Edukacja 2015, 3(134), 141–156
ISSN 0239-6858

Analiza efektów zastosowania pakietu edukacyjnego *Gramy w piktogramy*

Mirosław Dąbrowski
Instytut Matematyki, Uniwersytet Warszawski*

Bartosz Kondratek
Instytut Badań Edukacyjnych

W artykule przedstawiono analizę wyników badania sprawdzającego skuteczność zastosowania pakietu edukacyjnego *Gramy w piktogramy*. Jest on przeznaczony do wspierania rozwoju umiejętności posługiwania się językiem symbolicznym na etapie edukacji początkowej. Ma on na celu uruchomienie procesu zmiany sposobu nauczania matematyki. Badanie zostało przeprowadzone w schemacie eksperymentalnym z pomiarem powtarzanym i grupą kontrolną, przy randomizacji przeprowadzonej oddziałami wewnątrz szkół. Głównym problemem badawczym poddanym analizie była zmiana w poziomie ogólnego wskaźnika umiejętności posługiwania się językiem symbolicznym uczniów, związana z wdrożeniem pomocy dydaktycznej. W analizach wykorzystano modelowanie IRT oraz regresję wielopoziomową. Wyniki wskazują na istotny statystycznie wzrost umiejętności posługiwania się językiem symbolicznym w grupie eksperymentalnej, związany specyficznie z zastosowaniem badanego pakietu edukacyjnego.

Słowa kluczowe: edukacja matematyczna, styl pracy nauczyciela, modelowanie IRT, regresja wielopoziomowa, eksperyment randomizowany grupami.

Hugo Steinhaus (1887–1972), wybitny polski matematyk i współtwórca polskiej szkoły matematycznej, był znany z celnych powiedzonek i aforyzmów. Na długo przez uruchomieniem przez Unię Europejską programu budowania społeczeństwa wiedzy przestrzegał, że „Kraj bez matematyki nie wytrzyma współzawodnictwa z tymi, którzy uprawiają matematykę", zachęcał także do podnoszenia kompetencji matematycznych i zapewniał: „Po matematyce zrobisz to lepiej". Dlaczego? Przede wszystkim dlatego, że matematyka może, i powinna, uczyć dostrzegania związków, zależności i prawidłowości, wyciągania wniosków i przewidywania, argumentowania i przekonywania, czyli może, i powinna, uczyć myśleć. Jednocześnie, prowadzone badania pokazują, że praktyka polskiej edukacji wczesnoszkolnej skupia się przede wszystkim na „trenowaniu" rozwiązywania typowych zadań wedle ściśle wskazanych przez nauczyciela instrukcji, przy dużym nacisku

Badanie wykonane w ramach projektu „Piktografia. Rozwijanie umiejętności posługiwania się językiem symbolicznym w edukacji z zakresu nauk matematycznych z zastosowaniem piktogramów Asylco" realizowanego w partnerstwie z Uniwersytetem Warszawskim i współfinansowanego ze środków Europejskiego Funduszu Społecznego (Program Operacyjny Kapitał Ludzki 2007–2013, priorytet III: Wysoka jakość systemu oświaty, działanie 3.3.4: Poprawa jakości kształcenia).

* Adres: Banacha 2, 02-097 Warszawa. E-mail: m.dabrowski@mimuw.edu.pl

na arytmetyczną poprawność wykonywanych działań (np. Dąbrowski, 2013). Należy sądzić, że taki styl pracy z najmłodszymi uczniami nie sprzyja optymalnemu rozwojowi ich umiejętności matematycznych, tłumi ich twórczość i odwagę w podejmowaniu wcześniej nienapotkanych problemów matematycznych (Binkowska-Wójcik, Boroń, Brzyska, Cikorska i Fiertek, 2014). Wiele przemawia za tym, że taka sytuacja jest pochodną tradycji edukacyjnej związanej z procesem nauczania matematyki, dominującej w naszym społeczeństwie i w polskiej szkole – nie tylko w początkowym okresie nauczania (Karpiński, Grudniewska i Zambrowska, 2013). Tradycji, zgodnie z którą dziecko nie jest w stanie samodzielnie do niczego dojść i wie tylko to, czego dowiedziało się od dorosłych. *Gramy w piktogramy* stworzono m.in. po to, aby przyczynić się do uruchomienia procesu odchodzenia od tej tradycji.

Gramy w piktogramy

Pakiet edukacyjny *Gramy w piktogramy* został tak zaprojektowany i skonstruowany, aby skłaniał wykorzystujących go nauczycieli do zmiany swojego dotychczasowego stylu pracy, dawał uczniom więcej autonomii i zachęcał ich do większego zaangażowania intelektualnego w proces kształcenia – także do częstszego uczenia się we współpracy – oraz wspierał umiejętność efektywnego posługiwania się językiem symbolicznym. Na potencjalną skuteczność tych zmian wskazują prowadzone współcześnie badania nad uczeniem się mózgu dziecięcego (Gopnik, Meltzoff i Kuhl, 2004), nad prawidłowościami uczenia się matematyki (Sfard, 2008), a także nad uczeniem się we współpracy (Kołodziejczyk, Salamon-Bobińska, Karaszewski i Bobula, 2014; Slavin, 2012). Także analizy prowadzone pod kierunkiem Johna Hattiego (2009; 2011) pokazują, że takie interwencje edukacyjne mają dużą skuteczność.

Aby dzieci mogły efektywnie i ze zrozumieniem uczyć się matematyki, powinny być aktywne intelektualnie, umieć współpracować i rozmawiać o tym, co robią, ponieważ uczenie się tego przedmiotu jest procesem społecznym. Pakiet edukacyjny ma skutecznie zachęcić uczniów będących na początkowych etapach edukacji do myślenia i działania, dlatego uwzględniono w nim ich edukacyjne potrzeby. Między innymi dlatego bardzo często *Gramy w piktogramy* wykorzystuje reprezentację enaktywną i ikoniczną, przygotowującą do rozumienia symboliki matematycznej. Z drugiej strony, pakiet miał zainteresować dzieci i pobudzić ich motywację do uczenia się. Aby to osiągnąć, m.in. konsekwentnie proponowano pracę w parach i grupach, sięgano także po zadania o zróżnicowanym poziomie trudności, niekiedy innym od tego, jaki na co dzień oferuje szkoła. Wśród tych zadań regularnie występowały zadania otwarte – w naturalny sposób uruchamiające wyjaśnianie i argumentowanie, zadania problemowe – niejednokrotnie o interdyscyplinarnym charakterze. Zarówno proponowana organizacja pracy, jak i typy zadań miały także na celu zmianę stylu komunikowania się nauczyciela i uczniów, uruchomienie rzeczywistego procesu rozmawiania o matematyce podczas zajęć – zarówno uczniów ze sobą, jak i nauczyciela z uczniami oraz uczniów z nauczycielem.

Cele założone na etapie projektowania narzędzia zrodziły potrzebę przekonania nauczycieli, którzy będą wykorzystywali pakiet, że powinni zmienić swój styl pracy i umożliwić dzieciom inny sposób funkcjonowania niż na typowych zajęciach. Dlatego na etapie testowania narzędzia prowadzono z nauczycielami szkolenia (pierwsze przed rozpoczęciem pracy z pakietem, drugie w połowie roku szkolnego), stworzono im także możliwość stałych konsultacji.

