Mirosław Dąbrowski

Bartosz Kondratek

Analiza efektów zastosowania pakietu edukacyjnego *Gramy w piktogramy*, czyli o tym, czy warto zachęcać nauczycieli do zmiany stylu pracy.

Abstrakt

Artykuł przedstawia analizę wyników badania weryfikującego skuteczność zastosowania pakietu edukacyjnego ***Gramy w piktogramy****, przeznaczonego do wspierania rozwoju umiejętności posługiwania się językiem symbolicznym na etapie edukacji początkowej. 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 matematycznych uczniów związana z wdrożeniem pomocy dydaktycznej. W analizach wykorzystano modelowanie IRT oraz regresję wielopoziomową. Wyniki wskazują na istotny statystycznie wzrost ogólnego wskaźnika umiejętności matematycznych w grupie eksperymentalnej związany specyficznie z zastosowaniem badanego pakietu edukacyjnego.*

Wprowadzenie

Hugo Steinhaus (1887-1972) wybitny polski matematyk, współtwórca polskiej szkoły matematycznej, był znany także jako autor 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*, oraz zachęcał do podnoszenia kompetencji matematycznych zapewniając, że *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 (por. np. Dąbrowski 2013), że praktyka polskiej edukacji wczesnoszkolnej skupia się przede wszystkim na „trenowaniu" w rozwiązywaniu typowych zadań, wedle ściśle wskazanych przez nauczyciela instrukcji, przy dużym nacisku na arytmetyczną poprawność wykonywanych działań. Należy sądzić, że taki styl pracy z najmłodszymi uczniami nie sprzyja optymalnemu rozwojowi ich umiejętności posługiwania się językiem symbolicznym, jednocześnie tłumiąc ich twórczość i odwagę w podejmowaniu wcześniej nie napotkanych problemów matematycznych. Wiele przemawia za tym, że taka sytuacja jest pochodną tradycji edukacyjnej dominującej w naszym społeczeństwie i w polskiej szkole nie tylko w początkowym okresie nauczania. Tradycji zgodnie z którą dziecko nie jest w stanie samodzielnie do niczego dojść i wie tylko to, czego dowiedziało się od dorosłych. Pakiet edukacyjny *Gramy w piktogramy* został stworzony m.in. po to, aby przyczynić się do uruchomienia procesu odchodzenia od tej, hamującej optymalny rozwój uczniów, tradycji.

Artykuł rozpocznie opis założeń teoretycznych leżących u podstaw pakietu edukacyjnego *Gramy w piktogramy* (GP), oraz sposobu, w jaki wspiera on nauczanie matematyki w pierwszych klasach szkoły podstawowej. W dalszej kolejności zostanie przedstawione pytanie badawcze, dla rozwiązania którego zaprojektowano raportowane badanie wraz z postawionym względem tego pytania hipotezą. Dość szczegółowo opisane zostaną różne metodologiczne aspekty przeprowadzonego badania oraz sposób statystycznej analizy zebranych danych. Uznano, że taki poziom szczegółowości jest koniecznym kontekstem dla oceny trafności oraz rzetelności prezentowanych później wyników. Ostatecznie, zostaną zaprezentowane wyniki analiz statystycznych oraz przeprowadzona na ich temat dyskusja.

Gramy w piktogramy

Pakiet edukacyjny *Gramy w piktogramy* stworzony w ramach projektu *PIKTOGRAFIA Rozwijanie umiejętności posługiwania się językiem symbolicznym w edukacji z zakresu nauk matematycznych z zastosowaniem piktogramów Asylco* to środek dydaktyczny, który powstał we współpracy Wydawnictwa Bohdan Orłowski (lider) oraz Uniwersytetu Warszawskiego Wydziału Pedagogicznego (partner).

U podstaw pakietu leży przekonanie, że żeby dzieci mogły efektywnie i ze zrozumieniem uczyć się matematyki muszą być aktywne intelektualnie, muszą współpracować i rozmawiać o tym, co robią i o samej matematyce, gdyż uczenie się matematyki jest procesem o charakterze społecznym. Zatem już „na starcie" mamy dokładne zaprzeczenie tego, co na co dzień można obserwować na lekcjach matematyki w szkole i co jako społeczeństwo przyjmujemy jako „edukacyjną matematyczną normę".

W celu uruchomienia aktywności dzieci trzeba było z jednej strony uwzględnić ich edukacyjne potrzeby, stąd m.in. bardzo częste wykorzystywanie w pakiecie reprezentacji enaktywnej i ikonicznej, z drugiej zaś zainteresować ich i pobudzić motywację do uczenia się, co spróbowano osiągnąć m.in. konsekwentnie proponując pracę w parach i grupach oraz sięgając po zadania różniące się od tych, jakie na co dzień oferuje szkoła. Wśród tych zadań regularnie występowały zadania otwarte, w naturalny sposób uruchamiające wyjaśnianie i argumentowanie, czy zadania problemowe, niekiedy o interdyscyplinarnym charakterze. Zarówno proponowana organizacja pracy, jak i typy zadań miały na celu uruchomienie podczas zajęć procesu rozmawiania o matematyce – zarówno uczniów z sobą, jak i nauczyciela z uczniami oraz uczniów z nauczycielem.

