

dr Dorota Węziak-Białowska

Szkoła Główna Handlowa w Warszawie

Analiza skal punktowania zadań – matura z matematyki 2007

Wprowadzenie

Tworzenie skali punktacji, która jest integralną częścią schematu oceniania, jest bardzo ważnym zadaniem, gdyż pozwala prawidłowo klasyfikować odpowiedzi uczniów. Z jednej strony skala powinna być dostatecznie rozległa, żeby umożliwić zróżnicowanie umiejętności uczniów, z drugiej zaś powinna być dostatecznie krótka, aby umożliwić egzaminatorom prawidłową i szybką ocenę. Przeprowadzone analizy miały dać odpowiedź na następujące pytanie badawcze: czy punktacja przewidziana dla każdego z zadań egzaminu maturalnego z matematyki na poziomie rozszerzonym jest optymalna, czyli czy nie jest zbyt długa lub zbyt krótka. Aby odpowiedzieć na to pytanie, sprawdzono, jakie były rozkłady przyznanych punktów za poszczególne zadania. Następnie, korzystając z możliwości, które daje model Rascha, sprawdzono, jak kształtowały się progi w punktacji poszczególnych zadań. Na końcu przeprowadzono analizę obciążenia (stronniczości) uzyskanych rezultatów. W tym celu przeanalizowano interakcje między surowością oceniania egzaminatorów a poziomami trudności zadań.

Zgodnie ze schematem oceniania arkuszy egzaminacyjnych matury z matematyki na poziomie rozszerzonym w roku 2007 egzaminatorzy mogli przyznać za każde z zadań od zera do nawet siedmiu punktów. Liczba punktów zależała od tego, ile czynności składało się na zadanie. Punktując zadanie, egzaminator sprawdzał, czy dana czynność została wykonana poprawnie i wtedy przyznawał za nią przewidzianą w schemacie oceniania liczbę punktów. W trakcie pilotażu e-oceniania punktowanie zorganizowano podobnie z tą różnicą, że egzaminatorzy zapisywali jedynie sumaryczną liczbę punktów przyznanych za każde z zadań. Ponadto należy podkreślić, że egzaminatorzy widzieli jedynie obraz jednego zadania (i tylko to zadanie w danym momencie punktowali), a nie całej pracy ucznia.

Dobór prac egzaminacyjnych do pilotażu e-oceniania i proces oceniania

Do pilotażu e-oceniania egzaminu maturalnego z matematyki na poziomie rozszerzonym wybrano 522 prace. Prace te zostały wybrane z powiatów, w których średni wynik był najbardziej zbliżony do średniej krajowej. Były to powiaty: puławski, ropczycko-sędziszowski, myślenicki, Kraków, Nowy Sącz, Biała Podlaska. Konstruując w ten sposób próbę, założono, że wybrane prace stanowić będą reprezentację całej populacji prac z tego egzaminu. Konieczność odrzucenia części prac z powodów technicznych (prace z wysoką punktacją,

w których rozwiązania uczniów wykraczały poza obszar skanowania¹⁾ wpłynęła na obniżenie średniej liczby punktów w próbie zrealizowanej w stosunku do próby planowanej.

Tabela 1. Statystyki dla populacji i próby wybranej do pilotażu dla egzaminu maturalnego z matematyki na poziomie rozszerzonym

Statystyki	Kraj	Prace zakwalifikowane do e-ocenia
Liczba	27538	522
Średnia	24,9	21,95
Mediana	25	21
Dominanta	15	15
Odchylenie standardowe	11,88	12,43

Źródło: Szalencic, Węziak-Białowolska, *Czy e-ocenia może zapewnić większą rzetelność punktowania?*, [w:] *Uczenie się i egzamin w oczach nauczycieli*, Niemierko B., Szmigel M. K. (red.), XIV Konferencja Diagnostyki Edukacyjnej, Polskie Towarzystwo Diagnostyki Edukacyjnej, Opole 2008, s. 49.

Rozkłady przyznanych punktów

W pierwszym kroku analizie poddano liczby punktów przyznanych za poszczególne zadania. Uzyskane rozkłady (liczebności i odsetki) liczby przyznanych punktów za każde z zadań zestawiono w kolumnach: *Liczebność i Odsetek* w tabeli 2. oraz zilustrowano na wykresie 1.

