Paulina Skórska
Karolina Świst
dr Artur Pokropek
Instytut Badań Edukacyjnych w Warszawie

Indywidualne i grupowe efekty motywacji testowej uczniów

Wprowadzenie

Motywacja testowa (*test-taking motivation*) odnosi się do stopnia, w jakim egzaminowani są w stanie poświęcić maksimum wysiłku i zaangażowania rozwiązywanemu testowi, tak by w sposób adekwatny odzwierciedlić swój poziom umiejętności (Wise i DeMars, 2005, s. 2). Motywacja testowa ma szczególnie znaczenie w przypadku **testów o niskiej doniosłości** (*low-stakes tests*), to jest takich, których wyniki nie wiążą się z poważnymi konsekwencjami dla ucznia (w przeciwieństwie do testów o wysokiej doniosłości, którymi są egzaminy zewnętrzne decydujące o przyszłości ucznia, np. rekrutacji na studia) (Brown, Barry, Horst, Finney i Kopp, 2009)[1]. Ze względu na brak konsekwencji uczeń może nie traktować testu o niskiej doniosłości jako ważnego oraz nie wkładać wystarczającego wysiłku w jego rozwiązywanie. W takiej sytuacji wyniki testu będą nietrafne, a poziom umiejętności ucznia niedoszacowany. Niski wynik nie będzie odzwierciedlał niskiego poziomu umiejętności, a jedynie niską motywację ucznia (Swerdzewski, Finney i Harmes, 2007). Wise, Ma, Kingsbury i Hauser (2010) wskazali, że trafność wyników testu zależy jednocześnie od poprawnie skonstruowanych zadań i od gotowości ucznia do włożenia wysiłku w jego rozwiązanie, jednak motywacja testowa jest czynnikiem trudnym do kontroli. Dodatkowo wpływa na statystyczne właściwości zadań składających się na test, na przykład na oszacowania ich parametru trudności (Barry i Finney, 2009; DeMars, 2000).

Teoretyczne modele motywacji testowej opierają się najczęściej na teorii Atkinsona (1964) zakładającej, że motywacja składa się z dwóch czynników: *oczekiwania (expectancy)* sukcesu (uzyskania wysokiego wyniku z danego testu) oraz *wartości* (*value*) zadania (znaczenia, jakie dla ucznia ma konkretny test). Pintrich i de Groot (1990) rozszerzyli ten model, dodając trzeci komponent – *afektywny (affective)*, który obejmuje reakcje emocjonalne (emocje ucznia związane z konkretnym testem). Wise i DeMars (2005) dodali do tego modelu czwarty czynnik, którym jest *ważność (importance)*, jaką uczeń przypisuje określonemu testowi. Różne teorie mają jeden czynnik wspólny – wskazują, że motywacja jest zjawiskiem złożonym, należy więc zastanowić się nad doborem adekwatnej koncepcji motywacji leżącej u podstaw danego narzędzia badawczego. W teoriach tych zakłada się także, że motywacja testowa jest zjawiskiem jednostkowym, z drugiej strony ludzie funkcjonują w ramach wielu

[1] Przykładami testów niskiej stawki są: TIMSS (*Trends in International Mathematics and Science Study*), PIRLS (*Progress in International Reading Literacy Study*), PISA (*Programme for International Student Assessment*) czy NAEP (*National Assessment of Educational Progress*).

grup społecznych (Wentzel, 1998; Dennis, Phinney i Chuateco, 2005; Ryan, 2001). Badania wskazują, że w wielu wypadkach indywidualne normy (Willis, 1977) i zachowania (Asch, 1951) są uzależnione od grupowych norm i zachowań. Ludzie dzielą ze sobą wiele definicji sytuacji społecznych, więc można przypuszczać, że sytuacja egzaminowania jest także jedną z nich. Dlatego też niniejsze badanie stawia sobie za zadanie analizę nie tylko motywacji z poziomu jednostkowego, ale także z poziomu grupowego.

