

RADOSŁAW KACZAN, PIOTR RYCIELSKI, KRZYSZTOF RZEŃCA

Instytut Badań Edukacyjnych

TEST UMIEJĘTNOŚCI NA STARCIE SZKOLNYM – CHARAKTERYSTYKA NARZĘDZIA I JEGO WYKORZYSTANIE W RAMACH PT IBE

Streszczenie: Test Umiejętności na Starcie Szkolnym (TUNSS) jest testem przeznaczonym do badania umiejętności z zakresu nauki pisania i czytania oraz umiejętności matematycznych dzieci rozpoczynających naukę w klasie pierwszej szkoły podstawowej. Pomiar umiejętności dziecka odbywa się przy wykorzystaniu rozległego banku zadań, które zostały przygotowane w procesie badań pilotażowych i normalizacyjnych na podstawie metodologii IRT. Algorytm testowania adaptatywnego (CAT) zapewnia możliwość doboru optymalnych, ze względu na poziom umiejętności, zadań dla każdego dziecka.

Artykuł przedstawia etapy tworzenia narzędzia oraz najważniejsze charakterystyki powstałego testu. Zaprezentowana została również Platforma Testów IBE (PT IBE), która jest przykładem nowoczesnego narzędzia umożliwiającego wykorzystanie właściwości TUNSS na potrzeby tworzenia i udostępniania diagnostom innych testów psychologicznych i pedagogicznych.

Słowa kluczowe: pomiar umiejętności na stracie szkolnym, CAT (*computerized adaptive testing*), umiejętności matematyczne, czytania i pisanie

WPROWADZENIE: POMIAR UMIEJĘTNOŚCI DZIECI NA PROGU SZKOLNYM

Problematyka umiejętności dzieci, które rozpoczynają naukę w szkole podstawowej, może być rozpatrywana z kilku punktów widzenia. Po pierwsze, w kontekście ich gotowości szkolnej w ogóle, a w szczególności gotowości do nauki czytania i pisania w klasie pierwszej (Brzezińska, 2014; Czub, Wiliński, 2014). Ten wątek jest szeroko dyskutowany w związku z obniżeniem wieku rozpoczynania nauki w klasie pierwszej. Przywołując różnego rodzaju argumenty zarówno za, jak i przeciw tej reformie, powoływano się na wyniki osiągane przez dzieci w różnego rodzaju testach, sprawdzianach czy badaniach międzynarodowych (por. Konarzewski, 2013). Analizowane są wyniki dotyczące poziomu umiejętności dzieci pod koniec edukacji przedszkolnej lub na początku szkoły (por. IBE, 2011; Kopik, 2006). Jak również przywołuje się wyniki starszych uczniów i odnosi się do tzw. efektów odroczonej związanej z wcześniejszym bądź późniejszym rozpoczęciem nauki – por. Trends in International Mathematics and Science Study (TIMS) i Progress in International Reading Literacy Study (PIRLS). Po drugie, problem pomiaru umiejętności na starcie szkolnym jest również

istotny z punktu widzenia efektywności wczesnej edukacji, np. jakości pracy przedszkola (por. Yoshikawa i in., 2013). Wyraźnie zatem rysuje się zapotrzebowanie na narzędzia, które mogłyby być wykorzystywane zarówno na szeroką skalę w badaniach naukowych i ewaluacyjnych, jak i w indywidualnej diagnozie dzieci. Narzędzia takie powinny być nie tylko rzetelnie przygotowane z zachowaniem wszelkich prawidłowości konstrukcji testów (por. Hornowska, 2003; Jasińska, Modzelwski, 2014), muszą również być na bieżąco aktualizowane zarówno w kontekście norm, jak i mierzonych umiejętności. Zmiany dotyczące podstawy programowej realizowanej w przedszkolu lub szkole, pojawianie się nowych programów nauczania, wprowadzanie nowych wymagań co do tego, z czym dziecko powinno rozpoczynać naukę w szkole podstawowej skutkuje koniecznością przygotowania nowych i aktualnych narzędzi badawczych. Testy tego typu w rozwiniętych gospodarkach dostępne są z tzw. półki – wyspecjalizowane instytuty przygotowują odpowiednie arkusze testowe i podręczniki. Następnie testy udostępniane są po przeszkoleniu odpowiednim specjalistom. Przykładem może być test PIAT-R (ang. *Peabody Individual Achievement Test*) Fredericka Markwarda – publikowany, z aktualizowanymi normami, od 1989 r. Test ten pozwala m.in. na standaryzowany, wysokiej jakości pomiar umiejętności czytania i pisanie u dzieci 5-, 6- czy 7-letnich. Natomiast wiele z dostępnych i wykorzystywanych w Polsce narzędzi do badania umiejętności przygotowano na potrzeby dzieci 6- i 7-letnich w zupełnie odmiennej sytuacji edukacyjnej (por. Frydrychowicz i in., 2004). Zarówno normy, jak i stawiane przed dzieckiem zadania mogą być już nieaktualne lub nie przystawać do nowych wymagań i doświadczeń dziecka (Krasowicz-Kupis, Wiejak, Gruszczyńska, 2014). W takiej sytuacji bardzo ważna wydała się potrzeba przygotowania rzetelnego i trafnego narzędzia do diagnozy umiejętności dzieci 5-, 6-, i 7-letnich. Dzięki temu możliwe stanie się oszacowanie poziomu kompetencji dzieci znajdujących się na różnych etapach edukacji – zarówno odbywających roczne przygotowanie przedszkolne, jak i rozpoczynających naukę w klasie pierwszej. Do przeprowadzenia takich badań konieczne stało się przygotowanie testu, który uwzględniałby wcześniej wymienione założenie, mierzyl szerokie spektrum umiejętności oraz posiadał aktualne normy przygotowane dla odpowiednich grup wiekowych (Kaczan, Rycielski, 2012).