Piktogramy Asylco, ze względu na swój specyficzny styl komunikowania znaczenia oraz symboliczną „wielopoziomowość”, dobrze pasowały do przedstawionych założeń. Pakiet edukacyjny *Gramy w piktogramy* został przygotowany w trzech wariantach: dla klas 1–3 i 4–6 szkoły podstawowej oraz dla gimnazjum. Składa się on z zestawu pomocy dla uczniów oraz materiałów dla nauczyciela. Zestaw pomocy przeznaczony jest dla czteroosobowej grupy uczniów (to kolejny zabieg zachęcający do współpracy) i zawiera m.in.: bogaty zestaw piktogramów o różnym poziomie umowności; stemple z piktogramami do wykorzystania np. podczas rozwiązywania i układania zadań; gry (plansze, pionki, kostki) rozwijające np. rozumienie systemu dziesiętnego; tabliczki suchościeralne i flamastry do zapisywania rozwiązań zadań, projektowania piktogramów. W skład pakietu dla nauczyciela wchodzą: przewodnik dla nauczycieli; zestaw proponowanych scenariuszy zajęć; zestaw kart pracy o trzech poziomach trudności, służący m.in. do indywidualizacji pracy uczniów; zestaw pomocy dla nauczyciela, który zawiera m.in. piktogramy demonstracyjne, naklejki z piktogramami, modele wag oraz programy komputerowe wspierające rozwój umiejętności matematycznych uczniów. Ponadto opracowano elektroniczną wersję z materiałami do pobrania oraz szkolenie internetowe dla nauczycieli chcących korzystać z pakietu (zob. www.piktografia.pl).

Pytania i hipotezy badawcze

Aby sprawdzić skalę efektów zastosowania tego narzędzia, należy odpowiedzieć na pytanie, czy zastosowanie tego pakietu wiąże się ze zwiększeniem uczniowskich kompetencji w rozwiązywaniu zadań matematycznych, wymagających posługiwania się językiem symbolicznym i porównać rozwój tych umiejętności z sytuacją, w której nauczyciele nie korzystają z pakietu, a w pracy z uczniami wykorzystują tradycyjne metody nauczania matematyki. Tak postawiony problem skłonił do przeprowadzenia badania w schemacie eksperymentalnym z grupą kontrolną, w której nauczyciele nie korzystają z pakietu edukacyjnego *Gramy w piktogramy*. Na podstawie danych zebranych w tym badaniu starano się zweryfikować hipotezę mówiącą o wzroście umiejętności posługiwania się językiem symbolicznym w klasach eksperymentalnych, który wynika z zastosowania *Gramy w piktogramy*.

Ze względu na opisaną charakterystykę pakietu należy rozważyć dwojaki mechanizm potencjalnego oddziaływania pakietu na uczniów:

- bezpośredni, poprzez wykorzystywanie na lekcjach konkretnych pomocy edukacyjnych i uczestniczenie w sytuacjach edukacyjnych, które skonstruowano z myślą o wspieraniu rozwoju umiejętności myślenia symbolicznego,
- pośredni, poprzez zmiany w postawach i stylu pracy nauczycieli, które przekładają się na wprowadzenie bardziej efektywnych metod nauczania matematyki i ogólnie rozumianego wspierania rozwoju umiejętności matematycznych uczniów.

Mimo że takie rozróżnienie nie jest ściśle rozłączne (decyzję nauczyciela o wykorzystaniu pomocy można przecież uznać za zmianę w jego stylu pracy), wydaje się istotne ze względu na zrozumienie natury czynników odpowiedzialnych za ewentualną wartość dodaną wynikającą ze stosowania pakietu. Ujmując to inaczej – istotnym uzupełnieniem postawionej hipotezy byłoby pytanie, na ile ewentualne zmiany w kompetencjach uczniów są bezpośrednim efektem zastosowania pakietu, a na ile efektem zapośredniczonym poprzez zmiany, jakie nastąpiły w postawach nauczycieli, dzięki stosowaniu pakietu oraz dzięki udziałowi w szkoleniach. Ze względu na niewielką liczbę klas (i nauczycieli) biorących udział w badaniu eksperymentalnym, rozstrzygająca ilościowa analiza hipotez związanych z pośrednim oddziaływaniem pakietu niestety nie była możliwa. Dlatego główny problem badawczy poddawany analizie dotyczy jedynie występowania wspomnianej wartości dodanej wynikającej z zastosowania *Gramy w piktogramy*. Jednak do tematu potencjalnych zmian w postawach nauczycieli z grupy eksperymentalnej jeszcze powrócimy.

Opis planu badania eksperymentalnego

Na schemat badania eksperymentalnego wpłynęły następujące czynniki organizacyjne i ekonomiczne:

- Podstawową jednostką poddawaną manipulacji eksperymentalnej była cała klasa (oddział szkolny); założono występowanie istotnej korelacji wewnątrzklasowej zmiennej zależnej rzędu 0,15;
- Ze względu na towarzyszący ilościowemu badaniu eksperymentalnemu rozbudowany plan kosztownych badań jakościowych, ustalono, że oddziaływaniem eksperymentalnym zostanie objętych jedynie 8 klas[1].
- Ograniczenie możliwości doboru uczniów do warunków eksperymentalnych jedynie całymi klasami, przy istotnej korelacji wewnątrzklasowej, redukuje efektywną wielkość próby. Do oszacowania stopnia, w jakim wielopoziomowa struktura danych redukuje moc wnioskowania statystycznego służy wzór na tzw. efekt planu (*design effect*):

$$D = 1 + (m - 1)\rho, \tag{1}$$

gdzie: m to liczba osób w grupie; ρ – współczynnik korelacji wewnątrzklasowej.

Zakładając m w zakresie 15–20 oraz $\rho = 0{,}15$, oszacowano, że efektywna wielkość próby, $n_{ef} = {}^{n}/_{D}$, wyniesie w ok. 40 uczniów na każdy warunek eksperymentalny.

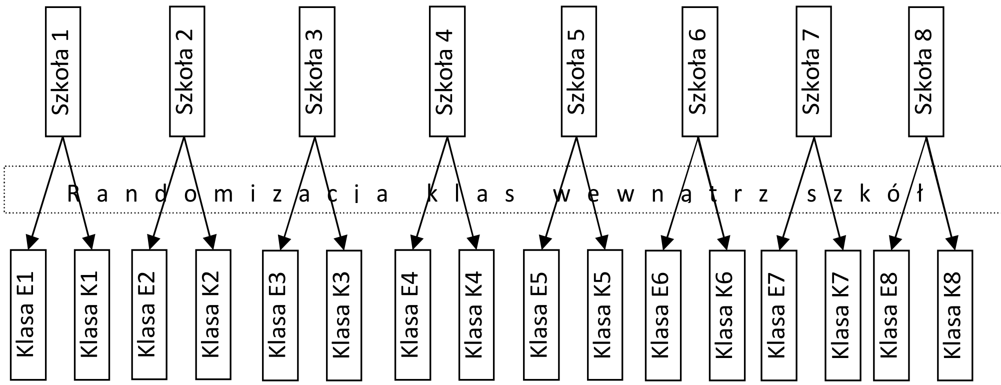
W obliczu nałożonych ograniczeń uznano, że najefektywniejszą (pod względem mocy statystycznej) strategią badania efektów oddziaływania będzie przeprowadzenie randomizowanego grupami eksperymentu z grupą kontrolną, z zastosowaniem zrównywania/dopasowania (*group/cluster-randomized experiment with pair matching*) oraz z pomiarem początkowym i końcowym zmiennej zależnej. Ogólny schemat takiego badania wpisuje się w postać eksperymentu z pomiarem powtarzanym oraz grupą kontrolną:

- grupa eksperymentalna: pretest → oddziaływanie eksperymentalne → posttest,
- grupa kontrolna: pretest → brak oddziaływania eksperymentalnego → posttest.

Przy czym rozdział uczniów do obu grup następuje całymi klasami, z randomizacją dokonaną wewnątrz szkoły (założono, że badaniem eksperymentalnym zostaną objęte szkoły mające przynajmniej dwa równoległe oddziały). Włączenie do badania grupy kontrolnej było konieczne ze względu na potrzebę kontroli efektu dojrzewania (*maturation*). W interwale między pierwszym a drugim pomiarem spodziewano się bowiem występowania pewnego „naturalnego" wzrostu poziomu umiejętności matematycznych uczniów, niezależnego od oddziaływania eksperymentalnego. Schemat rozdziału klas do grupy eksperymentalnej i kontrolnej w 8 szkołach objętych badaniem zilustrowano na Rysunku 1.

