

Bartosz Kondrątek

dr Henryk Szaleniec

Instytut Badań Edukacyjnych

Marek Legutko

Wydawnictwa Szkolne i Pedagogiczne

Rola diagnozy w Gimnazjalnym Programie Kształcenia Kompetencji Kluczowych

Wprowadzenie

Gimnazjalny Program Kształcenia Kompetencji Kluczowych GPKKK realizowany w latach 2010-2013 w ramach projektu e-Akademia Przyszłości obejmował cztery obszary aktywności, w których istotną rolę na poszczególnych etapach odgrywała diagnoza.



Rysunek 1. Schemat realizacji projektu e-Akademia Przyszłości. Zaciemnione pola oznaczają obszary związane z diagnozą prowadzą w szkole z wykorzystaniem ITC

Diagnoza obejmowała dwa etapy. Pierwszy etap to diagnoza wstępna, która przeprowadzona została na początku pierwszej klasy rocznika uczniów rozpoczynających naukę w gimnazjum w 2010 roku. Celem tej diagnozy było między innymi dostarczenie dodatkowych informacji przydatnych jako jedno z kryteriów do wyłonienia grupy uczniów, dla których zorganizowane zostały Szkolne Grupy Wyrównawcze (SGW), oraz uzdolnionych uczniów, którzy po półrocznej nauczycielskiej obserwacji w szkole zostali zgłoszeni do wirtualnych kół naukowych prowadzonych przez placówki o charakterze uniwersyteckim.

Szkolne Grupy Wyrównawcze to program działań obejmujący 60 godzin zajęć mających na celu wyrównanie szans na starcie w nowej szkole poprzez podniesienie samooceny, poznanie metod radzenia sobie w sytuacji kryzysowej, rozbudzenie aspiracji, uczenie rozwiązywania problemów w twórczy sposób.

Drugi etap to przeprowadzona trzykrotnie diagnoza obejmująca rocznik uczniów, którzy edukację w gimnazjum rozpoczęli w wrześniu 2010 roku. Badanie w tym etapie obejmowało wykorzystanie baterii testów sumujących:

1. na zakończenie I klasy (maj 2011 r.),
2. na zakończenie II klasy (maj 2012 r.),
3. na zakończenie III klasy (maj 2013 r.).

Testy sumujące zostały przygotowane w formie internetowej przez Ośrodek Diagnostyki Edukacyjnej WSiP. Testy sumujące na zakończenie każdej klasy obejmowały trzy obszary umiejętności: język angielski, matematykę oraz przedmioty przyrodnicze. Z języka angielskiego uczniowie brali udział w teście na poziomie podstawowym albo rozszerzonym. Testy ukierunkowane były na sprawdzenie opanowania kompetencji kluczowych w powiązaniu z wyżej wymienionymi przedmiotami. Prowadzone były on-line w szkolnych pracowniach komputerowych, w wyznaczonych dniach i pod opieką nauczycieli, których zadaniem było między innymi zapewnić rzetelność odpowiadania na pytania testowe.

W artykule omówione zostaną wyniki badań sumujących pod koniec klas 1-3. Wykorzystane w badaniach sumujących testy mierzyły te same umiejętności i zawierały zadania kotwiczące ze sobą sąsiadujące pomiary. Obecność zadań kotwiczących w testach umożliwiła sprowadzenie wyników z różnych badań sumujących do wspólnej skali, co, z kolei, umożliwiło analizę dynamiki zmian poziomu umiejętności uczniów uczestniczących w projekcie między końcem pierwszej a końcem trzeciej klasy. Przed prezentacją wyników analizy przybliżona zostanie ogólna charakterystyka testów, ich plany, sposób raportowania wyników oraz właściwości psychometryczne.

Ogólna charakterystyka testów

Zarówno na pierwszym, jak i na drugim etapie diagnozy, bateria zastosowanych narzędzi składała się z trzech testów:

1. matematyczny (matematyka, informatyka, przedsiębiorczość/WOS),
2. przyrodniczy (fizyka, chemia, biologia, geografia),
3. językowy (język angielski) na poziomie podstawowym i rozszerzonym, począwszy od maja 2011 r.

