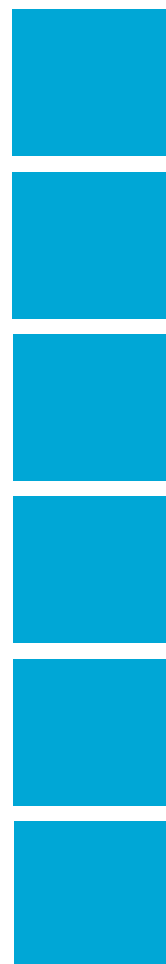


Podręcznik użytkownika zestawu testów osiągnięć szkolnych do klasy II i III szkoły podstawowej



Autor:
Iwona Pecyna

© Copyright by: Instytut Badań Edukacyjnych, Warszawa 2015

Wydawca:
Instytut Badań Edukacyjnych
ul. Górczewska 8
01-180 Warszawa
tel. (22) 241 71 00; www.ibe.edu.pl

Publikacja opracowana w ramach projektu systemowego: Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego współfinansowanego przez Unię Europejską ze środków Europejskiego Funduszu Społecznego, realizowanego przez Instytut Badań Edukacyjnych

Spis treści

WSTĘP	3
CZĘŚĆ 1. DLA NAUCZYCIELA	4
I.1. Charakterystyka testów osiągnięć szkolnych.....	4
Test świadomości językowej.....	4
Test umiejętności czytania	6
Test umiejętności matematycznych.....	8
I.2. Przeprowadzanie testowania.....	13
I.3. Interpretacja wyników.....	14
CZĘŚĆ 2. DLA BADACZA.....	17
II.1. Procedura tworzenia testów osiągnięć szkolnych	17
Badanie pilotażowe	17
II.2. Teoria odpowiedzi na zadanie testowe	19
Podstawy teoretyczne	19
Statystyki INFIT i OUTFIT	20
Krzywa charakterystyczna zadania.....	20
II.3. Analiza psychometryczna zadań	23
II.4. Rzetelność skal pomiarowych	29
Literatura:.....	31
ANEKS 2. TESTY OSIĄGNIĘĆ SZKOLNYCH	

WSTĘP

Edukacja wczesnoszkolna jest powszechnie uważana za bardzo ważny etap w życiu młodego człowieka. Ma ona znaczący wpływ na przyszłe jego sukcesy edukacyjne. Dlatego tak ważna jest możliwość rzetelnej oceny umiejętności ucznia w pierwszych latach jego nauki. Brakuje jednak testów, które mogłyby pomóc w tej sferze, a które charakteryzowałyby się odpowiednią jakością i posiadałyby szczegółową dokumentację potwierdzającą ich trafność. System egzaminów zewnętrznych poprzez przeprowadzany corocznie od 2002 roku *Sprawdzian w szóstej klasie* dostarcza pewnych informacji o osiągnięciach uczniów, jednak dopiero po sześciu latach nauki w szkole podstawowej.

Obserwujemy obecnie wzrost zainteresowania badaczy, jak i co może ważniejsze, samych szkół, możliwością oceny efektywności nauczania w pierwszych latach nauki w szkole podstawowej. Przejawem tego jest m.in. fakt, że w nieobowiązkowym Ogólnopolskim Badaniu Umiejętności Trzecioklasistów w 2013 roku wzięło udział 8837 szkół, tj. ponad 70% wszystkich szkół podstawowych w Polsce (Pregler, A., 2013), a rok później ponad 7600 szkół (Karpiński, M., Nowakowska, A., Orzechowska, M., Sosulska, D., Zambrowska, M., 2014). Wychodząc naprzeciw tym potrzebom, podjęto działania mające na celu przygotowanie ogólnodostępnych narzędzi do pomiaru umiejętności uczniów szkół podstawowych klas drugich i trzecich.

W 2009 roku w ramach projektu „Badania dotyczące rozwoju metodologii szacowania wskaźnika edukacyjnej wartości dodanej (EWD)” rozpoczęło się podłużne *Badanie uwarunkowań wyników nauczania w szkołach podstawowych*. W ramach tego badania zostały zebrane i poddane standaryzacji zadania, które następnie posłużyły do przygotowania trzech testów, które razem z niniejszym podręcznikiem Państwu udostępniamy.

Podręcznik stanowi dokumentację zestawu trzech testów: testu świadomości językowej, testu umiejętności czytania oraz testu umiejętności matematycznych. Został on podzielony na dwie części: pierwszą, bardziej ogólną z niezbędnymi informacjami dotyczącymi testów i interpretacji wyników, drugą – zawierającą bardziej szczegółowe informacje dotyczące metodologii konstrukcji testów i analizę psychometryczną zadań.

Część 1. jest w szczególności dedykowana nauczycielom edukacji wczesnoszkolnej, którzy chcieliby wykorzystać przygotowane testy jako dodatkowe narzędzie umożliwiające ocenę umiejętności swoich uczniów. Część 2. została przygotowany z myślą o osobach zajmujących się badaniami w diagnostyce edukacyjnej oraz konstruowaniem testów.

Integralną częścią podręcznika są aneksy. Pierwszy aneks zawiera podręczniki kodowania służące do oceny rozwiązań uczniowskich. Aneks drugi zawiera arkusze testowe.

CZĘŚĆ 1. DLA NAUCZYCIELA

I.1. Charakterystyka testów osiągnięć szkolnych

Mają Państwo do dyspozycji trzy testy osiągnięć szkolnych: test świadomości językowej, test umiejętności czytania oraz test umiejętności matematycznych. Testy te zostały przygotowane z myślą o uczniach klas II i III szkół podstawowych. Powinny być one przeprowadzone na początku drugiego półrocza (marzec – pierwsza połowa kwietnia). Podany termin przeprowadzenia badania uczniów jest istotny i ściśle związany z datami realizacji badania pilotażowego, które służyło przygotowaniu testów i norm wykonania (szczegóły w części 2, str.17).

Testy osiągnięć szkolnych są testami standaryzowanymi, w skład których weszły zadania przetestowane na próbie uczniów. Poniżej znajduje się szczegółowy opis każdego z trzech testów. W tabelach pokazano, ile zadań z poszczególnych bloków mierzących szczegółowe umiejętności i grupy treści znalazło się w ich ostatecznej wersji. Różnorodność wykorzystanych zadań miała na celu zapewnienie trafności treściowej testów.

Test świadomości językowej

Test świadomości językowej składa się z zadań, które możemy podzielić na trzy bloki. Pierwszy z nich zawiera zadania mierzące bogactwo słownikowe uczniów, a więc zadania na tworzenie i rozpoznawanie wyrazów o podobnym lub przeciwnym znaczeniu, tworzenie porównań, wyjaśnianie znaczenia słów, dobieranie słów do podanej definicji, rozpoznawanie błędnego i wskazanie poprawnego użycia słowa oraz wyjaśnianie znaczenia związków frazeologicznych.

Przykład zadania - zadanie nr 3 w teście świadomości językowej.

3. Sympatia to:

- A. uczucie niepokoju.
- B. przyjazny stosunek do kogoś.
- C. obojętność i zniechęcenie.

Drugi blok to zadania mierzące umiejętność pisania krótkich tekstów (redagowanie tekstu). W przygotowanym teście występuje tylko jedno zadanie z tego bloku.

Przykład zadania – zadanie nr 12 w teście świadomości językowej.

12. Odpowiedz pełnym zdaniem na pytanie dotyczące obrazka. Zdanie powinno mieć co najmniej cztery wyrazy.



Co robi kot?

Trzeci blok zawiera zadania sprawdzające elementy wiedzy o języku, czyli zadania z zakresu ortografii i interpunkcji oraz poprawności językowej, jak również zadania mierzące umiejętność tworzenia i uzupełniania zdań zgodnie z zasadami składni.

Przykład zadania – zadanie nr 13 w teście świadomości językowej.

13. Ułóż zdanie z podanymi wyrazami.

Na przykład:

książka czytać	Ewa czyta książkę.
-------------------	--------------------

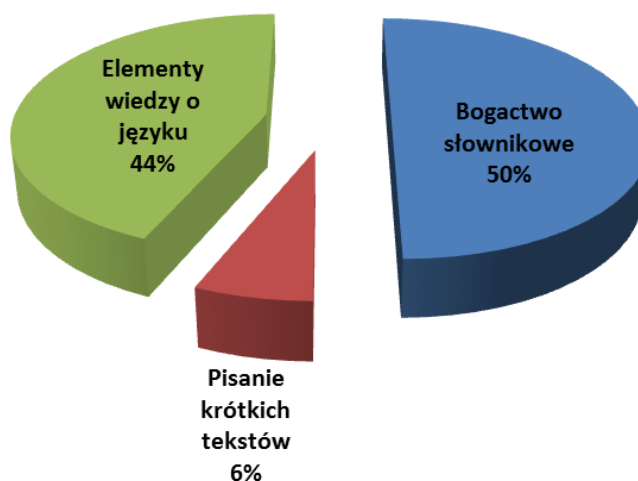
A teraz Ty:

Basia jezioro śpiewać	
-----------------------------	--

Plan testu świadomości językowej pokazujący, ile zadań mierzących poszczególne umiejętności weszło w skład testu, przedstawia poniższa tabela i wykres.