Oznaczało to, w konsekwencji, potrzebę przekonania nauczycieli pracujących z pakietem, m.in. w trakcie zorganizowanego dla nich szkolenia, że powinni zmienić swój styl pracy i pozwolić dzieciom na inny niż na typowych zajęciach sposób funkcjonowania. Piktogramy Asylco, ze względu na swój specyficzny styl komunikowania znaczenia oraz „wielopoziomowość" symboliczną, 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 gimnazjum (por. np. www.projekt-piktografia.pl). Składa się on z zestawu pomocy dla uczniów oraz pakietu 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 m.in. podczas rozwiązywania i układania zadań,
- gry (plansze, pionki, kostki) rozwijające m.in. rozumienie systemu dziesiętnego,
- tabliczki suchościeralne i mazaki do zapisywania rozwiązań zadań, projektowania piktogramów itp.

W skład pakietu dla nauczyciela wchodzą:

- ✓ przewodnik dla nauczycieli, który m.in. przedstawia filozofię edukacyjną pakietu,
- ✓ zestaw scenariuszy proponowanych zajęć,
- ✓ zestaw kart pracy o trzech poziomach trudności, służący m.in. 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.

Opracowano też wersję e-pakietu z materiałami do pobrania (www.piktografia.pl) oraz szkolenie e-learningowe dla nauczycieli chcących korzystać z pakietu.

Pytania i hipotezy badawcze

Pakiet edukacyjny *Gramy w piktogramy* został skonstruowany z myślą o wspieraniu rozwoju uczniów w sztuce efektywnego posługiwania się językiem symbolicznym. W związku z tym postawiono pytanie badawcze, czy zastosowanie tego pakietu edukacyjnego wiąże się ze zwiększeniem uczniowskich kompetencji w rozwiązywaniu zadań matematycznych wymagających posługiwania się językiem symbolicznym w porównaniu z sytuacją, w której nauczyciele nie korzystają z tego pakietu ograniczając się w pracy z uczniami do stosowanych przez siebie dotychczas metod nauczania matematyki. Tak postawiony problem badawczy zrodził konieczność przeprowadzenia badań w schemacie eksperymentalnym z grupą kontrolną, w której nauczyciele nie korzystają z pakietu edukacyjnego *Gramy w piktogramy*. Szczegółowy opis planu badań eksperymentalnych przedstawiono w dalszej części artykułu. Hipotezą badawczą, którą dane zebrane w badaniach eksperymentalnych miały na celu zweryfikować jest twierdzenie o wzroście umiejętności posługiwania się językiem symbolicznym w klasach eksperymentalnych, specyficznie związane z zastosowaniem w tych klasach pakietu edukacyjnego *Gramy w piktogramy*.

Opis planu badań eksperymentalnych

Następujące czynniki ekonomiczne oraz organizacyjne w istotny sposób wpłynęły na postać schematu eksperymentalnego zgodnie, z którym przeprowadzone zostało badanie:

1) Podstawową jednostką poddawaną manipulacji eksperymentalnej była cała klasa oraz założono występowanie istotnej korelacji wewnątrzklasowa zmiennej zależnej rzędu 0,15.
2) Ze względu na towarzyszący ilościowemu badaniu eksperymentalnemu rozbudowany plan kosztownych badań jakościowych, ustalono iż jedynie 8 oddziałów zostanie objętych odziaływaniem eksperymentalnym[1].

Ograniczenie możliwości doboru uczniów do warunków eksperymentalnych jedynie całymi klasami przy istotnej korelacji wewnątrzklasowej redukuje efektywną wielkość próby. Wykorzystując wzór na tzw. efekt planu (ang. *design effect*):

---

[1] Liczba ośmiu grup jest wskazywana w literaturze jako minimalna liczebność grup w eksperymencie z randomizowanymi grupami, zapewniająca moc statystyczną pozwalającą wykrywać istotne z praktycznego punktu widzenia efekty (Murray et al., 2004).