Prace nad przygotowaniem TUNSS

Najważniejsze założenia TUNSS

Prace mające na celu przygotowanie testu umiejętności na starcie szkolnym zostały zaplanowane w taki sposób, aby finalny test spełniał przyjęte założenia. Wśród tych założeń można wymienić te najważniejsze:

- test ma mieć charakter adaptacyjny;
- test ma mierzyć umiejętności typu szkolnego dzieci rozpoczynających naukę w klasie pierwszej będących po rocznym przygotowaniu przedszkolnym;
- mierzone umiejętności mają mieć szeroki zakres i obejmować umiejętności matematyczne oraz zbiór umiejętności z zakresu pisanie i czytanie, czyli umiejętności niezbędnych do nauki pisanie i czytanie oraz umiejętności samego pisanie i czytanie;
- test ma posiadać normy wiekowe dla dzieci w wieku 5,5–8,5 lat;
- wykonanie testu nie może wymagać od dziecka umiejętności czytanie oraz posługiwania się myszką lub klawiaturą komputera.

Konstrukcja testu

W celu przygotowania ostatecznej wersji TUNSS został sporządzony wyjściowy bank zadań, na który składało się 450 zadań, po 150 na każdą badaną umiejętność. Zadania te zostały przygotowane przez Dorotę Dziedziewicz na podstawie analiz już istniejących narzędzi diagnostycznych, w tym służących do badania gotowości/dojrzałości szkolnej oraz zadań zawartych w różnorodnych pakietach i podręcznikach kierowanych do uczniów klas pierwszych oraz dla 6-latków. Autorka zadań przeanalizowała również zapisy podstawy programowej i właściwych rozporządzeń Ministra Edukacji Narodowej odnoszących się do wychowania przedszkolnego oraz edukacji w pierwszych trzech klasach szkoły podstawowej. Należy zauważyć, że analizy tych zapisów odnosiły się do stanu prawnego obowiązującego na listopad 2011. Wszystkie zadania, które weszły do pierwszej wersji testu, zostały narysowane i opracowane graficznie przez Aleksandrę Gajdę.

Ze względu na przyjęte założenie o adaptatywności testu oraz konieczność dostosowania go do możliwości dzieci w wieku 6 i 7 lat test został zaprojektowany na urządzenie z ekranem dotykowym (tablety). Wszystkie zadania, które zostały przygotowane do testu poza oryginalnym i spójnym opracowaniem graficznym, mają również nagrane przez profesjonalnego lektora polecenia, które pozwalają w sposób wystandaryzowany prowadzić badanie. Natomiast zadaniem odpowiednio przeszkolonego badacza jest przygotowanie badania, nawiązanie kontaktu z dzieckiem oraz obsługa urządzenia i ocena zadań otwartych. TUNSS składa się zatem z banku zadań oraz aplikacji testującej, która zarządza bankiem zadań oraz pozwala na obsługę i prowadzenie testu. Aplikacja umożliwia wprowadzenie danych dziecka, w tym jego płci i daty urodzenia, jak również nadanie mu odpowiedniego ID i – jeśli taka jest potrzeba – przypisanie go do placówki, do której uczęszcza. Szerzej na temat obsługi TUNSS poprzez Platformę Testów IBE można przeczytać w dalszej części artykułu.

Zadania w teście umiejętności na starcie szkolnym odnoszą się zarówno do umiejętności nabywanych przez dzieci podczas rocznego przygotowania przedszkolnego, jak i tych, które rozwijane są podczas nauki w klasie pierwszej (Karwowski, Dziedziewicz, 2012, s. 22). Składają się one na trzy skale: umiejętności matematycznych, umiejętności z zakresu pisania oraz umiejętności z zakresu czytania. W wyniku przeprowadzania testu otrzymujemy oszacowanie poziomu umiejętności badanego dziecka dla tych trzech skal. Jednakże każde z zadań, które znalazło się w teście, jest opisane nie tylko przez przynależność do jednej ze skal, ale również zostało scharakteryzowane poprzez obszar treści, do jakich się odnosi oraz umiejętności ogólne i szczegółowe, które mierzy. W skali umiejętności matematycznych są to cztery obszary. Zostały one scharakteryzowane wraz z przypisanymi do nich umiejętnościami ogólnymi w tabeli 1.

TABELA 1. Skala umiejętności matematycznych – obszary treści oraz umiejętności ogólne

Obszar treści	Umiejętność ogólna
Liczby	(1) liczenie (2) sprawność rachunkowa (3) zapis graficzny liczb (4) działania na zbiorach (5) obliczenia pieniężne
Pomiar	(1) pomiar wielkości (2) pomiar długości (3) pomiar wysokości (4) pomiar czasu
Przestrzeń i kształt	(1) geometria na płaszczyźnie (2) orientacja w przestrzeni
Związki i zależności	(1) regularności (układy rytmiczne) (2) grupowanie obiektów (3) rozumowanie

W skali umiejętności z zakresu pisania są to również cztery obszary treści i wraz z przypisanymi do nich umiejętnościami ogólnymi zostały opisane w tabeli 2.