Autorzy narzędzi starali się przygotować testy tak, aby były one ukierunkowane na badanie kompetencji kluczowych uczniów w powiązaniu z przedmiotami gimnazjalnymi. Do każdego testu (z wyjątkiem testu na zakończenie klasy trzeciej) zostały przygotowane dodatkowe zestawy zadań przeznaczone do rozwiązania w ramach doskonalenia umiejętności przez uczniów po zakończeniu testu. Zadania te swoim zakresem obejmowały **trzy poziomy umiejętności i kompetencji gimnazjalistów**. Zakres zadań zarówno tych wchodzących w skład testów, jak i dodatkowych zestawów był zgodny z założeniami nowej formuły egzaminu gimnazjalnego, do którego uczniowie przystąpili w kwietniu 2013 roku jako już drugi rocznik uczący się zgodnie z wymaganiami nowej podstawy programowej.

Przy konstruowaniu testów przyjęto założenie, że wymagania określone w podstawie programowej obowiązującej na trzecim etapie edukacyjnym od 2009 roku są w swojej treści zgodne z kompetencjami kluczowymi sformułowanymi w Zaleceniu Parlamentu Europejskiego i Rady 2006/962/WE z dnia 18 grudnia 2006 r. Zapewniono także zachowanie zgodności sprawdzanych umiejętności z wymaganiami obowiązującej podstawy programowej, jednocześnie wyróżniając w planach testu odniesienia do kompetencji kluczowych, które są kształcone na lekcjach poszczególnych przedmiotów nauczania.

Ze względu na szeroki zakres kompetencji kluczowych i ograniczenia w możliwości zaobserwowania niektórych umiejętności w sytuacji testowej w żadnym teście nie był badany pełny zakres kompetencji kluczowych. Wprawdzie na każdym z przedmiotów nauczania mogą być kształtowane wszystkie kompetencje kluczowe to jednak na jednym przedmiocie jest wiele możliwości ich kształcenia, a na drugim – w znikomym zakresie.

Wszystkie testy zostały przeprowadzone on-line na platformie WSiP. Testy zawierały zadania zróżnicowane pod względem formy. Między innymi zawierały zadania: wielokrotnego wyboru, typu prawda – fałsz, na dobieranie, z luką oraz krótkiej odpowiedzi. Zadania wymagały czytania tekstów źródłowych i wskazówek, układania elementów/ szeregowania/ porównywania, animacji obrazków/ map/ wykresów, odsłuchiwania nagrań, tworzenia diagramów i wykresów.

Ramowe plany trzech testów każdej z czterech baterii

Wszystkie testy budowane były według takiego samego planu wiążącego treści przedmiotowe i kompetencje kluczowe. Tworzyły tym samym panel złożony z czterech powiązanych części (na rozpoczęcie gimnazjum oraz na koniec każdej klasy). W obrębie każdego testu (matematyka, przedmioty przyrodnicze i język angielski) występowały zadania kotwiczące kolejne testy (4 zadania w przypadku matematyki i przedmiotów przyrodniczych i 3 zadania z języka angielskiego na poziomie podstawowym i rozszerzonym).

Tabela 1. Ramowy plan baterii testów

Test/kompetencje kluczowe	1	2	3	4	5	6	7	8
Matematyka	+	+	+	+	+	+	+	
Przedmioty przyrodnicze	+		+	+	+		+	
Język angielski		+		+	+	+	+	+