Wariancja między klasami tej samej szkoły jest istotnie niższa od ogólnej wariancji między klasami. Pary oddziałów wewnątrz szkół są bardziej do siebie podobne niż pary losowo dobrane z całej populacji. To uzasadniało przeprowadzanie randomizacji wewnątrz szkoły. Zastosowanie zrównywania parami zredukuje tym samym prawdopodobieństwo uzyskania podziału na grupy (kontrolną i eksperymentalną),

[1] W literaturze wskazuje się, że jest to minimalna liczebność, na której można przeprowadzić eksperyment z randomizowanymi grupami, i jednocześnie zapewnić moc statystyczną wystarczającą do wykrywania istotnych z praktycznego punktu widzenia efektów (Murray, Varnell i Blitstein, 2004).



Rysunek 1. Założony przydział oddziałów do grupy kontrolnej i eksperymentalnej w badaniu.

które różniłyby się znacznie pod względem wyjściowego poziomu zmiennej zależnej. Dodatkowo randomizacja wewnątrz szkół zapewniła zrównoważenie obu badanych grup ze względu na zmienne środowiskowe związane z lokalizacją szkoły. Jest to procedura często zalecana w eksperymentach z randomizowanymi grupami (Lipsey i Hurley, 2009). Poza zrównywaniem ze względu na istotne dla badania zmienne pozwala ona częściowo zniwelować obniżenie efektywności wynikającej z istotnej korelacji wewnątrzklasowej (Imai, King i Nall, 2009). Warto zauważyć, że randomizacja wewnątrz tych samych szkół wprowadza pewne zagrożenie dla trafności wewnętrznej eksperymentu (*contamination*), wynikające z możliwości przepływu informacji między nauczycielem objętym oddziaływaniem eksperymentalnym a nauczycielem z grupy kontrolnej. Uznano jednak, że potencjalne straty będą mniejsze od zysków wynikających z takiego rozwiązania.

Innym istotnym aspektem proponowanego planu eksperymentalnego jest możliwość przeprowadzenia powtarzanego pomiaru na tych samych uczniach (pretest i posttest). Takie rozwiązanie niesie za sobą ewidentne korzyści metodologiczne (możliwość kontroli wyjściowych różnic między grupą eksperymentalną a kontrolną), jak i czysto statystyczne, polegające na zwiększeniu mocy statystycznej przy wnioskowaniu z pomiarów zależnych. Stosując pomiar powtarzany, należy jednak zadbać o kontrolę efektu pamięci, jaki by wystąpił, gdyby

Tabela 1
Schemat podziału zadań między narzędzia wykorzystane w preteście i postteście, umożliwiający połączenie (link) wyników uczniów z różnych grup

Próba\zbiór zadań	CKE blok zad. A	CKE blok zad. B	CKE blok zad. C	Zadanie dodatkowe 1	Zadanie dodatkowe 2
Badanie CKE	✓	✓	✓		
Pretest (grupy E i K)		✓		✓	
Posttest (grupy E i K)			✓		✓

uczniowie rozwiązywali dwukrotnie ten sam test umiejętności matematycznych. Problem ten rozwiązano, wykorzystując w pierwszym i drugim pomiarze testy umiejętności matematycznych złożone z różnych zadań. Stworzyło to jednak kolejne wyzwanie – konieczność kontrolowania trudności odmiennych narzędzi. W tym celu w preteście i postteście umieszczono porcję zadań z ogólnopolskiego *Badania umiejętności podstawowych uczniów trzecich klas szkół podstawowych* przeprowadzanego w 2008 r. w Centralnej Komisji Egzaminacyjnej (Dąbrowski, 2009). Badanie to zostało przeprowadzone na licznej ($N = 3965$), reprezentatywnej i ogólnopolskiej próbie uczniów oraz uwzględniało dużą liczbę zadań matematycznych (8 różnych arkuszy), których wyniki następnie wykalibrowano w modelach IRT (Kondratek, 2009). W Tabeli 1 przedstawiono ogólny schemat konstrukcji obu narzędzi wykorzystanych w preteście i postteście. Nie zawierają one wspólnych zadań, ale ze względu na zakotwiczenie w badaniu przeprowadzonym w CKE umożliwiły oszacowanie poziomu umiejętności matematycznych uczniów na wspólnej skali.

Zadania wykorzystane w badaniu CKE w tym schemacie podzielono na trzy rozłączne zbiory: „blok zadań A", które nie zostały wykorzystane w eksperymencie, ale zostały wykorzystane w badaniu CKE; „blok zadań B", które zostały włączone do pretestu; oraz „blok zadań C", które zostały włączone do posttestu. W celu zwiększenia rzetelności i trafności narzędzi wykorzystanych w badaniu eksperymentalnym, do pretestu, oprócz zadań z bloku B, włączono „zadanie dodatkowe 1"; analogicznie postąpiono w wypadku postestu – oprócz zadań z bloku C, włączono „zadanie dodatkowe 2".

Przyjmując warunki założone w trakcie projektowania eksperymentu (efektywna liczebność 40 uczniów na warunek eksperymentalny; poziom istotności $\alpha = 0{,}05$; kierunkowa hipoteza alternatywna; współczynnik ICC = 0,15 oraz, dodatkowo, korelacja między dwoma pomiarami wynosząca 0,5) oszacowano, że w tak zaplanowanym eksperymencie zakresowi mocy statystycznej z przedziału od 0,8 do 0,95 odpowiada wykrywalność efektów $f(V)$ w zakresie odpowiednio od 0,28 do 0,37. Taka czułość zaplanowanego eksperymentu została uznana za zadowalającą. Do obliczeń wykorzystano moduł do analizy mocy testu F dla efektu interakcyjnego ANOVA z pomiarami powtarzanymi dostępny w programie G*power (Faul Erdfelder, Lang i Buchner, 2007). Należy zauważyć, że te szacunki miały charakter orientacyjny ze względu na wiele uproszczeń (np. efektywna liczebność upraszczająca problem skupień przy obliczeniach związanych z mocą statystyczną czy zignorowanie problemu nierzetelności miar) oraz założeń dokonywanych a priori (wartość ICC, korelacja między pomiarami, liczebność oddziałów). Po przeprowadzeniu badania okazało się, że oszacowane wartości były bardzo zbliżone do przyjętych przed analizą: ICC w pierwszym pomiarze 0,08, w drugim pomiarze 0,17 (wzrost ICC odzwierciedla zwiększenie zróżnicowania między grupą eksperymentalną a kontrolną); korelacja między pomiarami wyniosła 0,44; liczebność oddziałów w pierwszym pomiarze wynosiła od 10 do 26 ($M = 20{,}6$), w drugim pomiarze od 13 do 25 ($M = 19{,}5$).

Zastosowane narzędzia badawcze

Podczas testowania skupiono się na trzech obszarach, które uznano za najistotniejsze z punktu widzenia weryfikacji skuteczności wykorzystania pakietu: modelowaniu matematycznym, rozumieniu pojęć i umiejętności posługiwania się nimi oraz rozwiązywaniu problemów z wykorzystaniem procesów poznawczych istotnych dla myślenia matematycznego. Rozległość tych obszarów zmusiła do dokonania ich egzemplifikacji.