TABELA 2. Skala umiejętności z zakresu pisania – obszary treści oraz umiejętności ogólne

Obszar treści	Umiejętność ogólna
Sprawność wzrokowo-motoryczna	(1) sprawność manualna (2) percepcja wzrokowa (3) uwaga i pamięć wzrokowa
Sprawność wzrokowo-przestrzenna	(1) orientacja na kartce papieru (2) orientacja w przestrzeni
Sprawność słuchowo-językowa	(1) synteza i analiza głoskowa (2) synteza i analiza sylabowa (3) uwaga i pamięć słuchowa
Kaligrafia i pisanie	(1) znajomość liter (2) kaligrafia (3) pisanie

W ostatniej skali, czyli związanej z umiejętnościami czytania, wyróżniono trzy obszary treści wraz z umiejętnościami ogólnymi (patrz tabela 3).

TABELA 3. Skala umiejętności z zakresu czytania – obszary treści oraz umiejętności ogólne

Obszar treści	Umiejętność ogólna
Sprawność słuchowo-językowa	(1) synteza i analiza głoskowa (2) synteza i analiza sylabowa (3) analiza zdaniowa (4) uwaga i pamięć słuchowa
Sprawność wzrokowo-słuchowa	(1) percepcja wzrokowa (2) słuchanie ze zrozumieniem
Umiejętność czytania	(1) znajomość liter (2) czytanie

Zadania w TUNSS mogą mieć charakter zadań zamkniętych wymagających od dziecka wybrania właściwej odpowiedzi i zaznaczenia jej na ekranie, np. poprzez dotknięcie danego elementu lub wymagających ułożenia czy uporządkowania jakiegoś zbioru poprzez przesuwanie na ekranie określonych elementów. Drugi typ zadań, czyli zadania otwarte, mogą wymagać od dziecka udzielenia krótkiej odpowiedzi, np. podanie liczby lub litery, które są widoczne na ekranie lub udzielenia dłuższej odpowiedzi, np. po wysłuchaniu historyjki dziecko ma powiedzieć ile lat i jak na imię miał jej bohater (Karwowski, Dziedziewicz, 2012 s. 31–33).

Algorytm i aplikacja testująca

Aplikacja testująca obsługuje „serce” całego testu, czyli algorytm testowania adaptatywnego, odpowiadający za wybór pierwszego i kolejnych zadań, które otrzymuje dziecko. Każde zadanie, które znalazło się w ostatecznej wersji TUNSS, opisywane jest według metodologii IRT przez dwa parametry: parametr trudność zdania (b) oraz dyskryminacja (a) (Kondrutek, Pokropek, 2013). Po wprowadzeniu do aplikacji informacji na temat płci i wieku dziecka aplikacja wybiera zadanie, które jest najlepiej dopasowane do prawdopodobnego poziomu umiejętności dziecka. Informacje o tym, jaki poziom umiejętności prezentują dzieci w poszczególnych grupach wiekowych, zostały przygotowane na podstawie badań normalizacyjnych i wprowadzono je do algorytmu. Zatem tak przygotowane normy wiekowe oraz parametry zadań (ich funkcja informacyjna) dają podstawę do określenia punktów startowych testu dla każdej ze skal. Po wyborze pierwszego

zadania na podstawie informacji o tym, jak zostało ono rozwiązane, algorytm wybiera kolejne zadania, które mają najlepsze właściwości, to znaczy dostarczają najwięcej informacji o poziomie umiejętności dziecka. Zadania dobierane są dla każdej ze skal oddzielnie i naprzemiennie. Badanie kończy się, gdy zostanie osiągnięty założony poziom błędu standardowego oszacowania dla danej umiejętności. Dziecko wykonuje nie mniej niż pięć i nie więcej niż 10 zadań na każdą z trzech skal.

Etapy pracy nad testem

W celu przygotowania testu, który spełniałby stawiane przed nim wymagania oraz posiadał odpowiednie właściwości psychometryczne, przeprowadzono trzy etapy badania, które prowadziły do powstania ostatecznej wersji testu. Każdy z tych etapów miał swoje odrębne cele i zostanie oddzielnie scharakteryzowany.

Badanie prepiłotażowe miało trzy podstawowe cele. Po pierwsze, miało ono przynieść informacje na temat tego, jak dzieci w wieku 5–7 lat będą rozwiązywać zadania na tablecie. Czy forma zadań, sposób ich prezentacji oraz aktywność wymagana od dziecka nie będą stanowiły przeszkody w rozwiązywaniu zadań. Po drugie, na tym etapie sprawdzano, czy instrukcje, polecenia oraz sposoby prezentacji zadań są zrozumiałe dla dzieci. Po trzecie, etap ten dał możliwość pierwszego przetestowania działania aplikacji i wprowadzenia daleko idących zmian i modyfikacji w konstrukcji aplikacji.