W przypadku zadania 1., za które uczeń mógł uzyskać od 0 do 5 punktów, najczęściej przyznawanymi liczbami punktów były 0 punktów (32,8%) lub 5 punktów (28,8%). Natomiast najrzadziej przyznawano 2 lub 3 punkty. Sytuacja taka miała miejsce w mniej niż 10% przypadków. Taki rozkład punktacji sugerował istnienie raczej cztero- (np. 0 – 1 – 2 – 3 punkty lub 0 – 1 – 4 – 5 punktów²⁾ niż przewidzianej schematem oceniania sześciostopniowej (0 – 1 – 3 – 4 – 5 punktów) skali.

Zadanie 2. charakteryzowało się najbardziej równomiernym rozkładem przyznanych punktów spośród wszystkich zadań. Żadna z liczby punktów nie była przyznawana w mniej niż 10% przypadków, co przyjmuje się za warunek konieczny do uznania danej kategorii punktowej za działającą prawidłowo³. Taki rozkład punktacji potwierdzał istnienie sześciostopniowej skali.

Zadanie 3., które zgodnie ze schematem punktowania miało sześciostopniową skalę punktacji, również nie miało kategorii punktowej przyznawanej w mniej niż 10% przypadków. Jednak rozkład przyznanych za to zadanie punktów był bardziej nierównomierny niż w przypadku zadania 2. Chociaż zdecydowanie dominowała w przypadku tego zadania maksymalna liczba możliwych do zdobycia punktów, czyli 5, to taki rozkład liczby punktów nie zaprzeczał istnieniu sześciostopniowej skali.

¹ rozwiązania umieszczone zostały na marginesach lub w obszarze przeznaczonym na rozwiązania innych zadań

² określenie sposobu punktowania zadania wymaga analizy jakościowej jego treści

³ J. M. Linacre, *Guidelines for Rating Scales*, MESA Research Note #2, <http://www.rasch.org/rn2.htm>

W przypadku zadania 4. dominujące liczby przyznanych punktów to liczba maksymalna (3 punkty) i minimalna (0 punktów). Uwagę zwraca to, że 1 punkt przyznany został jedynie w 5,5% przypadków. Taki rozkład punktacji sugeruje istnienie trzystopniowej skali.

W przypadku zadania 5. analiza rozkładu przyznanych punktów pokazała, że egzaminatorzy zwykle przyznawali od 0 do 3 punktów. Spośród rzadziej przyznawanych punktów, tj. od 4 do 7 punktów, przyznali najczęściej maksymalną ich liczbę (w 5,1% przypadków). Taki rozkład punktacji sugeruje istnienie pięciostopniowej skali.

W przypadku zadania 6 egzaminatorzy najczęściej przyznawali 0 punktów (60,7%), a następnie 1 lub 4 punkty (w około 10% przypadków). Pozostałe liczby punktów przyznawane były w mniej niż 6% przypadków. Taki rozkład punktacji sugeruje istnienie trzystopniowej skali.

Rozkład punktów za zadanie 7. był silnie asymetryczny prawostronnie. Oznaczało to, że egzaminatorzy przyznawali za to zadanie najczęściej 0 lub 1 punkt. Pozostałe liczby punktów przyznawane były zdecydowanie rzadziej. Na tym tle wyróżnia się maksymalna liczba punktów, którą przyznano w około 7% przypadków. Taki rozkład punktacji sugeruje istnienie trzystopniowej skali.

