dr Monika Jakubowska-Mirek
Uniwersytet Warszawski, Wydział Pedagogiczny

dr Ewa Stożek

Zrozumieć zadanie, czyli opowieść o tym, jak próbowano zmierzyć niewiedzę ucznia

Streszczenie

W tym artykule dokonujemy krytycznego osądu zadania zamkniętego użytego w matematycznym teście diagnostycznym uczniów klas siódmych. Zastanawiamy się nad użytecznością złożonego zadania, którego rozwiązanie wymaga dobrego zrozumienia problemu matematycznego. Zadania o słabych parametrach psychometrycznych nie są pożądane w teście, ale prowokują do stawiania wielu pytań na drodze do zrozumienia, dlaczego one nie działają. Dążenie do zapewnienia optymalnego funkcjonowania zadania w teście przy badaniu zróżnicowanej pod względem wiedzy i umiejętności populacji uczniów wymusza nadmierne, naszym zdaniem, uproszczenie zadań.

Wydział Pedagogiczny Uniwersytetu Warszawskiego od wielu lat wspiera lokalne samorządy w Kwidzynie i Ostrołęce w monitorowaniu osiągnięć uczniów szkół podstawowych. W ostatnich latach wraz z Urzędem Miasta Ostrołęki realizowany jest projekt „Co już umiemy?”[1], w ramach którego prowadzana jest, między innymi, diagnoza umiejętności matematycznych uczniów na rok przed egzaminem ósmoklasisty. Uczniowie rozwiązują test, który powinien dać odpowiedź, w jakim stopniu są oni gotowi do osiągnięcia sukcesu na egzaminie ósmoklasisty oraz jakie braki wymagają dalszej pracy. Tegoroczne (2023 r.) badanie było przeprowadzone 19 maja na grupie 328 uczniów klas siódmych ostrołęckich szkół podstawowych, włączając w to uczniów ze specjalnymi potrzebami edukacyjnymi oraz grupę uczniów – obywateli Ukrainy. Narzędzie diagnostyczne składało się z 20 zadań: 10 zadań zamkniętych wielokrotnej odpowiedzi (należało wybrać jedną poprawną odpowiedź: A, B, C lub D) oraz 10 zadań otwartych. Za cały test uczeń mógł maksymalnie uzyskać 30 punktów. W analizach uwzględniono wyniki 300 uczniów, którzy na egzaminie ósmoklasisty rozwiązywać będą zadania w arkuszu standardowym. Od 2022 roku w zadaniach zamkniętych dodatkowo prosimy uczniów o wybór jednego z trzech emotikonów na potwierdzenie tego, jak bardzo są pewni wyboru swojej odpowiedzi (Jakubowska-Mirek i Stożek, 2022). Pewność wyboru odpowiedzi w zadaniu zamkniętym może być charakterystyką ucznia, a może być też zinterpretowana jako charakterystyka zadania.

[1] https://www.pedagog.uw.edu.pl/ostroleka/projekty/co-juz-umiemy/

Koncepcja testu i strategia jego przygotowania

Przygotowanie testu co roku jest wyzwaniem. Uczniowie klasy siódmej nie są przyzwyczajeni do przekrojowych testów. Napotykamy trudności związane z tym, że treści realizowane w młodszych klasach są zapomniane, nieaktywne. Uczniowie nie pamiętają pojęć matematycznych, podstawowych reguł i wzorów. Badacze zajmujący się zagadnieniem pamięci i zapominania wskazują, że szybciej zapomina się *szczegóły niż ogólny sens materiału* (Jagodzińska 2008, s. 280). W odniesieniu do nauczania matematyki kluczowe znaczenie ma więc to, czy uczniowie poznali i zrozumieli ogólny sens tego, czego się nauczyli. Jeśli poznany przez nich materiał stanowi zestaw niemających sensu szczegółów, to nie będzie on długo przechowywany w pamięci. Wychodzimy z założenia, że zadania testowe, w miarę możliwości, powinny dotyczyć właśnie ogólnego sensu problemów i zagadnień matematycznych.