Narzędzia pomiaru motywacji testowej

Narzędzia pomiaru motywacji testowej można podzielić na dwie grupy. Pierwsza z nich obejmuje **miary samoopisowe**. Najczęściej wykorzystywane w praktyce to **Kwestionariusz Motywacji** (*Motivation Questionnaire)* (Wolf i Smith, 1993) oraz **Skala Opinii Uczniów (*Student Opinion Scale*)** (Sundre, 2007). Pierwszy kwestionariusz składa się z dwóch skal: *ważności (importance)* testu (pięć pytań kwestionariusza) oraz wysiłku *(effort)* (trzy pytania). Skala zaproponowana przez Sundre (1999) jest rozszerzeniem tej pierwszej, wzbogaconej o dwa dodatkowe pytania. Zaleca się interpretację wyników w powyżej wskazanych podskalach, ze względu na potwierdzoną empirycznie dwuczynnikowość struktury motywacji (Sundre i Finney, 2002). Miary samoopisowe zakładają równocześnie, że osoba wypełniająca jest świadoma swojego poziomu motywacji oraz potrafi go rzetelnie odzwierciedlić na skali (Swerdzewski i in., 2011).

Drugim typem narzędzi pomiaru motywacji testowej są **narzędzia behawioralne**, skupiające się na analizie zachowania uczniów. Miara oparta na **pomiarze czasu odpowiedzi** ucznia (*Response Time Effort*, RTE) zakłada, że niezmotywowani uczniowie nie poświęcają wystarczającej ilości czasu na przeczytanie zadania i udzielenie przemyślanej odpowiedzi (Wise i Kong, 2005). Innym sposobem pomiaru motywacji jest **analiza braków danych** w odpowiedziach uczniów. Niezmotywowani uczniowie mogą wykazywać **tendencję do omijania zadań** w teście. Niestety trudno rozstrzygnąć, czy ominięcia (i odpowiedzi wskazane przez zgadywanie) są spowodowane niską motywacją czy niskim poziomem umiejętności ucznia, który powinien być kontrolowany w każdym badaniu analizującym motywację testową (Dolata i Pokropek, 2010). Badania wskazują, że obydwa typy narzędzi pomiaru motywacji są ze sobą skorelowane. Swerdzewski i in. (2011) wskazali, że zgodność wyników metody opartej na pomiarze czasu odpowiedzi ucznia (*Response Time Effort*, RTE) i Skali Opinii Uczniów jest wysoka (mieści się w zakresie od 64,69% do 68,61%). Podobnie Wise i Kong (2005) wykazali korelację między RTE i podskalą wysiłku SOS na poziomie r=0,34.

Większość badań wskazuje na pozytywną korelację między poziomem motywacji testowej a wynikiem ucznia uzyskiwanym w teście (Wolf i Smith, 1995; Sundre i Kitsantas, 2004; Wise i DeMars, 2005; Thelk i in., 2009). Oznacza to, że zarówno uczniowie o niskim, jak i wysokim poziomie umiejętności mogą mieć obniżony poziom motywacji w testach o niskiej doniosłości. Metaanaliza Wise i DeMars (2005) wskazała, że **wielkość efektu** (standaryzowana różnica średnich) związku między wynikami testów i poziomem motywacji wynosi 0,59. Oznacza to, że zmotywowani uczniowie wypadają w testach o ponad pół odchylenia standardowego lepiej niż niezmotywowani uczniowie.

Cel badania, pytania badawcze i zastosowane metody

Celem badania było określenie, **w jaki sposób** motywacja na poziomie indywidualnym oraz na poziomie klasy wpływa na osiągnięcia uczniów oraz liczbę opuszczonych zadań otwartych. Zadaniem w badaniu była odpowiedź na następujące pytania badawcze:

1. W jakim stopniu motywacja testowa mierzona kwestionariuszem wyjaśnia wariancję wyników testu i liczbę opuszczonych zadań?
2. Czy poszczególne wymiary motywacji (czynniki) wyjaśniają wariancję wyników w teście bezpośrednio, czy też interakcja między czynnikami motywacji ma znaczenie?

Odpowiedź na pierwsze pytanie dostarczyła informacji o **bezpośredniej relacji pomiędzy motywacją testową a wynikami testu/ liczbą opuszczonych zadań otwartych**. W przypadku drugiego pytania określono wpływ interakcji pomiędzy czynnikami motywacji na wyniki/liczbę braków danych w zadaniach otwartych. Choć interakcje są postulowane w szeregu teorii motywacji, ich bezpośredni wpływ na wyniki jest rzadko empirycznie sprawdzany.