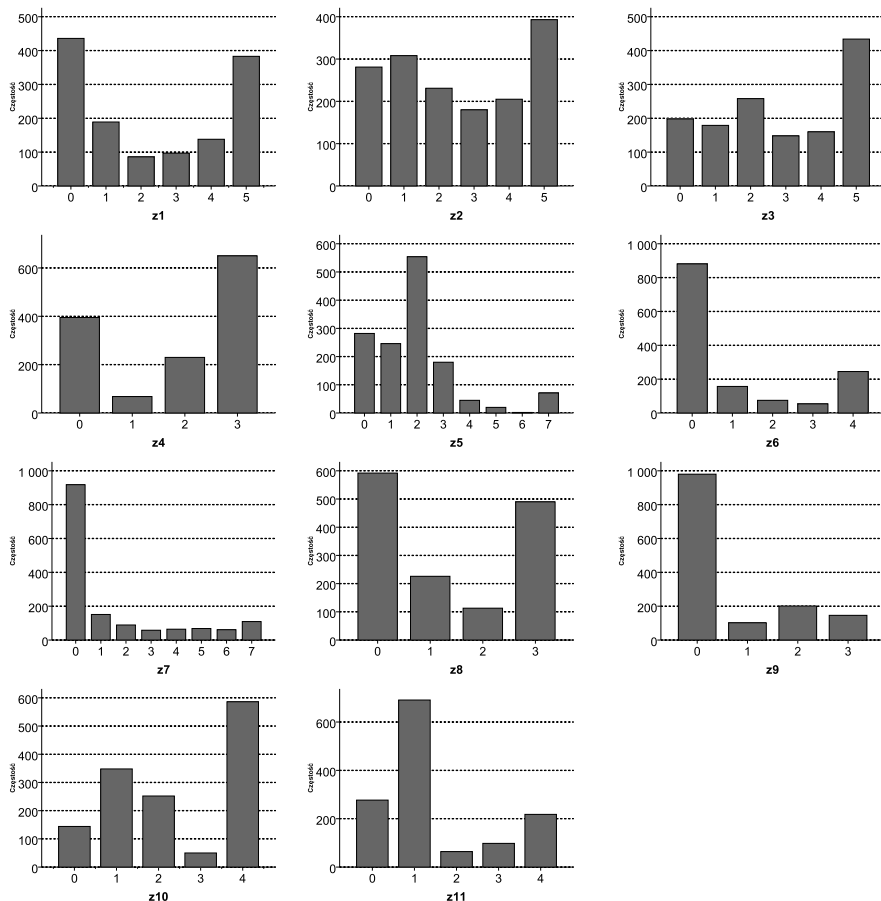
Za zadanie 8. egzaminatorzy mogli przyznać od 0 do 3 punktów. Rozkład przyznanych w pilotażu punktów był u-kształtny z najrzadziej przyznawaną liczbą punktów na poziomie 2 (8% przypadków). Taki rozkład punktacji sugeruje istnienie naturalnej trzystopniowej skali.

W przypadku zadania 9. najczęściej egzaminatorzy przyznawali 0 punktów (około 70% przypadków). Pozostałe trzy kategorie punktów przyznawane były w miarę równomiernie. Zatem chociaż 1 punkt przyznany został w mniej niż 10% przypadków, to biorąc pod uwagę, że do rozdysponowania między trzy kategorie punktowe zostało około 30% przypadków punktowania, można uznać, że skala punktacji zadania 9. była zgodna z percepcją egzaminatorów.

Cechą charakterystyczną zadania 10. było to, że 3 punkty za to zadanie zostały przyznane jedynie w 3,6% przypadków punktowania tego zadania. Pozostałe kategorie punktowe „wykorzystywane” były zdecydowanie częściej. Rozkład punktacji tego zadania sugeruje istnienie czterostopniowej skali.

W przypadku zadania 11. dwie spośród czterech kategorii punktowych, „wykorzystywane” były rzadko. Były to 2 punkty (4,7% przypadków) oraz 3 punkty (7,2% przypadków). Rozkład punktacji tego zadania wskazuje na istnienie trzystopniowej skali.

Podsumowując, w przypadku zadań: 1., 4. i 8. rozkład przyznanych punktów miał postać litery „u”, co oznaczało, że w tych zadaniach zdecydowanie najczęściej wykorzystywane w punktowaniu były skrajne kategorie punktowe. Ponadto w przypadku 9 z 11 zadań wystąpiły kategorie punktów niedoreprezentowane (wykorzystane w mniej niż 10% przypadków oceniania).



Wykres 1. Rozkłady przyznanych punktów⁴

Funkcjonowanie schematu oceniania

W kolejnym kroku sprawdzono, czy liczba możliwych do przyznania punktów za poszczególne zadania została ustalona odpowiednio. Do tego celu wykorzystano model Rascha. Dla każdego z zadań oszacowano progi⁵ dla skali punktacji i sprawdzono, czy ich kolejność nie jest zaburzona. Postępowanie takie miało swoje uzasadnienie w fakcie, że wyższa liczba punktów przyznanych za zadanie powinna współwystępować z wyższym poziomem umiejętności ucznia.

⁴ Arkusz egzaminacyjny dostępny jest pod adresem http://www.cke.edu.pl/images/stories/mat2_07/mat_pr.pdf

⁵ Próg to punkt, w którym uczeń ma 50% szans uzyskania jednej z dwóch sąsiadujących liczb punktów.