Zwiększenie odporności na zapominanie jest także uwarunkowane rozłożeniem treningu w czasie (Jagodzińska, 2008, s. 291). Powtarzanie materiału, wplatanie do zadań treści z poprzednich działów stanowią trening pozwalający na podtrzymanie dotychczasowej wiedzy i umiejętności. Jednak taka praktyka jest niezwykle rzadka. Programy do nauczania matematyki nie zakładają takich rozwiązań, dlatego uczniowie dość szybko zapominają zrealizowany materiał. Powtórki do egzaminu ósmoklasisty zaczynają się dopiero w ósmej klasie, dlatego uczniowie kończący klasą siódmą mogą mieć problem z odtworzeniem treści z poprzednich lat, w szczególności jeśli nie były one dostatecznie zrozumiałe. Rozbieżność umiejętności uczniów klas siódmych i ósmych pokazały nasze badania pilotażowe.

Test budujemy na planie założeń teoretycznych z badania TIMSS (Mullis i in., 2020): zadania reprezentują treści z działów *Liczby* (40% punktów tegorocznego testu), *Geometria* (40%), *Algebra* (15%), *Elementy statystyki* (5%) oraz sprawdzają umiejętności rozwiązywania zadań w sytuacjach typowych (70% punktów tegorocznego testu) i w sytuacjach nietypowych (30%). Tworząc test, oprócz zadań autorskich, inspirujemy się zadaniami ze zbiorów zadań, podręczników, innych badań, wcześniejszych edycji egzaminu ósmoklasisty. Zdarza nam się do testu włączyć zadanie z egzaminu maturalnego (poziom podstawowy), którego rozwiązanie jest w zasięgu możliwości matematycznych siódmoklasisty. Zadania są dyskutowane i modyfikowane, w czym pomaga próbne zastosowanie testu, choć nie daje pełnej gwarancji, że zadania testowe będą bez wad. W tegorocznej edycji pilotaż był dwustopniowy: pierwsza wersja była testowana na 21-osobowej grupie ósmoklasistów, a kolejna – na 78-osobowej grupie siódmoklasistów.

Skoro test przede wszystkim ma służyć diagnozie, to już na etapie tworzenia/włączania zadania do testu należałoby przewidzieć, jakiego rodzaju informację zwrotną dla nauczyciela chcemy sformułować. Jak bardzo złożony jest to problem, pokażemy na przykładzie jednego zadania zamkniętego.

Sześć krawędzi sześcianu, czyli o tym, czy da się przewidzieć problemy i błędne przekonania uczniów

Bohaterem naszej opowieści będzie następujące zadanie:

Zadanie 5. (0–1) ☺ 😐 ☹

Objętość sześcianu jest równa 27 cm^3. Jaka jest suma długości wszystkich krawędzi tego sześcianu?

A. 18 cm B. 36 cm C. 24 cm D. 12 cm

Uczeń wybiera jedną odpowiedź z czterech podanych oraz zaznacza jeden z emotikonów według zasady:

☺ – jestem pewny/pewna, że to poprawna odpowiedź,

😐 – mam wątpliwości, czy wybrałem/wybrałam poprawną odpowiedź,

☹ – odpowiedź wybrałem/wybrałam na chybił trafił.

Włączając to zadanie do testu, spodziewałyśmy się, że uczniowskim rozwiązaniom będą towarzyszyć wizualizacje – bo jak inaczej ustalić, ile krawędzi ma sześcian? Nie spodziewałyśmy się kłopotów z określeniem długości krawędzi sześcianu, bo przecież uczeń wie, że $3^3 = 27$. Dystraktory dobrane są tak, że odpowiadają różnej liczbie krawędzi sześcianu: A. 6 krawędzi; B. 12 krawędzi; C. 8 krawędzi; D. 4 krawędzie. Nasze oczekiwania spełniła badana grupa ósmoklasistów (poziom wykonania zadania 75%, moc różnicująca zadania 0,67). Jednak dla siódmoklasistów to zadanie okazało się zbyt trudne. Prześledźmy, co jest potrzebne, by poprawnie rozwiązać zadanie 5.