Dane

W analizach wykorzystano dane zebrane w ramach **badań zrównujących** wyniki egzaminów zewnętrznych w Polsce (Szaleniec i in., 2012). Badania zostały zrealizowane z wykorzystaniem ośmiu różnych testów rozwiązywanych przez trzy kohorty uczniów (uczniów ostatnich klas szkół podstawowych, gimnazjalnych i ponadgimnazjalnych). Całkowita liczba przebadanych uczniów wynosiła około 50 tys. W badaniu zrównującym, próbka uczniów wypełniała test złożony z zadań występujących we wszystkich poprzednich edycjach egzaminu i dodatkowych zadań kotwiczących. Dane wykorzystane w tych analizach zostały zebrane między 2011 a 2013 rokiem. W niniejszym opracowaniu prezentujemy wyłącznie wyniki dla etapów badań, w których wzięło udział powyżej 100 szkół, w celu uzyskania bardziej stabilnych oszacowań. Poniższa tabela wskazuje na wielkość prób (liczba klas – w nawiasie) w podziale na poszczególne typy szkół i typy egzaminu.

Tabela 1. Liczebność prób (uczniowie i klasy) w badaniu

Typ szkoły	Typ egzaminu	Badanie zrównujące 2011	Badanie zrównujące 2012	Badanie zrównujące 2013
ponadgimnazjalna	matematyka	-	-	4494 (202)
gimnazjum	humanistyczny	9593 (442)	-	-
	matematyczno-przyrodniczy	9551 (454)	-	-
podstawowa	-	-	9086 (449)	-

Kwestionariusz motywacji

Każdy uczeń, który brał udział w badaniach zrównujących, rozwiązywał test i wypełniał krótki kwestionariusz motywacji, zawierający 10 pytań dotyczących motywacji testowej. Po wstępnej eksploracyjnej analizie czynnikowej najlepiej dopasowane do danych okazały się: model jednoczynnikowy dla szkoły podstawowej oraz trzyczynnikowy dla gimnazjum i szkoły średniej. Poniżej przedstawiono pytania dotyczące motywacji oraz ich przypisanie do poszczególnych czynników.

Tabela 2. Struktura czynnikowa kwestionariusza motywacji

Pytania kwestionariusza	Model 1D	Model 3D		
	motywacja	oczekiwanie	afekt	wartość
1. Byłem(-am) dobrze przygotowany(-a) do tego testu.	✓	✓		
2. Starałbym(-abym) się bardziej, gdyby ten test był właściwym egzaminem.	✓			✓
3. Uczyłem(-am) się kilka dni specjalnie do rozwiązywanego dzisiaj testu.	✓	✓		
4. Jestem dobrze przygotowany(-a) do końcowego egzaminu.	✓	✓		
5. Część zadań była nudna, nie rozwiązywałem(-am) ich wcale.	✓		✓	
6. Dzisiaj nie czułem(-am) się na siłach, aby dobrze rozwiązać ten test.	✓			✓
7. Gdyby ten test był prawdziwym egzaminem, byłbym(-abym) lepiej przygotowany(-a).	✓			✓
8. Rozwiązywałem(-am) przede wszystkim pytania zamknięte, na resztę nie miałem(-am) ochoty.	✓		✓	
9. Zadania z tego testu rozwiązałem(-am) najlepiej, jak potrafiłem(-am).	✓		✓	
10. Dzięki temu testowi sprawdzę swoje przygotowanie do egzaminu.	✓		✓	

Do walidacji kwestionariusza motywacji wykorzystano dwa typy zmiennych zależnych: a) sumę punktów z każdego testu (zsumowano punkty za wszystkie zadania poprawnie rozwiązane przez danego ucznia, a następnie poddano zmienną standaryzacji – średnia 0 i odchylenie standardowe 1) i b) liczbę braków danych (ominięcia) w zadaniach otwartych (zadania zamknięte mogą być podatne na zgadywanie odpowiedzi). Poniższa tabela wskazuje statystyki opisowe dla braków danych w zadaniach otwartych: średnią liczbę braków danych, liczbę zadań otwartych w danym teście oraz średni odsetek braków danych w ramach zadań otwartych.