Tabela 1. Plan testu świadomości językowej.

Umiejętność ogólna	Umiejętność szczegółowa	Liczba zadań
Bogactwo słownikowe	Tworzenie synonimów lub antonimów	3
	Tworzenie wyrażen porównawczych	1
	Wyjaśnianie znaczenia słowa	1
	Rozpoznanie niepoprawnego użycia słowa (ze względu na znaczenie)	1
	Wyjaśnianie znaczenia związków frazeologicznych	2
Pisanie krótkich tekstów	Redagowanie tekstu	1
Elementy wiedzy o języku	Poprawność składniowa i gramatyczna zdań	3
	Ortografia i interpunkcja	2
	Rozpoznawanie i poprawianie błędów językowych	2



Rysunek 1. Podział zadań ze względu na sprawdzaną umiejętność ogólną.

Test umiejętności czytania

Test umiejętności czytania mierzy poziom rozumienia przez ucznia czytanych samodzielnie tekstów różnego typu: tekstu literackiego, popularnonaukowego oraz krótkiego tekstu sprawdzającego umiejętność czytania poleceń. Pytania do każdego tekstu sprawdzają różne kompetencje. Najwięcej zadań umieszczonych w teście mierzy umiejętność interpretacji, a więc umiejętność zrozumienia tekstu, wydobycia i wyjaśnienia jego sensu, połączenia wybranych faktów lub informacji, dostrzegania i wyjaśniania przyczyn i skutków opisanych zdarzeń, zjawisk, sytuacji, podania przypuszczalnych motywów działania, zachowania lub postawy bohaterów.

Przykład zadania – zadanie nr 8 w teście umiejętności czytania (do tekstu „Ptasznik i ptaki”).

8. Dlaczego czapla uważała, że postąpiła sprytnie, nie popisując się przed królem swoimi umiejętnościami? Napisz odpowiedź własnymi słowami.

Kolejna pula zadań to zadania wymagające od uczniów wyszukania informacji zawartej w tekście, podjęcia decyzji, które informacje są ważne, a które nie, oraz oceny czy podane informacje spełniają określone warunki.

Przykład zadania – zadanie nr 2 w teście umiejętności czytania (do tekstu „Ptasznik i ptaki”).

2. Znajdź i podkreśl w tekście fragment opisujący wygląd czapli.

Trzecią sprawdzaną kompetencją jest umiejętność refleksji i oceny tekstu. Zadania z tej grupy sprawdzają, czy uczeń potrafi odnieść tekst do własnego doświadczenia i wiedzy o świecie, czy potrafi dokonać oceny zdarzeń, postaci i poglądów w kontekście własnego doświadczenia czytelniczego i pozaszkolnego, jak również ocenić kompletność i spójność tekstu (w przygotowanym teście znajduje się jedno zadanie z tej grupy zadań).

Przykład zadania – zadanie nr 16 w teście umiejętności czytania (do tekstu o delfinach).

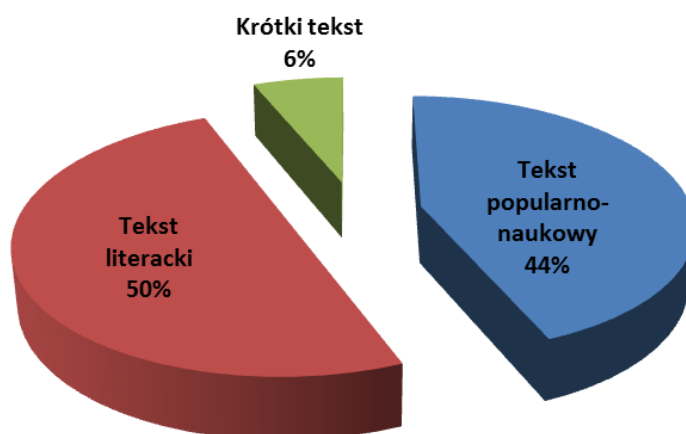
16. Jaką cechę można przypisać delfinom?

- A. Niezaradność
- B. Opiekuńczość
- C. Samolubność
- D. Płochliwość

Plan testu umiejętności czytania, pokazujący ile zadań mierzących poszczególne umiejętności weszło w skład testu, przedstawia poniższa tabela i wykres.

Tabela 2. Plan testu umiejętności czytania.

Rodzaj tekstu	Umiejętność	Liczba zadań
Literacki (proza)	Interpretacja	6
	Wyszukiwanie informacji	2
Popularnonaukowy	Interpretacja	3
	Wyszukiwanie informacji	3
	Refleksja i ocena	1
Krótki tekst	Czytanie poleceń	1



Rysunek 2. Podział zadań ze względu na przynależność do danego rodzaju tekstu.

Test umiejętności matematycznych

Test umiejętności matematycznych składa się z zadań, które można opisać za pomocą dwóch kategorii: umiejętności, którymi należy się posłużyć, by dane zadanie rozwiązać, oraz treści matematycznych, do których się ono odwołuje.

Umiejętności matematyczne

Wyróżniamy trzy umiejętności matematyczne. Pierwsza to odtwarzanie wiadomości i dobrze wyćwiczonych schematów, a więc umiejętność rozwiązywania zadań typowych, które wymagają użycia wyćwiczonych technik i posłużenia się dobrze znanymi, prostymi obiektami. W praktyce oznacza to korzystania przez ucznia z zapamiętanych przez niego pojęć lub schematów.

Przykład zadania – zadanie nr 10 w teście umiejętności matematycznych.

10. Opisz rysunek. Uzupełnij każde zdanie wyrazem *wyżej* lub *niżej*.



Trójkąt jest _____ niż kwadrat.

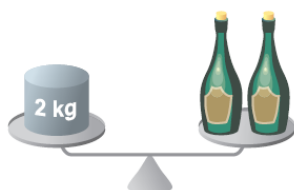
Koło jest _____ niż trójkąt.

Kwadrat jest _____ niż koło.

Drugą umiejętność polega na to wiązaniu różnych wiadomości i dobrze wyćwiczonych schematów na potrzeby rozwiązania zadań mniej rutynowych, ale niezbyt odległych od zadań typowych. Uczeń musi zwykle wykonać większą liczbą kroków, aby rozwiązać zadanie. Musi wybrać pojęcia i/lub procedury matematyczne odpowiednie dla rozwiązania danego problemu. Od ucznia oczekuje się, że będzie potrafił wykorzystać posiadane wiadomości do rozwiązania zadań, z których nie wynika wprost, jakie pojęcia, czy procedury należy zastosować.

Przykład zadania – zadanie nr 11 w teście umiejętności matematycznych.

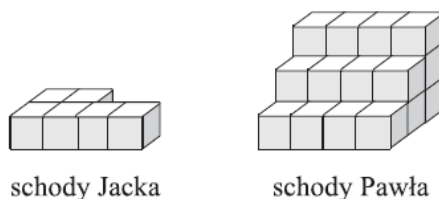
11. Na jednej szalce wagi stoi odważnik, a na drugiej dwie butelki. Waga jest w równowadze. Ile waży jedna butelka?



Trzecia grupa zadań sprawdza umiejętność przeprowadzenia prostego rozumowania matematycznego składającego się zwykle z kilku kroków. Od ucznia oczekuje się, że wykaże się twórczym podejściem do problemu, stworzy model lub wykona rozumowanie (np. uogólnienia). Zadania te wymagają ustalenia kolejności czynności prowadzących do rozwiązania problemu oraz wyciągania wniosków z kilku informacji podanych w różnej postaci.

Przykład zadania – zadanie nr 8 w teście umiejętności matematycznych.

8. Paweł i Jacek budowali takie same schody. Schody Pawła zostały skończone. Jacek jeszcze buduje. Ile klocków potrzebuje, żeby skończyć schody?



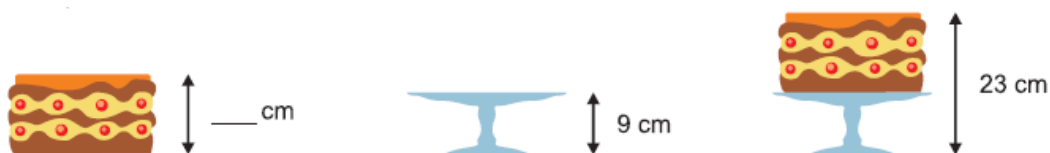
Treści matematyczne

Zadania możemy podzielić dodatkowo na trzy grupy: ilość, przestrzeń i kształt oraz zmiana, związki i zależności.

Obszar *ilość* dotyczy rozumienia przez uczniów pojęcia liczby i relacji między liczbami (np. rachunki na liczbach rzeczywistych). Do tego obszaru należą również zadania związane z pomiarem właściwości fizycznych przedmiotów, tj. długości, ciężaru, temperatury, czasu czy objętości.

Przykład zadania – zadanie nr 12 w teście umiejętności matematycznych.

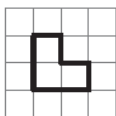
12. Jaką wysokość ma tort?



Obszar *przestrzeń i kształt* dotyczy zadań o problematyce geometrycznej oraz związków przestrzennych między obiektami. Obejmuje on umiejętność rozpoznawania i rysowania figur geometrycznych, dostrzegania symetrii i regularności oraz wymaga wykorzystania wyobraźni przestrzennej.