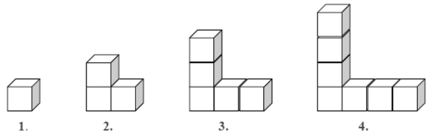
W przypadku modelowania matematycznego, czyli stosowania narzędzi matematycznych, w tym języka symbolicznego do opisywania badanego zjawiska, postanowiono skupić się na rozwiązywaniu nietypowych zadań tekstowych. Rozwiązywanie zadań tekstowych to najbardziej zaawansowany przejaw modelowania matematycznego w początkowym okresie edukacji. Wybór zadań nietypowych, czyli tradycyjnie nieobecnych podczas lekcji, miał wyeliminować ewentualny efekt wcześniejszego „wytrenowania". Oto jedno z wykorzystanych zadań tego typu:

> Jacek i Wojtek mieli po tyle samo lizaków. Wojtek oddał Jackowi dwa swoje lizaki. Teraz więc Jacek ma więcej lizaków niż Wojtek. O ile więcej?

Jednym z najważniejszych celów edukacji matematycznej jest budowanie rozumienia systemu dziesiętnego. Jego relacyjne, a nie zdegenerowane[2], rozumienie jest kluczowe dla całej pojawiającej się w szkole arytmetyki, z posługiwaniem się algorytmami obliczeniowymi włącznie. Dodatkowo zapis liczb w systemie dziesiętnym to najczęstszy szkolny przykład praktycznego zastosowania języka symbolicznego. W celu zbadania, w jakim stopniu uczniowie rozumieją system dziesiętny, sięgnięto po następujące zadanie:

> W tych liczbach dwucyfrowych zamazano niektóre cyfry. Tam, gdzie to możliwe, wstaw w okienko znak „>" albo „<". W pozostałe okienka wstaw znak zapytania: „?".
> a) 7●□48 b) ● 6□33 c) 6●□●2

Na koniec przytaczamy jedno z wykorzystanych rozbudowanych zadań o charakterze problemowym:



> Te budowle powstały z identycznych drewnianych klocków. Zbudowano je zgodnie z pewną regułą. Odgadnij, jaka to reguła.
> a) Z ilu klocków powinna się składać następna taka budowla?
> b) Ile klocków potrzeba do zbudowania dziesiątej takiej budowli?
> c) A ile potrzeba do zbudowania dwudziestej budowli z tej serii?
> d) Opisz, jak można szybko obliczyć, ile klocków potrzeba do zbudowania dwudziestej budowli z tej serii.

Zadanie tego typu wymaga zauważenia prawidłowości definiującej sekwencję brył, wykorzystania jej w prostej sytuacji oraz w sytuacjach stopniowo coraz bardziej skomplikowanych, zmuszających do dokonania np. generalizacji, i – wreszcie – zbudowania możliwie jednoznacznego wyjaśnienia, czy nawet argumentacji, także z wykorzystaniem języka symbolicznego.

W Tabeli 2 zestawiono zadania z reprezentatywnego badania umiejętności trzecioklasistów przeprowadzonego przez CKE w 2008 r., które wykorzystano do konstrukcji testu w pierwszym oraz drugim badaniu umiejętności uczniów, w podziale na trzy opisane powyżej obszary umiejętności. Zadania te odpowiadają blokom „B" oraz „C" w schemacie przedstawionym w Tabeli 1. Przytoczonym powyżej przykładowym zadaniom w Tabeli 1 odpowiadają odpowiednio oznaczenia: M1B_6, M2B_5a–c oraz M2B_7s.

Dobór próby i czas przeprowadzenia badania

Badanie zostało przeprowadzone na poziomie klas trzecich i trwało rok. Wybór poziomu wiekowego był podyktowany zakresem wykorzystania pakietu edukacyjnego *Gramy w piktogramy* w procesie kształcenia – w klasie trzeciej możliwości jego zastosowania są największe. Pierwszy pomiar umiejętności uczniów przeprowadzono we wrześniu 2012 r., a drugi na przełomie maja i czerwca 2013 r.

---
2 W sensie zdegenerowanego formalizmu, który przejawia się tym, że uczeń traktuje zapis symboliczny, np. liczby dwucyfrowej, jako „obrazek" i odnosi się do jego wyglądu, a nie sensu.

Tabela 2
Wykorzystanie zadań z reprezentatywnego badania CKE do konstrukcji pretestu oraz posttestu

Informacje o zadaniach			Liczba obserwacji w pomiarach				
Kod zadania w badaniu CKE	Obszar umiejętności*	Maks. liczba punktów	Badanie 2008	Pretest (grupa K)	Pretest (grupa E)	Posttest (grupa K)	Posttest (grupa E)
M2C_7b	DPW	1	940	170	160		
M2C_7as	DPW	4	940	170	160		
M2B_7s	DPW	4	1 046			163	149
M2B_5a	SD	1	1 046	170	160		
M2B_5b	SD	1	1 046			163	149
M2B_5c	SD	1	1 046			163	149
M2B_5d	SD	1	1 046	170	160	163	149
M1B_6	NZT	1	1 046	170	160		
M1A_6	NZT	1	1 078			163	149
M2A_6	NZT	1	1 077			163	149
M2C_6	NZT	1	940	170	160		

* DPW – dostrzeganie prawidłowości i wyjaśnianie; SD – system dziesiętny; NZT – nietypowe zadania tekstowe.

Wiele czynników organizacyjnych, ekonomicznych oraz merytorycznych ograniczało możliwość losowego doboru szkół do opisywanego badania eksperymentalnego. Wśród czynników organizacyjnych i ekonomicznych należy wymienić konieczność ograniczenia realizacji badania do trzech województw (małopolskie, mazowieckie, pomorskie) oraz do szkół mających co najmniej dwa oddziały uczniów w klasach trzecich. Możliwość doboru szkół do badania w dalszej kolejności ograniczało również przyjęcie założenia, że ze względu na konieczność znacznego zaangażowania szkół w proces badawczy, zostaną one wybrane spośród listy szkół dobrowolnie zgłaszających chęć udziału w badaniu. Ponadto, przy tak niewielkiej liczbie szkół przewidzianej do badania (8 placówek), zdecydowano, że przeprowadzenie prostego losowego doboru szkół do próby badawczej niosłoby duże ryzyko uzyskania próby szkół znacznie odbiegającej od charakterystyk ogółu populacji szkół prowadzących oddziały klas trzecich w Polsce. Mając na względzie powyższe warunki brzegowe, aby zmniejszyć zagrożenia dla trafności zewnętrznej badania, przyjęto, że spośród szkół zgłaszających swoje uczestnictwo w programie badawczym, 8 szkół zostanie dobranych do badania w taki sposób, aby wypełnić każdą z kombinacji powstałych przez skrzyżowanie następujących dwóch zmiennych:

- lokalizacja szkoły, która przyjmuje dwie wartości: (a) wieś i miasta poniżej 10 tys. (b) miasta powyżej 10 tys.,
- średni poziom szkoły, który przyjmuje cztery wartości powstałe przez podział średnich wyników szkół uzyskanych w *Badaniu umiejętności podstawowych uczniów trzecich klas szkół podstawowych* na równoliczne ćwiartki za pomocą kwartyli.

Decyzja o podziale lokalizacji szkoły na takie dwie warstwy była podyktowana bardzo zbliżonymi wynikami szkół wiejskich i szkół usytuowanych w miastach poniżej 10 tys. mieszkańców (w porównaniu z większymi miastami), jaką zaobserwowano m.in. w *Ogólnopolskim badaniu umiejętności trzecioklasistów* (OBUT; Pregler i Wiatrak, 2011). Ponadto podział lokalizacji szkół na takie dwie warstwy dzieli populację uczniów w Polsce w przybliżeniu na połowę, dzięki czemu każde z 8 pól powstałych przez skrzyżowanie obu zdefiniowanych zmiennych odpowiada w przybliżeniu 1/8 populacji uczniów klas trzecich w Polsce.