Uczeń:

a. wie, co to jest sześcian oraz krawędzie sześcianu, potrafi go sobie wyobrazić oraz narysować;
b. rozumie pojęcie objętości sześcianu i wie, jak obliczyć objętość sześcianu;
c. potrafi wyznaczyć długość krawędzi sześcianu na podstawie jego objętości;
d. poprawne określa liczbę krawędzi sześcianu;
e. sprawnie oblicza sumę długości wszystkich krawędzi sześcianu.

Na każdym z tych etapów uczniowi mogło brakować niezbędnej wiedzy, umiejętności, mógł się też zwyczajnie pomylić. Niewielu uczniów wspierało się rysunkami w brudnopisie czy na marginesie arkusza testowego. Dysponując jedynie informacją o tym, jaki uczeń wybrał dystraktor, nie mamy wglądu w jego tok rozumowania, ani nie znamy charakteru błędu, który popełnił.

Nasze doświadczenia z poprzednich edycji projektu sugerują również, że uczeń klasy VII może poczuć się przytłoczony językiem matematyki tego zadania. W tym krótkim poleceniu mamy nagromadzenie terminów matematycznych: sześcian, objętość, krawędź, długość, suma. Jak pokazują uczniowskie rozwiązania innych zadań, każde z nich w większym lub mniejszym stopniu może być problemem dla ucznia klasy VII.

Thomas M. Haladyna, Steven M. Downing oraz Michael C. Rodriguez dokonali przeglądu podręczników zawierających wskazówki do pisania testów, w szczególności zadań wielokrotnego wyboru (2002). Wymienili oni 31 wskazówek w obrębie 5 aspektów testu, które pojawiały się najczęściej w analizowanych podręcznikach. Między innymi wskazali na zasadę stosowania prostego słownictwa – adekwatnego do możliwości badanej grupy. Coraz częściej mówi się o potrzebie odwołania do doświadczeń uczniowskich, konstruowaniu treści zadań z wykorzystaniem fabuły, która jest im bliska. W ten sposób unika się negatywnego wpływu nieistotnych z perspektywy nauczania matematyki zmiennych, jak sprawność czytania czy zasób leksykalny. Co jednak w przypadku terminologii *stricte* matematycznej? Język matematyki jest precyzyjny – pozwala na jednoznaczne określenia i zapobiega błędom, wynikającym z interpretacji. Dlatego posługiwanie się nim jest tak istotne. Jednak, jak pokazują nasze doświadczenia w projekcie „Co już umiemy?", uczniowie mają problem w posługiwaniu się językiem matematycznym. Uważamy, że uczeń kończący szkołę podstawową powinien swobodnie posługiwać się terminologią matematyczną, natomiast mamy świadomość, że w odniesieniu do omawianego tu zadania i aspektu diagnostycznego należy potraktować uczniowskie kłopoty z terminologią jako zmienną zakłócającą trafność informacji zwrotnej.

Niezależnie od przyczyny kłopotów z tym zadaniem, uczniowie wyraźnie deklarują swoją bezradność wobec postawionego problemu, zaznaczając emotikon ☹.

Pewność wyboru odpowiedzi[2] dla całego zadania jest ujemna i wynosi −0,19. Poziom wykonania zadania to 30%, a jego moc różnicująca to 0,36. Zatem według przyjętej interpretacji (Niemierko, 1999) to zadanie jest trudne i słabo różnicuje, mogłoby więc zostać zastąpione zadaniem o lepszych parametrach psychometrycznych. Z drugiej strony prowokuje ono do poszukiwania odpowiedzi związanych z trudnościami, jakie napotkali w nim uczniowie.