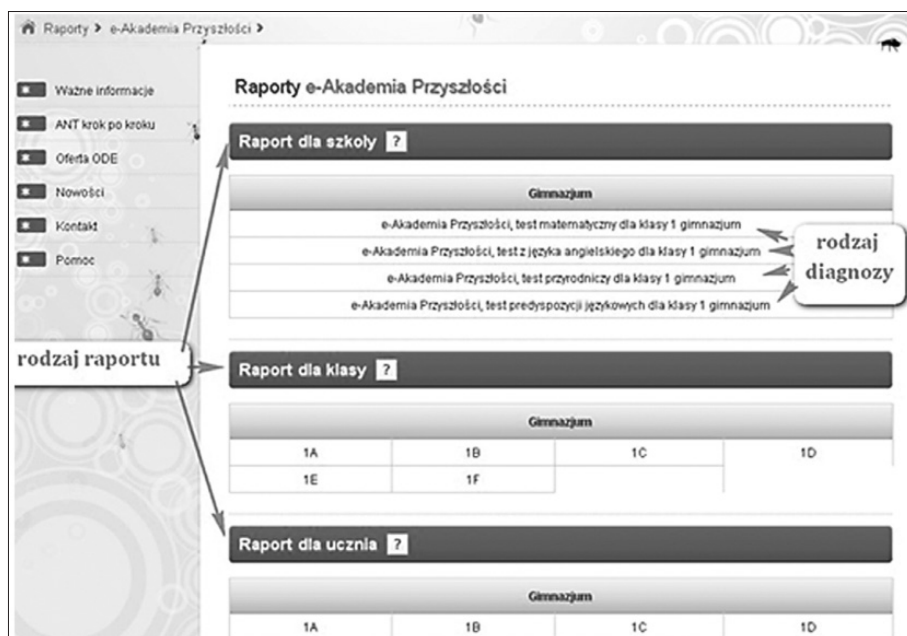
1. Porozumiewanie się w języku ojczystym
2. Porozumiewanie się w językach obcych
3. Matematyczne i podstawowe kompetencje naukowo-techniczne
4. Kompetencje informatyczne
5. Umiejętność uczenia się
6. Kompetencje społeczne i obywatelskie
7. Inicjatywność i przedsiębiorczość
8. Świadomość i ekspresja kulturalna

Przykładowo dla testu matematycznego obejmującego treści z matematyki, informatyki, przedsiębiorczości /WOS założono przedstawione poniżej powiązanie z kompetencjami kluczowymi.

- Porozumiewanie się w języku ojczystym
 - Rozumienie różnych tekstów o charakterze matematycznym
 - Tworzenie różnych tekstów o charakterze matematycznym
- Porozumiewanie się w języku obcym
 - Rozumienie podstawowych określeń z języka angielskiego przy korzystaniu z różnych programów komputerowych
 - Rozumienie podstawowych określeń z języka angielskiego przy korzystaniu z komputera, kalkulatora i innych urządzeń technicznych
- Kompetencje matematyczne i podstawowe kompetencje naukowo-techniczne
 - Przetwarzanie informacji o charakterze matematycznym
 - Stosowanie reguł matematycznych w codziennych sytuacjach życiowych i do rozwiązywania problemów
- Kompetencje informatyczne
 - Posługiwanie się technologiami TI do rozwiązywania problemów matematycznych
 - Posługiwanie się technologiami TI do rozwiązywania problemów z zakresu przedsiębiorczości
- Umiejętność uczenia się
 - Zdolność do zdobywania nowych umiejętności matematycznych
- Kompetencje społeczne i obywatelskie
 - Posługiwanie się terminami matematycznymi w komunikacji międzyludzkiej
- Inicjatywność i przedsiębiorczość
 - Twórcze rozwiązywanie problemów matematycznych
 - Prezentowanie odpowiednich postaw w zakresie przedsiębiorczości.

Komunikowanie wyników uczniom i nauczycielom

Przez wszystkie trzy diagnozy (pod koniec I, II i III klasy) uczniowie posługiwali się takimi samymi loginami, co w konsekwencji umożliwiło połączenie wyników poszczególnych uczniów w kolejnych latach. Wyniki diagnozy były dostępne po zalogowaniu się w systemie informatycznym obsługującym diagnozę, dla trzech głównych adresatów: uczniów, dyrektorów szkół i nauczycieli, których dyrektor wskazał jako odbiorców informacji o wynikach. Uczniowie najwcześniej mieli okazję poznać wyniki, które uzyskali podczas diagnozy. Wynik sumaryczny za cały test każdy uczeń mógł poznać bezpośrednio po zakończeniu rozwiązywania zadań i wysłaniu wyników na serwer. Kilka dni później, po zakończeniu diagnozy, we wszystkich szkołach uczniowie mogli podglądać, jak odpowiadał na poszczególne zadania. To pierwsza bardzo ważna informacja zwrotna z diagnozy, szczególnie że dostępna, po zalogowaniu się, dla każdego ucznia prawie na miesiąc przed zakończeniem roku szkolnego. Uczeń mógł poznać swój wynik sumaryczny w kontekście wyników klasy, szkoły i wszystkich uczniów uczestniczących w projekcie. Nauczyciele i dyrektorzy szkół (również po zalogowaniu się) mieli dostęp do wyników zagregowanych dla szkoły i poszczególnych klas, a także do wyników osiągniętych za poszczególne zadania przez swoich uczniów. Poniżej na rysunku 2 przedstawiona jest organizacja komunikatów o wynikach dostępna na platformie przez wyróżnionych adresatów.