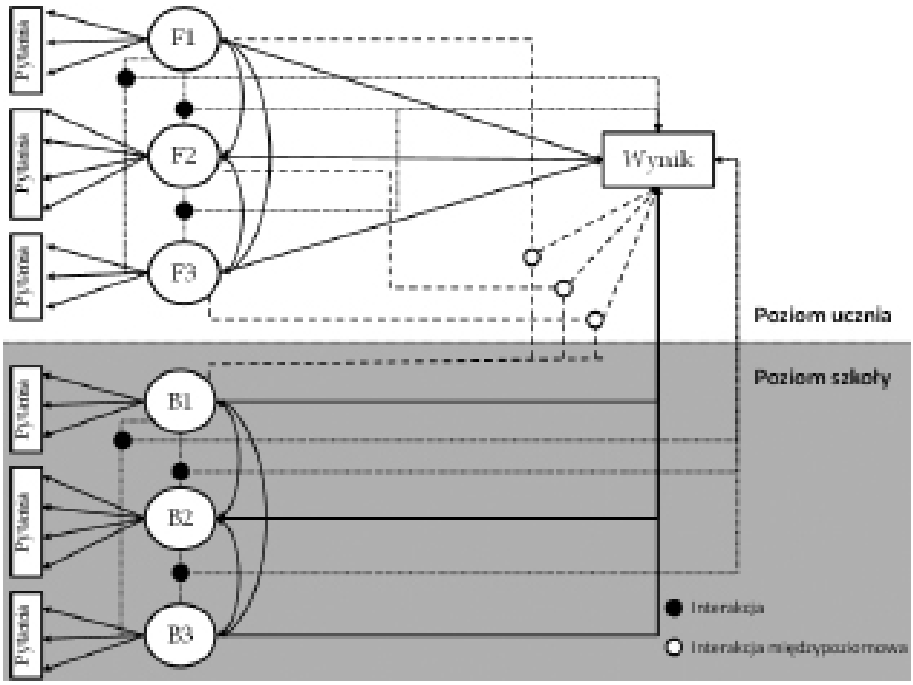
Tabela 3. Statystyki opisowe braków danych w zadaniach otwartych według typu badania

Typ szkoły i rok badania	Typ egzaminu	Średnia liczba braków danych w zadaniach otwartych	Średni odsetek braków danych w zadaniach otwartych
szkoła ponadgimnazjalna 2013	matematyka	1,600	16,970
gimnazjum 2011	humanistyczny	0,888	5,220
gimnazjum 2011	matematyka i przyroda	3,490	65,110
szkoła podstawowa 2012	-	1,150	19,500

Model strukturalny

W modelu wykorzystano dwie zmienne zależne: **wyniki testów** ucznia wypełniającego kwestionariusz motywacji oraz **liczbę braków danych w zadaniach otwartych**. W pierwszym wypadku zmienna zależna była traktowana jako zmienna ciągła, a w drugim jako zmienna o ujemnym rozkładzie dwumianowym (Muthén i Muthén, 2010, s. 28-29).

Dla każdej próby w analizie został oszacowany wielopoziomowy model z interakcjami na poziomie ucznia, szkoły oraz interakcjami międzypoziomowymi. Przykład modelu 3-czynnikowego znajduje się na poniższym rysunku. Czynniki ukryte oznaczono dużymi kółkami, a linie ciągłe wskazują na związki między czynnikami, czynnikami a pytaniami oraz czynnikami a wynikami. Interakcje, efekty (oraz ich interakcje) na I poziomie (białe tło) odnoszą się do wpływu motywacji ucznia na indywidualne wyniki, podczas, gdy efekty na II poziomie (szare tło) wskazują na efekty wpływu motywacji grupowej na indywidualne wyniki. Małe czarne kółka na rysunku reprezentują interakcje między czynnikami, a przerywane linie związki między interakcjami i czynnikami oraz interakcjami i wynikami. Interakcje międzypoziomowe, oznaczone przez białe kółka, zostały oszacowane i włączone w model regresji dla zmiennej wynikowej (kropkowane linie). Model pozwolił określić, czy wymiary motywacji z różnych poziomów mogą ze sobą wchodzić w interakcję. Rysunek 1. pokazuje tylko wycinek międzypoziomowych interakcji dla zobrazowania modelu (F1*FB1, F2*FB1, F3*FB1), lecz w praktyce estymowano model ze wszystkimi możliwymi interakcjami między poziomami (F1*FB1, F2*FB1, F3*FB1, F1*FB2, F2*FB2, F3*FB2, F1*FB3, F2*FB3, F3*FB3). W przypadku jednoczynnikowego modelu nie estymowano interakcji ani na I, ani na II poziomie, lecz tylko interakcje między poziomami.