W badaniach prepiłotażowych wzięło udział 117 dzieci. Charakterystyka przebadanej próby znajduje się w tabeli 4.

TABELA. 4. Charakterystyki próby w badaniu prepiłotażowym

Wiek badanych dzieci	5,5–8,5 roku			
Placówka, do której uczęszczało dziecko	Ostatnia grupa w przedszkolu	„Zerówka” w przedszkolu	„Zerówka” w szkole	Klasa I SP
Wielkość miejscowości	Małe miasto ($N = 30$) i duże miasto ($N = 30$)	Małe miasto ($N = 10$) i duże miasto ($N = 10$)	Małe miasto ($N = 10$) i duże miasto ($N = 10$)	Małe miasto ($N = 10$) i duże miasto ($N = 10$)
Płeć	Równoliczne grupy dziewczynek i chłopców	Równoliczne grupy dziewczynek i chłopców	Równoliczne grupy dziewczynek i chłopców	Równoliczne grupy dziewczynek i chłopców
Liczba zbadanych dzieci	60	18	20	19

Na tym etapie prepiłotażu test składał się z 450 zadań po 150 na każdą skalę. W tabeli 5 przedstawiono liczbę zadań w każdej. Jak widać po prepiłotażu, ze skali matematyki usunięto najmniejszą liczbę zadań, natomiast ze skal umiejętności z zakresu pisania i czytania zadań, które były problematyczne i nie spełniały wymogów nieco więcej.

Zadania usuwano ze względu na dwa kryteria. Po pierwsze, te zdania, których konstrukcja była wadliwa i w trakcie badania badacze odnotowali, że dzieci nie rozumiały poleceń, nie wiedziały, czego się od nich oczekuje lub konstrukcja zadania okazała się niedostosowana do możliwości badania na tablecie, np. część zadań bada-

TABELA 5. Liczba zadań w banku zadań etapy prepilotażowy i pilotażowy

Skala	Ogółem	
	Prepilotaż	Pula zadań, które weszły do pilotażu
matematyka	150	146
pisanie	150	136
czytanie	150	138

jących umiejętność wskazywania położenia przedmiotów w przestrzeni. Po drugie, zadania, które sprawiały trudność większości dzieci ze względu na to, że wymagały posługiwania się umiejętnościami, które nie są ćwiczone ani w przedszkolu, ani w szkole, np. umiejętności wyróżniania sylab w nagłosie, śródgłosie i wygłosie. W stosunku do tego typu zadań poddano je ponownej ocenie eksperckiej i konsultowano ich trudność z nauczycielami uczącymi w klasach 1–3. Na tej podstawie podjęto decyzje o niewłączaniu ich do pilotażowej wersji testu.

Na etapie badania prepilotażowego niezmiernie ważnym elementem była obserwacja zachowań dzieci, ich reakcji na tablet oraz prezentowane na nim zadania. Test właściwy poprzedzany jest tzw. samouczkiem, podczas którego dziecko ma możliwość zapoznania się z tabletem oraz różnego rodzaju czynnościami, jakich będzie od niego wymagał test, np. przeciąganie i układanie elementów, zaznaczanie elementów poprzez dotknięcie go na ekranie. Podczas wykonywania z dzieckiem samouczka badacze mieli okazję obserwować oraz dopytywać dzieci o to, co jest dla nich trudne lub niezrozumiałe. Większość dzieci nie miała problemów z wykonywaniem zadań, choć dla sporej grupy (badanie prepilotażowe odbywało się w styczniu 2012 r.) był to pierwszy kontakt z tabletem. W większości dzieci reagowały bardzo pozytywnie na test i zadania, które traktowały jako rodzaj zabawy. Doświadczenia i informacje zebrane podczas prepilotażu pozwoliły na udoskonalenie zadań i poleceń oraz dalsze dopracowanie aplikacji. Jakościowy charakter prepilotażu sprawił, że dla zadań zostały obliczone tylko statystyki trudności, ale nie były one podstawą podejmowania decyzji co do dalszych losów zadania.

Badanie pilotażowe, na podstawie którego określono parametry zadań, wykorzystane do przygotowania algorytmu testowania adaptatywnego (CAT) zostało przeprowadzone na takich samych grupach wiekowych dzieci jak badanie prepilotażowe. Były to dzieci uczęszczające do przedszkoli oraz do szkół. Szczegółowe liczebności zbadanej grupy zaprezentowano w tabeli 6.

TABELA 6. Liczebności zbadanych dzieci w badaniu pilotażowym w podziale na wiek dziecka i ścieżkę edukacyjną

Grupa:	7-latki	6-latki	5-latki	Suma:
przedszkole	0	281	2286	2567
zerówka w przedszkolu	2	636	384	1022
zerówka w szkole	33	520	270	823
klasa I szkoły podstawowej	666	160	1	827
Suma:	701	1597	2941	5239

Na tym etapie badania wszystkie zadania, które weszły do puli po prepiłotażu, zostały podzielone na 73 zestawy. Każdy z zestawów składał się z 36 zadań po 12 zadań na jedną skalę. Zeszyty były przydzielane dzieciom w taki sposób, który zapewniał, że każde zadanie zostało rozwiązane przez co najmniej 400 dzieci. Uzyskane wyniki pozwoliły określić parametry każdego zadania. Było to możliwe dzięki zastosowaniu metodologii IRT. Wszystkie zadania zostały poddane procesowi skalowania, przy czym dla zadań dychotomicznych zastosowano logistyczny model dwuparametryczny IRT (2PL) (Hambleton, Swaminathan, 1985). Szacowana była moc dyskryminacyjna (współczynnik a) oraz trudność (współczynnik b) każdego z zadań. W przypadku zadań o szerszej skali oceny ocenianych 0–1–2 zastosowano tzw. *generalized partial credit model* (GPCM) (Penfield, Bergeron, 2005). Skalowanie zrealizowane zostało przy wykorzystaniu program XCalibre w wersji 4.6.1 (Karwowski, Dziedziewicz, 2012).