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

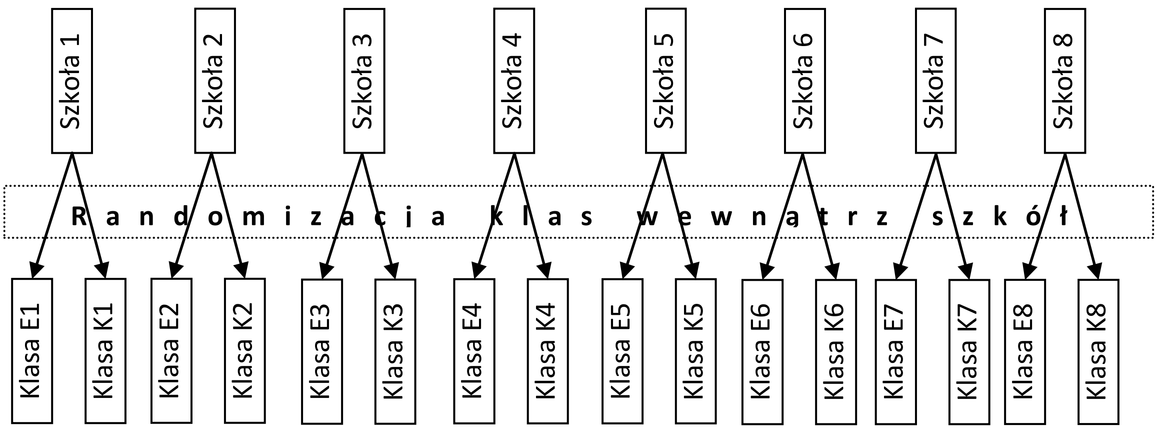
gdzie m – liczba osób w grupie, ρ – współczynnik korelacji wewnątrzklasowej, efektywną wielkość próby $n_{ef} = {}^{n}/_{D}$ w badaniach oszacowano na około 40 uczniów na każdy warunek eksperymentalny (zakładając m w zakresie 15-20 oraz ρ=0,15).

W obliczu nałożonych ograniczeń uznano, że najefektywniejszą (w rozumieniu mocy wnioskowania statystycznego) strategią badania efektów oddziaływania będzie przeprowadzenie randomizowanego grupami eksperymentu z grupą kontrolną ze zrównywaniem/dopasowaniem (ang. *group/cluster-randomized experiment with pair matching*) z pomiarem początkowym oraz końcowym zmiennej zależnej. Ogólny schemat takiego badania wpisuje się w postać eksperymentu z pomiarem powtarzanym oraz grupą kontrolną:

E: Pretest -> oddziaływanie eksperymentalne -> Posttest

K: Pretest -> brak oddziaływania eksperymentalnego -> Posttest

gdzie E oraz K oznaczają odpowiednio: grupę kontrolną oraz eksperymentalną. Przy czym rozdział uczniów do grup E oraz K następuje całymi klasami, z randomizacją dokonaną wewnątrz szkoły (założono, że badaniem eksperymentalnym zostaną objęte klasy przynajmniej dwuoddziałowe). Włączenie do badania grupy kontrolnej było konieczne ze względu na potrzebę kontroli efektu dojrzewania (ang. *maturation*), tj. spodziewano się występowania w interwale pomiędzy pierwszym a drugim pomiarem pewnego „naturalnego” wzrostu poziomu umiejętności matematycznych uczniów niezależnie od oddziaływania eksperymentalnego. Dla ośmiu szkół[2] objętych badaniem eksperymentalnym rozdział klas do grup E oraz K ilustruje schemat przedstawiony na Rysunku 1.



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

[2] Na etapie doboru szkół do badania okazało się, że – ze względu na specyfikę szkół wiejskich – udało się zrekrutować tylko 7 szkół prowadzących przynajmniej dwa oddziały klas trzecich, które jednocześnie wpadałyby w założone przedziały lokalizacji oraz średnich wyników szkoły. Dwa oddziały z ósmej szkoły przewidzianej na etapie projektowania eksperymentu zostały podczas realizacji badania dobrane z dwóch różnych szkół, pochodzących z tej samej lokalizacji oraz wpadających w tę samą kategorię wyników punktowych. Zob. także fragment odnoszący się do doboru próby.

Przeprowadzanie randomizacji wewnątrz szkoły wychodziło z założenia, że wariancja między klasami tej samej szkoły będzie istotnie niższa niż wariancja między klasami w całej populacji. Zastosowanie zrównywania parami zredukuje tym samym prawdopodobieństwo uzyskania podziału na grupy, kontrolną i eksperymentalną, które różniłyby się znacznie ze względu na wyjściowy poziom zmiennej zależnej. Dodatkowo, randomizacja wewnątrz szkół zapewniła zrównoważenie grup E oraz K ze względu na zmienne środowiskowe związane z lokalizacją szkoły. Jest to procedura często zalecana w eksperymentach z randomizowanymi grupami (Lipsey & Hurley 2008), która poza zrównywaniem ze względu na istotne dla badania zmienne pozwala po części zniwelować obniżenie efektywności badania wynikające z istotnej korelacji wewnątrzklasowej (Imai et. al. 2009). Warto zauważyć, że randomizacja wewnątrz tych samych szkół wprowadza pewne zagrożenie dla trafności wewnętrznej eksperymentu wynikające z możliwości przepływu informacji między nauczycielem objętym oddziaływaniem eksperymentalnym, a nauczycielem z grupy kontrolnej (ang. *contamination*) – uznano jednak, że potencjalne straty nie przeważały płynących z takiego rozwiązania zysków.

