do konkretnego testu, przynajmniej w sytuacji, w której mamy niewiele zadań mierzących poszczególne umiejętności.

Na Rysunku 9 przedstawiono prawdopodobieństwo posiadania poszczególnych umiejętności przez dwóch wybranych uczniów z badanego zbioru danych. O pierwszym z nich (górny panel) możemy powiedzieć, że opanował umiejętność odczytywania informacji z wykresów oraz posiadł znajomość figur płaskich i potrafi wykorzystywać ich właściwości. Niskie prawdopodobieństwa posiadania bardziej złożonych umiejętności wskazują, że należy podjąć z nim pracę na wielu frontach wiedzy matematycznej, próbując wyrównać braki powstałe na wcześniejszym etapie nauki

Inną sytuację przedstawiają wyniki ucznia, które zostały przedstawione w dolnym panelu Rysunku 9. Ma on wyraźne braki w geometrii, które wymagają nadrobienia. Prawdopodobieństwo opanowania materiału związanego z geometrią płaską i bryłami jest niewielkie, bo bliskie 40% (oczywiście „geometrię" musimy tutaj zawęzić do geometrii mierzonej na omawianym teście).

Warto zwrócić uwagę na to, jak bogate informacje dostajemy za pomocą modelowania opartego na podejściu diagnostycznym. W podejściu różnicującym jedyną informacją byłaby taka, że uczeń pierwszy jest nieznacznie słabszy od ucznia drugiego. Nieznacznie, ponieważ liczba poprawnych odpowiedzi (punktów dla KTT) dla pierwszego ucznia wynosi 14, dla drugiego zaś 15. Różnica jednego punktu to bardzo niewiele i z całą pewnością zostałaby uznana za nieistotną statystycznie.

Podsumowanie i dyskusja

Jak zostało pokazane, podejście diagnostycznie wraz z wykorzystaniem odpowiednich modeli może być źródłem niezwykle ważnych, z punktu widzenia edukacji, informacji. Zarówno na poziomie całej populacji, jak i konkretnych uczniów. Modele diagnostyczne stanowią ciekawą alternatywę dla szeroko stosowanych modeli IRT. Należy jednak podkreślić, że w pewnych sytuacjach stosowanie modeli IRT jest bardziej uzasadnione, w innych modele diagnostyczne wydają się bardziej odpowiednim wyborem.

Stosowanie modeli diagnostycznych wiąże się z kilkoma problemami, które nie pozwalają zdobyć im odpowiedniej pozycji w pomiarze edukacyjnym i diagnostyce. Podstawowy problem w zastosowaniu przedstawionych narzędzi polega na trudności konstrukcji narzędzia pomiarowego dopasowanego do potrzeb diagnozy i współgrającego z możliwościami, jakie daje nam modelowanie statystyczne. Konstrukcja dobrego narzędzia statystycznego wymaga współpracy specjalistów przedmiotowych wyszkolonych w dziedzinie pomiaru edukacyjnego wraz ze specjalistami posiadającymi odpowiednią wiedzę z zakresu psychometrii. Konstrukcja instrumentu pomiarowego wymaga głębokiej wiedzy na temat konstruktów, które chce się mierzyć, jak również ograniczeń modeli statystycznych oraz narzędzi pomiarowych, za pomocą których chce się dokonać pomiaru. Specyfikacja macierzy *Q*, wybór reguły kondensacyjnej określającej, w jaki sposób relacje między umiejętnościami wpływają na prawdopodobieństwo poprawnej odpowiedzi, czy decyzja, z jakim charakterem zmiennej ukrytej (ciągłym czy dyskretnym) mamy do czynienia, nie są nigdy decyzjami banalnymi.

Dodatkowo modele diagnostycznie nie mają w pełni opracowanego zaplecza statystycznego. Chodzi tutaj przede wszystkie o miary dopasowania modeli oraz statystyki, które ułatwiałyby pracę nad konstrukcją macierzy *Q* (por. Wilhelm i Robitzsch, 2009).

Pomimo przedstawionych problemów wydaje się, że praca nad pomiarem diagnostycznym z wykorzystaniem modeli zmiennych ukrytych, a szczególnie modeli diagnostycznych (CDM), może być inwestycją, która przyniesie znaczący zysk nie tylko dla metodologii pomiaru edukacyjnego i diagnostycznego, lecz dla całej edukacji.

Literatura

Asparouhov, T. i Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397–438.

Bolt, D. M. i Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo. *Applied Psychological Measurement*, *27*(6), 395–414.

Centralna Komisja Egzaminacyjna (2012). *Osiągnięcia uczniów kończących gimnazjumw roku 2012*. Warszawa: Centralna Komisja Egzaminacyjna.

Gibbons, R. D. i Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*(3), 423–436.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*(4), 301–321.

Hagenaars, J. A. i McCutcheon, A. L. (red.). (2002). *Applied latent class analysis*. Cambridge: University Press.