Rysunek 1. Model strukturalny motywacji testowej szacowany w badaniu

Wyniki

Tabela 4. wskazuje wyniki wielopoziomowego modelu strukturalnego. Lewa część tabeli pokazuje **wyniki ucznia** w teście umiejętności, prawa część dotyczy **liczby braków danych**. Jeżeli chodzi o wyniki odnoszące się do testu umiejętności, jak wynika z tabeli, wzrost o jedno odchylenie standardowe czynnika *oczekiwanie* (f1) wiąże się ze wzrostem wyników z egzaminu w części humanistycznej o 0,074, przy kontroli pozostałych czynników. Grupowy efekt dla czynnika *oczekiwanie* (fb1) jest także istotny i równy 0,262. Jeśli więc średnia klasowa dla czynnika *oczekiwanie* (fb1) rośnie o jedno odchylenie standardowe, to wyniki danego ucznia z części humanistycznej wzrastają o 0,262. Grupowy efekt *afektu* (fb2) jest dodatnio związany z wynikami testu, podczas gdy efekty grupowe czynnika *wartość* (fb3) – ujemnie (choć należy pamiętać o odwróconej skali tego czynnika). Wyniki w części humanistycznej wskazują, że interakcje między czynnikami: *oczekiwanie* (f1) i *wartość* (f3) oraz *afekt* (f2) i *wartość* (f3) są istotne statystycznie. W obu przypadkach czynniki wzmacniają się nawzajem, co oznacza, że różne czynniki motywacji wpływają na wyniki łącznie, a nie niezależnie. Dodatkowo jedna interakcja na poziomie II jest istotna – między grupowym *oczekiwaniem* (fb1) oraz grupową *wartością* (fb3). W efektach międzypoziomowych ujawnia się interakcja pomiędzy *afektem* na poziomie ucznia (f2) i na poziomie klasy (fb2) oraz między *afektem* na poziomie ucznia (f2) i *wartością* na poziomie klasy (fb3). Podobne rezultaty otrzymano w przypadku braków danych, choć interpretacja przy użyciu ujemnego

rozkładu dwumianowego jest nieco inna (modeluje się logarytm oczekiwanych częstości jako funkcję zmiennych niezależnych). W przypadku matematyki dla szkoły ponadpodstawowej czynniki *oczekiwanie* (f1) i *afekt* (f2) na poziomie indywidualnym są ujemnie związane z liczbą opuszczonych zadań. Czynnik *wartość* (f3) jest bliski zeru i nieistotny, wskazując na brak związku z opuszczeniami w zadaniach otwartych. W przypadku poziomu klasy istotne są efekty czynnika *afekt* (fb2) – ujemny związek z liczbą ominięć oraz czynnika *wartość* (fb3) – dodatni (należy jednak pamiętać o odwróceniu skali).

Tabela 4. Wyniki modelu strukturalnego dla wyników testów oraz liczby braków danych (w przypadku braków danych ładunki są niestandaryzowane)

	Wynik testu				Braki danych w zadaniach otwartych			
Czynnik	MAT	HUM	MATiP	SP	MAT	HUM	MATiP	SP
f1	0,102**	0,074**	0,087**		-0,159**	-0,106**	-0,130**	
f2	0,106**	0,128**	0,111**	0,222**	-0,208**	-0,576**	-0,406**	-0,393**
f3	0,000	-0,004	-0,010**		-0,010	-0,017	0,032**	
fb1	0,379**	0,262**	0,178**		-0,224	-0,103	-0,060	
fb2	0,453**	0,161*	0,452**	0,131**	-2,209**	-1,361**	-1,333**	-0,380**
fb3	-0,299*	-0,378**	-0,400**		1,749**	1,071	0,712**	
Istotne statystyczne interakcje	f1f2; f2f3; f1fb1 f1fb2; f2fb3; fb1fb2 fb2fb3	f1f3; f2f3; fb1fb3 f2fb2; f2fb3	f1f2; f1fb3; fb1fb2; fb1fb3	brak	fb1fb2, fb1fb3	f1f2, f2fb2, fb1fb2, fb1fb3, fb2fb3	f1f2, f1fb2, f2fb2, f2fb3, fb1fb2, fb1fb3, fb2fb3	f1fb1

MAT – Matematyka Matura (2013); HUM – Egzamin humanistyczny, gimnazjum (2011); MATiP – Egzamin matematyczno-przyrodniczy (2011); SP = Sprawdzian w szkole podstawowej (2012); **p<0,05 *p<0,1

Podsumowując efekty dwóch pierwszych czynników motywacji – *oczekiwanie* i *afekt* są istotne statystycznie i mogą wyjaśnić wariancję wyników uczniów i ominięć zadań na poziomie indywidualnym. W przypadku czynnika *wartość* efekt praktycznie nie różni się od zera dla wyników egzaminacyjnych i jest istotny statystycznie tylko w przypadku testu z matematyki i przyrody 2011. Dla liczby opuszczeń efekt ten wygląda podobnie.