W kontekście omawianego badania eksperymentalnego przyjęto średni wynik szkoły na skali umiejętności matematycznych, jaki był przedstawiony w raportach dla szkół w badaniach OBUT 2011, jako najlepszą dostępną miarę „średniego poziomu szkoły". Dotyczy on umiejętności matematycznych uczniów klas trzecich. Alternatywnie rozważano wykorzystanie w tym celu wyników ze sprawdzianu po klasie szóstej szkoły podstawowej. Zaletą wykorzystania wyników sprawdzianu do warstwowania szkół byłaby powszechność tego egzaminu (w badaniach OBUT szkoły uczestniczą dobrowolnie – nie obejmuje ono całej populacji). Jednak zadecydowano się na warstwowanie ze względu na wyniki części matematycznej OBUT, ponieważ sprawdzian miał charakter ponadprzedmiotowy i dotyczył umiejętności uczniów w klasie szóstej. W Tabeli 3 przedstawiono zakresy średnich wyników szkół w badaniu OBUT 2011, jakie uzyskano przy opisanym podziale szkół na 8 wartości.

Spośród 34 szkół, które zgłosiły chęć udziału w badaniu i prowadziły co najmniej dwa oddziały klas trzecich, udało się wypełnić jedynie 7 komórek Tabeli 3. Dla warunku: I ćwiartka + wieś i miasta poniżej 10 tys. mieszkańców nie zgłosiła się żadna taka szkoła. Dlatego zdecydowano się w jej miejsce zrekrutować dwie różne szkoły jednooddziałowe, które razem spełniały ten warunek. Jedną szkołę wylosowano do grupy kontrolnej, drugą do grupy eksperymentalnej.

Zastosowane metody statystycznej analizy danych

Przeprowadzoną analizę danych eksperymentalnych można podzielić na trzy etapy:

- dopasowanie do danych wielogrupowego modelu IRT,
- wygenerowanie dla każdego ucznia w każdym z pomiarów kompletu wartości możliwych (*plausible values*, PV) do późniejszego wykorzystania jako wskaźnik poziomu umiejętności,
- oszacowanie parametrów modelu wielopoziomowej regresji liniowej, w której zmienną zależną był poziom umiejętno-

Tabela 3

Zakres średnich wyników szkół podstawowych na ogólnej skali umiejętności matematycznych w badaniach OBUT 2011 (skala: 100; 15) w ćwiartkach podzielonych ze względu na lokalizację[*]

	Zakres średnich wyników szkół			
Lokalizacja szkół	I ćwiartka	II ćwiartka	III ćwiartka	IV ćwiartka
Wieś i miasta poniżej 10 tys. mieszk.	poniżej 93,8	[93,8;98,2)	[98,2;103,2]	powyżej 103,2
Miasta powyżej 10 tys. mieszk.	poniżej 97,5	[97,5;101,1)	[101,1;105,0]	powyżej 105,0

[*] Przy obliczaniu rozkładu średnich wyników szkół, wyłączono szkoły z mniejszą liczbą uczniów niż 5.

ści a zmiennymi niezależnymi: przynależność do poszczególnych warunków eksperymentalnych oraz zagnieżdżenie uczniów w pomiarach i w szkołach.

Każdy z wymienionych etapów zostanie pokrótce opisany. Przyglądając się Tabelom 1 i 2, widzimy, że pełna macierz danych zawiera odpowiedzi na zadania, które zostały udzielone przez uczniów pochodzących z pięciu różnych grup: uczniów z reprezentatywnego badania CKE w 2008 r. oraz z czterech grup powstałych przez skrzyżowanie dychotomicznych warunków: „grupa kontrolna–eksperymentalna” oraz „przed–po oddziaływaniu eksperymantalnym”. Jednocześnie narzędzia wykorzystane w badaniu 2008 r., w preteście i w postteście składają się z różnych zadań. Aby uwzględnić w modelu statystycznym zmiany w poziomie umiejętności pomiędzy badanymi grupami w takim schemacie, wykorzystano metody charakterystyczne dla zrównywania wyników testowych (Kolen i Brennan, 2004, Pokropek i Kondratek, 2012, Szaleniec, Grudniewska, Kondratek, Kulon i Pokropek, 2012).

Do zbioru danych dopasowano wielogrupowy jednowymiarowy model IRT (*item response theory*), który ma postać:

$$P(\boldsymbol{U}=\boldsymbol{u}|P)=\int f(\boldsymbol{u},\theta,\boldsymbol{\beta})\ \psi_P(\theta)\ d\theta, \qquad (2)$$

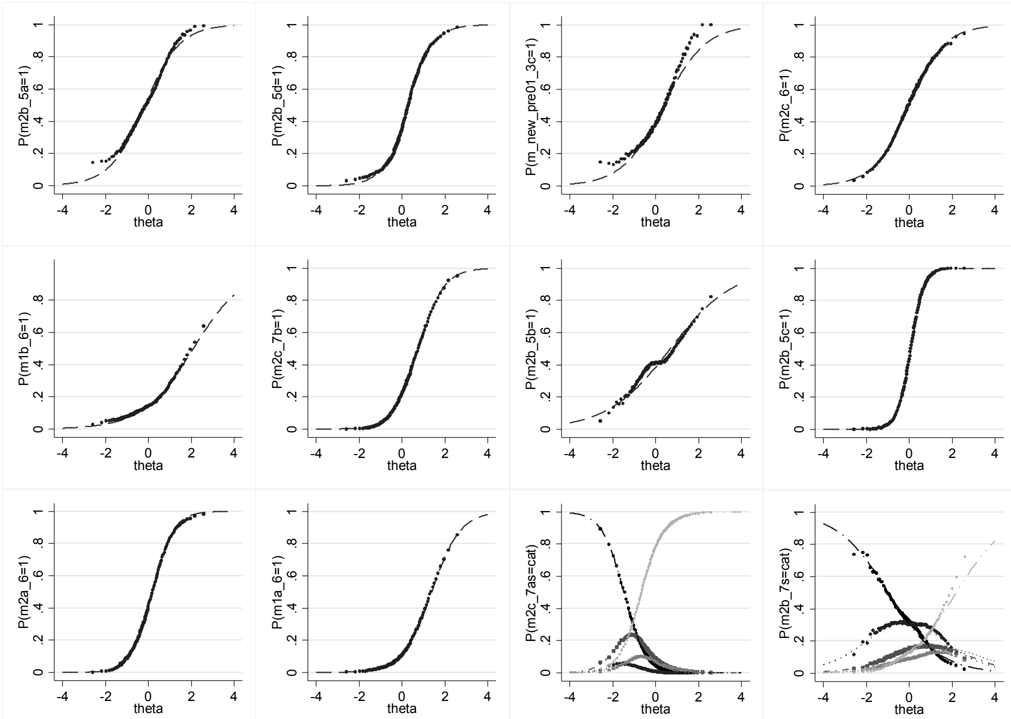
gdzie: θ jest losową zmienną ukrytą opisującą poziom umiejętności uczniów; $\psi_P(\theta)$ jest funkcją gęstości prawdopodobieństwa określającą rozkład zmiennej θ w populacji P; $f(\boldsymbol{u},\theta,\boldsymbol{\beta})$ jest funkcją, która określa prawdopodobieństwo zaobserwowania konkretnej wartości $\boldsymbol{u}$ wektora odpowiedzi $\boldsymbol{U}$, w zależności od poziomu umiejętności θ oraz wektora parametrów $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_n)$, gdzie parametry zadania $\boldsymbol{\beta}_i$ również mogą być wektorami.

Do zadań ocenianych dychotomicznie dopasowano dwuparametryczny model logistyczny (*two-parameter logistic model*, 2PLM), natomiast pozostałe zadania modelowano za pomocą modelu oceny stopniowanej (*graded reponse model*, GRM).