Ponadto na podstawie oszacowań progów wyznaczono wykresy krzywych prawdopodobieństwa (*ang. category probability curves*) dla każdego z zadań⁶.

Do oceny sprawdzenia poprawności funkcjonowania schematu oceniania zastosowano wskazówki opracowane przez Linacre'a⁷. Dla przejrzystości formułowanych na tej podstawie wniosków wskazówki te przytoczono w załączniku. Wyniki analizy poprawności funkcjonowania schematu oceniania przedstawiono w tabeli 2. W analizie szczególną uwagę zwrócono na wskazówki nr 1, 3, 5.

Analiza liczebności przyznanych punktów za poszczególne zadania pokazała, że kryterium dziesięciu obserwacji przypadających na każdą liczbę możliwych do przyznania punktów nie zostało spełnione jedynie w przypadku zadania 5. W tym zadaniu 6 punktów zostało przyznanych tylko dwukrotnie na ponad 1300 przypadków oceniania. Jest to dosyć silną przesłanką za zredukowaniem (zmniejszenie o jeden) liczby ocenianych czynności. Kwestią do ustalenia w toku dalszych analiz pozostaje rozstrzygnięcie, która to powinna być czynność.

Kształt rozkładu przyznanych punktów został omówiony w punkcie *Rozkłady przyznanych punktów*. Wyniki tej analizy przemawiają za zmniejszeniem liczby przyznawanych punktów, co zwykle jest równoważne ze zredukowaniem liczby ocenianych czynności, w przypadku wszystkich zadań z wyjątkiem zadania 9. Redukcja liczby ocenianych czynności może oznaczać połączenie niektórych z nich. Jednak, aby wskazać, które to z nich, niezbędna jest analiza jakościowa treści zadań oraz schematu oceniania.

Kryterium wzrostu wartości średnich umiejętności wraz ze wzrostem liczby punktów (przeciętny poziom umiejętności w tabeli 2.) zostało spełnione jedynie w przypadku zadania 9. Po jednym zaburzeniu monotoniczności zanotowano w zadaniach: 3., 4., 8. oraz 10. W przypadku zadań: 1., 2., 6., 7. i 11. wystąpiły po dwa zaburzenia monotoniczności przeciętnych poziomów umiejętności. Natomiast w przypadku zadania 5. takich zaburzeń było aż pięć. Wystąpienie zaburzenia kolejności oznacza, że nie zawsze za określone zadanie więcej punktów otrzymują uczniowie o wyższych umiejętnościach.

⁶ Wykresy te konstruuje się w dwuwymiarowym układzie współrzędnych, gdzie na osi pionowej mierzone jest prawdopodobieństwo przyznania danej liczby punktów, natomiast na osi poziomej odmierza się różnice między oszacowaniami umiejętności ucznia i oszacowaniami poziomu trudności zadania ($B_n - D_n$). Dla każdej z liczby możliwych do przyznania punktów za dane zadania wyznacza się odpowiadającą jej krzywą prawdopodobieństwa. Krzywa ta pokazuje, z jakim prawdopodobieństwem zostanie przyznana odpowiadająca jej liczba punktów przy danym poziomie umiejętności ucznia. Punkty przecięcia się kolejnych krzywych prawdopodobieństwa odpowiadają progom wyznaczonym dla każdej pary sąsiednich liczb punktów.

⁷ J. M. Linacre, *Guidelines for Rating Scales*, MESA Research Note #2, <http://www.rasch.org/rn2.htm>; T. G. Bond, Ch. M. Fox, *Applying The Rasch Model. Fundamental Measurement in the Human Science*, Lawrence Erlbaum Associates, Publishers Mahaw, New Jersey 2001, s.158 – 172; J. M. Linacre, *Optimizing Rating Scale Category Effectiveness*, [w:] *Introduction to Rasch Measurement. Theory, Models and Applications*, red. E. V. Smith, R. M. Smith, , Jam Press, Maple Grove, Minesota 2004, s.258-278.