W badaniu ujawniły się **silne efekty grupowe motywacji**. Efekty wszystkich trzech czynników okazały się istotne i (poza sprawdzianem) wyższe niż na poziomie indywidualnym. Co ciekawe, nawet czynnik *wartość*, nieistotny na poziomie indywidualnym, jest istotny na poziomie grupowym. Te same tendencje charakteryzują modele z liczbą opuszczonych zadań (poza grupowym efektem czynnika *wartość* w próbie uczniów gimnazjum rozwiązującej test humanistyczny w 2011 roku). We wszystkich próbach (dla braków danych) grupowy efekt czynnika *oczekiwanie* nie był istotny, przy jednoczesnej istotności dla czynnika *afekt* i *wartość* (poza testem humanistycznym z 2011 roku). Jest to prawdopodobnie efekt małej mocy testów przy założeniu hierarchicznej struktury danych.

Analizując interakcje na poziomie indywidualnym, można zauważyć istotne efekty między czynnikami *oczekiwanie – afekt* w większości prób. Te efekty wydają się wzmacniać. Co ciekawe, kiedy wyniki testu są zmienną zależną, czynnik *oczekiwanie* wchodzi w interakcję z czynnikiem *wartość* (który samodzielnie jest nieistotny). Wydaje się, że chociaż czynnik *wartość* nie wpływa bezpośrednio na wyniki, jest mediatorem związku między czynnikiem *oczekiwanie* a wynikami testu.

Podsumowując, można stwierdzić, że interakcje na poziomie grupowym przypominają te na poziomie indywidualnym, np. występują istotne interakcje między czynnikami *oczekiwanie* i *afekt* (poza częścią humanistyczną egzaminu w 2011). Wyniki wskazują na dodatkowe istotne interakcje międzypoziomowe. W większości tych interakcji nie ujawnia się jakiś wyraźny wzorzec, ale efekt międzypoziomowej interakcji między *afektem* na poziomie indywidualnym (f2) i *wartością* na poziomie grupowym (fb3) są istotne dla 2 prób (matura 2013 i test humanistyczny 2011) w przypadku analizy wyników testów osiągnięć oraz 2 prób (test humanistyczny i matematyczno-przyrodniczy) w przypadku analizy braków danych.

Podsumowanie i dyskusja wyników

Zarówno efekty indywidualne motywacji, jak i grupowe są istotnymi predyktorami wyników uczniów oraz liczby braków danych w zadaniach otwartych (poza czynnikiem *wartość* na poziomie indywidualnym). Podczas gdy efekty indywidualne motywacji były wykazywane w wielu badaniach empirycznych (np. Wolf i Smith, 1995; Sundre i Kitsantas, 2004; Wise i DeMars, 2005; Thelk i in., 2009), efekty grupowe nie były częstym przedmiotem analiz. Artykuł wskazuje, że efekty te powinny być brane pod uwagę za każdym razem, gdy analizuje się motywację testową. Testowanie może zwiększyć motywację zewnętrzną uczniów (Benmansour, 1999), a społeczne porównania między uczniami wpływają na poziom motywacji w klasie. Nie ma jednak jasności, czy efekty te są wywołane wpływem rówieśników, strategiami nauczycieli czy interakcją obu tych czynników. Biorąc pod uwagę wyniki dla modeli (2-4), uwzględniających interakcje między czynnikami, można stwierdzić, że wielopoziomowe modele związku między motywacją a obiema zmiennymi wynikowymi są znacznie bardziej skomplikowane niż można by oczekiwać na podstawie poprzednich badań. W predykcji wyników uczniów istotne znaczenie mają interakcje na I, II poziomie oraz międzypoziomowe – proste interakcje na poziomie indywidualnym są niewystarczające do opisu relacji między motywacją testową i wynikami uczniów.