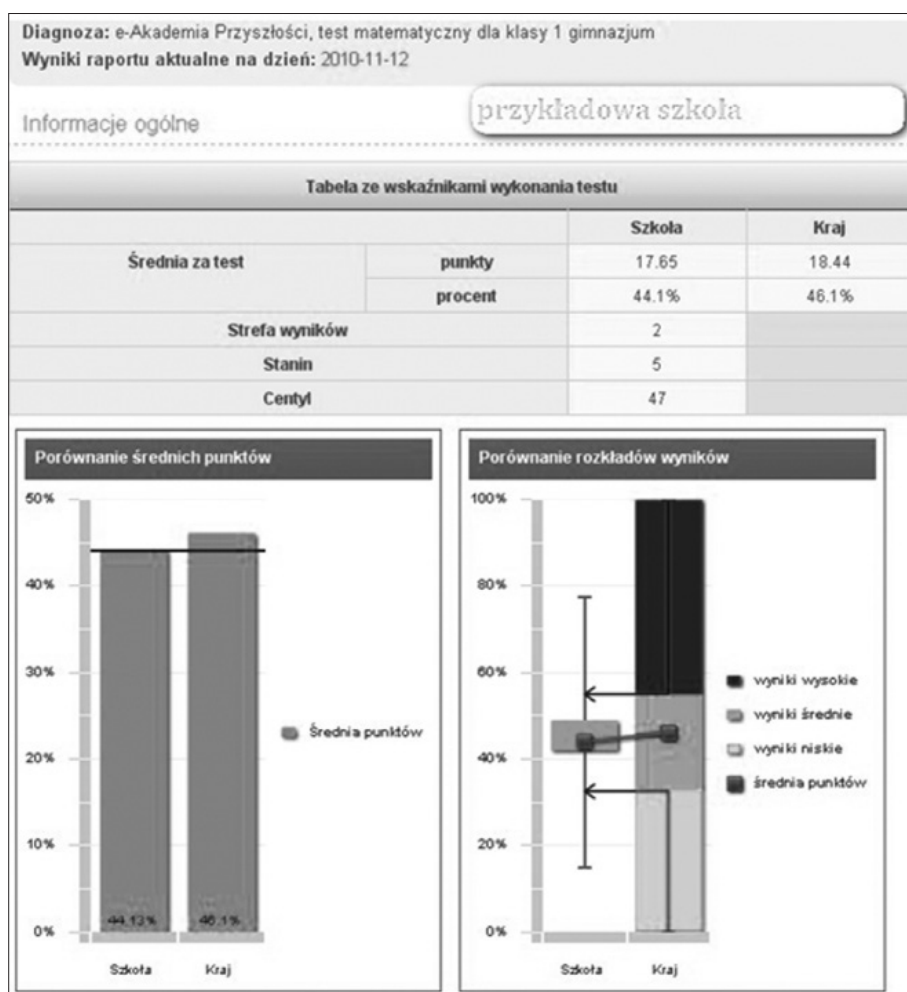


Rysunek 2. Organizacja komunikatów na temat wyników diagnozy

Informacja o wynikach diagnozy dla szkoły, klasy i ucznia podzielona była na pięć bloków. Pierwszy prezentował ogólny wynik diagnozy, drugi wyniki poszczególnych zadań, trzeci wyniki według wymagań ogólnych (kompetencji), czwarty – wyniki według wymagań szczegółowych, piąty – wyniki według typów zadań zastosowanych w teście on-line.

Ogólny wynik diagnozy przedstawiany był tak jak na rysunku 3.