Tabela 2. Analiza punktacji

Zadanie	Liczba punktów	Liczebność	Odsetek	Przeciętny poziom umiejętności	Oczekiwany poziom umiejętności	Próg
Z1	0	436	33%	-0,22	-0,89	
	1	189	14%	-0,28*	-0,4	0,11
	2	86	6%	0,16	-0,07	0,51
	3	97	7%	-0,25*	0,2	-0,1
	4	138	10%	0,23	0,49	-0,07
	5	383	29%	0,25	0,92	-0,44
Z2	0	281	18%	-0,24	-1,25	
	1	308	19%	-0,09	-0,51	-0,86
	2	231	14%	-0,16*	-0,06	0,05
	3	180	11%	0,04	0,26	0,38
	4	205	13%	0,01*	0,54	0,31
	5	393	25%	0,28	0,89	0,11
Z3	0	198	14%	0,09	-1,07	
	1	179	13%	0,21	-0,39	-0,6
	2	258	19%	0,23	0,05	-0,52
	3	148	11%	0,06*	0,39	0,78
	4	160	12%	0,28	0,69	0,46
	5	434	32%	0,43	1,06	-0,13
Z4	0	395	29%	0,19	-0,62	
	1	68	5%	-0,10*	-0,02	1,42
	2	230	17%	0,39	0,45	-1,02
	3	650	48%	0,46	0,93	-0,4
Z5	0	282	20%	-0,82	-1,72	
	1	246	18%	-0,82*	-1,04	-1,29
	2	554	40%	-0,69	-0,58	-1,64
	3	180	13%	-0,72*	-0,22	0,71
	4	45	3%	-0,6	0,09	1,3
	5	20	1%	-0,82*	0,38	1,01
	6	2	0%	-1,02*	0,67	2,76
	7	71	5%	-0,35	1,07	-2,84

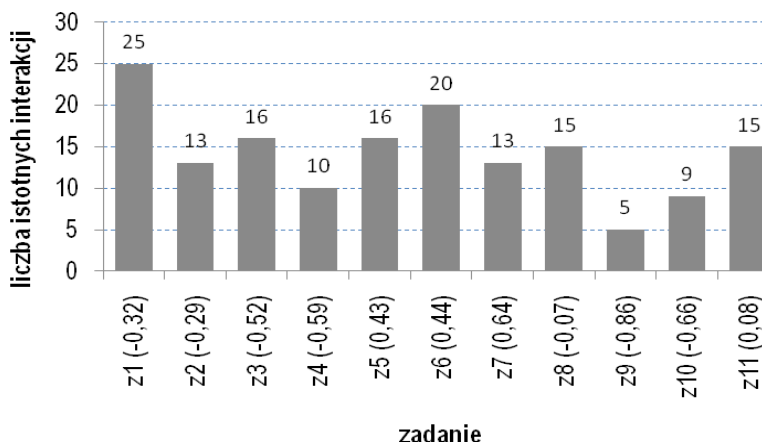
Zadanie	Liczba punktów	Liczebność	Odsetek	Przeciętny poziom umiejętności	Oczekiwany poziom umiejętności	Próg
Z6	0	881	62%	-0,84	-1,17	
	1	157	11%	-0,55	-0,62	0,85
	2	75	5%	-0,60*	-0,27	0,29
	3	54	4%	-0,74*	0,02	0,2
	4	245	17%	-0,58	0,39	-1,34
Z7	0	918	60%	-1,1	-1,46	
	1	151	10%	-0,61	-0,89	0,68
	2	89	6%	-1,27*	-0,58	-0,19
	3	58	4%	-0,89	-0,36	-0,04
	4	64	4%	-0,6	-0,17	-0,36
	5	68	4%	-0,94*	0,02	-0,13
	6	61	4%	-0,71	0,22	0,23
Z8	0	592	42%	-0,35	-0,95	
	1	226	16%	-0,16	-0,27	0,42
	2	113	8%	-0,26*	0,15	0,66
	3	490	34%	-0,1	0,59	-1,08
Z9	0	980	69%	-1,23	-1,49	
	1	102	7%	-1,16	-0,84	1,14
	2	201	14%	-0,85	-0,41	-1,29
	3	146	10%	-0,81	0,08	0,14
Z10	0	144	10%	0,28	-0,98	
	1	348	25%	0,34	-0,12	-1,31
	2	252	18%	0,45	0,4	0,53
	3	50	4%	0,43*	0,74	2,25
	4	586	42%	0,5	1,08	-1,46
Z11	0	277	21%	-0,39	-1,24	
	1	691	51%	-0,35	-0,45	-1,7
	2	64	5%	-0,56*	0	2,18
	3	98	7%	0,06	0,35	-0,24
	4	218	16%	-0,34*	0,76	-0,24