Szacowanie poziomu umiejętności uczniów bez uwzględnienia motywacji testowej wprowadza zagrożenie dla trafności testów oraz podejmowanych na ich podstawie decyzji. Zwłaszcza biorąc pod uwagę fakt, że wyniki uczniów niezmotywowanych mogą odzwierciedlać wariancję nieistotną dla motywacji (*construct-irrelevant variance*, CIV) (Messick, 1989; Messick 1995). Powinno to być uwzględnione w praktyce w procesie podejmowania jakichkolwiek decyzji na podstawie wyników testów niskiej doniosłości. Jedną z możliwości poradzenia sobie z problemem CIV jest usunięcie wyników uczniów

niezmotywowanych (filtrowanie) z dalszych analiz (Setzer, Wise, van den Heuvel i Ling, 2013). W takim wypadku należy się jednak upewnić, że kryterium usuwania danych nie jest systematycznie związane z poziomem umiejętności uczniów (może to prowadzić do zmiany rozkładu umiejętności) (Sundre i Wise, 2003). W związku z tym najlepszą strategią jest zachęcanie uczniów do zaangażowania się w rozwiązywanie testów o niskiej doniosłości, np. poprzez przygotowanie ciekawych zadań przez autorów testu (Thelk i in., 2009).

Kolejne badania powinny uwzględniać inne cechy mogące wpływać na motywację testową, takie jak: cechy osobowości (Barry i Finney, 2009), uprzednia wiedza uczniów (Brown, Barry, Horst, Finney i Kopp, 2009) czy zmęczenie przy rozwiązywaniu kolejnych testów lub kolejnych zadań w teście (Barry i Finney, 2009). Warto także w kolejnych badaniach sprawdzić spójność wyników, gdy mierzy się motywację za pomocą kwestionariusza i miar behawioralnych. Warto skupić się na pogłębieniu refleksji nad znaczeniem efektów grupowych oddziaływania motywacji, co nie znalazło do tej pory kompleksowego wyjaśnienia w literaturze. Nie ma jasności, czy efekty grupowe są spowodowane przez wpływ rówieśników (np. Sacerdote, 2011) i/lub wpływ nauczycieli. W naszym badaniu nauczyciele mogli (ale nie musieli) wykorzystać wyniki uczniów podczas oceniania szkolnego. Nie wiadomo, w ilu przypadkach nauczyciel podjął taką decyzję oraz czy poinformował o niej uczniów, co mogło podwyższyć ich motywację. W związku z tym, że w kolejnej (2014) edycji badań zrównujących uwzględniono w kwestionariuszu pytanie o wykorzystanie wyników testu w ocenianiu, pogłębiona zostanie także analiza specyfiki efektów grupowych.

Dalsze analizy powinny skontrolować także, czy wyniki badania zrównującego nie mają znamion testów doniosłych. Kwestionariusz użyty w kolejnych edycjach badania zawiera bowiem pytania dotyczące tego, czy wyniki tych testów nie są wykorzystywane przez nauczycieli w ocenianiu wewnątrzszkolnym i czy uczniowie są świadomi takiego użycia ich wyników.

Bibliografia

1. Asch, S.E. (1951). Effects of group pressure upon the modification and distortion of judgments [w:] H. Guetzkow (red.), *Groups, leadership, and men* (177190). Pittsburgh: Carnegie.
2. Atkinson, J.W. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand.
3. Barry, C.L. i Finney, S.J. (2009). *Exploring Change in Test-Taking Motivation*. Paper presented at 2009 annual meeting of the Northeastern Educational Research Association, Denver, CO.
4. Benmansour, N. (1999). Motivational orientations, self-efficacy, anxiety and strategy use in learning high school mathematics in Morocco. *Mediterranean Journal of Educational Studies*, *4*, 1-15.
5. Brown, A.R., Barry, C.L., Horst, S.J., Finney, S.J., i Kopp, J.P. (2009). *Identifying types of test takers in low-stakes conditions: A mixture modeling approach.* Paper presented at the annual meeting of the Northeastern Educational Research Association, Rocky Hill, CT.