Do wykorzystania w procesie dydaktycznym, zdaniem nauczycieli, istotny był zwłaszcza komunikat, jak konkretne zadania rozwiązywane były przez poszczególnych uczniów, klasy w kontekście wyników uzyskiwanych w całej szkole. Stopień, w jakim dostępna pod koniec roku szkolnego diagnoza była wykorzystywana przez nauczycieli do doskonalenia procesu kształcenia, zapewne w znacznym stopniu zależał od zaangażowania nauczycieli w analizę prezentowanych wyników. Na spotkaniach szkoleniowych nauczyciele deklarowali, że korzystają z raportów z diagnozy.



Rysunek 3. Przykład komunikowania ogólnego wyniku diagnozy dla szkoły

Właściwości psychometryczne narzędzi wykorzystanych w diagnozie

W tabeli 2 przedstawiono informacje charakteryzujące właściwości testów zastosowanych w badaniach sumujących po klasach 1-3. Ze względu na znaczną liczbę wszystkich wykorzystanych w badaniu zadań ograniczono się do podania opisu na poziomie całego testu. Wyszczególniono jedynie liczbę zadań odznaczających się bardzo niską korelacją z resztą testu, tj. zadań o niskiej dyskryminacji. Wszystkie narzędzia składały się z około 20 zadań i można było w nich zdobyć od 30 do 45 punktów.

Najwyższą rzetelnością (α -Cronbacha) charakteryzowały się testy z języka angielskiego – jej wartość była w przedziale od 0,84 do 0,90. Rzetelność testów matematycznych wzrastała od 0,80 w pierwszym pomiarze do 0,86

w pomiarze pod koniec klasy 3. Z największym błędem pomiarowym należy się liczyć w przypadku testów z przedmiotów przyrodniczych pod koniec dwóch pierwszych klas ($\alpha=0,76$). Relatywnie obniżona rzetelność niektórych testów idzie w parze z liczbą zadań, których korelacja z wynikiem w reszcie testu przyjmowała bardzo niskie wartości (przyjęto próg 0,15). Tak nisko dyskryminujące zadania wyłączono z później przeprowadzonego modelowania IRT (*item response theory*).

Tabela 2. Parametry psychometryczne oraz rozkład zadań kotwiczących w testach wykorzystanych w badaniach sumujących po klasach 1-3 gimnazjum

umiejętność	badanie sumujące na koniec:	liczba:		maks. pkt	średnia	SD	α -Cronbacha	liczba zadań:		
		uczniów	zadań					korel. z resztą <0,15	linkujących w dół	linkujących w górę
przedmioty przyrodnicze	1 klasy	18541	23	41	15,98	6,15	0,76	4	---	4
	2 klasy	17435	20	40	16,79	6,03	0,76	1	4	4
	3 klasy	15207	20	40	21,84	7,01	0,84	0	4	---
matematyka	1 klasy	19025	21	41	17,41	7,60	0,80	1	---	3
	2 klasy	18115	21	30	14,29	5,97	0,81	1	3	4
	3 klasy	16055	25	30	16,59	6,60	0,86	0	4	---
język angielski - poziom podstawowy	1 klasy	6983	17	45	20,06	8,21	0,86	0	---	3
	2 klasy	6270	18	30	13,39	7,41	0,89	0	3	3
	3 klasy	5151	20	30	14,07	7,60	0,90	0	3	---
język angielski - poziom rozszerzony	1 klasy	11401	16	40	20,20	9,16	0,88	0	---	3
	2 klasy	10565	18	30	14,45	6,10	0,84	1	3	3
	3 klasy	9233	20	31	11,84	6,24	0,86	0	3	---