Kondratek, B. i Pokropek, A. (2013). IRT i pomiar edukacyjny. *Edukacja*, *124*(4), 42–66.

Maris, E. (1995). Psychometric latent response models. *Psychometrika*, *60*(4), 523–547.

Podstawa programowa kształcenia ogólnego dla gimnazjów i szkół ponadgimnazjalnych, których ukończenie umożliwia uzyskanie świadectwa dojrzałości po zdaniu egzaminu maturalnego. W: *Rozporządzenie Ministra Edukacji Narodowej z dnia 23 grudnia 2008 r. w sprawie podstawy programowej wychowania przedszkolnego oraz kształcenia ogólnego w poszczególnych typach szkół* (Załącznik 4). Dz. U. z 15 stycznia 2009 r., nr 4, poz. 17, Warszawa.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, *9*(4), 401–412.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, *21*(1), 25–36.

Robitzsch, A., Kiefer, T., George, A. C., Uenlue, A. i Robitzsch, M. A. (2013). Pakiet statystyczny CDM. Pobrano z: https://sites.google.com/site/alexanderrobitzsch/software

Rozporządzenie Ministra Edukacji Narodowej z dnia 27 sierpnia 2012 r. w sprawie podstawy programowej wychowania przedszkolnego oraz kształcenia ogólnego w poszczególnych typach szkół. Dz. U. z 30 sierpnia 2012 r. poz. 977, Warszawa.

Rupp, A. A., Templin, J. i Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.

Sympson, J. B. (1978). A model for testing with multidimensional items. W: D. J. Weiss (red.), *Proceedings of the 1977 Computerized Adaptive Testing Conference* (s. 82–98). Minneapolis: University of Minnesota.

Templin, J. L. i Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*(3), 287.

Torre, J. de la (2009). DINA model and parameter estimation: a didactic. *Journal of Educational and Behavioral Statistics*, *34*(1), 115–130.

Torre, J., de la i Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333–353.

Wilhelm, O. i Robitzsch, A. (2009). Have cognitive diagnostic models delivered their goods? Some substantial and methodological concerns. *Measurement: interdisciplinary research and perspectives*, *7*(1), 53–57.

Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, *45*(4), 479–494.

Xu, X. (2012). Review of diagnostic measurement: theories, methods, and applications. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(2), 320–327.

Aneks – parametry zadań dla modeli DINA

	Model DINA – podstawa programowa				Model 2 DINA – wiedza ekspercka			
Zadanie	Zgadywanie	Potknięcie	Dyskry-minacja	RMSEA	Zgadywanie	Potknięcie	Dyskry-minacja	RMSEA
1	0,60	0,10	0,30	0,020	0,59	0,11	0,30	0,023
2	0,28	0,25	0,47	0,026	0,26	0,26	0,48	0,030
3	0,41	0,14	0,45	0,023	0,43	0,14	0,43	0,030
4	0,15	0,45	0,40	0,052	0,14	0,46	0,40	0,062
5	0,56	0,17	0,27	0,019	0,55	0,18	0,28	0,025
6	0,70	0,01	0,29	0,018	0,63	0,01	0,36	0,015
7	0,39	0,14	0,47	0,035	0,39	0,12	0,49	0,031
8	0,62	0,04	0,35	0,053	0,60	0,04	0,36	0,061
9	0,46	0,09	0,45	0,025	0,46	0,09	0,44	0,041
10	0,45	0,11	0,44	0,030	0,46	0,11	0,43	0,055
11	0,76	0,01	0,23	0,047	0,65	0,01	0,34	0,031
12	0,31	0,17	0,52	0,020	0,28	0,19	0,53	0,022
13	0,23	0,25	0,52	0,022	0,21	0,25	0,54	0,022
14	0,10	0,39	0,51	0,035	0,10	0,41	0,49	0,039
15	0,37	0,29	0,35	0,040	0,36	0,28	0,36	0,030
16	0,37	0,36	0,27	0,026	0,37	0,36	0,26	0,031
17	0,63	0,03	0,34	0,043	0,65	0,03	0,32	0,068
18	0,43	0,17	0,40	0,049	0,43	0,16	0,41	0,048
19	0,38	0,15	0,48	0,031	0,38	0,12	0,50	0,024
20	0,38	0,17	0,45	0,031	0,36	0,13	0,51	0,027
21	0,10	0,20	0,70	0,026	0,10	0,26	0,65	0,042
22	0,03	0,45	0,52	0,028	0,05	0,45	0,49	0,032
23	0,05	0,28	0,67	0,026	0,06	0,27	0,67	0,028
	Miary dopasowania modelu				Miary dopasowania modelu			
	Deviance	508080,8	BIC	508843,4	Deviance	507813,1	BIC	509526,4
	AIC	508234,8	RMSEA	0,031	AIC	508159,1	RMSEA	0,035