

Sławomir Sapanowski

Okręgowa Komisja Egzaminacyjna w Łodzi

Różnicowanie uczniów o tych samych osiągnięciach egzaminacyjnych

Wydaje się oczywiste, że wynik¹ wyższy uzyskany na egzaminie jest bardziej wartościowy od wyniku niskiego. Ale postawmy odwrotne pytanie: czy ten sam wynik u dwóch różnych zdających ma tę samą wartość?

a) wynik a osiągnięcie

Wyniki uzyskane w ramach testowania uczniów powinny być uzyskane za pomocą zadań o wysokiej trafności², a test powinien cechować się wysoką rzetelnością³ i co za tym idzie małym błędem pomiaru (Sapanowski, 2013). Jeśli te warunki są spełnione, możemy z dużą dozą prawdopodobieństwa analizować, które treści programowe nie stanowiły dla zdających większego problemu i zostały opanowane na poziomie zadowalającym⁴, a które obszary wiedzy są dla uczniów niczym *terra incognita*. Wymienione zagadnienia są oczywiście bardzo istotne w procesie dydaktycznym, ale chcemy tutaj uniknąć domeny szkolnej i nauczycielskiej i powtarzamy pytanie: czy identyczne wyniki punktowe oznaczają taki sam poziom osiągnięć?

Odpowiedź, wbrew pozorom, nie jest zbyt oczywista. Wynik to suma uzyskanych punktów za wykonanie pewnych określonych czynności. A przecież zadania, które stawiamy przed testowanym uczniem, mają różny poziom trudności. Absurdem byłoby żądać od geniusza matematycznego umiejętności wbijania gwoździ i podobnie – zlecić rozwiązanie równania różniczkowego sprzątaczkę (oczywiście z całym szacunkiem dla pracy sprzątaczk). Oczekujemy od tych osób rozwiązywania problemów na zupełnie różnych poziomach. Cóż z tego, że potrafią wykonywać podobną liczbę zadań, skoro są to rozłączne obszary działalności człowieka? Wynik punktowy (czytaj: liczba rozwiązanych zadań) jest taki sam, ale osiągnięcia różne.

Posłużmy się innym przykładem, może nie tak abstrakcyjnym. Załóżmy, że chcemy przyjąć kogoś do pracy. W celu wybrania jednego z dwóch kandydatów zlecamy im wykonanie dwóch zadań. Okazuje się, że obaj wykonali tylko po jednym, nie radząc sobie z drugim. Wiemy, że postawione przed kandydatami prace miały różną trudność. Powstaje pytanie: którego kandydata zatrudnić?

¹ Przez wynik rozumiemy liczbę uzyskanych punktów.

² Trafność (ang. *validity*) to cecha testu, która oznacza, że test mierzy to, co chcemy (opierając się na planie testu) zmierzyć.

³ Rzetelność to, mówiąc inaczej, powtarzalność wyników testowania. Test o wysokiej rzetelności stosowany wielokrotnie daje podobne rezultaty. Rzetelność wyraża się liczbą z przedziału (0;1). Istnieje co najmniej kilka sposobów estymowania wskaźnika rzetelności.

⁴ Zadowalający poziom osiągnięć przyjmujemy w sposób arbitralny.

Czy tego, który wykonał zadanie trudne, a nie podjął się wykonania zadania łatwiejszego? Czy może odwrotnie? Przyjmiemy tę osobę, która proste zlecenie przeprowadziła bezbłędnie, a trudne ją przerosło? Z pewnym wahaniem wybieramy kandydata, który skutecznie zrealizował trudniejsze zadanie.

Podsumowując powyższe rozważania, należy stwierdzić, że wynik wynikowi nierówny, nawet jeśli jest taki sam!

b) różnicowanie uczniów o tych samych wynikach

Wyniki egzaminacyjne wyrażone w skali punktowej (lub procentowej) powodują, co zresztą jest nieuniknione⁵, że pojawiają się bardzo liczne grupy uczniów legitymujących się tym samym rezultatem. Intuicyjnie czujemy jednak, że w ramach tych samych osiągnięć egzaminacyjnych uczniowie mogą się znacznie różnić pod względem opanowania umiejętności matematycznych.

Proponowany niżej model zróżnicowania wyników nie jest nowy. Podobny system zastosowano w „Lidze zadaniowej – Klub 44” prezentowanej na łamach miesięcznika matematyczno-fizyczno-astronomicznego „Delta”. Fragment regulaminu ligi zadaniowej w zupełności wyjaśnia zasady przyznawania oceny dla danego rozwiązania. Ocenę mnożymy przez współczynnik trudności danego zadania: $WT = 4 - 3S/N$, gdzie S oznacza sumę ocen za rozwiązania tego zadania, a N – liczbę osób, które nadesłały rozwiązanie choćby jednego zadania z danego numeru [...] – i tyle punktów otrzymuje nadsyłający.