Przykład zadania – zadanie nr 6 w teście umiejętności matematycznych.

6. Ta figura powstała z trzech kwadratów:



Jeśli będziemy ją rysować po liniach kratki, możemy ją narysować w czterech różnych ułożeniach. Dwa z nich przedstawiono na poniższym rysunku.

Narysuj dwa inne ułożenia tej figury.



Trzeci obszar *zmiana, związki i zależności* dotyczy umiejętności dostrzegania przez ucznia związków i zależności przedstawionych w sposób graficzny, słowny, tabelaryczny lub symboliczny.

Przykład zadania – zadanie nr 5 w teście umiejętności matematycznych.

5. Ośrodek wczasowy mieści się w dziewięciopiętrowym budynku.

Pokoje są numerowane według reguły:

- 1 piętro ➡ pokoje od 100 do 130
2 piętro ➡ pokoje od 200 do 230
3 piętro ➡ pokoje od 300 do 330 i tak dalej.

1) Na którym piętrze znajduje się pokój 617?

Odpowiedź: Na piętrze.

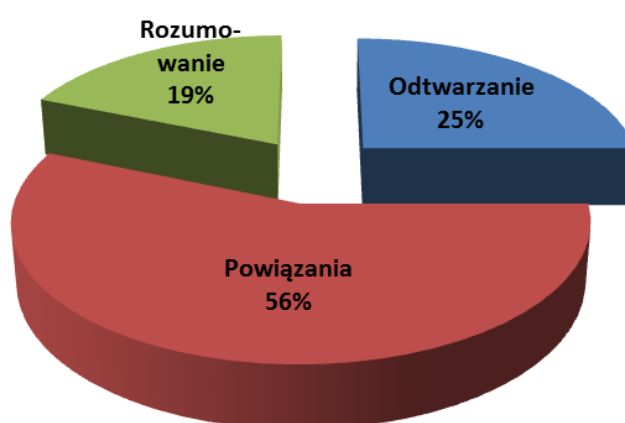
2) Jaki pokój na V piętrze ma najniższy numer?

Odpowiedź: Pokój z numerem

Plan testu umiejętności matematycznych pokazujący, ile zadań mierzących poszczególne umiejętności weszło w skład testu, przedstawia poniższa tabela i wykres.

Tabela 3. Plan testu umiejętności matematycznych.

Umiejętność	Grupa treści	Treści	Liczba zadań
Odtwarzanie	Ilość	Pomiar czasu	2
		Rozumienie pojęcia liczby	1
	Przestrzeń i kształt	Położenie obiektów względem siebie	1
Powiązania	Ilość	Dodawanie i odejmowanie liczb naturalnych	1
		Zadania tekstowe (rozwiązywanie i układanie)	2
		Pomiar długości i ciężaru	2
	Przestrzeń i kształt	Figury płaskie	1
		Symetria i regularność	1
	Zmiana, związki, zależności	Dostrzeganie związków i prawidłowości	2
Rozumowanie	Ilość	Dodawanie i odejmowanie liczb naturalnych	1
		Zadania tekstowe (rozwiązywanie i układanie)	1
	Przestrzeń i kształt	Wyobrażenia przestrzenne	1



Rysunek 3. Podział zadań ze względu na sprawdzaną umiejętność.

I.2. Przeprowadzanie testowania

Testy są przeznaczone dla uczniów klas II i III szkół podstawowych. Wyniki badania pilotażowego pokazały, że optymalny czas przeznaczony na rozwiązanie testów przez uczniów klas II wynosi 45 minut, a przez uczniów klas III wynosi 35 minut.

Termin, w którym należy badanie przeprowadzić to marzec – pierwsza połowa kwietnia. Jest on ściśle związany z terminem przeprowadzenia badania pilotażowego, a w związku z tym możliwością prawidłowej interpretacji wyników w odniesieniu do próby badawczej (korzystanie ze skali centylowej – str. 15).

Informacje jakie należy przekazać uczniom przed przystąpieniem do rozwiązywania testu:

- ✓ W teście jest 13 (test świadomości językowych), 14 (test umiejętności matematycznych) lub 16 zadań (test umiejętności czytania).
- ✓ Zadania są ponumerowane i każde zaczyna się od liczby, po której jest kropka i dalej jest tekst zadania.
- ✓ Niektóre zadania składają się z kilku podpunktów, z których każdy jest oznaczony cyfrą z nawiasem. Każdy podpunkt należy rozwiązać oddzielnie.
- ✓ W teście są różne zadania – takie, w których trzeba coś samemu napisać, lub takie, w których trzeba wybrać i zaznaczyć odpowiedź z podanych odpowiedzi. Tam gdzie trzeba coś samemu napisać, to pod tekstem zadania znajdują się trzy linie, w których trzeba wpisać odpowiedź, tak jak pisze się w zeszycie. W zadaniach, w których podanych jest kilka odpowiedzi do wyboru, należy zaznaczyć **tylko** jedną. Odpowiedzi są oznaczone dużymi literami A, B, C i D. Trzeba otoczyć kółkiem literę stojącą przy wybranej odpowiedzi. Jeśli nastąpi pomyłka, to przekreślamy złą odpowiedź krzyżykiem i zaznaczamy kółkiem inną – poprawną.
- ✓ *Dodatkowe uwagi tylko do testu matematycznego* – w teście są zadania, w których oprócz miejsca na odpowiedź jest jeszcze miejsce przeznaczone na obliczenia - kratki pod tekstem zadania. Można tam zapisać działania, które pozwolą obliczyć prawidłową odpowiedź. Są również zadania, w których na końcu jest zdanie poprzedzone słowem „Odpowiedź”. Trzeba wówczas uzupełnić odpowiedź wpisując uzyskany wynik w kratkę lub trzylinię.
- ✓ Należy uważnie czytać tekst poleceń i zawsze dobrze się zastanowić nad odpowiedzią.

I.3. Interpretacja wyników

Rozwiązania uczniów muszą zostać ocenione zgodnie z zasadami opisanymi w podręcznikach kodowania (aneks 1.). Tylko wtedy wyniki uczniów będą mogły być interpretowane w świetle przedstawionych norm.

Maksymalna liczba punktów jaką uczeń może uzyskać za poprawne rozwiązanie wszystkich zadań (zgodnie z załączonymi podręcznikami kodowania – aneks 1.) to:

- w teście świadomości językowej – 20 punktów;
- w teście umiejętności czytania – 16 punktów;
- w teście umiejętności matematycznych – 19 punktów.

Wyniki uczniów można interpretować na tle:

- wyników uczniów danej klasy, w której był przeprowadzony test (kto najsłabiej napisał test, a kto najlepiej, kto uzyskał wynik powyżej średniej w klasie, a kto wynik poniżej średniej w klasie);
- wyników uczniów biorących udział w badaniu pilotażowym (próba ogólnopolska), wykorzystując w tym celu skalę centylową.

Skala centylowa jest skalą pozycyjną wyników tworzoną na podstawie wyników wszystkich uczniów rozwiązujących dany test. W celu ustalenia wartości centyli porządkujemy wyniki rosnąco (od najmniejszych do największych) i dzielimy tak powstały zbiór po kolejnych możliwych do uzyskania punktach z danego testu (np. od 0 do 20). Dla każdej wartości ustalamy, ile procent uczniów uzyskało dany wynik lub niższy (procent skumulowany). Wartość ta jest równa wartości centyla. Jeśli uczeń klasy II uzyskał 12 punktów z testu sprawdzającego umiejętności matematyczne, co odpowiada 76 centylowi (tabela 4), oznacza to, że 76% uczniów rozwiązujących dany test w badaniu pilotażowym uzyskało wynik niższy lub taki sam.

Na kolejnej stronie zamieszczona została tabela pozwalająca odczytać pozycję wyniku ucznia na skali centylowej. Należy zwrócić uwagę, że dla każdego testu i dla każdej klasy (II i III) jest osobna skala centylowa. Możemy zauważyć, że przy takiej samej liczbie punktów w obrębie tego samego testu, wartość centyla (poza skrajnymi wartościami sumy punktów) jest wyższa w klasie II. Na przykład, w teście świadomości językowej 10 punktów uzyskane przez ucznia klasy drugiej odpowiada 57 centylowi, a klasy trzeciej – 32 centylowi. Czyli więcej uczniów uzyskało wynik równy 10 punktom lub niższy w klasie II niż w klasie III. Jest to zgodne z intuicją, że ten sam zestaw zadań jest trudniejszy dla uczniów klas drugich szkoły podstawowej.

Aby móc w pełni poprawnie przenosić wyniki uczniów na skalę centylową bardzo ważne jest, aby ocena rozwiązań uczniowskich odbywała się w taki sam sposób, jak to miało miejsce podczas badania pilotażowego. Dlatego należy przyznawać punkty zgodnie z wytycznymi zawartymi w podręcznikach kodowania zawartymi w aneksie 1 (uwaga ta dotyczy w szczególności zadań otwartych).

Tabela 4. Wyniki uczniów w punktach i odpowiadające im wartości centyli dla trzech testów osiągnięć szkolnych.