Dopasowanie modelu do danych oceniono poprzez analizę ułożenia empirycznych proporcji odpowiedzi z każdej kategorii punktowej w centylach umiejętności względem krzywych charakterystycznych wykreślonych zgodnie z oszacowanymi parametrami modelu IRT. Dopasowanie modelu IRT uznano za zadowalające (Rysunek 2). Zakończone sukcesem dopasowanie do danych jednowymiarowego modelu IRT można rozpatrzeć jako element pozytywnie weryfikujący trafność teoretyczną narzędzia – uzyskujemy potwierdzenie występowania pojedynczego głównego czynnika umiejętności matematycznych odpowiedzialnego za obserwowaną współzmienność odpowiedzi na wykorzystane w badaniu zadania.

Wielogrupowy model IRT został dopasowany do danych w taki sposób, aby średnia oraz odchylenie standardowe rozkładu umiejętności uczniów biorących udział w badaniu CKE były ustalone na wartościach odpowiednio: 0 oraz 1. Dzięki takiemu zabiegowi wspólna skala zmiennej ukrytej, na której przedstawiane są wyniki pretestu i posttestu, jest odniesiona do wystandaryzowanego rozkładu wyników z reprezentatywnej i licznej próby uczniów, którzy brali udział w badaniu CKE. Umożliwia to odniesienie wyników uczniów biorących udział w eksperymencie do ogółu populacji i pozwala na ocenę trafności zewnętrznej badania.

Ze względu na wykorzystanie do pomiaru umiejętności matematycznych narzędzi psychometrycznych, w analizie wyników konieczne było uwzględnienie ich nierzetelności. W przeciwnym razie oszacowania efektów i błędów standardowych byłoby obciążone. Dopasowanie do danych modelu IRT dało taką możliwość. Zamiast korzystać z punktowych oszacowań wyników uczniów do weryfikacji postawionej hipotezy badawczej, analizę przeprowadzono, korzystając z wygenerowanych na podstawie wektora



Rysunek 2. Dopasowanie modelu IRT do zadań wykorzystanych do pomiaru umiejętności matematycznych uczniów.

udzielonych przez każdego ucznia odpowiedzi zestawów dwustu PV. Są one realizacjami z rozkładu a posteriori umiejętności ucznia, warunkowanego poprzez zmienne niezależne wykorzystywane w późniejszych analizach. Podstawy przeprowadzania analiz statystycznych z wykorzystaniem PV można znaleźć w pracy Margaret Wu (2005), natomiast pogłębionego opracowania uzasadniającego takie podejście dostarcza monografia Rodericka Little'a oraz Donalda Rubina (2002). Podczas generowania PV, oprócz wektora odpowiedzi udzielonych przez ucznia, uwzględniono jako dodatkowe zmienne warunkujące wszystkie zmienne zastosowane w późniejszej analizie – efekty stałe dla warunków eksperymentalnych oraz zagnieżdżenie uczniów w pomiarach i w szkołach. Dla oszacowania rzetelności pomiaru zmiennej zależnej w pierwszym oraz w drugim pomiarze policzono średnią korelację pomiędzy wszystkimi PV uczniów w tych pomiarach. Uzyskano odpowiednio wartości: 0,72 oraz 0,67.

W celu weryfikacji postawionej hipotezy o pozytywnym wpływie pakietu edukacyjnego *Gramy w piktogramy* na badane umiejętności uczniów, zastosowano model regresji trójpoziomowej:

$$Y_{ijk} = \gamma_{000} + \gamma_1\, {}^1X_{ijk} + \gamma_2\, {}^2X_{ijk} + \gamma_{1x2}\, {}^1X_{ijk}\, {}^2X_{ijk} + \varepsilon_{00k} + \varepsilon_{0jk} + \varepsilon_{ijk}, \quad (3)$$

gdzie:

- Y_{ijk} – wartość zmiennej zależnej, tj. poziomu umiejętności matematycznych ucznia i zagnieżdżonego w pomiarze j (pretest–posttest), zagnieżdżonego w klasie k;

- ${}^{1}X_{ijk}$, ${}^{2}X_{ijk}$, ${}^{1}X_{ijk}\,{}^{2}X_{ijk}$ – wartości zmiennych niezależnych określających warunki eksperymentalne, odpowiednio:
  - ${}^{1}X_{ijk}$ – zmienna wskazująca na grupę (0 = kontrolna; 1 = eksperymentalna);
  - ${}^{2}X_{ijk}$ – zmienna wskazująca drugi pomiar (0 = pretest; 1 = posttest);
  - ${}^{1}X_{ijk}\,{}^{2}X_{ijk}$ – iloczyn powyższych zmiennych, tj. zmienna interakcyjna wskaźnika grupy i wskaźnika pomiaru (0 = inne przypadki niż „1"; 1 = posttest i grupa eksperymentalna);
- γ_{000}, γ_1, γ_2, γ_{1x2} – stałe współczynniki regresji (efekty stałe), których wartości szacujemy z danych; odpowiednio: wyraz wolny; efekt dla zmiennej ${}^{1}X_{ijk}$; efekt dla zmiennej ${}^{2}X_{ijk}$; oraz efekt dla interakcji obu zmiennych ${}^{1}X_{ijk}$ oraz ${}^{2}X_{ijk}$;
- ε_{ijk} – wartość losowego błędu z poziomu ucznia dla ucznia *i* zagnieżdżonego w pomiarze *j*, zagnieżdżonego w klasie *k*; przyjmujemy, że ten składnik losowy ma rozkład $N(0, \rho_1)$;
- ε_{0jk} – wartość losowego błędu z poziomu pomiaru, dla pomiaru *j* zagnieżdżonego w klasie *k*; przyjmujemy, że ten składnik losowy ma rozkład $N(0, \rho_2)$;
- ε_{00k} – wartość losowego błędu z poziomu klasy, dla klasy *k*; przyjmujemy, że ten składnik losowy ma rozkład $N(0, \rho_3)$;

Istotą powyższej regresji trzypoziomowej jest rozróżnienie składnika resztowego na: błąd ε_{ijk} odpowiadający za niewyjaśnioną wariancję uczniowskich wyników zagnieżdżonych w pomiarach (czyli wewnątrz pojedynczego ucznia) ρ_1; błąd ε_{0jk} odpowiadający za niewyjaśnioną wariancję między pomiarami wewnątrz klasy ρ_2 oraz błąd ε_{00k} odpowiadający za niewyjaśnioną wariancję między klasami ρ_3. Jest to model dla pomiarów powtarzanych, zatem zwiększa moc wnioskowania statystycznego poprzez uwzględnienie korelacji między wynikami tych samych uczniów z pierwszego i drugiego pomiaru, a jednocześnie uwzględnia istotną korelację wewnątrzklasową. Przykłady zastosowania regresji wielopoziomowej do analizy danych z powtarzanych pomiarów można znaleźć u Sophii Rabe-Hesketh i Andersa Skrondala (2008). Warto zaznaczyć, że takiego samego modelu regresji użyto również przy warunkowaniu podczas wcześniej opisanego generowania PV.

Włączenie trzech zmiennych niezależnych do modelu regresji, jak w równaniu (3), to klasyczne podejście analizy danych zebranych w schemacie eksperymentu z pomiarami powtarzanymi i grupą kontrolną. Współczynniki regresji przy opisanym powyżej zakodowaniu zmiennych mają następującą interpretację:

- wyraz wolny γ_{000} odpowiada średniemu poziomowi umiejętności w grupie kontrolnej w momencie pierwszego pomiaru,
- γ_1 określa, o ile wyższy jest poziom umiejętności grupy eksperymentalnej od grupy kontrolnej, niezależnie od tego, czy uwzględniamy pierwszy czy drugi pomiar,

Tabela 4
*Oszacowania współczynników regresji w modelu wielopoziomowym weryfikującym istotność statystyczną efektów w badaniu skuteczności pakietu Gramy w piktogramy**

			95% przedział ufności			
Zmienne w modelu regresji	Efekt (γ)	*SE*	dolna gr.	górna gr.	*z*	*p*
Wyraz wolny	-0,158	0,159	-0,470	0,154	-0,991	0,322
Wskaźnik grupy E (${}^{1}X_{ijk}$)	-0,016	0,226	-0,459	0,428	-0,069	0,945
Wskaźnik drugiego badania (${}^{2}X_{ijk}$)	0,657	0,120	0,422	0,891	5,494	0,000
Interakcja (${}^{1}X_{ijk}\,{}^{2}X_{ijk}$)	0,314	0,178	-0,035	0,662	1,762	0,078

* *p*-wartości dla bezkierunkowych hipotez alternatywnych $H_1: \gamma \neq 0$.