Proponowana poniżej analiza oparta była na wynikach osób przystępujących w ramach projektu „E-matura” do egzaminu z matematyki na poziomie podstawowym. W projekcie organizowanym przez Politechnikę Łódzką uczestniczyło ponad 2500 uczniów.

Ocena zadania w poniższym przykładzie jest silnie uzależniona od jego trudności. Za rozwiązanie zadania trudnego zyskuje się więcej niż za podobnie punktowane zadanie łatwe. Pomnożmy wynik punktowy dla danego zadania przez jego trudność, a następnie policzmy sumę iloczynów zgodnie ze wzorem:

$$WZ = \sum_i^n x_i \cdot q_i$$

gdzie: WZ – wynik zróżnicowany, x_i – ocena i -tego zadania, q_i – trudność i -tego zadania wyrażona liczbą od 0 do 1.

Efektem tego zabiegu jest skala wyników, która obejmuje zakres od 0 do liczby nieznacznie przekraczającej 23. W porównaniu z pierwotną skalą punktową jest ona zdecydowanie krótsza, ale dlatego, że obecnie wynik ucznia może przyjmować prawie dowolne wartości z przedziału $\langle 0; 23,18 \rangle$, prezentowana skala jawi nam się jako pseudociągła. Na 2508 uczestników e-matury otrzymujemy 2466 unikatowych wyników zróżnicowanych (WZ). Inaczej mówiąc, wśród zdających wystąpiło 2466 różnych wzorców uzyskanych punktów⁶.

⁵ Wynika to z tego, że komunikowanie wyników odbywa się na skali dyskretnej.

⁶ Uczniowie ze skrajnymi wynikami (0 i 50) nie będą i nie mogą być różnicowani. Ich wzorce odpowiedzi są dokładnie takie same.

Praktycznie każdemu uczniowi możemy przypisać unikalny wynik powiązany ze wzorcem udzielonych poprawnych rozwiązań. W poprzedniej skali, gdzie wynik uzyskany był sumą punktów za wszystkie zadania, różnych, unikatowych wyników mogliśmy odnotować jedynie 47⁷.

Weźmy pod uwagę wyniki dwóch uczniów, którzy przystąpili do e-matury, i ograniczmy się tylko do siedmiu pierwszych zadań. Obaj w tym zakresie uzyskali po 3 pkt (rozwiązali poprawnie trzy zadania oceniane dychotomicznie). Ich wyniki zestawiono w poniższej tabeli.

Numer zadania	1	2	3	4	5	6	7	
Trudność zadania	0,02	0,19	0,11	0,36	0,59	0,25	0,05	suma
Uczeń A	0	0	1	0	1	0	1	3
Uczeń B	1	0	1	0	0	1	0	3

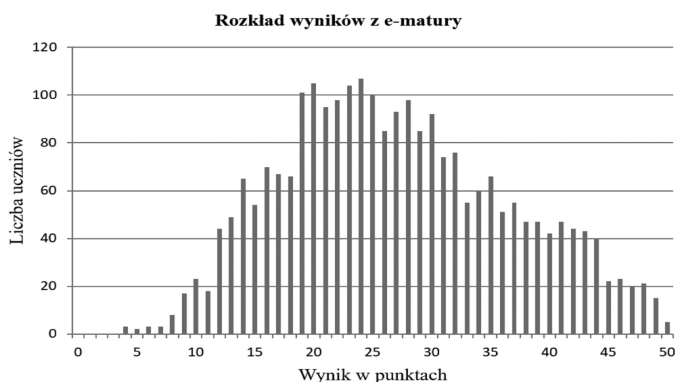
Po wymnożeniu wyniku punktowego przez trudność zadania otrzymujemy:

Numer zadania	Wyniki zróżnicowane							suma
	1	2	3	4	5	6	7	
Uczeń A	0	0	0,11	0	0,59	0	0,05	0,75
Uczeń B	0,02	0	0,11	0	0	0,25	0	0,38

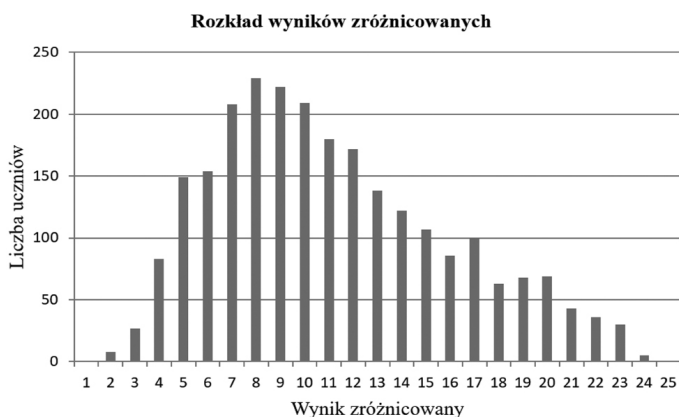
„Wartość” wyniku ucznia A jest równa 0,75 i wyższa od „wartości” wyniku ucznia B (0,38).

c) rozkład wyników zróżnicowanych

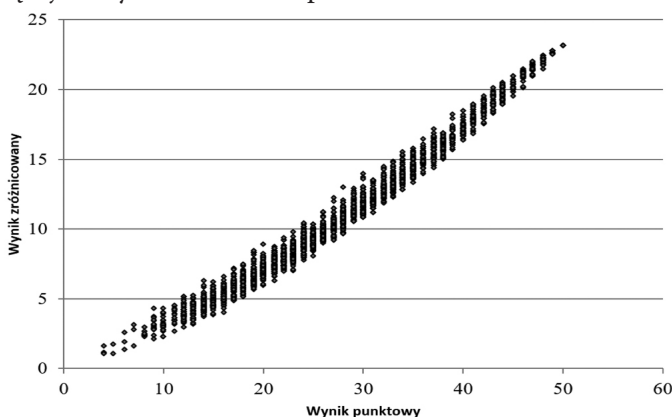
Rozkład wyników punktowych jest nieregularny. Cechuje się prawoskośnością i ujemną kurtozą, w związku z czym nie możemy mówić o rozkładzie normalnym. A to stało się, o czym już wspomniano, prawidłowością w egzaminach zewnętrznych z matematyki. Rozkład wyników zróżnicowanych podlega dokładnie tym samym ograniczeniom (dodatnia skośność i ujemna kurtoza). Tak więc przekształcenie wyników na zróżnicowane nie zakłóca rozkładu ani radykalnie go nie zmienia.



⁷ Właściwie możliwe jest uzyskanie 51 różnych wyników (od 0 do 50), ale ze względu na to, że nikt nie uzyskał rezultatu z zakresu 0–3 pkt, skala pomiarowa skurczyła się do zakresu 4–50 pkt.



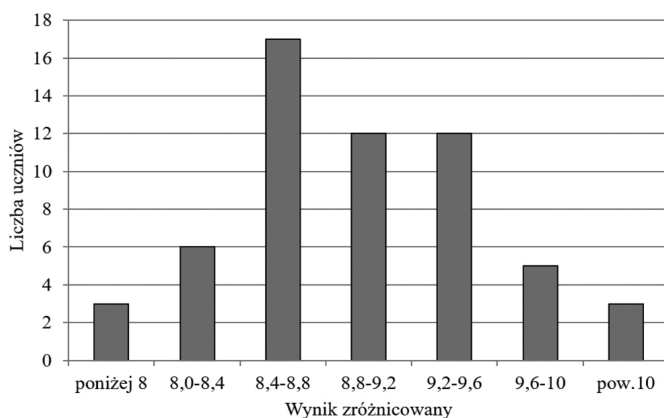
Na kolejnym wykresie przedstawiono wykres rozrzutu, który obrazuje wielkość „rozmycia”, a może lepiej zróżnicowania w grupach zdających o takim samym wyniku punktowym. Zróżnicowanie to jest największe w środku skali pomiarowej, co zresztą nie powinno dziwić, ponieważ dla tych wyników istnieje najwięcej różnych wzorców odpowiedzi.



Przyjrzyjmy się dokładniej uczniom, którzy uzyskali w czasie e-matury 24 punkty. Wybór taki podyktowany został największą liczebnością tej grupy (107 osób). Okazuje się, że rozkład wyników zróżnicowanych ma podobne cechy do rozkładu całej populacji.

Wyniki zróżnicowane⁸ mieszczą się w przedziale (7,77; 10,39), a ich miary położenia są zbliżone i odpowiednio równe – średnia 8,97 i mediana 8,92. Najciekawsze jest to, że miary skupienia, czyli skośność i kurtoza, są zbieżne z miarami skupienia dla całej badanej populacji. Rozkład jest prawoskośny, a kurtoza – ujemna.

⁸ Mówimy tutaj tylko i wyłącznie o grupie zdających, którzy uzyskali 24 pkt na egzaminie e-matura.



Czy ta prawidłowość powtarza się dla pozostałych kategorii wyników? Odpowiedź nie jest jednoznaczna. Dla 75% grup wyników skośność jest dodatnia, rozkład jest prawoskośny. A średnia skośność dla wszystkich kategorii wyniku równa się 0,28. Wydaje się, że taka zbieżność nie jest przypadkowa. Jednakże potwierdzenie tej hipotezy lub jej zaprzeczenie wymaga dodatkowych badań, w tym w szczególności na rozkładach lewoskośnych. Kurtosa zachowuje się jeszcze mniej przewidywalnie, a to dlatego, że dla różnych grup istnieją różne liczby możliwych wzorców odpowiedzi. W związku z tym kurtosa nie jest obliczana dla zmiennej o charakterze losowym, a dla zdeterminowanej przez wzorce udzielanych odpowiedzi.