

Ewa Stożek

Okręgowa Komisja Egzaminacyjna w Łodzi

Zadania „rozrywające” w testach

Na podstawie analizy danych empirycznych ze sprawdzianu 2004 i 2006 roku wyodrębniono zadania „odpowiedzialne” za dwumodalność rozkładu wyników tych testów. Takie zadania nazwano zadaniami „rozrywającymi”. Podjęto próbę opisu cech charakterystycznych tych zadań oraz określenia ich parametrów.

1. Co to jest zadanie „rozrywające”?

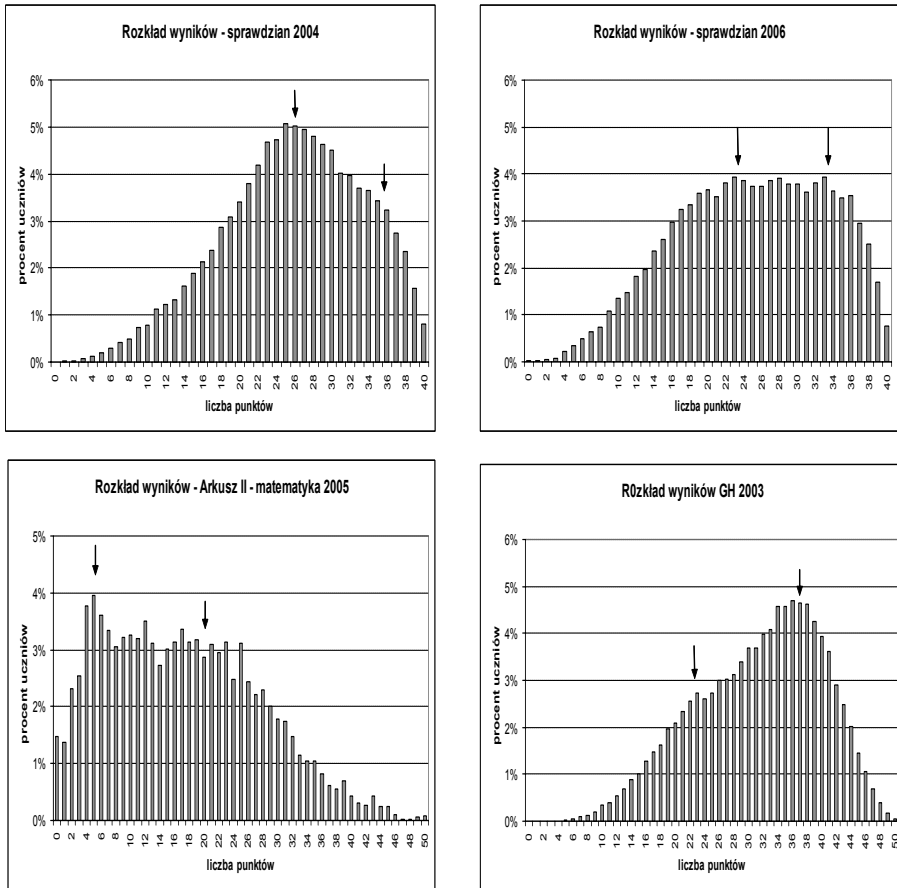
Rozkłady wyników z egzaminów zewnętrznych przeprowadzanych w Polsce po każdym etapie kształcenia rzadko kiedy są rozkładami normalnymi. Co więcej, często są to rozkłady z wieloma dominantami (rozkłady wielomodalne). Taki kształt rozkładu sugeruje duże zróżnicowanie populacji zdających. O ile na egzaminie maturalnym na poziomie rozszerzonym jest to zrozumiałe, to na sprawdzianie, który z założenia miał być testem sprawdzającym, a nie różnicującym, taki charakter wyników może budzić niepokój.