Analiza wyników diagnozy w wielogrupowym modelu IRT

Porównywanie średnich wyników w testach sumujących w tabeli 2 nie pozwala na wyciągnięcie wniosków o zmianach w poziomie umiejętności rozwiązyjących je uczniów, gdyż nie jest kontrolowana potencjalnie istniejąca różnica łatwości testów między poszczególnymi pomiarami. Aby rozdzielić efekt zmian w poziomie umiejętności od efektu zmian w poziomie łatwości testów, do danych dopasowano wielogrupowy model IRT (*item response theory*):

$$P(U = u | \mathcal{P}) = \int f(u, \theta, \beta) \psi_{\mathcal{P}}(\theta) d\theta,$$

gdzie θ jest losową zmienną ukrytą opisującą poziom umiejętności uczniów; $\Psi_{\mathcal{P}}(\theta)$ jest funkcją gęstości prawdopodobieństwa określającą rozkład zmiennej θ w populacji \mathcal{P} ; $f(u, \theta, \beta)$ jest funkcją, która określa prawdopodobieństwo zaobserwowania konkretnej wartości u wektora odpowiedzi U , w zależności od poziomu umiejętności θ oraz wektora parametrów $\beta_i = (\beta_1, \beta_2, \dots, \beta_n)$, gdzie parametry zadania β_i również mogą być wektorami (np. dla dwuparametrycznego

modelu logistycznego $\beta_i = (a_p, b_p)$). Różnice w poziomie umiejętności są rozdzielone od różnic w łatwości testów w takim modelu poprzez bezpośrednie wprowadzenie parametrów opisujących te dwa zjawiska, tj. odpowiednio: parametrów rozkładu Ψ_p oraz parametrów zadań β_i . Więcej informacji o wykorzystaniu modelu wielogrupowego IRT do sprowadzania wyników z różnych pomiarów do wspólnej skali można przeczytać np. w: Kolen & Brennan, (2004).

Do zadań ocenianych 0-1 dopasowano krzywe dwuparametrycznego modelu logistycznego (*two-parameter logistic model*, 2PLM), natomiast do zadań ocenianych na większą liczbę punktów dopasowano krzywe modelu oceny częściowej (*graded response model*, GRM). Szczegółowe wytłumaczenie parametryzacji tych modeli oraz zależności kształtu krzywych charakterystycznych od parametrów można znaleźć np. w pionierskich pracach je wprowadzających (Birnbaum, 1968 oraz Samejima, 1969) lub w nowszych opracowaniach tematu IRT (np. de Ayala 2009). W modelowaniu nie uwzględniono zadań, które korelowały z resztą testu poniżej 0,15 (tabela 2). Niestety na liście wykluczonych zadań znalazło się jedno zadanie przyrodnicze i jedno zadanie matematyczne, które kotwiczyło testy sumujące po klasie I z testami sumującymi po klasie II. Oznacza to, że sprowadzenie do wspólnej skali wyników z pierwszego pomiaru do wyników z drugiego pomiaru zostało przeprowadzone jedynie na podstawie odpowiedzi na 3 zadania dla umiejętności przyrodniczych i 2 zadania dla umiejętności matematycznych.

Parametry modelu IRT zostały oszacowane z wykorzystaniem algorytmu *marginalized maximum likelihood* (MML, Book & Aitkin, 1981). Dopasowanie modelu do danych na poziomie zadań oceniono poprzez analizę empirycznych proporcji odpowiedzi z danej kategorii punktowej w zestawieniu z prawdopodobieństwami przewidywanymi przez krzywe charakterystyczne w każdym centylu poziomu umiejętności θ . Empiryczne proporcje były zgodne z prawdopodobieństwami przewidywanymi przez dopasowane krzywe. Ponieważ parametry krzywych charakterystycznych dla zadań w kontekście postawionego problemu badawczego (analiza zmian w poziomie umiejętności) mają znaczenie czysto techniczne oraz zadań było wiele ich omówienie zostanie tu pominięte.

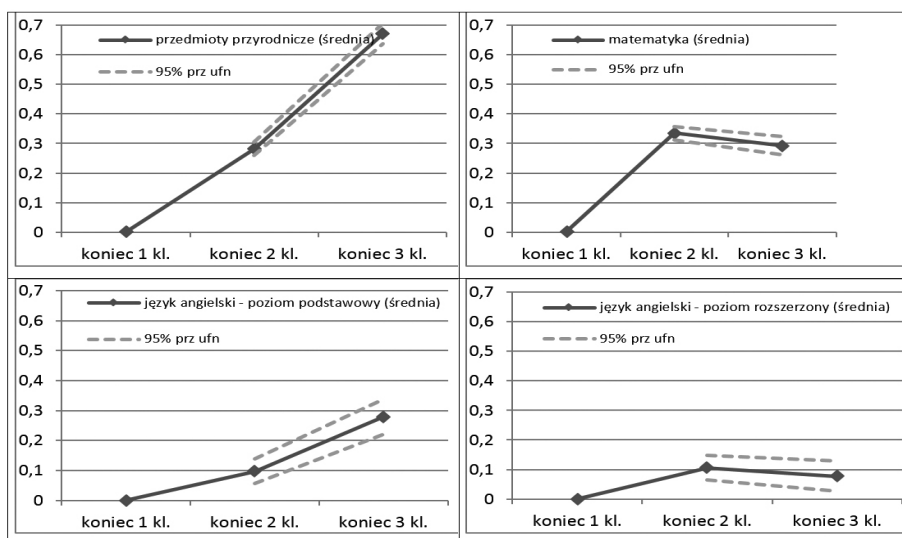
W tabeli 3 zestawiono oszacowane w modelu wielogrupowym średnie dla poziomu umiejętności uczniów na koniec każdej klasy gimnazjum. Rozkład umiejętności uczniów zakotwiczone ze średnią 0 oraz odchyleniem standardowym 1 na pierwszym pomiarze. W związku z tym zakotwiczeniem średnie uzyskane w pomiarze na zakończenie klas 2 oraz 3 informują, o jaką część odchylenia standardowego rozkładu wyników w pierwszym pomiarze zmienił się poziom umiejętności uczniów w porównaniu do stanu na zakończenie klasy 1. Błędy standardowe obliczono jako odchylenie standardowe z 200-krotnej replikacji dopasowania modelu IRT na próbach bootstrap uczniów biorących udział w badaniu. Na wykorzystanie techniki bootstrap zdecydowano się ze względu na bardzo małą liczbę zadań kotwiczących pomiary ze sobą. Analogiczne informacje do tych z tabeli 3 przedstawiono graficznie na rysunku 4.

Tabela 3. Wartości średniego poziomu umiejętności w trzech badaniach sumujących oszacowane w wielogrupowym modelu IRT wraz z błędem standardowym oraz 95% przedziałem ufności

badanie sumujące na koniec:	przedmioty przyrodnicze				badanie sumujące na koniec:	matematyka			
	średnia	SE	95% przedział ufn.			średnia	SE	95% przedział ufn.	
			dolna gr.	górną gr.				dolna gr.	górną gr.
1 klasy	0,000	----	----	----	1 klasy	0,000	----	----	----
2 klasy	0,282	0,012	0,259	0,304	2 klasy	0,335	0,011	0,313	0,356
3 klasy	0,672	0,017	0,638	0,705	3 klasy	0,293	0,016	0,262	0,324
badanie sumujące na koniec:	język angielski - poziom podstawowy				badanie sumujące na koniec:	język angielski - poziom rozszerzony			
	średnia	SE	95% przedział ufn.			średnia	SE	95% przedział ufn.	
			dolna gr.	górną gr.				dolna gr.	górną gr.
1 klasy	0,000	----	----	----	1 klasy	0,000	----	----	----
2 klasy	0,098	0,021	0,057	0,139	2 klasy	0,106	0,021	0,065	0,147
3 klasy	0,280	0,029	0,222	0,337	3 klasy	0,077	0,026	0,026	0,129

Dla testu przyrodniczego obserwujemy wyraźny wzrost poziomu umiejętności – do 0,28 pod koniec klasy drugiej i aż do 0,67 pod koniec klasy trzeciej. Również jednostajny trend wzrostowy, aczkolwiek wyraźnie słabszy, zaobserwowano dla testu z języka angielskiego na poziomie podstawowym – wzrost do 0,10 na koniec klasy drugiej i do 0,28 na koniec klasy trzeciej. Dla języka angielskiego na poziomie rozszerzonym oraz dla matematyki obserwujemy tylko wzrost wyników między pomiarem na koniec klasy 1 a pomiarem na koniec klasy 2 (odpowiednio: 0,33 oraz 0,10). W pomiarze na koniec klasy trzeciej natomiast dla tych dwóch przedmiotów zaobserwowano nieznaczny spadek w porównaniu z wynikami na koniec klasy drugiej (średnie odpowiednio – 0,29 oraz 0,07).

Obniżenie (brak wzrostu) między końcem drugiej a trzeciej klasy dla umiejętności mierzonych testami matematycznym oraz z języka angielskiego na poziomie rozszerzonym jest zastanawiający. W przypadku testu matematycznego pewnym wytłumaczeniem takiego wzorca wyników może być to, że cały test zastosowany w 2012 roku był tematycznie osadzony w Mistrzostwach Europy w Piłce Nożnej 2012, których Polska była współorganizatorem. Mogło to spowodować zwiększoną motywację uczniów piszących test w 2012 roku i zmniejszoną na zadaniach kotwiczących z tego testu wśród uczniów piszących test na zakończenie trzeciej klasy. Taką konstrukcją testu w kontekście pomiaru podłużnego należy uznać za błędną i podważa ona wiarygodność wyników analizy dla umiejętności matematycznych. W przypadku języka angielskiego na poziomie rozszerzonym nie doszukano się wytłumaczenia po stronie konstrukcji testów. Aczkolwiek ze względu na niewielką liczbę zadań kotwiczących do przedstawionych wyników należy podejść z pewną ostrożnością. Możliwe, że inny dobór zadań wykorzystywanych do sprowadzenia wyników z trzech pomiarów do wspólnej skali zaskutkowałby odmiennymi oszacowaniami średnich.



Rysunek 4. Średni poziom umiejętności w trzech badaniach sumujących oszacowany w wielogrupowym modelu IRT wraz z 95% przedziałem ufności

Podsumowanie

Przeprowadzone analizy pozwalają na sformułowanie wniosków o przyroście poziomu kompetencji w odniesieniu do przedmiotów przyrodniczych i języka angielskiego na poziomie podstawowym. W przypadku matematyki i języka angielskiego nie zaobserwowano przyrostu w trzeciej klasie gimnazjalnej. Tak jak już wspomniano wcześniej, w przypadku matematyki można to zjawisko próbować wyjaśnić usterkami konstrukcyjnymi testu zastosowanego w pomiarze podłużnym pomiędzy drugim i trzecim rokiem edukacji. Analiza pomiaru przyrostu umiejętności na poziomie rozszerzonym z języka angielskiego mogłaby być pełniejsza, gdyby kontrolowane były dodatkowe zmienne kontekstowe związane z dosyć powszechnym uczeniem się języka obcego na dodatkowych lekcjach finansowanych przez rodziców.

Trzeba również wziąć pod uwagę niewielką liczbę zadań kotwiczących kolejne pomiary, dlatego wnioski wynikające z badań należy interpretować z pewną ostrożnością.

Jakkolwiek narzędzia zastosowane w prezentowanej diagnozie uczniów uczestniczących w projekcie nie były wolne od metodologicznych niedociągnięć (niewiele zadań kotwiczących, wykorzystanie tych samych zadań w pomiarze powtarzanym, zadania kotwiczące w jednym z pomiarów tematycznie związane z odbywającą się w danym roku imprezą sportową), to uzyskane wyniki można potraktować jako cenną przesłankę w analizie zjawiska dynamiki zmian umiejętności uczniów w trakcie nauki w gimnazjum, gdyż zjawisko to nie było wcześniej w Polsce w sposób systematyczny badane.

Projekty, podobne jak w analizowanym przypadku, mogą w przyszłości stanowić obszar umożliwiający prowadzenie badań w tej dziedzinie.

Bibliografia

1. de Ayala, R. J. (2009), *The Theory and Practice of Item Response Theory*, The Guilford Press: New York - London.
2. Birnbaum, A. (1968), *Some latent trait models*. W: Lord, F., M. & Novick, M., R. (Red.). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
3. Book, R. D. & Aitkin, M. (1981), *Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm*. *Psychometrika* 46(4): 443-459.
4. Kolen, M. J., & Brennan R. L. (2004), *Test equating, scaling, and linking: Methods and practice* (2nd ed.). New York, NY: Springer-Verlag.
5. Samejima, F. (1969), *Estimation of Latent Ability Using a Response Pattern of Graded Scores* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society.