Spróbujmy dociec, jak działa to zadanie. W tabeli 1 zebrano charakterystyki poszczególnych dystraktorów, by zobaczyć, na ile uczniowie byli pewni swoich wyborów, jak często wybierali dany dystraktor (atrakcyjność) i jak to było skorelowane z ogólnym wynikiem testu (moc różnicująca). Najczęściej uczniowie wybierali poprawną odpowiedź B, choć wskaźnik pewności wyboru równy 0,14 pokazuje, że wielu uczniów nie było przekonanych, że jest to poprawna odpowiedź. Bardzo popularna wśród uczniów była też odpowiedź D, w której uczniowie ograniczyli się do obliczenia sumy krawędzi bocznych (lub krawędzi jednej podstawy). Przy wyborze tej odpowiedzi uczniowie byli najmniej przekonani o słuszności swojego wyboru.

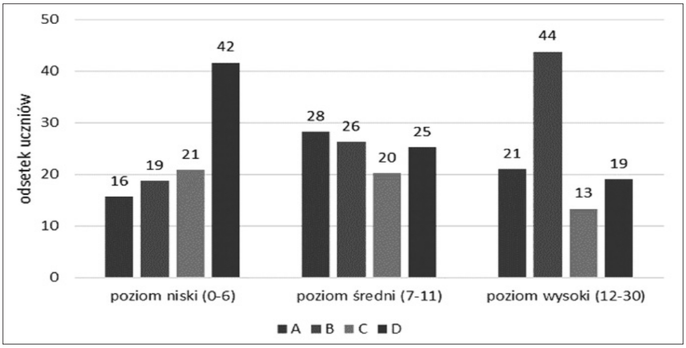
[2] Emotikonom zostały przypisane wartości liczbowe ☺ = 1, 😐 = 0, ☹ = −1. Jako syntetyczny wskaźnik pewności wyboru odpowiedzi dla zadania przyjęto średnią wartość „emotikonów" dla wszystkich uczniów. Wskaźnik ten przyjmuje wartości od −1 do 1 (−1 oznacza, że wszyscy uczniowie w tym zadaniu wybierali odpowiedź na chybił trafił, 1 oznacza, że wszyscy uczniowie byli pewni swojego wyboru odpowiedzi w zadaniu; ujemna wartość wskaźnika oznacza, że uczniowie w danym zadaniu częściej „strzelali", niż byli pewni odpowiedzi, dodatnia wartość – częściej byli pewni odpowiedzi niż „strzelali").

Tabela 1. Charakterystyki dystraktorów dla zadania 5.*

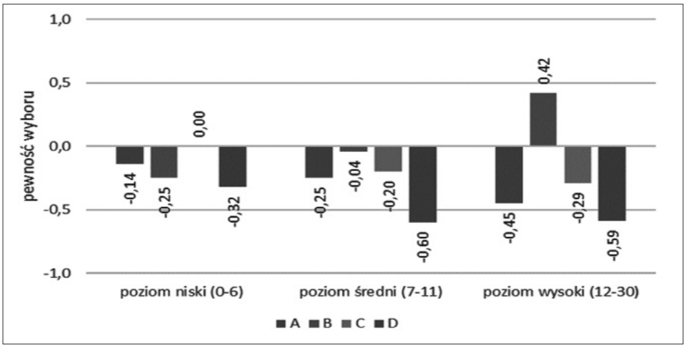
	A	B	C	D
pewność wyboru dystraktora	-0,30	**0,14**	-0,15	-0,46
atrakcyjność dystraktora	22%	**30%**	18%	28%
moc różnicująca dystraktora	-0,06	**0,36**	-0,10	-0,22