Zróżnicowanie wyników wewnątrz populacji może być następstwem wpływu różnych czynników, o różnej sile oddziaływania, np.:

1. zróżnicowanie populacji ze względu na lokalizację szkoły (miasto-wieś),
2. zróżnicowanie populacji ze względu na rodzaj szkoły (publiczna-niepubliczna),
3. zróżnicowanie populacji ze względu na typ szkoły ponadgimnazjalnej (LO, LP na egzaminie maturalnym),
4. zróżnicowanie wynikające z cech zadań występujących w teście.

Ostatni z wymienionych czynników: pojedyncze zadanie w teście, odpowiedzi na które tak silnie różnicują populację, że prowadzi to do rozkładu wyników o dwóch modach, będziemy nazywać **zadaniem „rozrywającym”**.

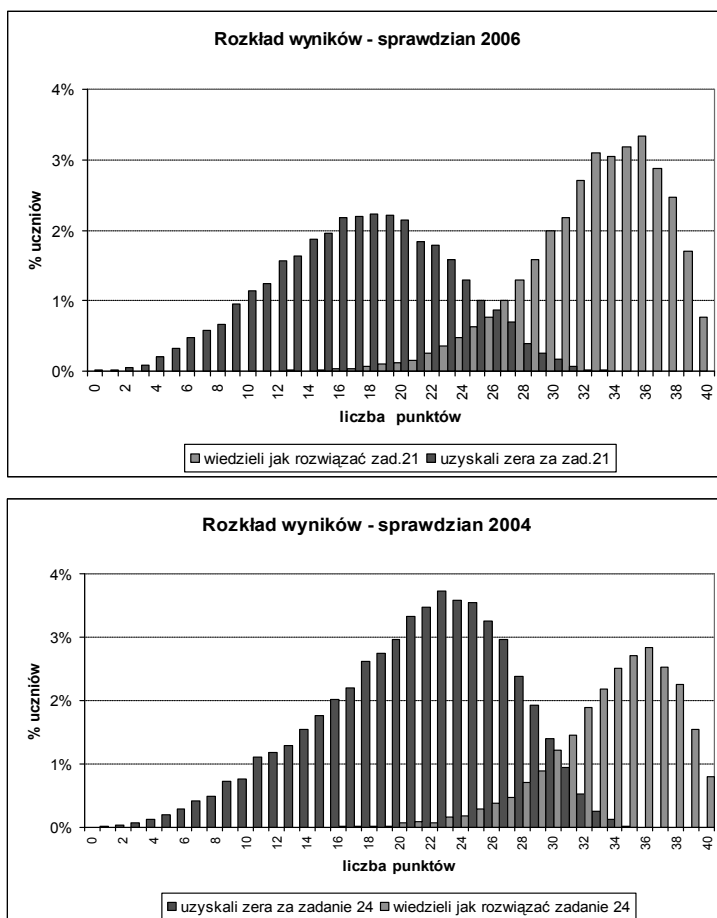
Na rysunku 1. przedstawiono przykłady dwumodalnych rozkładów wyników sprawdzianu 2004, sprawdzianu 2006, egzaminu gimnazjalnym części humanistycznej 2003 oraz wielomodalnego rozkładu wyników egzaminu maturalnego (matematyka, Arkusz II) 2005.



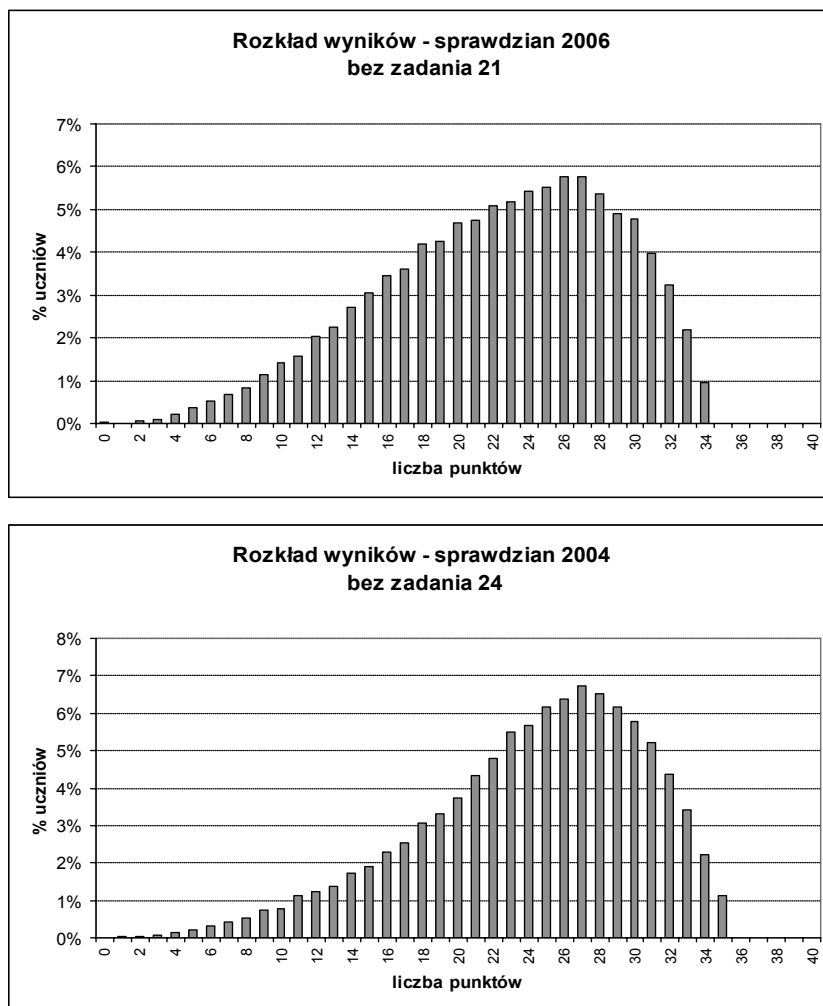
Rys. 1. Przykłady wielomodalnych rozkładów wyników na różnych egzaminach zewnętrznych w latach 2003-2006. Dane OKE Łódź.

Czy rzeczywiście jedno zadanie może powodować powstawanie drugiej mody w rozkładzie wyników? Na sprawdzianie w 2004 i w 2006 roku takimi zadaniami okazały się zadania matematyczne – zadanie 24. oraz zadanie 21. odpowiednio.

Rysunki poniżej pokazują, jak zadanie matematyczne podzieliło populację uczniów: lewy rozkład to wyniki uczniów, którzy uzyskali zero punktów za zadanie, a prawy rozkład – wyniki uczniów, którzy wiedzieli, jak rozwiązać zadanie (uzyskali punkty za metodę, być może stracili punkty za błędy rachunkowe). Duża odległość między wierzchołkami (modami) tych rozkładów oraz podobne liczebności podgrup są przyczyną powstawania dwumodalnego rozkładu wyników.



Rys.2. Zadanie matematyczne podzieliło populację zdających na sprawdzianie 2004 i 2006.



Rys.3. Rozkłady wyników na sprawdzianie 2006 i 2004 bez zadań matematycznych.

Zwróćmy uwagę na fakt, że rozkłady wyników sprawdzianu bez tych zadań są rozkładami z jedną modą (rys.3), zatem uzasadnione jest twierdzenie, że za dwumodalność rozkładów wyników są odpowiedzialne właśnie te zadania. Porównajmy podstawowe parametry tych zadań: ¹

¹ Opis zadań ze względu na łatwość i moc różnicującą za: D.Sołtys, M.K.Szmigiel, *Doskonalenie kompetencji nauczycieli w zakresie diagnostyki edukacyjnej*, Zamiast korepetycji, Kraków 1997, s.64.