- γ_2 określa, o ile wyższy jest poziom umiejętności uczniów w drugim pomiarze w porównaniu z pomiarem pierwszym, niezależnie od tego, czy uwzględniamy grupę kontrolną czy eksperymentalną,
- γ_{1x2} określa, o ile jest wyższy poziom umiejętności grupy eksperymentalnej w drugim pomiarze, jeżeli uwzględnimy już informację z wcześniejszych dwóch czynników, czyli po uwzględnieniu średniej zmiany poziomu umiejętności między pomiarami oraz po uwzględnieniu średniej różnicy w poziomie umiejętności między grupami.

Ostatni parametr, γ_{1x2}, jest zatem w kontekście oceny efektów stosowania pakietu *Gramy w piktogramy* parametrem najistotniejszym. To on określa, ile wynosi wzrost umiejętności w grupie eksperymentalnej w drugim pomiarze – specyficznie związanym z oddziaływaniem eksperymentalnym.

Wyniki

Wyniki dopasowania opisanego modelu regresji do zebranych danych przedstawiono w Tabeli 4. Przypomnijmy, że zgodnie ze sposobem zakotwiczenia parametrów modelu wielogrupowego IRT (który posłużył do wygerowania PV), rozkład umiejętności matematycznych ma średnią 0 oraz odchylenie standardowe 1 dla wyników uczniów uzyskanych w badaniu reprezentatywnym z 2008 r. Zatem przedstawione w Tabeli 4 efekty uzyskane w badaniu eksperymentalnym są wyrażone na skali odchylenia standardowego w badaniach reprezentatywnych.

Oszacowana wartość dla wyrazu wolnego w modelu regresji (γ_{000}) jest ujemna, co wskazuje na niższy poziom mierzonych umiejętności podczas pierwszego pomiaru w grupie kontrolnej niż wśród uczniów biorących udział w badaniu reprezentatywnym w 2008 r. Ten kierunek różnicy nie zaskakuje, ponieważ pretest został przeprowadzony na poczatku roku szkolnego w klasie trzeciej, natomiast badanie reprezentatywne CKE przeprowadzono pod koniec tej klasy. Ze względu na duży błąd oszacowania parametru brakuje podstaw do stwierdzenia, że ten efekt jest istotny statystycznie.

Oszacowanie parametru γ_1 przy zmiennej wskazującej na grupę eksperymentalną jest bardzo bliskie zeru i nieistotne statystycznie, co oznacza, że między grupą eksperymentalną a kontrolną praktycznie nie było różnic w poziomie umiejętności podczas pierwszego badania (chociaż 95% przedział ufności wokół parametru jest dość szeroki: od -0,46 do 0,43 odchylenia standardowego). Jest to spodziewany efekt randomizacji oddziałów do grupy eksperymentalnej i kontrolnej.

Oszacowanie parametru γ_2 przy zmiennej wskazującej na drugi pomiar (posttest), wyniosło +0,657, a błąd standardowy był kilkukrotnie mniejszy (0,120). Oznacza to, że poziom mierzonych umiejętności uczniów między pierwszym a drugim pomiarem, niezależnie od oddziaływania eksperymentalnego, wzrósł o przeszło połowę odchylenia standardowego, istotnie statystycznie. Wzrost poziomu umiejętności uczniów między pierwszym a drugim pomiarem również jest spodziewanym wynikiem, który ukazuje słuszność przeprowadzenia badania w schemacie eksperymentalnym z grupą kontrolną. Bez włączenia do badania grupy kontrolnej nie sposób byłoby oddzielić efektu stosowania pakietu edukacyjnego *Gramy w piktogramy* od naturalnie występującego na przestrzeni klasy trzeciej znacznego wzrostu kompetencji matematycznych uczniów.

Oszacowanie parametru γ_{1x2}, a zatem parametru interakcji mierzącego wzrost umiejętności matematycznych specyficznie związany z oddziaływaniem eksperymentalnym, wyniosło 0,314. A zatem wzrost poziomu umiejętności matematycznych punktowo oszacowano na przeszło 0,3 odchylenia standardowego, co stanowi ok. 48% wzrostu, jaki nastąpił ze względu na upływ czasu między pierwszym a drugim pomiarem. Przy dość dużym błędzie standardowym 95% przedział ufności wokół tego oszacowania wyniósł od -0,03 do 0,66 odchylenia standardowego, zawiera więc w sobie nawet wartości ujemne. Przy bezkierunkowej hipotezie alternatywnej oznaczałoby to efekt nieistotny statystycznie (*p*-wartość = 0,078). Dla oceny efektywności oddziaływania eksperymentalnego przy hipotezie badawczej zakładającej określony kierunek wpływu, hipotezę zerową dla parametru interakcji testuje się względem kierunkowej hipotezy alternatywnej, zakładającej zmianę zgodną z zamierzonym kierunkiem. W naszej sytuacji oznacza to parę:

$$H_0 : \gamma_{1x2} = 0$$

$$H_1 : \gamma_{1x2} > 0$$

Przy tak postawionej hipotezie alternatywnej oraz dodatniej wartości oszacowanego efektu oznacza to *p*-wartość równą połowie *p*-wartość dla bezkierunkowej hipotezy alternatywnej, czyli 0,039. Przy poziomie istotności $\alpha = 0{,}05$ oznacza to efekt istotny statystycznie. Ostatecznie zebrane dane pozwalają na odrzucenie hipotezy zerowej o braku efektu oddziaływania eksperymentalnego na poziom umiejętności uczniów na korzyść hipotezy alternatywnej, mówiącej o wzroście poziomu umiejętności specyficznie związanym z zastosowaniem pakietu edukacyjnego *Gramy w piktogramy*.

Dyskusja

Badanie pokazało, że wzrost umiejętności posługiwania się językiem symbolicznym wśród uczniów z klas korzystających przez rok z pakietu edukacyjnego był istotnie statystycznie wyższy niż w klasach kontrolnych. Pakiet *Gramy w piktogramy* został tak skonstruowany, aby skłonić pracujących z nim nauczycieli do refleksji nad swoim warsztatem zawodowym i zachęcić ich do modyfikacji stylu swojej codziennej pracy. Można więc przypuszczać, że zaobserwowany istotny efekt jest bezpośrednią konsekwencją wystąpienia zmian w zakresie czynników powiązanych ze sposobem pracy nauczyciela.

Choć opisane badanie eksperymentalne nie zostało zaprojektowane do weryfikacji takiej hipotezy, pomiarowi uczniowskich umiejętności towarzyszyło również przeprowadzenie badania poglądów edukacyjnych nauczycieli w klasach kontrolnych i eksperymentalnych. Wśród nauczycieli z grupy eksperymentalnej zaobserwowano znaczne, co do wartości, zmiany w poziomie trzech mierzonych wymiarów poglądów edukacyjnych: zmalał ich pesymizm edukacyjny oraz formalizm edukacyjny, natomiast wzrósł wynik na skali promowania samodzielności (Dąbrowski i Żytko, 2013). Kierunek zmian jest zgodny z oczekiwanym, ponieważ z innych badań (np. Kondratek, 2011) wiadomo, że wymienione trzy wskaźniki istotnie korelowały z umiejętnościami matematycznymi uczniów (dwa pierwsze ujemnie, ostatni dodatnio). Niestety, ze względu na bardzo małą próbę nauczycieli, brakowało mocy statystycznej do uznania wymienionych zmian za statystycznie znamienne. Należy mieć również na względzie, że wykryte we wcześniejszym badaniu zależności miedzy poglądami edukacyjnymi nauczycieli a wynikami uczniów, mają charakter korelacyjny i nie można na ich podstawie w sposób rozstrzygający wnioskować o kierunku przyczynowo-skutkowym, np. między spadkiem pesymizmu edukacyjnego a wzrostem umiejętności uczniów.