***wyniki dla 300 uczniów z badania głównego**

Następnie podzielono uczniów ze względu na wynik testu na trzy grupy o porównywalnej liczebności: uczniowie o niskich wynikach z przedziału 0–6 punktów to 96 osób, uczniowie o średnich wynikach z przedziału 7–11 punktów to 99 uczniów oraz uczniowie o wysokich wynikach z przedziału 12–30 punktów to 105 uczniów. Okazuje się, wśród słabszych uczniów najbardziej popularny był dystraktor D, uczniowie o średnich wynikach mieli kłopot z ustaleniem, która z odpowiedzi jest najbardziej wiarygodna (z lekkim wskazaniem na odpowiedź A – być może uczniowie sądzili, że skoro to sześcian, to powinien mieć też sześć krawędzi), dopiero wśród uczniów o wysokich wynikach z testu poprawna odpowiedź B była najbardziej popularna (por. rys. 1).



Rysunek 1. Atrakcyjność dystraktorów (odsetek uczniów wybierających dany dystraktor) w grupach uczniów o różnym poziomie umiejętności mierzonym ogólnym wynikiem testu



Rysunek 2. Pewność wyboru dystraktorów w grupach uczniów o różnym poziomie umiejętności mierzonym ogólnym wynikiem testu

Choć uczniowie o niskim wyniku za cały test najchętniej wybierali odpowiedź D, to jednak byli jej najmniej pewni. Najbardziej byli pewni odpowiedzi C, która odpowiada sytuacji, gdy uczeń w obliczeniach uwzględnia krawędzie podstaw, ale nie uwzględnia krawędzi bocznych sześcianu. Wyniki zarówno dla grupy uczniów o niskich, jak i średnich wynikach są przypadkowe (por. rys. 2). Pokazują to wskaźniki pewności wyboru odpowiedzi i potwierdzają niskie wartości mocy różnicującej. Dopiero w grupie uczniów o wysokich wynikach w teście zadanie zaczyna funkcjonować w sposób zadowalający: zadanie jest nadal trudne, ale umiarkowanie różnicuje (por. tab. 2)

Tabela 2. Parametry zadania 5. w podziale na grupy ze względu na wynik w teście

poziom wyniku w całym teście	niski	średni	wysoki
poziom wykonania zadania 5.	19%	26%	44%
moc różnicująca zadania 5.	0,13	0,05	0,52

Zatem dla zróżnicowanej pod względem umiejętności matematycznych populacji zadanie jest wyraźnie za trudne i badaczowi nie pozostaje nic innego, jak je uprościć. Można zaproponować różne warianty tego uproszczenia, poniżej przedstawiono trzy propozycje. Zadanie 5.1 sprawdza tylko umiejętność obliczenia długości krawędzi sześcianu, jeśli dana jest jego objętość. Zadanie 5.2 sprawdza tylko to, czy uczeń poprawnie ustala liczbę krawędzi sześcianu.

Zadanie 5.1. (0–1) ☺ 😐 ☹

Objętość sześcianu jest równa 27 cm^3. Długość krawędzi tego sześcianu jest równa

A. 9 cm B. 3 cm C. 8 cm D. 6 cm

Zadanie 5.2. (0–1) ☺ 😐 ☹

Długość krawędzi sześcianu jest równa 3 cm. Jaka jest suma długości wszystkich krawędzi tego sześcianu?

A. 18 cm B. 36 cm C. 24 cm D. 12 cm

Natomiast w zadaniu 5.3 treść pozostaje bez zmian, jedynie zmieniamy dystraktory, tak by uwzględnić najczęstsze błędy uczniów w obliczaniu długości krawędzi przy zadanej objętości oraz ustalaniu liczby krawędzi sześcianu. Najczęściej popełnianym przez uczniów błędem w wyznaczaniu długości krawędzi jest a = 9 (uczniowie utożsamiają potęgowanie z mnożeniem, pierwiastkowanie z dzieleniem (Rushton, 2014)), a najczęstszym błędem w określaniu liczby krawędzi n = 6. To nam daje cztery prawdopodobne odpowiedzi[3].

Zadanie 5.3. (0–1) ☺ 😐 ☹

Objętość sześcianu jest równa 27 cm^3. Jaka jest suma długości wszystkich krawędzi tego sześcianu?

A. 108 cm B. 36 cm C. 18 cm D. 54 cm

[3] Odpowiednio dystraktor **A**: a = 9 cm, n = 12; **B**: a = 3 cm, n = 12; **C**: a = 3 cm, n = 6; **D**: a = 9 cm, n = 6.

Nowa, uproszczona forma zadania pozwoliłaby na bardziej jednoznaczne wnioski dotyczące uczniowskich umiejętności i popełnianych błędów. Propozycje 5.1 i 5.2 pozwalają określić jedynie poziom wiedzy ucznia, dopiero propozycja 5.3 daje szansę na poznanie rozumowania ucznia. Jednak żeby ocenić, czy ta wersja zadania charakteryzowałaby się lepszymi parametrami psychometrycznymi, należałoby przeprowadzić kolejne badanie.

Podsumowanie

Przeprowadzona krytyczna analiza złożonego zadania zamkniętego pokazała dylematy związane z jego użyciem w testach diagnostycznych. Przestrzeganie zasad tworzenia zadań zamkniętych oraz dążenie do optymalizacji odpowiednich parametrów zadania zachęca do stosowania w formie zadań zamkniętych prostych pytań dotyczących wiedzy i schematów oraz jednowątkowych problemów. Tymczasem zasadniczym celem nauczania matematyki jest rozumowanie matematyczne, u którego podstaw leży logiczne myślenie, umiejętność zastosowania poznanych reguł i zasad matematycznych.

W omawianym zadaniu założono opanowanie przez uczniów określonych podstaw (w tym opanowanie terminologii matematycznej). Poziom wykonania zadania może sugerować, że wielu uczniów nie posiada podstawowej wiedzy matematycznej i podstawowych nawyków rozwiązywania problemów matematycznych. Może też sugerować, że dla części uczniów zadanie było na tyle złożone w swojej strukturze i w warstwie językowej, że nawet nie podejmowali próby jego rozwiązania, o czym może świadczyć duża liczba deklarowanych wyborów odpowiedzi na chybił trafił. Jednocześnie analiza tak udzielonych odpowiedzi uczniowskich nie daje wglądu w sposób rozumowania ucznia, a więc uniemożliwia poznanie blokad hamujących rozwój myślenia matematycznego.

Nasze doświadczenia pokazują, że warto na etapie pilotażu zadań (testu) pozwolić uczniom w zadaniach zamkniętych zadeklarować pewność swojej odpowiedzi. Ujemna wartość wskaźnika pewności wyboru odpowiedzi jest ważnym sygnałem uczniowskich problemów z tym zadaniem.

Bibliografia

Jagodzińska, M. (2008). *Psychologia pamięci. Badania, teorie, zastosowania.* Gliwice: Wydawnictwo Helion.

Jakubowska-Mirek, M., Stożek, E. (2022). *Pewność wyboru odpowiedzi w zadaniach zamkniętych*. W: Niemierko, B., Szmigel, K.M. (red.), *Diagnozowanie kształcenia w edukacji zdalnej i stacjonarnej*, Kraków: PTDE, 327-335.

Haladyna T., Downing S., Rodriguez M. (2002), A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment Applied Measurement in Education, *Applied Measurement in Education*, 15(3), 309-334.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science. Retrieved from Boston College*, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-results/

Niemierko, B. (1999), *Pomiar wyników kształcenia*, Warszawa: WSiP

Rushton, N. (2014). Common errors in Mathematics. *Research Matters: A Cambridge Assessment publication, 17, 8-17.* https://www.cambridgeassessment.org.uk/our-research/all-published-resources/research-matters/rm-17/