	łatwość	moc różnicująca
zadanie 21. (2006)	0,42 (trudne)	0,81 (dobrze różnicuje)
zadanie 24. (2004)	0,31 (trudne)	0,72 (dobrze różnicuje)

Czy można na tej podstawie wnioskować, że wszystkie zadania trudne i dobrze różnicujące będą zadaniami „rozrywającymi”?

2. Krzywe charakterystyczne zadania

W teorii odpowiedzi na pozycje testu (*item response theory, IRT*) do opisu cech zadania stosuje się trójparametryczne krzywe charakterystyczne².

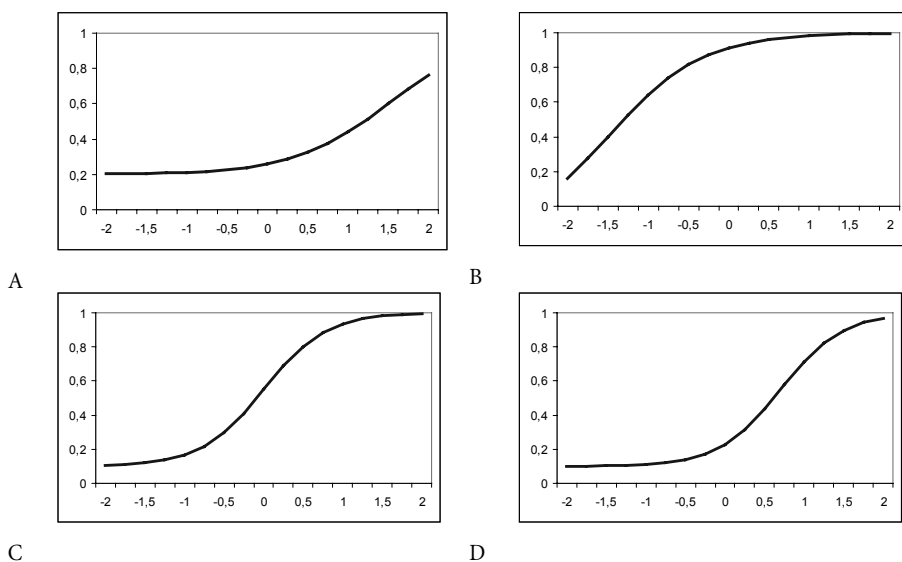
Krzywą charakterystyczną zadania opisuje trójparametryczna funkcja logistyczna postaci:

$$p(\theta) = c + \frac{1 - c}{1 + e^{-1,702 \cdot a(\theta - b)}}$$

gdzie (oznacza poziom pewnej cechy latentnej (np. poziomu osiągnięć, poziomu zdolności), p – prawdopodobieństwo udzielenia poprawnej odpowiedzi na zadanie, natomiast a , b , c są parametrami. Parametr a jest związany z mocą różnicującą zadania, a parametr b – z łatwością zadania. Parametr c w zadaniach wielokrotnego wyboru jest nazywany parametrem zgadywania i jest związany z podatnością zadania na zgadywanie. Geometryczna interpretacja tych parametrów jest następująca: im większe a , tym większy kąt nachylenia stycznej do krzywej logistycznej w punkcie przegięcia b , a $y = c$ jest asymptotą poziomą wykresu funkcji $p(\theta)$.

Zadanie z krzywą charakterystyczną A jest zadaniem trudnym, które nie różnicuje słabych uczniów, różnicuje natomiast najlepszych. Zadanie z krzywą charakterystyczną B jest zadaniem łatwym, które nie różnicuje najlepszych uczniów. Zadania z krzywymi charakterystycznymi C i D są zadaniami dzielącymi populację na dwie grupy, wewnątrz których nie różnicują. Im bardziej stroma będzie krzywa charakterystyczna i im bliżej punkt przegięcia będzie środkowej wartości cechy θ , tym wyraźniejszy będzie podział – zadanie będzie zadaniem „rozrywającym”.