Decyzja co do tego, czy zadanie zostanie włączone do wersji testu, którą wykorzystano w badaniu normalizacyjnym, była podejmowana na podstawie dwóch kryteriów. Po pierwsze, brano pod uwagę parametry każdego z zadań, czyli związki ze skalą, w skład której wchodziło, oraz dopasowanie do modelu teoretycznego IRT. Po drugie, wszystkie zadania poddano ocenie zewnętrznego niezależnego recenzenta, który posiadał kompetencje w ocenie zadań przeznaczonych do diagnozy umiejętności dzieci w tym wieku. Na podstawie tych dwóch kryteriów do dalszych badań na etapie normalizacyjnym przeszło 336 zadań.

W badaniu normalizacyjnym wzięło udział 3769 dzieci w wieku od 5 do 7 lat. Podobnie jak w badaniu pilotażowym były one rekrutowane z oddziałów przedszkolnych, szkolnych oddziałów przedszkolnych (zerówek) i pierwszych klas szkół podstawowych (patrz tabela 7). Badanie zostało przeprowadzone w okresie pomiędzy majem a czerwcem 2012 r. Na tym etapie przygotowania testu zastosowano algorytm testowanie adaptatywnego i pełen bank zadań.

TABELA 7. Liczebności grup w badaniu normalizacyjnym ze względu na wiek dziecka i status edukacyjny

Grupa:	7-latki	6-latki	5-latki	Suma:
przedszkole	3	110	1138	1251
zerówka w przedszkolu	0	742	224	966
zerówka w szkole	6	366	294	666
klasa I szkoły podstawowej	703	183	0	886
Suma:	712	1401	1656	3769

W badaniu normalizacyjnym uzyskano wyniki, które pozwoliły na ocenę poprawności pracy algorytmu oraz na stworzenie na podstawie uzyskanych wyników norm wiekowych dla trzech skal testu. Normy te zostały przygotowane osobno dla dziewczynek i chłopców w wieku od 5,5 do 8,5 roku w przedziałach co pół roku. Przy zastosowaniu wielopoziomowej regresji uwzględniającej pogrupowanie uczniów w klasach stworzono również dodatkowy zestaw norm osobno dla trzech grup wiekowych 5,6–6,5, 6,6–7,5 oraz 7,6–8,5 lat z uwzględnieniem wieku wyrażonego w miesiącach życia. Ten zestaw norm został zaimplementowany w teście. Na podstawie tych norm oraz informacji, do jakiego typu placówki uczęszcza dziecko, prezentowane są raporty z wynikami badania i grupami odniesienia.

Informacje o rzetelności i trafności testu

Test umiejętności na starcie szkolnym został opracowany w taki sposób – aby rzetelność pomiaru w podstawowym zakresie mierzonej cechy nie spadała poniżej odpowiednika alfa Cronbacha = 0,95. Ponieważ do wytworzenia testu zastosowano metodologię probabilistycznej teorii odpowiedzi na zadanie testowe oraz algorytm testowania adaptatywnego – dla każdego dokonanego pomiaru znany jest jego błąd. Błąd pomiaru w prosty sposób można przekształcić na popularną miarę rzetelności alfa Cronbacha. Błąd pomiaru zgodnie z założeniami konstrukcji testu jest związany nieliniowo z poziomem mierzonej cechy badanego. Dla wartości średnich błąd jest najniższy – w większości przypadków zachowuje też dość niski poziom w zakresie mierzonej cechy od -2 do +2 odchylenia standardowego. Powyżej drugiego i poniżej minus drugiego odchylenia standardowego błąd pomiaru istotnie wzrasta. Jest to zjawisko normalne – szczególnie dla testów rekomendowanych do stosowania przesiewowego. Rzetelność pomiaru dla 95% badanych uczniów została określona w próbie normalizacyjnej na akceptowalnym poziomie alfa = 0,95.

Test umiejętności na starcie szkolnym stworzono do pomiaru umiejętności myślenia matematycznego, umiejętności związanych z nauką pisania i umiejętności związanych z nauką czytania. Test składa się zatem z trzech podskal, które pozwalają na obliczanie trzech niezależnych wyników badanego dziecka. Skale te są oczywiście ze sobą skorelowane, jednak nie na tyle, aby można było mówić o tożsamości mierzonych konstruktów. Zależności między analizowanymi skalami przedstawiono w tabeli 8.