Innym istotnym aspektem proponowanego planu eksperymentalnego jest przeprowadzenie powtarzanego pomiaru na tych samych uczniach (pretest i posttest). Takie rozwiązanie niesie za sobą ewidentne korzyści metodologiczne (możliwość kontroli wyjściowych różnic między grupą E i K), jak i czysto statystyczne, polegające na zwiększeniu mocy statystycznej przy wnioskowaniu z pomiarów zależnych. Stosując pomiar powtarzany należy jednak należycie zadbać o kontrolę efektu pamięci, jaki by wystąpił, gdyby uczniowie rozwiązywali na preteście i postteście ten sam test umiejętności matematycznych. Zastosowanym w badaniu rozwiązaniem tego problemu było wykorzystanie w pierwszym i drugim pomiarze testów umiejętności matematycznych zbudowanych z różnych zadań, co jednak stworzyło kolejne wyzwanie, mianowicie konieczność kontroli trudności takich różnych narzędzi.

Aby skontrolować trudność dwóch różnych narzędzi wykorzystanych w preteście i postteście, w obu narzędziach umieszczono porcję zadań wziętych z przeprowadzonego w 2008 roku ogólnopolskiego *Badania umiejętności podstawowych uczniów trzecich klas szkół podstawowych* przeprowadzanego przez Centralną Komisję Egzaminacyjną (Dąbrowski, 2009). Wymienione badania były przeprowadzone na licznej (N=3965) reprezentatywnej ogólnopolskiej próbie uczniów oraz uwzględniały dużą liczbę zadań matematycznych (8 różnych arkuszy), których wyniki następnie wykalibrowano w modelach IRT (Kondratek, 2009). Ogólny schemat konstrukcji dwóch różnych narzędzi wykorzystanych w preteście oraz postteście, które nie zawierają wspólnych zadań, ale ze względu na zakotwiczenie w badaniach przeprowadzonych przez CKE umożliwiły oszacowanie poziomu umiejętności matematycznych uczniów na wspólnej skali przestawiono w Tabeli 1.

Tabela 1. Schemat podziału zadań pomiędzy narzędzia wykorzystane w preteście i postteście umożliwiający połączenie (ang. *link*) wyników uczniów z różnych grup.

Próba\zbiór zad.	CKE blok zad. A	CKE blok zad. B	CKE blok zad. C	zad. dodatkowe 1	zad. dodatkowe 2
Badanie CKE	✓	✓	✓		
E&K pretest		✓		✓	
E&K posttest			✓		✓

Zadania wykorzystane w badaniach CKE w tym schemacie podzielono na trzy rozłączne zbiory: „blok zadań A”, które nie były wykorzystane w eksperymencie, ale były 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 oraz trafności narzędzi wykorzystanych w badaniach do pretestu, oprócz zadań z bloku B, włączono również „zadania dodatkowe 1”; analogicznie do postestu, oprócz zadań z bloku C, włączono „zadania dodatkowe 2”. Więcej o treściowej zawartości narzędzi wykorzystanych w badaniach można przeczytać w następnych akapitach.

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.

I tak, w przypadku modelowania matematycznego, czyli stosowania narzędzi matematycznych 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 nie pojawiających się podczas lekcji, miał wyeliminować ewentualny efekt wcześniejszego „wytrenowania”. Oto jedno z wykorzystanych zadań tego typu:

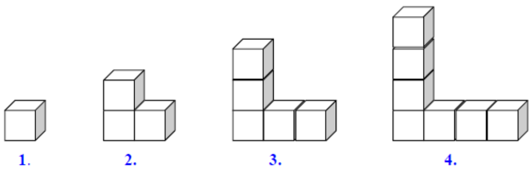
> Jacek i Wojtek mieli po tyle samo lizaków. Wojtek oddał Jackowi dwa swoje lizaki.
> Teraz więc Jacek ma więcej lizaków niż Wojtek. O ile więcej?

Jednym z najważniejszych celów edukacji matematycznej jest budowanie rozumienia systemu dziesiętnego. Jego relacyjne, a nie zdegenerowanie, 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 stosowania w praktyce języka symbolicznego. W celu zbadania, na ile 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

I na koniec 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.

W Tabeli 2 zestawiono zadania z reprezentatywnego badania umiejętności trzecioklasistów przeprowadzonego przez CKE w 2008, które wykorzystano do konstrukcji testu wykorzystanego w pierwszym oraz drugim badaniu umiejętności uczniów w rozbiciu 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.

Tabela 2. Wykorzystanie zadań z reprezentatywnego badania CKE do konstrukcji pre-testu oraz post-testu. „DPW" - dostrzeganie prawidłowości i wyjaśnianie; „SD" – system dziesiętny; „NZT" – nietypowe zadania tekstowe.

Informacje o zadaniach			Liczba obserwacji w pomiarach				
kod zadania w badaniach CKE	obszar umie-jętności	maksym. liczba punktów	badanie 2008	pre-test (grupa kontrolna)	pre-test (grupa eksperym.)	post-test (grupa kontrolna)	post-test (grupa eksperym.)
M2C_7b	DPW	1	940	170	160		
M2C_7as	DPW	4	940	170	160		
M2B_7s	DPW	4	1046			163	149
M2B_5a	SD	1	1046	170	160		
M2B_5b	SD	1	1046			163	149
M2B_5c	SD	1	1046			163	149
M2B_5d	SD	1	1046	170	160	163	149
M1B_6	NZT	1	1046	170	160		
M1A_6	NZT	1	1078			163	149
M2A_6	NZT	1	1077			163	149
M2C_6	NZT	1	940	170	160		

Dobór próby i czas przeprowadzenia badania.