Kryterium wzrostu wartości progu wraz ze wzrostem liczby przyznanych punktów nie było spełnione dla żadnego z zadań z wyjątkiem zadania 9. (por. kolumna *Próg* w tabeli 2.). Oznaczało to, że tylko dla tego zadania przyznawane

liczby punktów były dodatkowo skorelowane z poziomem umiejętności uczniów. Natomiast w przypadku zadań „z zaburzeniami” wystąpiły liczby punktów, których przyznanie nie było nigdy najbardziej prawdopodobne, bez względu na poziom umiejętności ucznia.

Analiza interakcji egzaminator-zadanie

W ostatnim etapie przeprowadzono analizę obciążenia wyników, czyli zbadano, czy określone egzaminatory mają „predyspozycje” do punktowania w specyficzny sposób (np. zawsze bardziej surowo lub łagodniej niż wynikałoby to z założeń modelu) określonych zadań. W tym celu sprawdzono, czy występują statystycznie istotne ($\alpha = 0,05$) interakcje między poziomem surowości oceniania i poziomem trudności zadania. Okazało się, że interakcje takie wystąpiły dla 68 z 81 egzaminatorów, przy czym maksymalna liczba takich interakcji wyniosła sześć i dotyczyła dwójki egzaminatorów. Sprawdzono również, jakich zadań dotyczyły te interakcje. Wyniki tej analizy zilustrowano na wykresie 2.



Wykres 2. Zadania, dla których wystąpiły statystycznie istotne interakcje.

*w nawiasach podano poziomy trudności zadań (w logitach)

Statystycznie istotne interakcje najczęściej dotyczyły zadania 1. (31,2% wszystkich obciążeń) oraz zadania 6. (12,7%) i następnie kolejno zadań 3. i 4. (10,2%) oraz 8. i 11. (9,6%). Natomiast zdecydowanie najrzadziej dotyczyły zadania 9. (3,2% wszystkich obciążeń). Taki wynik jest zgodny z wnioskami płynącymi z wcześniejszych analiz – zadanie 9. jest najmniej problematycznym zadaniem.

Zakończenie

Przeprowadzone analizy pokazały, że skale punktowania poszczególnych zadań nie funkcjonują właściwie. W przypadku niemal każdego zadania występowały liczby punktów (kategorie punktowe), które przyznawane były

bardzo rzadko. Ponadto w przypadku żadnego z zadań nie stwierdzono, aby wyższa liczba przyznanych za odpowiednie zadanie punktów odpowiadała wyższemu przeciętnemu poziomowi umiejętności.

Analogiczne analizy przeprowadzone na całej populacji prac (około 43 tys.) potwierdziły powyżej sformułowane wnioski. Jest to tym ważniejsze, że badając całą populację uczniów, badamy całe spektrum poziomu umiejętności. Zatem tym bardziej należy oczekiwać, że wyższa liczba przyznanych za odpowiednie zadanie punktów będzie odpowiadać wyższemu przeciętnemu poziomowi umiejętności.

Wydaje się, że aby uniknąć takich nieprawidłowości, niezbędne jest podjęcie prac nad jakościową analizą schematu oceniania oraz wprowadzenie analizy funkcjonowania skali punktowania zadań już na etapie standaryzacji (próbne go zastosowania) testu.

Załącznik – Wskazówki⁸

1. Liczba obserwacji przypadająca na daną kategorię punktów powinna wynosić przynajmniej 10 – jest to niezbędne w celu uzyskania dobrej jakości oszacowań dla progów F_k . W przypadku niskiej liczby odpowiedzi przypadających na daną kategorię odpowiedzi dla wszystkich pozycji skali razem istnieje niebezpieczeństwo uzyskania nieprecyzyjnych i niestabilnych oszacowań progów. Gdy ten warunek nie jest spełniony, zaleca się połączenie sąsiednich kategorii lub potraktowanie pewnych obserwacji jako braków danych.