Dodatkowych informacji o potencjalnym wpływie zmiany stylu pracy nauczyciela na poziom umiejętności uczniów dostarczyły dane jakościowe, zebrane podczas systematycznych obserwacji prowadzonych przez cały okres prowadzenia eksperymentu, ze sprawozdań pisanych przez nauczycieli w trakcie pracy z pakietem (Dąbrowski i Żytko, 2013) oraz z wywiadów przeprowadzonych z nauczycielami i uczniami na koniec rocznego testowania pakietu. Warto w tym kontekście przytoczyć kilka wypowiedzi[3]:

> Spojrzałam na uczniów z innej strony, np. uczniowie słabsi mnie zaskoczyli.
>
> Wcześniej chyba nie miałam takiej świadomości, że pytania są takie cenne. Teraz się przyzwyczaiłam i jak wprowadzam jakiś temat, to zadaję im mnóstwo pytań, na które muszą sami odpowiadać i ukierunkowywać się.
>
> Jako nauczycielka z wieloletnim stażem pedagogicznym dużo się nauczyłam. Nie podpowiadam już dzieciom, oczekuję cierpliwie na odpowiedź. Nie zakładam, że dziecko nie poradzi sobie z zadaniem tak, jak to czasami wcześniej się zdarzało. Wiem, że podczas nauczania musi być aktywny uczeń, nie nauczyciel.

Opisaną w artykule analizę przeprowadzono dla ogólnego wskaźnika umiejętności matematycznych, tak aby uzyskać możliwie najbardziej rzetelną miarę umiejętności. Analiza przeprowadzona na poziomie pojedynczych zadań (por. Dąbrowski i Żytko, 2013) wskazuje, że wzrost wyników w grupie eksperymentalnej nastąpił we wszystkich założonych obszarach, czyli w zakresie rozwiązywania nietypowych zadań tekstowych, rozumienia systemu dziesiętnego i posługiwania się nim oraz rozwiązywania problemów. Przytoczone wyniki w podziale na podobszary umiejętności są jednak obciążone z powodu niewielkiej próbki zadań przypadających na każdą skalę. Należy więc traktować je jako wskazówkę dla możliwych kierunków badań w przyszłości.

[3] Wypowiedzi zaczerpnięte są z *Raportu z ewaluacji innowacyjnej pomocy dydaktycznej: Pakiet edukacyjny Gramy w piktogramy i efektów jego stosowania na etapie testowania* (www.projekt-piktografia.pl). Autorzy tego raportu podkreślają, że dobrze oddają one poglądy całej grupy nauczycieli testujących pomoc.

Literatura

Binkowska-Wójcik, W., Boroń, I., Brzyska, S., Cikorska, M., Fiertek R. i in. (2014). *Bydgoski bąbel matematyczny. O wprowadzaniu zmian w nauczaniu matematyki w klasach I–III*. Warszawa: Instytut Badań Edukacyjnych.

Dąbrowski, M. (2013). *(Za)trudne, bo trzeba myśleć*. Warszawa: Instytut Badań Edukacyjnych.

Dąbrowski, M. (red.). (2009). *Badanie umiejętności podstawowych uczniów trzecich klas szkoły podstawowej. Raport z badań ilościowych 2008*. Warszawa: Centralna Komisja Egzaminacyjna.

Dąbrowski, M. i Żytko M. (red.). (2013). *Raport z testowania innowacyjnej pomocy dydaktycznej: Pakiet edukacyjny Gramy w piktogramy*. Warszawa: Wydawnictwo Bohdan Orłowski.

Faul, F., Erdfelder, E., Lang, A.-G. i Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Gopnik, A., Meltzoff, A. N. i Kuhl, P. K. (2004). *Naukowiec w kołysce. Czego o umyśle uczą nas małe dzieci*. Poznań: Media Rodzina.

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement.* Routledge: London–New York.

Hattie, J. (2011). *Visible learning for teachers: maximizing impact on learning.* Routledge: London–New York.

Imai, K., King, G. i Nall, C. (2009). The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation. *Statistical Science*, *24*(1), 29–53.

Karpiński, M., Grudniewska, M i Zambrowska, M. (2013). *Nauczanie matematyki w gimnazjum. Raport z badania*. Warszawa: Instytut Badań Edukacyjnych.

Kolen, M. J. i Brennan R. L. (2004). *Test equating, scaling, and linking: methods and practice* (wyd. 2). New York: Springer.

Kołodziejczyk, J., Salamon-Bobińska, K., Karaszewski, N. i Bobula, S. (2014). Nauczanie kooperatywne (uczenie się we współpracy). W: G. Mazurkiewicz (red.), *Edukacja jako odpowiedź. Odpowiedzialni nauczyciele w zmieniającym się* *świecie* (s. 163–177). Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Kondratek, B. (2009). Konstrukcja skal mierzących umiejętności językowe i matematyczne uczniów oraz poglądy edukacyjne nauczycieli. W: M. Dąbrowski (red), *Badanie umiejętności podstawowych uczniów trzecich klas szkoły podstawowej. Część III: trzecioklasista i jego nauczyciel* (s. 186–215). Warszawa: Centralna Komisja Edukacyjna.

Kondratek, B. (2011). Poglądy edukacyjne nauczycieli klas 1–3. W: M. Dąbrowski (red.), *Trzecioklasiści 2010. Badanie umiejętności podstawowych uczniów trzecich klas szkoły podstawowej. Raport z badań ilościowych* (s. 230–241.) Warszawa: Centralna Komisja Edukacyjna.

Lipsey, M. W. i Hurley S. M. (2009). Design sensitivity: statistical power for applied experimental research. W: L. Bickman i D. J. Rog (red.), *The SAGE handbook of applied social research methods* (wyd. 2, s. 44–76). Thousand Oaks: Sage.

Little, R. J. A. i Rubin, D. B. (2002). *Statistical analysis with missing data* (wyd. 2). New York: Wiley.

Murray, D. M., Varnell, S. P. i Blitstein, J. L. (2004). Design and analysis of group-randomized trials: a review of recent methodological developments. *American Journal of Public Health*, *94*(3), 423–432.

Pokropek, A., Kondratek, B. (2012), Zrównywanie wyników testowania. Definicje i przykłady zastosowania. *Edukacja*, *120*(4), 52–71.

Pregler, A. i Wiatrak, E. (red.). (2011). *Ogólnopolskie badanie umiejętności trzecioklasistów. Raport OBUT 2011*. Warszawa: Centralna Komisja Edukacyjna.

Rabe-Hesketh, S. i Skrondal, A. (2008). *Multilevel and longitudinal modelling using Stata*. College Station: Stata Press.

Sfard, A. (2008). *Thinking as communicating: human development, the growth of discourses, and mathematizing.* Cambridge: Cambridge University Press.

Slavin, R. E. (2012). Uczenie się oparte na współpracy. Dlaczego praca w grupach jest skuteczna. W: F. Benavides, H. Dumont i D. Istance (red.), *Istota uczenia się. Wykorzystanie wyników badań w praktyce* (s. 248–276). Warszawa: Wolters Kluwer.

Szaleniec, H., Grudniewska, M., Kondratek, B., Kulon, F. i Pokropek, A. (2012). Wyniki egzaminu gimnazjalnego 2002–2010 na wspólnej skali. *Edukacja*, *119*(3), 9–30.

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation* *31*(2–3), 114–128.