Badanie zostało zrealizowane na poziomie klas trzecich. Wybór poziomu wiekowego był podyktowany zakresem wykorzystania pakietu edukacyjnego GP w procesie kształcenia – w klasie trzeciej możliwości jego zastosowania są największe. Pierwszy pomiar umiejętności uczniów został przeprowadzony we wrześniu 2012 roku, drugi pomiar umiejętności został przeprowadzony na przełomie maja i czerwca 2013 roku.

Wiele czynników organizacyjnych, ekonomicznych oraz merytorycznych ograniczało możliwość losowego doboru szkół do opisywanych badań eksperymentalnych. Wśród czynników organizacyjnych oraz ekonomicznych należy wymienić konieczność ograniczenia realizacji badania do trzech województw (małopolskie, mazowieckie oraz 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 procesie badawczym będą one wybrane spośród listy szkół dobrowolnie zgłaszających chęć udziału w badaniu. Ponadto, przy tak niewielkiej liczby szkół przewidzianej do badania jak 8, zdecydowano, że przeprowadzenie prostego losowego doboru szkół do próby badawczej niosłoby znaczne ryzyko uzyskania próby szkół odbiegającej w znaczny sposób ze względu na istotne charakterystyki od 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 ósemkę szkół do udziału w badaniu eksperymentalnym (Rysunek 1) zostanie dobrane w taki sposób, aby wypełnić każdą z kombinacji powstałych przez skrzyżowanie następujących dwóch zmiennych:

- *lokalizacja szkoły*, która przyjmuje dwie wartości: (i) wieś i miasta poniżej 10 tys. (ii) miasta powyżej 10 tys.
- *średni poziom szkoły*, który przyjmuje cztery wartości powstałe przez podział średnich wyników szkół w ogólnopolskich badaniach umiejętności na równoliczne ćwiartki za pomocą kwartyli rozkładu średnich wyników szkół.

Decyzja o podziale lokalizacji szkoły na takie dwie warstwy jest podyktowana bardzo zbliżonymi wynikami szkół wiejskich oraz szkół w miastach poniżej 10 tys. mieszkańców, w porównaniu z większymi miastami, jaką zaobserwowano między innymi w badaniach OBUT (Pregler & Wiatrak, 2011). Ponadto podział lokalizacji szkół na takie dwie warstwy dzieli populację uczniów w Polsce w przybliżeniu na połowę, dzięki czemu każde z ośmiu pól powstające przez skrzyżowanie tak zdefiniowanych zmiennych *lokalizacja szkoły* oraz *średni poziom szkoły* odpowiada w przybliżeniu 1/8 populacji uczniów klas trzecich w Polsce.

Za najlepszą dostępną miarą *średniego poziomu szkoły* w kontekście omawianego badania eksperymentalnego przyjęto średni wynik szkoły na skali umiejętności matematycznych, jaki był przedstawiony w raportach dla szkół w badaniach OBUT 2011, gdyż dotyczy on umiejętności matematycznych uczniów klas trzecich. Alternatywnie rozważano wykorzystanie w tym celu wyników ze sprawdzianu po klasie VI szkoły podstawowej. Zaletą wykorzystania wyników sprawdzianu do warstwowania szkół byłaby powszechność tego egzaminu (w badaniach OBUT szkoły uczestniczą dobrowolnie i badanie nie obejmuje całej populacji), jednak zdecydowano się na warstwowanie ze względu na wyniki części matematycznej OBUT, ponieważ sprawdzian miał charakter ponad-przedmiotowy oraz dotyczył umiejętności uczniów w klasie VI. W Tabeli 3 przedstawiono zakresy średnich wyników szkół w badaniu OBUT 2011, jakie uzyskuje się przy opisanym podziale szkół na osiem wartości.

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

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

Spośród 34 szkół zgłaszających swój udział w badaniu prowadzących dwa oddziały klas trzecich udało wypełnić jedynie 7 komórek Tabeli 3. Dla warunku: I ćwiartka + wieś & miasta poniżej 10 tys. mieszkańców nie zgłosiła się żadna szkoła prowadząca dwa oddziały klas trzecich, w związku z czym zdecydowano się w jej miejsce zrekrutować dwie różne szkoły jednoodziałowe spełniające 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: (i) dopasowanie do danych wielogrupowego modelu IRT, (ii) wygenerowanie dla każdego ucznia w każdym z pomiarów kompletu *plausible values* (PV) do późniejszego wykorzystania jako wskaźnik poziomu umiejętności, (iii) oszacowanie parametrów modelu wielopoziomowej regresji liniowej, w której zmienną zależną był poziom umiejętności a zmiennymi niezależnymi była przynależność do poszczególnych warunków eksperymentalnych oraz zagnieżdżenie uczniów w pomiarach oraz w szkołach. Każdy z wymienionych etapów zostanie pokrótce opisany.