TABELA 8. Współczynniki korelacji Pearsona dla skal TUNSS w próbie normalizacyjnej $N = 3061$, wszystkie zależności istotne na poziomie $p < 0,001$

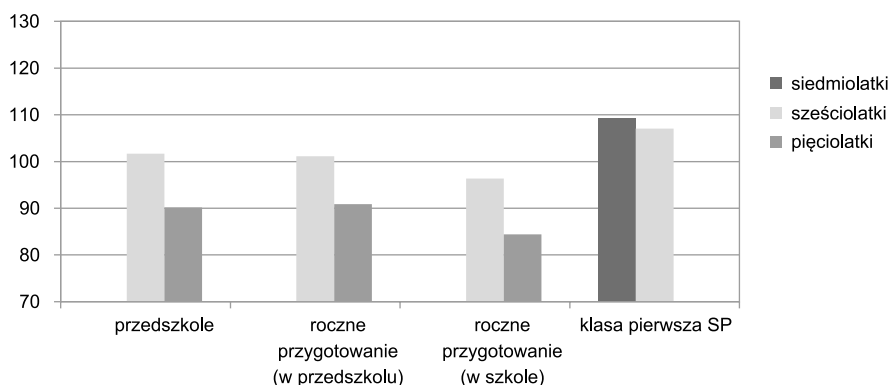
	Umiejętności matematyczne	Umiejętności związane z pisanie	Umiejętności związane z czytaniem
Umiejętności matematyczne	1	0,578**	0,622**
Umiejętności związane z pisanie		1	0,673**
Umiejętności związane z czytaniem			1

Uzyskane wyniki wskazują na około 30% wspólnej zmienności między skalą myślenia matematycznego a skalami językowymi (pisanie i czytanie) oraz na około 40% wspólnej zmienności między skalami językowymi. Wyniki takie *wskazują na znaczną i istotną korelację* badanych konstruktów jednak zdecydowanie nie można twierdzić o ich tożsamości. Wydaje się to cechą dobrego i trafnego narzędzia do pomiaru tego typu umiejętności u dzieci 6- i 7-siedmioletnich.

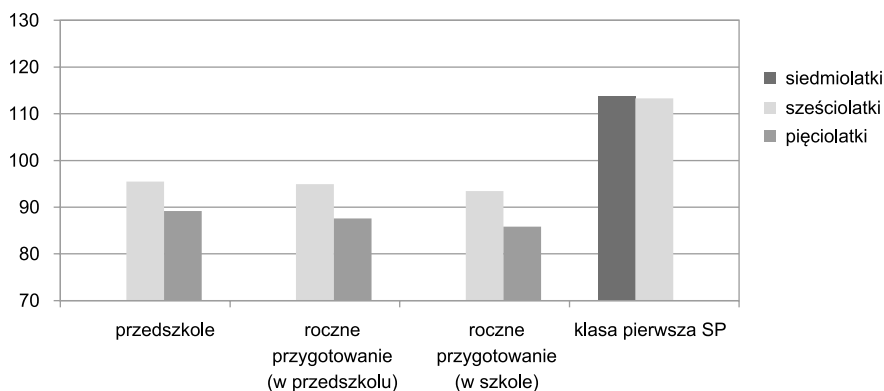
Badania normalizacyjne – wyniki

Trzy podstawowe wyniki uzyskane w badaniu ostateczną wersją testu, to punkty na skali umiejętności matematycznych, czytania i pisanie. Uzyskane wartości, wraz z podziałem na wiek i ścieżkę edukacyjną dziecka, przedstawiają kolejno rysunki 1, 2 i 3.

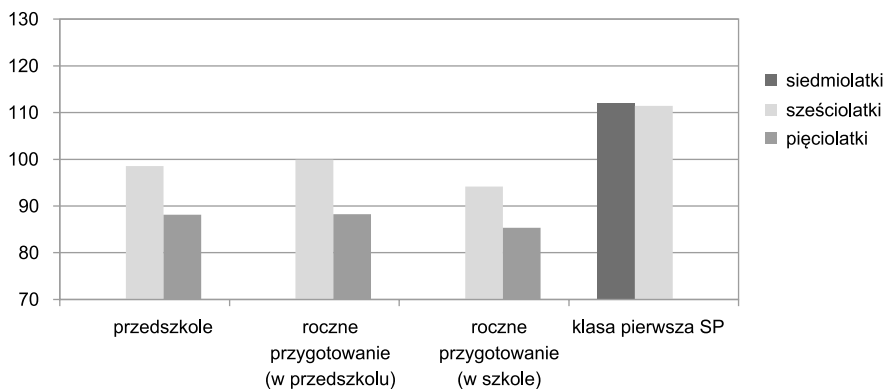
Jak widać na rysunku 1 umiejętności matematyczne badanych dzieci zależne są od ich wieku, ale również od placówki, do której dziecko uczęszcza. Efekt wieku jest liniowy – im dzieci starsze, tym wyższy poziom umiejętności matematycznych prezen-



RYSUNEK 1. Średnie wartości dla skali umiejętności matematycznych w podziale na wiek badanych dzieci i wybraną przez rodziców ścieżkę edukacyjną



RYSUNEK 2. Średnie wartości dla skali umiejętności związanych z pisanem w podziale na wiek badanych dzieci i wybraną przez rodziców ścieżkę edukacyjną



RYSUNEK 3. Średnie wartości dla skali czytania w podziale na wiek badanych dzieci i wybraną przez rodziców ścieżkę edukacyjną

tują. W przypadku porównań między placówkami na uwagę zasługuje niższy poziom umiejętności w tzw. zerówkach szkolnych – czyli szkolnych oddziałach przedszkolnych. Efekt ten jest niezależny od wieku.