Wynik w punktach	Test świadomości językowej		Test umiejętności czytania		Test umiejętności matematycznych	
	Klasa II	Klasa III	Klasa II	Klasa III	Klasa II	Klasa III
	wartość centyla	wartość centyla	wartość centyla	wartość centyla	wartość centyla	wartość centyla
0	1	1	1	1	1	1
1	1	1	1	1	2	1
2	3	1	3	1	4	1
3	6	2	5	2	8	2
4	10	3	8	3	13	3
5	16	5	12	5	20	6
6	22	9	17	8	28	10
7	30	13	23	12	36	15
8	39	18	30	17	44	21
9	48	25	38	22	53	29
10	57	32	47	29	62	37
11	65	40	57	38	69	46
12	73	49	67	49	76	56
13	80	58	77	61	83	66
14	86	67	87	74	89	75
15	91	76	95	89	93	84
16	95	84	100	100	97	91
17	98	90	-	-	99	96
18	99	96	-	-	100	99
19	100	99	-	-	100	100
20	100	100	-	-	-	-

Wynik każdego pomiaru jest obarczony niepewnością. Nie należy jej kojarzyć z pomyłką ucznia lub autorów zadań. Jest to nieodłączny element procesu pomiarowego. W przypadku testów osiągnięć szkolnych głównymi źródłami niepewności pomiarowej są: subiektywna ocena osoby sprawdzającej wykonanie zadań otwartych, zgadywanie przez ucznia w zadaniach zamkniętych, dyspozycja intelektualna w dniu pisania testu czy ściąganie. Błąd pomiaru może również wynikać z samej konstrukcji testu, tj. arbitralności wyboru zadań testowych czy stroniczości testu. W związku z tym, wynik jaki uzyskał uczeń nie oddaje w doskonały sposób jego poziomu umiejętności.

Przeprowadzone analizy pokazują, że błąd pomiaru w opisywanych testach zmienia się wraz z mierzonym poziomem umiejętności uczniów i mieści się w granicy 1 punktu - wyłączając skrajne wartości poziomu umiejętności. Oznacza to, że uczeń mógł uzyskać jeden punkt więcej lub jeden punkt mniej, przy posiadanym poziomie umiejętności. Jeśli więc jeden uczeń uzyskał 11 punktów, a drugi 10 lub 9 punktów, to mówienie o tym, że pierwszy z nich lepiej opanował umiejętności sprawdzane danym testem nie jest uprawnione. W przypadku testu umiejętności czytania wartość błędu gwałtownie rośnie w zakresie wysokich poziomów umiejętności, a więc najmniej informacji test niesie o uczniach najlepszych. Nieco szerzej ten temat został omówiony w części 2 podręcznika (str. 29). Ważne jest, aby nie przeceniać niewielkich różnic w wynikach uzyskanych przez poszczególnych uczniów z danego testu.

Nie należy też bezpośrednio porównywać sumy punktów jaką uzyskał ten sam uczeń z trzech różnych testów i na tej podstawie wyciągać wniosków, które umiejętności opanował on lepiej, a które gorzej. Należy zwrócić uwagę na różną długość skali (maksymalną liczbę punktów możliwych do uzyskania z poszczególnych testów) oraz różną łatwość testów. Porównanie to jest tylko możliwe w odniesieniu do wyników z badania pilotażowego (jak wynik ucznia w danym zakresie plasuje się na tle wyników z próby).

Oprócz analizy ilościowej wyników można również dokonać ich analizy jakościowej. W części drugiej podręcznika, w tabelach nr 5., 6., 7. znajduje się charakterystyka zadań wchodzących w skład trzech testów, zawierająca m.in. informacje, jakie umiejętności ogólne i szczegółowe sprawdza dane zadanie. Porównując łatwości poszczególnych zadań, możemy wnioskować o tym, z jakim typem zadań nasi uczniowie poradzili sobie najlepiej, a jakie umiejętności opanowali słabiej. Taka analiza wymaga dokładniejszego przyjrzenia się testom i ustalenia, które odpowiedzi złożyły się na końcowy wynik punktowy ucznia.

CZĘŚĆ 2. DLA BADACZA

W tej części znajdują się informacje dotyczące metodologii przygotowywania testów, wybranej do analiz teorii budowy testów oraz przedstawiona została analiza psychometryczna zadań.

II.1. Procedura tworzenia testów osiągnięć szkolnych

Konstrukcja rzetelnych testów osiągnięć szkolnych jest procesem bardzo wymagającym i czasochłonnym. Prace nad przygotowaniem testów rozpoczęły się od ustalenia głównych założeń dotyczących koncepcji pomiaru osiągnięć szkolnych, określenia umiejętności, które mają być badane oraz opracowania procedur i narzędzi badawczych. Zdefiniowana została populacja i przyjęto teorię pomiaru - teorię odpowiedzi na zadania (*item response theory*, IRT), wykorzystaną do analizy psychometrycznej zadań. Jej opis znajdują Państwo w rozdziale II.2 niniejszego podręcznika.

Na etapie projektowania narzędzi (testów) szczegółowo określono treści i umiejętności definiujące badane konstrukty. W związku z tym analizie została poddana podstawa programowa kształcenia ogólnego, wzięto pod uwagę wnioski płynące z międzynarodowych badań umiejętności oraz z nowej formuły sprawdzianu dla klasy VI. Opracowana koncepcja skal była przedmiotem konsultacji merytorycznych z ekspertami oraz poddana została zewnętrznej recenzji. Na tej podstawie zostały przygotowane plany testów precyzujące, ile zadań mierzących poszczególne umiejętności i odnoszących się do poszczególnych treści powinno się znaleźć w ostatecznych wersjach testów. Wszystko to miało na celu zapewnienie trafności treściowej przygotowanych narzędzi. Kolejnym etapem było więc ułożenie, a następnie poddanie badaniu pilotażowemu zadań, z których wybrano najlepsze do ostatecznej wersji testów. Przyjęte standardy tworzenia testów zostały opisane w artykule A. Jasińskiej i M. Modzelewskiego (Edukacja, 2014).

Badanie pilotażowe

We współpracy z osobami mającymi doświadczenie w układaniu zadań przygotowano pulę nowych zadań zgodną z koncepcją pomiaru, z których następnie wybrano zadania do badania pilotażowego. W badaniu tym wykorzystano 36 zeszytów testowych, 12 rodzajów zeszytów testowych testu umiejętności czytania, 12 rodzajów zeszytów testu umiejętności pisania i świadomości językowej oraz 12 rodzajów zeszytów testu matematycznego. Każdy z zeszytów zawierał średnio po 10 zadań rozlokowanych na czterech stronach formatu A4. Dodatkowo wykorzystano 1-stronicowy szkoleniowy test, służący poinstruowaniu uczniów, jak wypełniać testy.

Zadania od siebie niezależne zostały przydzielone do poszczególnych zeszytów losowo. Zadania układające się w wiązki (jak w przypadku zadań do tekstów w teście umiejętności czytania) zostały rozlokowane do zeszytów w taki sposób, by w co najwyżej dwóch zeszytach testowych wystąpił ten sam tekst. Przydzielając zadania do zeszytów starano się także

zachować równowagę ze względu na czas potrzebny do wypełnienia poszczególnych zeszytów.

Populacją badaną byli uczniowie klas II i III szkół podstawowych dla dzieci i młodzieży, ogólnodostępnych (czyli z wykluczeniem szkół specjalnych, integracyjnych, wyrównawczych itp.), bez specyfiki (czyli z wykluczeniem szkół sportowych itp.), nie związanych (czyli z wykluczeniem szkół przyszpitalnych itp.). Operatem pierwotnym losowania była lista szkół sporządzona przez SIO z roku szkolnego 2009/2010, przy czym populację docelową szkół ograniczono do placówek posiadających co najmniej dwa oddziały na poziomie klasy I i II, co pozwalało wnioskować, że będą to szkoły, które w kolejnym roku szkolnym będą posiadały co najmniej dwa oddziały na poziomie klasy II i III.

Dla testu matematycznego, z uwagi na większy związek testowanych zadań z treściami nauczania, przyjęto założenie, że uczniowie klas III nie będą rozwiązywali zadań najłatwiejszych, a uczniowie klas II – najtrudniejszych (według przewidywań autorów). Operacja ta miała na celu podniesienie trafności fasadowej testu. W związku z tym założeniem z zeszytów testowych przeznaczonych dla uczniów klas III wykluczono ok. 30 zadań najłatwiejszych, a z zeszytów testowych przeznaczonych dla uczniów klas II wykluczono ok. 30 zadań najtrudniejszych. Dzięki temu zminimalizowane zostało ryzyko, że na sposób odpowiedzi uczniów na zadania miała wpływ niedostosowana ich trudność (zniechęcenie, gdy zadania są zbyt trudne lub zlekceważenie testu, jeśli zadania są zbyt łatwe).

Realizacja badania w terenie została powierzona firmie *Instytut Badawczy ARC Rynek i Opinia*, wyłonionej w wyniku przetargu publicznego .

Pierwsze sesje testowe rozpoczęły się 08.03.2011, a całe badanie zakończyło się 13.04.2011. Zgodnie z założeniami udało się je zrealizować w 45 szkołach losowo wybranych w skali kraju. Każdy uczeń miał do rozwiązania cztery zeszyty testowe, które wypełniał podczas czterech kolejnych dni. Pierwszego dnia uczniowie rozwiązywali testy podczas 60 minutowej sesji testowej, drugiego, trzeciego i czwartego dnia - podczas 45 minutowej sesji testowej. Plan dystrybucji poszczególnych zeszytów wśród uczniów zapewnił, że każde zadanie zostało rozwiązane przez ok. 340 dzieci.

Kodowanie rozwiązanych przez uczniów zadań i wprowadzanie danych do elektronicznych baz przeprowadzone zostało we współpracy z osobami zatrudnionymi przez Centralną Komisję Egzaminacyjną, pod kierunkiem koordynatora badania pilotażowego z ramienia CKE. Kodowanie odbyło się zgodnie z przekazanymi koderom przygotowanymi kluczami kodowania oraz ogólnymi zasadami kodowania (aneks 1.). Klucze kodowe zostały doprecyzowane na podstawie próbnego kodowania faktycznych rozwiązań uczniowskich. Spośród wszystkich zakodowanych zeszytów testowych zostały wylosowane zeszyty testowe do zakodowania powtórnego. Operacja ta miała na celu analizę jakości testowanych zadań, a dokładniej rzetelności kodowania zadań otwartych. Powtórny kodowaniem zostały objęte zeszyty testowe 20% przebadanych uczniów.

Uzyskane odpowiedzi na testowane zadania pozwoliły na dopasowanie do danych modelu IRT, dzięki czemu właściwości wszystkich zadań można było opisać na wspólnej skali. To z kolei pozwoliło na porównanie zadań między sobą i ocenę ich jakości. Ponadto, dzięki przyjętej teorii odpowiedzi na zadania możliwe okazało się oszacowanie właściwości psychometrycznych testów powstałych z wyselekcjonowanych zadań.

II.2. Teoria odpowiedzi na zadanie testowe

Na potrzeby konstrukcji testów osiągnięć szkolnych dla uczniów klas II i III szkoły podstawowej została wybrana teoria odpowiedzi na zadanie testowe (item response theory, IRT), będąca alternatywą dla klasycznej teorii testu. W ostatnich latach koncepcja ta zdobywa coraz większą popularność i obserwujemy jej dynamiczny rozwój. Modele IRT stały się już właściwie podstawowym narzędziem statystycznym dla osoby zajmującej się badaniami edukacyjnymi. Dzięki tej teorii możemy precyzyjniej określić parametry zadań i narzędzi z nich stworzonych. Pozwala ona pełniej opisać zależność między poziomem umiejętności ucznia a prawdopodobieństwem udzielenia przez niego poprawnej odpowiedzi na zadanie. Należy pamiętać jednak, że teoria ta posiada również swoje wady i pewne ograniczenia.

Podstawy teoretyczne

U podstaw teorii odpowiedzi na zadanie testowe leżą założenia o jednowymiarowości konstruktu oraz lokalnej niezależności zadań. Zgodnie z założeniem o jednowymiarowości, otrzymany przez ucznia wynik z danego testu może być wyjaśniony przez jedną „ukrytą cechę” ucznia, zwaną też cechą utajoną lub latentną. W naszym przypadku są to umiejętności ucznia rozwiązującego test. Założenie to jest bardzo trudne do spełnienia w 100%. Po pierwsze, możemy sobie wyobrazić, że rozwiązanie zestawu zadań składających się na dany test wymaga od ucznia posłużenia się wieloma procesami intelektualnymi, jednak dopóki są one powiązane i zgodne ze sobą, to możemy mówić o jednowymiarowości. Po drugie, na wynik który otrzymuje uczeń wpływ mają również czynniki osobowościowe, np. poziom motywacji czy skłonność do podejmowania ryzyka oraz związane chociażby z faktem wcześniejszego rozwiązywania różnych testów przez ucznia, przez co jest on bardziej lub mniej obyty z taką sytuacją.

Drugie założenie związane z lokalną niezależnością zadań jest spełnione, gdy odpowiedź ucznia na rozwiązywane zadanie nie zależy od tego, jak odpowiedział on wcześniej na inne zadania, czyli pomiędzy zadaniami nie ma żadnej zależności i powiązań, poza wynikającymi z poziomu mierzonych umiejętności. Problem pojawia się np. w przypadku zadań odnoszących się do jednego tekstu (co ma miejsce w przypadku testu sprawdzającego umiejętność czytania), gdyż wydaje się, że ich korelacja między sobą może być większa niż z zadaniami opartymi na innym tekście.

Chociaż założenia o jednowymiarowości i lokalnej niezależności nie są spełniane w 100% to nie wyklucza to jednak korzystania z modeli IRT. Do oceny, czy wybrany model jest odpowiedni do analizy posiadanych danych, służą statystyki dopasowania, które obrazowo

mówiąc porównują otrzymane wyniki z wynikami teoretycznymi (oczekiwanymi) zgodnie z założeniami wybranego modelu. Jeśli to dopasowanie jest dobre, wówczas wiemy, że wybrany model jest odpowiedni do analiz.

Do naszych analiz wykorzystaliśmy model jednoparametryczny Rascha, który jest najbardziej restrykcyjny (spośród dostępnych modeli IRT) co do założeń o właściwościach zadań, ale jednocześnie dzięki temu jest mniej skomplikowany i bardziej przystępny. Zakłada on, że zadania różnią się jedynie trudnością (Rasch, G., 1960). Zadania oceniane wielopunktowo były analizowane w ramach modelu odpowiedzi częściowej (*partial credit model*) Geoffa Mastersa (1982), będącym uogólnieniem modelu Rascha dla zadań ocenianych kategorialnie.

Więcej informacji na temat teorii odpowiedzi na zadania znajdą Państwo w publikacji M. Jakubowskiego i A. Pokropka (Badając egzaminy, 2009). Podstawy modelowania IRT wraz z praktycznymi wskazówkami, pomagającymi w wyborze modelu oraz interpretacji wyników zostały opisane w artykule B. Kondratka i A. Pokropka (Edukacja, 2013).

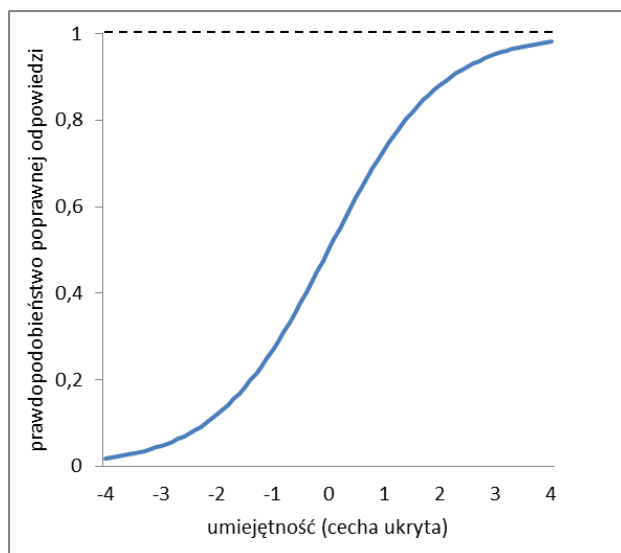
Statystyki INFIT i OUTFIT

Aby określić, jak dobrze wybrany model jest dopasowany do zebranych danych możemy posłużyć się dwoma współczynnikami: INFIT i OUTFIT, zwracanymi przez program statystyczny. Współczynniki te oparte są na wariancji standaryzowanych reszt, przy czym współczynnik INFIT (weighted fit) jest ważony, a OUTFIT (unweighted fit) - nie.

Współczynniki INFIT i OUTFIT mogą przyjmować wartość od 0 do $+\infty$, przy czym wartością optymalną (jaką chcielibyśmy obserwować) jest 1. Jeśli rozpatrywane współczynniki przyjmują taką wartość (równą jeden), to możemy powiedzieć, że uzyskane odpowiedzi na dane zadanie pasują do przewidywań modelu. Zazwyczaj mamy pewne odchylenia od wartości pożądanej. W literaturze nie ma jednoznacznej odpowiedzi, jakie są dopuszczalne wartości statystyk INFIT i OUTFIT. Dla przykładu, Wilson (2005) przyjmuje za akceptowalne wartości tych współczynników mieszczące się w przedziale od 0,75 do 1,33. Bond i Fox (2001) podają, że dla testów wielokrotnego wyboru dopuszczalne wartości statystyk dopasowania to 0,7 – 1,3. W przygotowanych przez nas testach współczynniki INFIT i OUTFIT opisujące poszczególne zadania mieszczą się w przedziale od 0,75 do 1,07 (tabele 5-7 str. 26-28). Takie wartości współczynników pozwalają sądzić o prawidłowym wyborze modelu.

Krzywa charakterystyczna zadania

Korzystając z teorii IRT dla każdego zadania możemy wyznaczyć krzywą charakterystyczną, która obrazuje związek między poziomem umiejętności testowanej osoby, a prawdopodobieństwem udzielenia przez nią poprawnej odpowiedzi (ang. item characteristic curve – ICC). Krzywą taką można opisać jako krzywą logistyczną. Jest ona rosnąca w całym swoim przedziale, co oznacza, że im wyższa jest wartość umiejętności ucznia, tym wyższe jest prawdopodobieństwo udzielenia przez niego prawidłowej odpowiedzi na dane zadanie (rysunek 4).

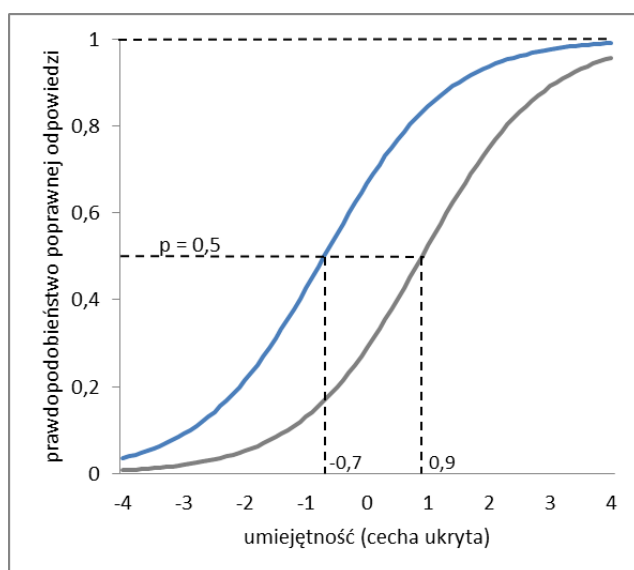


Rysunek 4. Krzywa charakterystyczna zadania

Jednocześnie prawdopodobieństwo poprawnej odpowiedzi jest zawsze większe od zera i mniejsze od jedności. Oznacza to, że nawet przy bardzo niskiej wartości umiejętności (cechy ukrytej) uczeń ma szansę odpowiedzieć poprawnie na zadanie, a pewne prawdopodobieństwo błędnego rozwiązania zadania istnieje nawet przy bardzo wysokiej wartości umiejętności. Takie założenie jest zgodne z ogólnymi odczuciami i obserwacjami dotyczącymi rozwiązywania testów.

Trudność zadania

Model Rascha zakłada, że trudność zadania testowego jest jedyną charakterystyką zadania wpływającą na wynik uzyskany przez egzaminowanego. Wskaźnik trudności zadania jest równy wartości umiejętności ucznia potrzebnej do uzyskania prawdopodobieństwa 0,5 udzielenia poprawnej odpowiedzi na zadanie, co graficznie przedstawia rysunek 5.



Rysunek 5. Wyznaczanie trudności zadań

Pozioma przerywana linia na rysunku 5 wyznacza prawdopodobieństwo udzielenia poprawnej odpowiedzi na zadanie równe 0,5. Przecięcie tej linii z krzywą charakterystyczną zadania wyznacza trudność tego zadania. Zadanie, którego krzywa charakterystyczna jest przesunięta bardziej w lewą stronę ma trudność równą -0,7, natomiast drugie zadanie charakteryzuje się trudnością 0,9 (pionowe przerywane linie). Drugie zadanie jest więc trudniejsze od pierwszego – wymaga wyższego poziomu umiejętności, aby uzyskać takie samo prawdopodobieństwo prawidłowej odpowiedzi, jak dla zadania pierwszego. Widzimy tutaj również, że trudność zadania, jak i umiejętność ucznia jest przedstawiana na jednej skali.

Skala trudności zadań może teoretycznie przybierać wartości od minus nieskończoności do plus nieskończoności, w praktyce jednak w zdecydowanej większości przypadków mieści się ona w przedziale od -4 do 4.

II.3. Analiza psychometryczna zadań

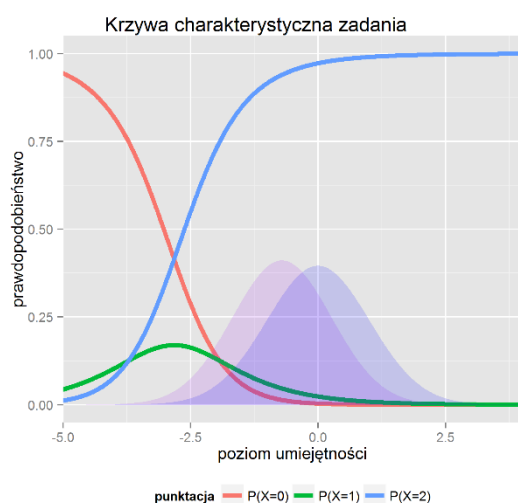
Poniżej przedstawiono krzywe charakterystyczne dla wybranych zadań z każdego z trzech testów osiągnięć szkolnych. Krzywe przedstawiono na tle rozkładów umiejętności uczniów klas II (różowy rozkład) oraz uczniów klas III (fioletowy rozkład). Dołączono treść zadania oraz informację o jego trudności. Do analiz wykorzystano model jednoparametryczny Rascha, w którym dla wszystkich zadań dyskryminacja została ustalona w jedności.

W teście świadomości językowej wśród zadań najłatwiejszych, jak i najtrudniejszych znalazły się zadania sprawdzające u uczniów tą samą umiejętność - znajomość ortografii i interpunkcji. Oba zadania były punktowane na skali 0-1-2. Dzięki zastosowanej teorii odpowiedzi na zadania można dokonać analizy rozkładu prawdopodobieństwa każdego wyniku możliwego do uzyskania.

Zadanie nr 6

W miejsce kropek wstaw brakujące litery:
ń lub ni.

Wszeszło sło.....ce i powitaliśmy nowy dzie.....
Fra.....o obudził się i pomyślał:
„Oj,e lubię po.....edziałku”.

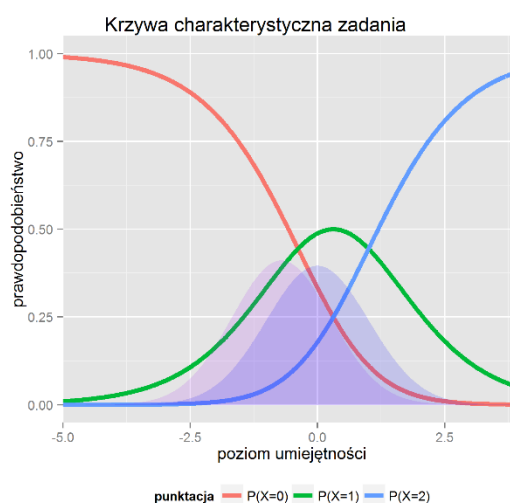


Trudność całego zadania wyniosła -2,82

Zadanie nr 9

W tym tekście brakuje przecinków, kropek i wielkich liter. Przepisz tekst poprawnie.

aneta chodzi do szkoły muzycznej dziewczynka potrafi grać na pianinie jej ulubionymi kompozytorami są fryderyk chopin karol szzymanowski i stanisław moniuszko aneta najbardziej lubi śpiewać dlatego zna wiele piosenek



Trudność całego zadania wyniosła 0,31

Na wykresie dla zadania nr 9 (trudniejszego) widzimy, że uczniowie o niskim poziomie umiejętności uzyskują z największym prawdopodobieństwem zero punktów. Dla wyższych wartości poziomu umiejętności (od -0,35 do 1,0) najbardziej prawdopodobne jest uzyskanie 1 punktu, a dla $\theta > 1,0$ najwyższe prawdopodobieństwo jest dla wyniku 2. Taki układ jest logiczny i prawidłowy.

W przypadku zadania nr 6. prawdopodobieństwo uzyskania 1 punktu (zielona krzywa) nigdy nie osiąga wartości najwyższej. Poniżej wartości poziomu umiejętności -2,6 najbardziej prawdopodobne jest, że uczeń nie otrzyma żadnego punktu za to zadanie, natomiast

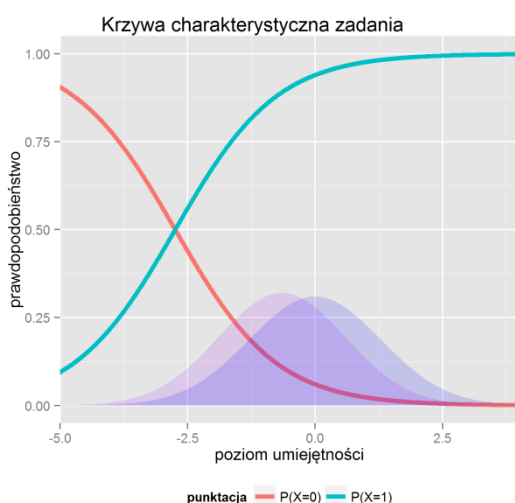
powyżej tej wartości – najbardziej prawdopodobne jest uzyskanie maksymalnej liczby punktów, czyli dwóch.

W teście umiejętności czytania wśród najłatwiejszych zadań największą grupę stanowiły zadania sprawdzające umiejętność wyszukiwania informacji z tekstu, najtrudniejsze okazało się natomiast zadanie sprawdzające umiejętność czytania krótkich poleceń.

Zadanie nr 13 (do tekstu o delfinach)

Podczas poruszania się w wodzie delfinom pomagają...

- A. płetwy boczne.
- B. płetwy ogonowe.
- C. płetwy brzuszne.
- D. płetwy grzbietowe.

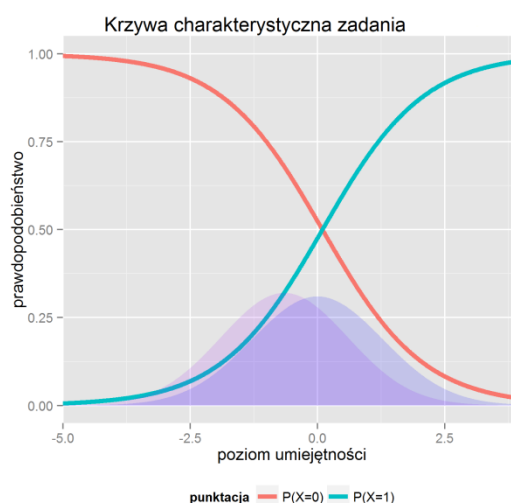


Trudność zadania wyniosła -2,74

Zadanie nr 9

Przeczytaj tekst. Podkreśl wyrazy oznaczające nazwy osób.

Mieszkańcy Wrocławia obchodzą dzisiaj 700-lecie powstania swojego miasta. Z tej okazji prezydent miasta wręczył orderzy zasłużonym żołnierzom. Podczas uroczystej gali wystąpiła światowej sławy piosenkarka.

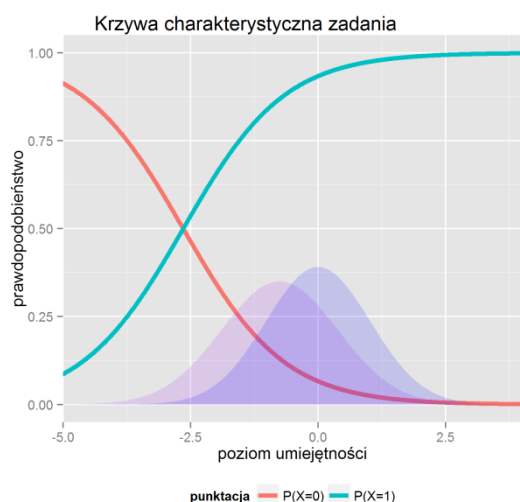


Trudność zadania wyniosła 0,10

W teście umiejętności matematycznych do najłatwiejszych zadań należały zadania sprawdzające umiejętność rozwiązywania zadań typowych, które wymagają użycia wyćwiczonych technik i posłużenia się dobrze znanymi, prostymi obiektami. W praktyce oznacza to korzystanie przez ucznia z zapamiętanych przez niego pojęć lub schematów (umiejętność odtwarzania wiadomości i dobrze wyćwiczonych schematów). Najtrudniejsze okazały się zadania (dwa z trzech umieszczonych w teście) sprawdzające umiejętność rozumowania, a więc przeprowadzenia prostego rozumowania matematycznego składającego się z kilku kroków. Krzywą charakterystyczną pokazano dla jedno z tych zadań – dotyczącego rozwiązywania zadań tekstowych.

Zadanie nr 11

Na jednej szalce wagi stoi odważnik, a na drugiej dwie butelki. Waga jest w równowadze. Ile waży jedna butelka? (rys. na str. 9)

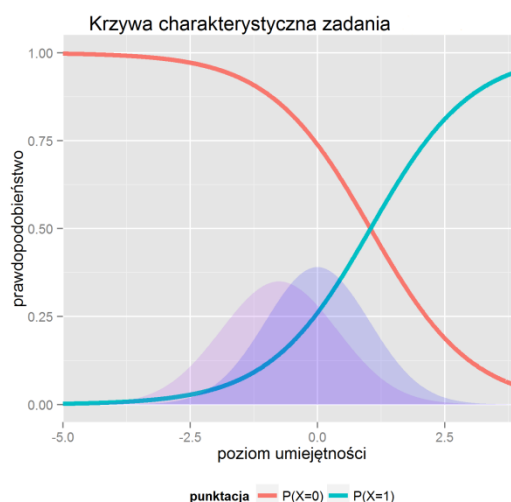


Trudność zadania wyniosła -2,65

Zadanie nr 7

Na rysunku pokazano ceny różnych pitek w sklepie Alfred.

Pani kupiła w sklepie Alfred kilka pitek na nagrody w konkursie sportowym. Zapłaciła za zakupy 41 zł. Jakie piłki kupiła?



Trudność zadania wyniosła 1,04

Na powyższych rysunkach zostały pokazane, obok krzywych charakterystycznych zadań, rozkłady umiejętności uczniów klas drugich (kolor różowy) i trzecich (kolor niebieski) – według założeń modelu. Zgodnie z oczekiwaniami, krzywa rozkładu umiejętności uczniów klas drugich jest przesunięta w lewą stronę, w zakres niższych wartości umiejętności (patrz również tabela 10).

Tabele 5.-7. zawierają opis oraz parametry wszystkich zadań, które weszły do testu świadomości językowej, umiejętności czytania oraz umiejętności matematycznych. W tabeli 7. występują braki danych w statystykach dopasowania spowodowane faktem, że w badaniu pilotażowym część zadań rozwiązywali tylko uczniowie klas drugich, a część – tylko uczniowie klas trzecich szkoły podstawowej.

Tabela 5. Charakterystyka zadań wchodzących w skład testu świadomości językowej.

Nr zadania w teście	Umiejętność ogólna	Umiejętność szczegółowa	Trudność	Statystyki dopasowania			
				II klasa		III klasa	
				OUTFIT	INFIT	OUTFIT	INFIT
1.1.	Bogactwo słownikowe	Wyjaśnianie znaczenia związków frazeologicznych	-1,73	1,02	1,01	0,88	0,93
1.2.	Bogactwo słownikowe	Wyjaśnianie znaczenia związków frazeologicznych	-1,19	0,95	0,96	0,96	0,99
2.	Elementy wiedzy o języku	Poprawność składniowa i gramatyczna zdań	0,02	0,97	0,98	0,94	0,96
3.	Bogactwo słownikowe	Wyjaśnianie znaczenia słowa	-2,36	0,98	1,01	0,99	0,93
4.1.	Elementy wiedzy o języku	Rozpoznawanie i poprawianie błędów językowych	0,81	0,90	0,94	0,95	0,98
4.2.	Elementy wiedzy o języku	Rozpoznawanie i poprawianie błędów językowych	-1,02	0,99	0,99	1,00	1,01
5.	Bogactwo słownikowe	Tworzenie synonimów lub antonimów	0,24	1,03	1,03	0,94	0,96
6.	Elementy wiedzy o języku	Ortografia i interpunkcja	-2,82	0,97	0,97	0,89	0,96
7.1.	Bogactwo słownikowe	Tworzenie synonimów lub antonimów	-1,57	0,94	0,96	0,98	1,01
7.2.	Bogactwo słownikowe	Tworzenie synonimów lub antonimów	1,11	0,93	0,96	0,99	1,00
8.	Elementy wiedzy o języku	Poprawność składniowa i gramatyczna zdań	-0,20	0,98	1,00	0,97	0,99
9.	Elementy wiedzy o języku	Ortografia i interpunkcja	0,31	0,90	0,92	0,96	0,96
10.	Bogactwo słownikowe	Tworzenie wyrażen porównawczych	-0,92	0,98	0,99	1,06	1,04
11.	Bogactwo słownikowe	Rozpoznanie niepoprawnego użycia słowa (ze względu na znaczenie)	-0,44	0,98	0,99	0,99	0,99
12.	Pisanie krótkich tekstów	Redagowanie tekstu	-0,74	0,95	0,96	1,04	1,03
13.	Elementy wiedzy o języku	Poprawność składniowa i gramatyczna zdań	-2,69	1,03	1,00	0,79	0,89

Tabela 6. Charakterystyka zadań wchodzących w skład testu umiejętności czytania.

Nr zadania w teście	Tekst	Umiejętność	Trudność	Statystyki dopasowania			
				II klasa		III klasa	
				OUTFIT	INFIT	OUTFIT	INFIT
1.	Literacki (proza)	Wyszukiwanie informacji	-1,59	0,88	0,91	0,88	0,93
2.	Literacki (proza)	Wyszukiwanie informacji	-2,53	0,89	0,96	1,04	1,11
3.	Literacki (proza)	Interpretacja	-0,71	0,88	0,90	0,94	0,98
4.	Literacki (proza)	Interpretacja	-1,72	0,91	0,93	0,92	0,98
5.	Literacki (proza)	Interpretacja	-1,06	0,90	0,92	0,90	0,95
6.	Literacki (proza)	Interpretacja	-1,82	0,91	0,96	0,83	0,92
7.	Literacki (proza)	Interpretacja	-0,91	0,91	0,92	0,90	0,93
8.	Literacki (proza)	Interpretacja	-0,72	0,91	0,92	0,86	0,90
9.	Krótki tekst	Czytanie poleceń	0,10	0,92	0,94	0,90	0,91
10.	Popularnonaukowy	Interpretacja	-0,47	0,95	0,96	0,90	0,92
11.	Popularnonaukowy	Wyszukiwanie informacji	-1,00	0,97	0,97	0,91	0,95
12.	Popularnonaukowy	Interpretacja	-2,15	0,96	0,99	0,91	0,99
13.	Popularnonaukowy	Wyszukiwanie informacji	-2,74	0,85	0,93	0,92	1,06
14.	Popularnonaukowy	Wyszukiwanie informacji	-2,67	0,82	0,92	0,96	1,03
15.	Popularnonaukowy	Interpretacja	-1,46	0,98	0,99	0,87	0,91
16.	Popularnonaukowy	Refleksja i ocena	-2,61	1,04	1,07	0,75	0,83

Tabela 7. Charakterystyka zadań wchodzących w skład testu umiejętności matematycznych.

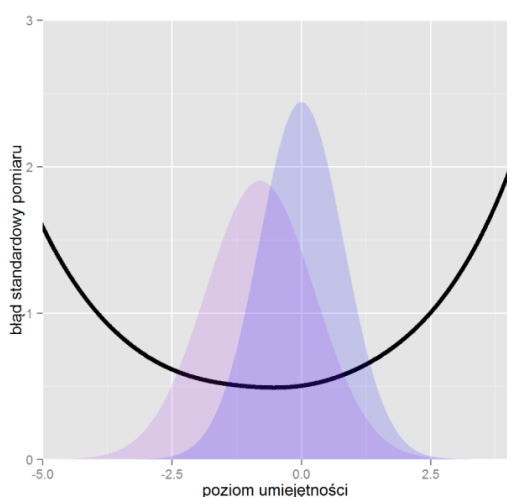
Nr zadania w teście	Grupa treści	Treści	Umiejętność	Trudność	Statystyki dopasowania*			
					II klasa		III klasa	
					OUTFIT	INFIT	OUTFIT	INFIT
1.	Ilość	Rozumienie pojęcia liczby	Odtwarzanie	-2,54	0,85	0,97		
2.1.	Ilość	Pomiar czasu	Odtwarzanie	-2,25	0,77	0,87		
2.2.	Ilość	Pomiar czasu	Odtwarzanie	-1,59	0,83	0,87		
3.	Ilość	Dodawanie i odejmowanie liczb naturalnych	Powiązania	0,70	0,93	0,92	0,95	0,97
4.	Zmiana, związki, zależności	Zadania tekstowe (rozwiązywanie i układanie)	Powiązania	-1,63	0,90	0,94	0,92	0,97
5.1.	Zmiana, związki, zależności	Dostrzeganie związków i prawdziwości	Powiązania	-0,80	0,91	0,93	0,90	0,92
5.2.	Zmiana, związki, zależności	Dostrzeganie związków i prawdziwości	Powiązania	0,26	0,92	0,95	0,92	0,93
6.	Przestrzeń i kształt	Figury płaskie	Powiązania	-0,73	0,95	0,96	0,97	0,97
7.	Zmiana, związki, zależności	Zadania tekstowe (rozwiązywanie i układanie)	Rozumowanie	1,04			0,90	0,93
8.	Przestrzeń i kształt	Wyobrażenia przestrzenna	Rozumowanie	1,01			0,90	0,93
9.	Ilość	Dodawanie i odejmowanie liczb naturalnych	Rozumowanie	-1,06	0,93	0,95	0,93	0,95
10.	Przestrzeń i kształt	Położenie obiektów względem siebie	Odtwarzanie	-2,30	0,83	0,92		
11.	Ilość	Pomiar długości i ciężaru	Powiązania	-2,65	0,89	0,95	0,93	0,98
12.	Ilość	Pomiar długości i ciężaru	Powiązania	-0,23	0,90	0,91	0,92	0,93
13.	Zmiana, związki, zależności	Zadania tekstowe (rozwiązywanie i układanie)	Powiązania	-0,66			0,89	0,91
14.	Przestrzeń i kształt	Symetria i regularność	Powiązania	-0,86			0,80	0,80

*Braki w danych dla statystyk dopasowania jest związany z faktem, że niektórych zadań nie rozwiązywali uczniowie klas II, a niektórych uczniowie klas III.

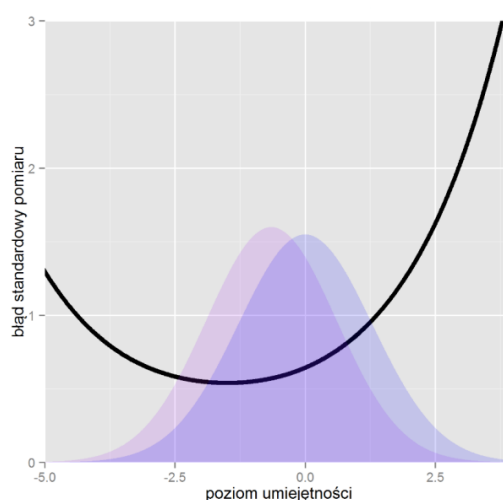
II.4. Rzetelność skal pomiarowych

Każdy pomiar, w tym również edukacyjny, obarczony jest niepewnością pomiarową. Precyzja, z jaką możemy oszacować wynik ucznia o różnym poziomie cechy ukrytej (umiejętności) mówi nam o rzetelności skal.

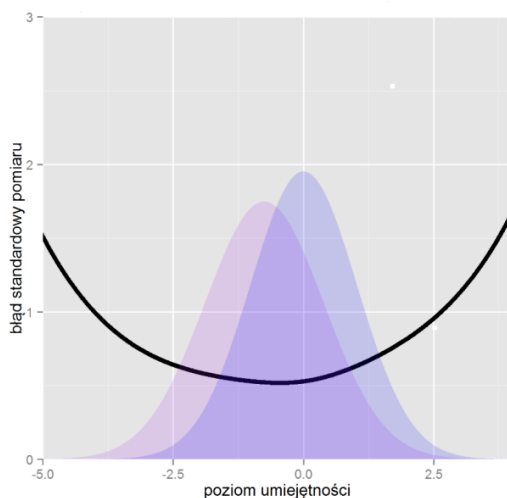
Na rysunkach 7-9 pokazane zostały wielkości błędów pomiaru dla trzech testów na tle rozkładów umiejętności uczniów klas II (różowy rozkład) oraz uczniów klas III (fioletowy rozkład). Widzimy, że wielkość błędu zmienia się dla różnych poziomów umiejętności. Dla testu świadomości językowej i umiejętności matematycznych wielkość błędu nie przekracza jednego punktu w szerokim zakresie poziomu umiejętności uczniów (od ok -4 do 2,5). W przypadku testu umiejętności czytania wartość błędu wyraźnie rośnie w zakresie wyższych poziomów umiejętności (od 1,5). Zadania, które weszły w zakres testu czytania okazały się więc trochę zbyt łatwe w stosunku do umiejętności badanych uczniów. Stąd test ten niesie najmniej informacji o uczniach o wysokim poziomie umiejętności w zakresie czytania. Należy bowiem pamiętać, że na błąd oszacowania wyniku ucznia ma wpływ liczba zadań przez niego rozwiązywanych oraz dopasowanie trudności zadań do poziomu jego umiejętności.



Rysunek 7. Błąd standardowy pomiaru dla testu świadomości językowej



Rysunek 8. Błąd standardowy pomiaru dla testu umiejętności czytania



Rysunek 9. Błąd standardowy pomiaru dla testu umiejętności matematycznych

Choć widzimy, że poziom dokładności pomiaru jest różny w zależności od poziomu umiejętności, to możemy wyznaczyć jedną liczbę dla danego testu opisującą precyzję pomiaru. Jest to współczynnik rzetelności EAP wyznaczony z wariancji wartości oczekiwanych rozkładów a'posteriori. W sytuacji doskonale rzetelnego pomiaru współczynnik rzetelności przyjmowałby wartość 1. W tabeli 10 zamieszczone są wartości współczynników rzetelności EAP dla trzech testów osiągnięć szkolnych.

Tabela 10. Parametry rozkładów umiejętności i współczynnik rzetelności.

Testy	klasa	Parametry rozkładów umiejętności w ramach klas		Współczynnik rzetelność EAP
		średnia	wariancja	
świadomości językowej	III	0,00	1,02	0,765
	II	-0,71	0,94	0,769
umiejętności czytania	III	0,00	1,66	0,756
	II	-0,65	1,56	0,787
umiejętności matematycznych	III	0,00	1,04	0,759
	II	-0,75	1,30	0,799

Wartości współczynników są na zadowalającym poziomie, zwłaszcza pamiętając, że przygotowane testy, są testami krótkimi, zawierającymi po 16 zadań (licząc osobno każdy podpunkt). Na potrzeby analiz średnia dla klasy III została ustalona w punkcie 0, stąd taka wartości w tabeli dla każdego testu.

Literatura:

- Bond, T.G., Fox, C.M. (2001). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Jakubowski, M., Pokropek, A. (2009). *Badając Egzaminy*. Warszawa: Centralna Komisja Egzaminacyjna.
- Jasińska, A., Modzelewski, M. (2014). Testy osiągnięć szkolnych TOS3: przykład narzędzia skonstruowanego z wykorzystaniem modelu Rascha. *Edukacja*, 2(127), 85-107.
- Karpiński, M., Nowakowska, A., Orzechowska, M., Sosulska, D., Zambrowska, M. (2014). Raport z ogólnopolskiego badania umiejętności trzecioklasistów OBUT^m 2014. Warszawa: Instytut Badań Edukacyjnych.
- Kondratek, B., Pokropek, M. (2013). IRT i pomiar edukacyjny. *Edukacja*, 4(124), 42-66.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174.
- Pregler, A. (red.) (2013). Ogólnopolskie badanie umiejętności trzecioklasistów. Raport z badania OBUT 2013. Warszawa: Instytut Badań Edukacyjnych.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Wilson, M. (2005). *Constructing measures: an Item Response Modelling approach*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.