Przyglądając się Tabeli 1. oraz 2. widzimy, że pełna macierz danych zawiera odpowiedzi na zadania, które zostały udzielone przez uczniów pochodzących z pięciu różnych grup: od uczniów z reprezentatywnego badania CKE w 2008 roku oraz z czterech grup powstałych przez skrzyżowanie dychotomicznych warunków „grupa kontrolna-eksperymentalna” oraz „przed-po oddziaływaniu eksperymantalnym”. Jednocześnie, narzędzia wykorzystane w badaniu 2008, w pre-teście i w post-teście składają się z różnych zadań. Aby zamodelować zmiany w poziomie umiejętności pomiędzy badanymi grupami w takim schemacie wykorzystano metody charakterystyczne dla zrównywania wyników testowych (Kolen & Brennan, 2004, Pokropek & Kondratek 2012, Szaleniec et al., 2012).

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

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

gdzie θ jest losową zmienną ukrytą opisującą poziom umiejętności uczniów; $\psi_{\mathcal{P}}(\theta)$ jest funkcją gęstości prawdopodobieństwa określającą rozkład zmiennej θ w populacji $\mathcal{P}$; $f(\boldsymbol{u}, \theta, \boldsymbol{\beta})$ jest funkcją, która określa prawdopodobieństwo zaobserwowania konkretnej wartości $\boldsymbol{u}$ wektora odpowiedzi $\boldsymbol{U}$, w zależności od poziomu umiejętności θ oraz wektora parametrów $\boldsymbol{\beta} = (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \dots, \boldsymbol{\beta_n})$, gdzie parametry zadania $\boldsymbol{\beta_i}$ również mogą być wektorami. Do zadań ocenianych 0-1 dopasowano dwuparametryczny model logistyczny (*two-parameter logistic model*, 2PLM), natomiast zadania oceniane na więcej punktów modelowano za pomocą modelu oceny stopniowanej (*graded reponse model*, GRM).

Wielogrupowy model IRT został do danych dopasowany w taki sposób, aby średnia oraz odchylenie standardowe rozkładu umiejętności uczniów biorących udział w badaniach 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 uczniów z reprezentatywnej i licznej próby, którzy brali udział w badaniach CKE. Umożliwia to odniesienie wyników uczniów biorących udział w eksperymencie do ogółu populacji, tym samym pozwalając na ocenę trafności zewnętrznej badania.

Ze względu na wykorzystanie narzędzi psychometrycznych do pomiaru umiejętności matematycznych, przy analizie wyników konieczne było uwzględnienie nierzetelności tych narzędzi. W przeciwnym razie oszacowania efektów oraz błędów standardowych byłoby obciążone. Dopasowanie do danych modelu IRT stworzyło możliwość uwzględnienia nierzetelności pomiarów przy analizie danych. Zamiast korzystać z punktowych oszacowań wyników uczniów do weryfikacji postawionej hipotezy badawczej analizy przeprowadzono korzystając z wygenerowanych na podstawie wektora udzielonych przez każdego ucznia odpowiedzi zestawów dwustu tzw. *plausible values* (PV). PV są ciągnięciami 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 teoretycznych podstaw uzasadniających takie podejście dostarcza monografia Roderick'a Little'a oraz Donalda Rubin'a (2002). Przy generowaniu PV oprócz wektora odpowiedzi udzielonych przez ucznia uwzględniono również jako dodatkowe zmienne warunkujące wszystkie zmienne uwzględnione w późniejszej analizie – efekty stałe dla warunków eksperymentalnych oraz zagnieżdżenie uczniów w pomiarach i w szkołach.

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

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

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 grupę eksperymentalną (0=grupa kontrolna; 1=grupa eksperymentalna)

  ${}^{2}X_{ijk}$ – zmienna wskazująca drugi pomiar (0=pre-test; 1=post-test)

  ${}^{1}X_{ijk}\,{}^{2}X_{ijk}$ – iloczyn powyższych zmiennych, tj. zmienna interakcyjna wskaźnika grupy i wskaźnika pomiaru (0=inne przypadki niż „1"; 1=post-test 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 $N(0, \sigma_1)$;
- ε_{0jk} – wartość losowego błędu z poziomu pomiaru, dla pomiaru j zagnieżdżonego w klasie k; przyjmujemy, że ten składnik losowy ma rozkład $N(0, \sigma_2)$;
- ε_{00k} – wartość losowego błędu z poziomu klasy, dla klasy k; przyjmujemy, że ten składnik losowy ma rozkład $N(0, \sigma_3)$.

Istotą powyższej regresji trójpoziomowej jest rozróżnienie składnika resztowego na błąd ε_{ijk} odpowiadający za niewyjaśnioną wariancję uczniowskich wyników zagnieżdżonych w pomiarach (czyli wewnątrz pojedynczego ucznia) σ_1, błąd ε_{0jk} odpowiadający za niewyjaśnioną wariancję między pomiarami wewnątrz klasy σ_2 oraz błąd ε_{00k} odpowiadający za niewyjaśnioną wariancję między klasami σ_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 Rabe-Hesketh & Skrondal (2008). Warto zaznaczyć, że takiego samego modelu regresji użyto również przy warunkowaniu podczas wcześniej opisanego generowania *plausible values*.