Na rysunku 2 przedstawiono średnie wartości skali umiejętności związanych z nauką pisania – efekty wieku i placówki są takie same jak w przypadku umiejętności matematycznych. Na uwagę zasługuje wyraźnie wyższy poziom umiejętności związanych z pisanem, obserwowany w pierwszych klasach szkół podstawowych. Efekt ten związany jest bezpośrednio z podstawą programową – pierwsza klasa to okres, kiedy dzieci rozpoczynają intensywny trening umiejętności pisania.

Rysunek 3 obrazuje średnie wartości na skali umiejętności czytania. Obserwowane tu efekty (wieku i placówki) mają ten sam charakter co w przypadku omawianych wcześniej skal. Umiejętności czytania przyrastają liniowo z wiekiem. Wyraźnie obserwowalny jest efekt niższych kompetencji (niezależnie od wieku) w „zerówkach szkolnych”.

Platforma Testów IBE

Przygotowanie TUNSS (pilotaż i normalizacja) oraz jego wykorzystanie w projekcie mającym na celu oszacowanie poziomu umiejętności dzieci 6- i 7-letnich na starcie szkolnym było okazją do sprawdzenia właściwości narzędzia i jego przydatności w badaniach naukowych (por. Kaczan, Rycielski, 2013, 2014). W kolejnym kroku postanowiono przystosować narzędzie do potrzeb diagnozy indywidualnej dzieci odbywających roczne przygotowanie przedszkolne oraz rozpoczynających naukę w klasie pierwszej. W tym celu jako pierwszy krok w kierunku udostępnienia narzędzia diagnostom, np. tym pracującym w poradniach psychologiczno-pedagogicznych, przygotowano Platformę Testów (PT IBE). PT IBE jest bezpłatną aplikacją dostępną w sklepie Google Play™ przeznaczoną do wykonywania testów psychologicznych, edukacyjnych oraz raportowania wyników. Aplikacja może obsługiwać banki zadań przedstawione jako linearna lista zadań. Jednakże jej główną zaletą jest możliwość zastosowania adaptatywnego algorytmu CAT (ang. *computerized adaptive testing*). Algorytm CAT dobiera zadania dynamicznie, wybór kolejnego zadania zależy od ocenianego na bieżąco poziomu umiejętności badanego. PT IBE daje możliwość wykonywania zadań bezpośrednio na urządzeniu zgodnie z poleceniami przygotowanymi przez autora testu, co zapewnia jednolite, wystandaryzowane warunki badania. Aplikacja obsługuje testy zarówno z automatyczną oceną zadań, jak również te oceniane na bieżąco przez badacza. Ta druga opcja oceny umożliwia projektowanie otwartych zadań, które mogą wymagać od badanego np. wypowiedzi lub wykonania rysunku. Po przeprowadzeniu badań na urządzeniu aplikacja umożliwia wygenerowanie interaktywnego raportu podsumowującego, który wizualizuje wyniki badania. Podobnie jak zadania testowe, szablony raportów są zaprojektowane przez autorów zestawu testów i dzięki zastosowaniu elastycznego języka HTML5 są wysoce konfigurowalne. Moduł raportowania jest zintegrowany z Google Drive™ oraz Google Cloud Print™. Kolejną zaletą PT IBE jest możliwość automatycznego wysyłania anonimowej kopii wyników badań do autorów zestawu testowego. Umożliwia to autorom np. bieżącą aktualizację norm czy też łatwą rekaliibrację testu. Aplikacja nie wysyła żadnych wyników, bez wyraźnej zgody użytkownika. Aplikacja zawiera kilka testów pokazowych (Demo), w tym prezentujące TUNSS. Chcąc wykonać badania, użytkownik musi uzyskać dostęp do zestawów od zewnętrznego dostawcy, który projektuje takie testy dla PT IBE lub też samodzielnie opracować test. Aplikacja została stworzona na system Android i zalecane jest używanie jej na tabletach o przekątnej minimum 10”.

PODSUMOWANIE

Test Umiejętności na Starcie Szkolnym jest w tej chwili najlepszym pod względem psychometrycznym narzędziem dostępnym na rynku polskim z przeznaczeniem do badań dzieci w wieku wczesnoszkolnym. Test ten zapewnia precyzyjne oszacowania umiejętności typu szkolnego w obszarze matematyki, czytania i pisania – traktując te trzy wymiary jako zmienne ciągle. Dzięki przyjęciu takich założeń oraz zastosowaniu algorytmu adaptatywnego możliwe jest uzyskanie wyników np. dla skali pisania również dla dzieci zupełnie nie potrafiących pisać. TUNSS zapewnia szybki i precyzyjny pomiar umiejętności na starcie szkolnym i może być z powodzeniem wykorzystany w szeroko rozumianych badaniach edukacyjnych. TUNSS może być również stosowany przez psychologów i pedagogów do diagnozy przesiewowej w celu identyfikacji uczniów mających specjalne potrzeby edukacyjne. Nie rekomendujemy stosowania testu w przypadku stwierdzonych zaburzeń procesów uczenia się lub w przypadkach problemów poznawczych czy emocjonalnych – konstrukcja testu nie pozwala, by na podstawie otrzymanych wyników stawiać szczegółową diagnozę. W takim przypadku odsyłamy do specjalistycznych lub wielowymiarowych testów zdolności, jak np. IDS.