² Hulin CH.L., Parsons C.K., *Wprowadzenie do teorii odpowiedzi na pozycje testu*, [w:] J.Brzeziński (red.) Trafność i rzetelność testów psychologicznych. Wybór tekstów. GWP, Gdańsk 2005, s.213-271.



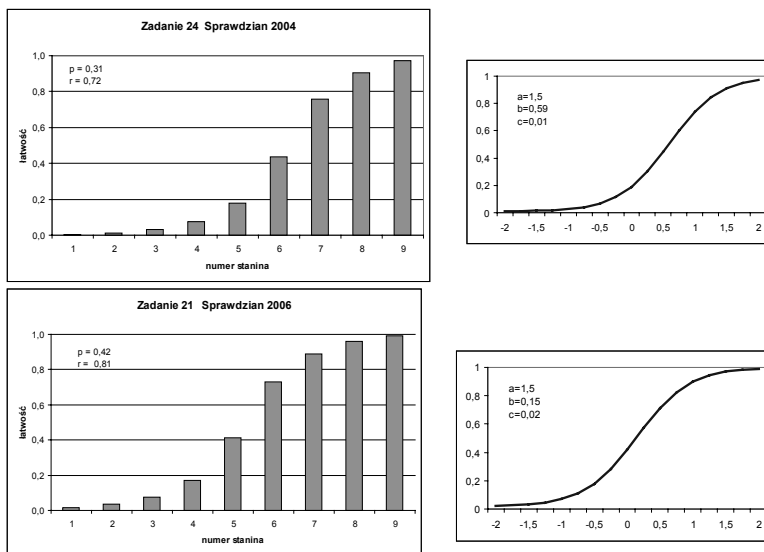
Rys.4. Przykłady krzywych charakterystycznych dla różnych zadań

3. Parametry zadań „rozrywających”

Spróbujmy wyrysować krzywe charakterystyczne dla zadań nr 21 i 24 ze sprawdzianu 2006 i 2004 odpowiednio.

Zrobimy to w nieco uproszczony sposób: obliczymy łatwości omawianych zadań dla grup osób, których wyniki znalazły się w kolejnych staninach. Kolejny stanin możemy traktować jako określony poziom umiejętności ucznia, a łatwość zadania, jako prawdopodobieństwo udzielenia poprawnej odpowiedzi. Krzywą charakterystyczną uzyskamy jako krzywą regresji nieliniowej (logistycznej) z trzema parametrami (w tym celu wykorzystamy dostępną w programie SPSS opcję regresji nieliniowej).

Krzywe charakterystyczne tych zadań wyraźnie wskazują na ich „rozrywający” charakter, w szczególności krzywa zadania 21. Na podstawie analizy danych empirycznych ze sprawdzianu (również z lat 2002-2005) można próbować wnioskować, że zadanie będzie zadaniem „rozrywającym”, jeśli parametr $a \geq 1,5$ i jednocześnie $b \in (-0,6; 0,6)$. Bardziej szczegółowe analizy większej ilości materiału empirycznego (szczególnie testów maturalnych) mogą przynieść wiele nowych informacji o charakterze zadań „rozrywających”. Charakterystyka zadania poprzez analizowanie parametrów krzywej charakterystycznej daje więcej informacji o zadaniu niż analiza współczynników łatwości i mocy różnicującej.



Rys.5. Krzywe charakterystyczne dla zadań 21. i 24.

Pytania na zakończenie:

1. Czy zadania „rozrywające” są zadaniami pożądanymi w teście, szczególnie w teście o charakterze sprawdzającym?
2. Jak odróżnić zadania, które są zadaniami „rozrywającymi” na skutek swojej wadliwej konstrukcji lub nieodpowiedniego schematu punktowania od zadań w sposób zamierzony różnicujących populację?
3. Czy na etapie konstrukcji i standaryzacji testu należałoby analizować krzywe charakterystyczne zadań? Jak duża wobec tego powinna być próba standaryzacyjna?