Trzy zmienne niezależne włączone do modelu regresji są klasycznym rozwiązaniem dla analizowania w podejściu regresyjnym 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
- 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 edukacyjnego *Gramy w piktogramy* parametrem najistotniejszym. To on określa, ile wynosi wzrost umiejętności w grupie eksperymentalnej w drugim pomiarze, który jest specyficznie związany z oddziaływaniem eksperymentalnym.

Zanim przejdziemy do omówienia wyników analizy konieczne jest jeszcze kilka słów wprowadzenia do metodologii przeprowadzania analiz statystycznych z wykorzystaniem *plausible values*, gdyż zamiast punktowych oszacowań poziomu umiejętności uczniów wykorzystano komplet 200 PV dla każdego ucznia (o generowaniu PV napisano wcześniej). Przy dokonywaniu analiz z wykorzystaniem PV konieczne jest przeprowadzenie obliczeń dla każdego zestawu PV niezależnie i odpowiednie uśrednienie wyników. Jeżeli $\tau(\theta)$ jest nieznanym parametrem, który jest funkcją poziomu umiejętności (czyli w naszym modelu regresji: $\gamma_{000}, \gamma_1, \gamma_2, \gamma_{1x2}$), to uzyskujemy w wyniku analiz na PV tyle oszacowań tego parametru, ile zestawów PV wygenerowano. U nas będzie ich 200: $\widehat{\tau_1(\theta)}$, $\widehat{\tau_2(\theta)}$,…, $\widehat{\tau_{200}(\theta)}$. Każdemu oszacowanu $\widehat{\tau_m(\theta)}$ będzie odpowiadało również oszacowanie wariancji błędu $\widehat{A_m}$ wynikającego z losowania uczniów do próby. Informacja z 200 niezależnych oszacowań parametru $\tau(\theta)$ jest agregowana w jedną statystykę poprzez uśrednienie:

$$\widehat{\tau(\theta)} = \frac{\sum_{m=1}^{200} \widehat{\tau_m(\theta)}}{200},$$

a wariancja błędu tego estymatora jest dana wzorem:

$$\hat{V} = \frac{\sum_{m=1}^{200} \widehat{A_m}}{200} + \left(1 + \frac{1}{200}\right) \frac{\sum_{m=1}^{200} \left(\widehat{\tau_m(\theta)} - \widehat{\tau(\theta)}\right)^2}{200 - 1}.$$

Jak widać, oszacowanie wariancji powstaje poprzez uśrednienie wariancji z każdej z PV oraz dodanie informacji o zróżnicowaniu oszacowań estymatora między PV – w ten sposób uwzględniamy błąd pomiaru we wnioskowaniu statystycznym.

Wyniki

Wyniki dopasowania opisanego modelu regresji do zebranych danych zestawiono w Tabeli 4. Przypomnijmy, że zgodnie z tym w jaki sposób zakotwiczono parametry modelu wielogrupowego IRT, który posłużył do wygerowania PV, rozkład umiejętności matematycznych ma średnią 0 oraz odchylenie standardowe 1 dla uczniów w badaniach reprezentatywnych z 2008 roku, zatem efekty w Tabeli 4. są wyrażone na skali odchylenia standardowego w badaniach reprezentatywnych.

Tabela XXX4. Oszacowania współczynników regresji w modelu wielopoziomowym weryfikującym istotność statystyczną efektów w badaniu skuteczności programu Piktografia; p-wartości w tabeli pokazane dla bezkierunkowych hipotez alternatywnych $H_1: \gamma \neq 0$.

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

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 badaniach reprezentatywnych z 2008 roku. Ten kierunek różnicy nie zaskakuje, gdyż pre-test został przeprowadzony na poczatku klasy 3, natomiast badania reprezentatywne CKE przeprowadzono pod koniec klasy 3. Ze względu na duży błąd oszacowania parametru nie ma jednak 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 i 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 grup eksperymentalnej i kontrolnej.

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

Oszacowanie parametru γ_{1x2}, a zatem parametru interakcji mierzącego wzrost umiejętności matematycznych związany specyficznie z odziaływaniem eksperymentalnym, wyniosło +0,314. A zatem wzrost poziomu umiejętności matematycznych punktowo oszacowano na przeszło 0,3 odchylenia standardowego, co stanowi około 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 wynosi od -0,03 do +0,66 odchylenia standardowego, czyli nawet zahacza nieznacznie o wartości ujemne. Przy bezkierunkowej hipotezie alternatywnej oznaczałoby to efekt nieistotny statystycznie (p-wartość równa 0,078). Dla oceny efektywności oddziaływania eksperymentalnego przy hipotezie badawczej zakładającej określony kierunek wpływu hipotezę zerową dla parametru interakcji testuje się względem kierunkowej hipotezie alternatywnej zakładającej zmianę zgodną z zamierzonym kierunkiem. Dla naszej sytuacji oznacza to parę:

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

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

Przy tak postawionej hipotezie alternatywnej oraz dodatniej wartości oszacowanego efektu γ_{1x2} oznacza to p-wartość równą połowie p-wartości dla bezkierunkowej hipotezy alternatywnej, czyli: 0,039. Przy poziomie istotności α=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 związanym specyficznie ze stosowaniem pakietu edukacyjnego *Gramy w piktogramy*.