LITERATURA CYTOWANA

- Brzezińska A. (2014). Gotowość dziecka do szkoły i gotowość szkoły do przyjęcia dziecka: interakcyjne ujęcie gotowości szkolnej. W: A.I. Brzezińska, J. Matejczuk, P. Jankowski, M. Rękosiewicz (red.). *Sześciolatki w szkole: rozwój i wspomaganie rozwoju* (s. 229–245). Poznań: Wydawnictwo Humaniora.
- Czub, M., Wiliński, M. (2014). *Diagnoza gotowości szkolnej w ujęciu interakcyjnym*. Warszawa: IBE.
- Dolata, R. (2009). *Szkoła, segregacje, nierówności*. Warszawa: Uniwersytet Warszawski.
- Frydrychowicz, A., Koźniewska, E., Sobolewska, M., Zwierzyńska, E. (2004). *Testy psychologiczne i pedagogiczne w poradnictwie. Przewodnik metodyczny*. Warszawa: CMPPP.
- Hambleton, R.K., Swaminathan, H. (1985). *Item Response Theory: Principles and applications*. Boston: Kluwer.
- Hornowska, E. (2003). *Testy psychologiczne. Teoria i praktyka*. Warszawa: Wydawnictwo Naukowe Scholar.
- Instytut Badań Edukacyjnych (2011). *Raport o stanie edukacji 2010*. Warszawa: IBE.
- Jasińska, A., Modzelewski, M. (2014). Testy osiągnięć szkolnych TOS3: przykład narzędzia skonstruowanego z wykorzystaniem modelu Rascha. *Edukacja*, 2 (127) 85–107.
- Kaczan, R., Rycielski, P. (2012). Diagnoza umiejętności dzieci 5-, 6- i 7-letnich za pomocą *Testu Umiejętności na Starcie Szkolnym TUNSS*. W: B. Niemierko, M.K. Szmigiel (red.), *Regionalne i lokalne diagnozy edukacyjne* (s. 170–177). Kraków: Wyd. grupa tomami.
- Kaczan, R., Rycielski, P. (2013). Zasoby rodziny a umiejętności szkolne dzieci 6- i 7-letnich. *Studia Edukacyjne*, 27, 159–171.
- Kaczan, R., Rycielski, P. (2014). Co wiemy o 6-latkach w szkole na podstawie badań? W: A.I. Brzezińska, J. Matejczuk, P. Jankowski, M. Rękosiewicz, M. (red.), *Sześciolatki w szkole: rozwój i wspomaganie rozwoju* (s. 247–261). Poznań: Wydawnictwo Humaniora.
- Karwowski, M., Dziedziewicz, D. (2012). *Test Umiejętności na Starcie Szkolnym. Podręcznik*. Warszawa: Instytut Badań Edukacyjnych.
- Konarszewski, K. (2013). Wiek startu szkolnego a osiągnięcia w nauce w okresie wczesnoszkolnym. *Edukacja*, 4, 5–19.
- Kondratek, B., Pokropek, A. (2013). IRT i pomiar edukacyjny. *Edukacja*, 4 (124), 42–66.

- Kondrątek, B., Pokropek, A. (2013). IRT i pomiar edukacyjny. *Edukacja*, 4(124), 42–66.
- Kopik, A. (red.). (2007). *Sześcioletki w Polsce. Raport 2006: Diagnoza badanych sfer rozwoju*. Kielce: Wydawnictwo Tekst.
- Krasowicz-Kupis, G., Wiejak, K., Gruszczyńska, K. (2014). *Katalog metod diagnozy rozwoju poznawczego dziecka na etapie edukacji przedszkolnej i wczesnoszkolnej. Tom I – narzędzia dostępne w poradniach psychologiczno-pedagogicznych i szkołach*. Warszawa: IBE.
- Penfield, R.D., Bergeron, J.M. (2005). Applying a Weighted Maximum Likelihood latent trait estimator to the Generalized Partial Credit Model. *Applied Psychological Measurement*, 29, 218–233.

SKILL TEST AT THE BEGINNING OF SCHOOL EDUCATION – CHARACTERISTIC OF THE TOOL AND ITS UTILITY IN TESTING PLATFORM OF THE EDUCATIONAL STUDIES INSTITUTE

Abstract: Skill test at the beginning of school education (STBSE) is a test designed for assessing writing, reading, and mathematical skills of children who are starting their primary school education. The assessment of a child's skills uses a versatile set of tasks, which were prepared in a process of pilot and normative studies based on IRT methodology. Computerized Adaptive Testing (CAT) allows optimal tasks for each child, based on the child's individual skill level. The article presents the stages of the process of test's

designing and the most important characteristics of the test. Also, the testing platform of the Educational Studies Institute is presented. This platform is an example of a modern tool allowing utilisation of the STBSE for the needs of creating and sharing other psychological and pedagogical tests with the diagnosticians.

Key words: skill assessment at the beginning of school education, CAT (computerized adaptive testing), mathematical, reading, and writing skills