2. Kształt rozkładu – sytuacją optymalną jest rozkład jednostajny. Dopuszcza się jednak również rozkłady jednomodalne symetryczne, jednomodalne z maksimum dla jednej z kategorii skrajnych i rozkłady bimodalne z maksimumami dla skrajnych kategorii. Za najbardziej problematyczne uważa się rozkłady wielomodalne, a także te z długimi ogonami odpowiadającymi stosunkowo rzadko wybieranym liczbom punktów. Należy jednak te wskazówki traktować ostrożnie, zawsze mając na uwadze charakter badanego zjawiska.

W przypadku wystąpienia rozkładów „problematycznych” zalecane jest łączenie sąsiednich kategorii punktów.

3. Wzrost wartości średnich umiejętności wraz ze wzrostem liczby punktów – średni poziom umiejętności obliczany jest dla wszystkich uczniów, którzy uzyskali daną liczbę punktów. Ponieważ większa liczba punktów powinna odpowiadać wyższemu poziomowi umiejętności, wartości średnich poziomów umiejętności powinny monotonicznie wzrastać. W przypadku zaburzeń zalecane jest opuszczenie problematycznych kategorii lub połączenie ich z kategoriami bezpośrednio je poprzedzającymi.

4. Wartość statystyki dopasowania OMS powinna być mniejsza niż 2,0 – wartości powyżej 2 wskazują, że dana liczba punktów wnosi więcej zaburzeń niż informacji. Zaleca się wtedy wnikliwe przyjrzenie się danej kategorii punktów,

⁸ Tekst załącznika pochodzi z D. Węziak, *Zastosowanie porządkowego skalowania Rascha do optymalizacji długości skali odpowiedzi*, [w:] *Ilościowe i jakościowe metody badania rynku. Pomiar i jego skuteczność*, (red.) J. Garczarczyk, Wydawnictwo Naukowe Akademii Ekonomicznej w Poznaniu, Poznań 2006, s.139-141.

a później ewentualnie jej opuszczenie, połączenie z kategorią sąsiednią lub rozdzielenie na dwie nowe kategorie. Należy pamiętać, że kategorie skrajne są bardziej narażone na występowanie OMS przekraczającego 2 niż kategorie centralne, dlatego decyzje na temat ich usunięcia lub połączenia trzeba podejmować ze szczególną rozważą.

5. Progi F_k powinny wzrastać wraz ze wzrostem liczby punktów – wraz ze wzrostem poziomu umiejętności uzyskanie większej liczby punktów powinno być coraz bardziej prawdopodobne, czyli krzywe prawdopodobieństw powinny wyglądać jak „pasma górskie”. Zaburzenie monotoniczności wskazuje na różnice w rozumieniu przez egzaminatorów odpowiednich kategorii punktowych. Zalecane jest w takim przypadku przekoderowanie odpowiedzi lub połączenie kategorii.

6. Wartości progów F_k powinny wzrastać o przynajmniej 1,4 logita. Wzrost wartości progów o przynajmniej 1,4 logita pozwala skalę punktacji o m kategoriach traktować jako $m-1$ podskal o punktacji 0-1. Wskazówka (6) ma charakter bardzo ogólny, ponieważ wraz ze wzrostem liczby kategorii odpowiedzi maleje wymagany minimalny poziom różnic między progami pozwalający na utożsamianie skali o m kategoriach odpowiedzi z serią $m-1$ pozycji dychotomicznych o odpowiedziach tak/nie. Przykładowo dla skali o trzech kategoriach odpowiedzi minimalny poziom różnic między progami wynosi 1,4 logita, a dla skali o pięciu już 1,0 logit.

Bardziej szczegółowo problem przedstawił H. Huynh (1994, s. 111-119). Dla skal o trzech, czterech i pięciu kategoriach odpowiedzi odległości między kolejnymi progami F_k definiuje wzór:

$$F_{k+1} - F_k \geq \ln \left[1 + \frac{n+1}{k(n-k)} \right]$$

gdzie: k – numer progów, n – liczba wszystkich progów.

Wystąpienie wartości mniejszych od zalecanych sugeruje, że kategorie są zbyt wąskie. Zalecane rozwiązanie to połączenie kategorii.

7. Wartości progów F_k powinny wzrastać o co najwyżej 5,0 logitów. Zbyt duże różnice występujące między progami wskazują na zbyt szeroko zdefiniowaną kategorię punktów i pociągają za sobą zmniejszenie precyzji pomiaru. W takiej sytuacji sugeruje się zdefiniowanie większej liczby punktów, co jest tożsame z podziałem najszerszej lub najszerszych kategorii na dwie węższe.