Dyskusja

Badanie pokazało, że wzrost umiejętności posługiwania się językiem symbolicznym wśród uczniów z klas korzystających z pakietu edukacyjnego *Gramy w piktogramy* 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. W związku z tym można postawić hipotezę, że zaobserwowany istotny efekt jest bezpośrednią konsekwencją wystąpienia zmian w zakresie czynników powiązanych ze sposobem pracy nauczyciela. Mimo, iż opisane w artykule 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 M., Żytko M. 2013). Kierunek zmian jest zgodny z oczekiwanym, gdyż 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 nieliczną próbę nauczycieli, brakowało mocy statystycznej do uznania wymienionych zmian za statystycznie znamienne.

Dodatkowych informacji w temacie zapośredniczenia zmian w umiejętnościach uczniów poprzez zmiany w stylu pracy nauczycieli dostarczają również dane o charakterze jakościowym zebrane podczas wywiadów przeprowadzonych z nauczycielami na koniec rocznego testowania pakietu. Szczególnie kilka wypowiedzi jest wartych przytoczenia w tym kontekście:[3]

- *Do tej pory uczono nas: działanie, działanie, działanie i to, że działanie musi być zapisane. Pakiet to zmienił. Nas uczono: działanie powinno być zapisane, a pakiet otworzył drzwi do innej formy.*
- *Dzięki uczestnictwu w projekcie zaczęłam bardziej wierzyć w uczniów.*
- *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****.*

Jeżeli przyjmiemy założenie, że zaobserwowana efektywność pomocy dydaktycznej w ogromnej mierze zależała od tego, na ile zachęciła ona nauczyciela do zmiany swojego stylu pracy, do faktycznego intelektualnego zaktywizowania uczniów, to należy uznać, że w przypadku pakietu edukacyjnego *Gramy w piktogramy* ten cel udało się osiągnąć.

Jednocześnie, zebrane dane wskazują na konieczność przeprowadzenia dalszych badań, które w bardziej kompleksowy sposób analizowałyby zależność pomiędzy sposobem nauczania matematyki, a umiejętnościami uczniów, szczególnie w kontekście możliwości wpływania na styl i przekonania zawodowe nauczyciela, tak aby jego praca była bardziej efektywna.

Opisaną w artykule analizę przeprowadzono dla ogólnego wskaźnika umiejętności matematycznych, tak aby uzyskać możliwie najbardziej rzetelną miarę umiejętności. Gdyby jednak przeanalizować wzrost zaobserwowany na tym ogólnym wskaźniku na poziomie zadań (por. Dąbrowski M., Żytko M. 2013) okazuje się, że specyficzny 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. Co ciekawe, także w obszarach nie objętych oddziaływaniami pakietu, a objętych pomiarem: wykonywanie obliczeń oraz rozwiązywanie typowych zadań tekstowych, przyrost umiejętności uczniów klas eksperymentalnych był zauważalnie wyższy niż w klasach kontrolnych. Ostatnie może być sygnałem, że w efekcie

[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).

wykorzystania pakietu nastąpił transfer na te typowo „szkolne” obszary matematyki. Przytoczone wyniki w rozbiciu na podobszary są jednak obarczone obciążeniem niewielkiej próbki zadań przypadających na każdą skalę, należy więc je traktować jako wskazówkę dla możliwych kierunków badań w przyszłości.

Bibliografia

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

Dąbrowski, M. (2013), *(Za)trudne, bo trzeba myśleć*. IBE, Warszawa.

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

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

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

Kolen, M. J., & Brennan R. L. (2004), *Test equating, scaling, and linking: Methods and practice* (2nd ed.). New York, NY: Springer-Verlag.

Imai K., King G., 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), s. 29-53.

Little, R.J.A., Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, 2nd edition, New York: John Wiley.

Lipsey M. W., Hurley S. M. (2008). Design Sensitivity: Statistical Power for Applied Experimental Research. W: Bickman L., Rog D. J. (red.), *The SAGE Handbook of Applied Social Research Methods 2nd ed.*(s. 44-76). SAGE Publications Ltd.

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

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

Pregler A., Wiatrak E., (red.), 2011. *Ogólnopolskie Badanie Umiejętności Trzecioklasistów Raport OBUT 2011*. Warszawa: Centralna Komisja Edukacyjna.

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

Szaleniec, H., Grudniewska, M., Kondratek, B., Kulon, F., 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. Elsevier: *Studies in Educational Evaluation* 31, 114-128.