6. DeMars, C.E. (2000). Test stakes and item format interactions. *Applied Measurement in Education*, *13*(1), 55-77.
7. Dennis, J.M., Phinney, J.S., i Chuateco, L.I. (2005). The role of motivation, parental support, and peer support in the academic success of ethnic minority first-generation college students. *Journal of College Student Development*,*46*(3), 223-236.
8. Dolata, R. i Pokropek, A. (2010). Motywacja a wynik testu z nauk przyrodniczych Studium na przykładzie PISA 2006 [w:] Niemierko B., Szmigel M.K. (red.). *Teraźniejszość i przyszłość oceniania szkolnego. XVI Krajowa Konferencja Diagnostyki Edukacyjnej, Toruń, 22-24 października 2010 r.,* Kraków: gRUPA TOMAMI, 86-97.
9. Messick, S. (1989). Validity [w:] R. Linn (Ed.), *Educational measurement* (wyd. 3,13-103). Washington, DC: American Council on Education.
10. Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741-749.
11. Muthén, L.K. i Muthén, B.O. (2010). *Mplus User's Guide. Sixth Edition*. Los Angeles, CA: Muthén & Muthén.
12. Pintrich, P.R i de Groot, E.V. (1990). Motivational and Self-Regulated Learning Components of Classroom Academic Performance. *Journal of Educational Psychology*, *82*(1), 33-40.
13. Ryan, A.M. (2001). The peer group as a context for the development of young adolescent motivation and achievement. *Child development*, *72*(4), 1135-1150.
14. Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? *Handbook of the Economics of Education*, *3*, 249-277.
15. Setzer, J.C., Wise, S.L., van den Heuvel, J.R. i Ling, G.(2013). An Investigation of Examinee Test-Taking Effort on a Large-Scale Assessment. *Applied Measurement in Education, 26*(1), 34-49.
16. Sundre, D.L. (1999). Does examinee motivation moderate the relationship between test consequences and test performance? (Report No. TM029964). Harrisonburg, Va.: James Madison University. (ERIC Documentation Reproduction Service No. ED432588.)
17. Sundre, D.L., i Finney, S.J. (2002). *Enhancing the validity and value of learning assessment: Furthering the development of a motivation scale*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
18. Sundre, D.L. i Kitsantas, A.L. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology*, *29*(1), 6-26.
19. Sundre D.L. i Wise S.L. (2003). *'Motivation filtering': An exploration of the impact of low examinee motivation on the psychometric quality of tests.* Paper presented at the annual meeting of National Council on Measurement in Education, Chicago.
20. Sundre, D.L. (2007). The Student Opinion Scale (SOS): A measure of examinee motivation. Test Manual. Dostępne http://www.jmu.edu/assessment/resources/Overview.html
21. Swerdzewski, P.J., Harmes, J.C. i Finney, S.J. (2011). Two approaches for identifying low-motivated students in a low-stakes assessment context. *Applied Measurement in Education, 24*, 162-188.
22. Swerdzewski, P., Finney, S.J. i Harmes, J.C. (2007). *Skipping the Test: Using Evidence to Inform Policy Related to Those Students Who Avoid Taking Low-Stakes Assessments in College.* Paper presented at the annual meeting of the Northeastern Education Research Association, Rocky Hill, CT.

23. Szaleniec, H., Grudniewska, M., Kondratek, B., Kulon, F., Pokropek, A., Stożek, E. i Żółtak, M. (2012). *Raport tematyczny z badania. Analiza porównawcza wyników egzaminów zewnętrznych – sprawdzian w szóstej klasie szkoły podstawowej i egzamin gimnazjalny*. Dostępne http://eduentuzjasci.pl/pl/publikacje-ee-lista/raporty/181-raport-z-badania/analiza-porownawcza-wynikow-egzaminow-zewnetrznych/946-analiza-porownawcza-wynikow-egzaminow-zewnetrznych-sprawdzian-w-szostej-klasie-szkoly-podstawowej-i-egzamin-gimnazjalny.html
24. Thelk, A.D., Sundre, D.L., Horst, S.J., i Finney, S.J. (2009). Motivation matters: Using the Student Opinion Scale (SOS) to make valid inferences about student performance. *Journal of General Education, 58,* 129-166.
25. Wentzel, K.R. (1998). Social relationships and motivation in middle school: The role of parents, teachers, and peers. *Journal of educational Psychology*, *90*(2), 202.
26. Willis P. (1977). *Learning to labour: how working class kids get working class jobs*. Farnborough: Saxon House.
27. Wise, S.L. i DeMars, C.E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*(1), 1-17.
28. Wise, S.L. i Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, *18*, 162- 183.
29. Wise, S.L., Ma, L., Kingsbury, G.G. i Hauser, C. (2010). *An investigation of the relationship between time of testing and test-taking effort.* Paper presented at the annual meeting of the National Council on Measurement in Education, Denver, CO.
30. Wolf, L.F. i Smith, J.K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education*, *8*, 227-242.
31. Wolf, L.F. i Smith, J.K. (1993). *The effects of motivation and anxiety on test performance*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta.