Paulina Skórska
Karolina Świst
Instytut Badań Edukacyjnych

Wielkość efektu płci w wewnątrzszkolnych i zewnątrzszkolnych wskaźnikach osiągnięć ucznia

Wprowadzenie

Zróżnicowanie wskaźników osiągnięć uczniów ze względu na płeć

Dwie najważniejsze grupy wskaźników osiągnięć uczniów to wyniki oceniania wewnątrzszkolnego i zewnątrzszkolnego. Wewnątrzszkolnymi wskaźnikami osiągnięć ucznia są przede wszystkim **oceny szkolne** oraz **ich średnia arytmetyczna**. Najważniejszym zewnątrzszkolnym wskaźnikiem osiągnięć są **wyniki standaryzowanych testów szerokiego zasięgu**, tj. powszechnych egzaminów zewnętrznych stosowanych w polskim systemie edukacyjnym. Wielkość tych obydwu rodzajów wskaźników osiągnięć jest różna w zależności od tego, czy dotyczy kobiet, czy mężczyzn (*gender gap*). Badania dowodzą, że kobiety osiągają zazwyczaj wyższe oceny w szkole niż mężczyźni (m.in. Konarzewski, 1996), a także kończą szkołę z wyższą **kumulatywną średnią ocen** (*cumulative grade point average*, CGPA), natomiast ich wyniki w standaryzowanych egzaminach zewnętrznych (takich jak amerykański *Scholastic Aptitude Test*, SAT) są zazwyczaj niższe (np. Perkins, Kleiner, Roey i Brown, 2004; Noftle i Robins, 2007). Jednocześnie badania wskazują, że iloraz inteligencji kobiet nie różni się od ilorazu inteligencji mężczyzn, różnice w dyspozycjach psychicznych i zdolnościach umysłowych są nieznaczące (Konarzewski, 1996; American Association of University Women Educational Foundation, 1998; Cole, 1997; Pomerantz, Altermatt i Saxon, 2002). Hyde i Kling (2001) nazywają to zjawisko **efektem niedoszacowania predykcji wyników egzaminacyjnych dla kobiet** (*female underprediction effect*, FUE) i **przeszacowania predykcji wyników egzaminacyjnych dla mężczyzn**. Efekt ten wykazano w szeregu badań empirycznych (Linn, 1973; Stricker, Rock i Burton, 1993; Leonard i Jiang, 1999; Young i Kobrin, 2001). Trend ten pojawia się również podczas dalszych etapów edukacyjnych, gdyż oceny kobiet z większości przedmiotów na studiach są wyższe niż oceny mężczyzn (Willingham i Cole, 1997; Mau i Lynn, 2001). Badania porównawcze wskazują też, że wśród 30 krajów będących członkami Organizacji Współpracy Gospodarczej i Rozwoju dominująca kiedyś przewaga mężczyzn w uzyskiwaniu dyplomów ukończenia studiów zniknęła we wszystkich krajach, oprócz Szwajcarii, Turcji, Japonii i Korei (OECD, 2006).

Należy pamiętać, że zagadnienie zróżnicowania wskaźników osiągnięć ze względu na płeć jest wieloaspektowe. Różnice w osiągnięciach pomiędzy chłopcami i dziewczętami mają w sobie komponent rozwojowy. Na początku edukacji szkolnej wyniki testów z matematyki i czytania są podobne w obydwu grupach,

natomiast później rośnie przewaga chłopców w osiągnięciach z matematyki i dziewcząt z czytania (Maccoby i Jacklin, 1974; Willingham i Cole, 1997). Nie wiadomo jednak, jakie jest zróżnicowanie wyników w obydwu tych grupach, ponieważ brane są pod uwagę różne miary osiągnięć uczniów (Willingham i Cole, 1997) i czy różnice międzypłciowe zmniejszają się w ciągu ostatnich lat (Feingold, 1988; Hyde i in., 1990), czy pozostają stabilne (np. Hedges i Nowell, 1995).

Czynniki warunkujące zróżnicowanie wskaźników osiągnięć uczniów u chłopców i dziewcząt

Zmienną osobowościową, która może wyjaśniać fakt, że dziewczynki uzyskują systematycznie wyższe oceny szkolne niż chłopcy, jest **sumienność** – jeden z najsilniejszych predyktorów ocen szkolnych (Noftle i Robins, 2007; O'Connor i Paunonen, 2007; Poropat, 2009). Uczniowie mający wysokie nasilenie tej cechy uzyskują wyższe oceny szkolne niż można by oczekiwać na podstawie ich wyników w standaryzowanych testach szerokiego zasięgu. Cecha ta występuje w większym nasileniu u dziewczynek niż u chłopców i choć efekt ten jest niewielki, to powtarzalny (Feingold, 1994; Srivastava, John, Gosling i Potter, 2003; Donnellan i Lucas, 2008; Schmitt, Realo, Voracek i Allik, 2008). Sumienność jest także związana z efektem niedoszacowania GPA – im większe nasilenie cechy sumienności, tym wyższe oceny na studiach w stosunku do tych, które przywidywał wynik SAT. Efekt wpływu płci na niedoszacowanie wyniku GPA był w sposób istotny statystycznie mediowany przez sumienność (Kling, Noftle i Robins, 2013). Na inną podobną cechę osobowości – **samodyscyplinę (samokontrolę)** wskazali Duckworth i Seligman (2006). Metaanaliza Silverman (2003) dowodzi, że poziom samodyscypliny u kobiet jest systematycznie wyższy w porównaniu z mężczyznami. Duckworth i Seligman (2006) wykazali, że to właśnie wyższy poziom samodyscypliny u kobiet mediuje związek płci z ocenami szkolnymi. Wyższy poziom samodyscypliny wiąże się ze stylem odrabiania pracy domowej – wskazuje się, że dziewczynki np. zaczynają wcześniej przygotowywanie pracy domowej w ciągu dnia i spędzają dwukrotnie więcej czasu niż chłopcy na jej wykonywaniu.

Niższe oceny szkolne chłopców w stosunku do ocen dziewczynek można powiązać z faktem, że chłopców częściej dotykają **zaburzenia i trudności rozwojowe** na wczesnych etapach nauczania, np. problemy z koncentracją uwagi (Zill i West, 2001). W związku z tym, że ocena szkolna zawiera w sobie coś więcej niż sam komponent dydaktyczny, pełni ona nie tylko funkcję informacyjną, ale także **społeczno-wychowawczą** (Niemierko, 2001). Dlatego też ocena może być traktowana jako nagroda lub kara za zachowanie ucznia (zgodne lub nie z oczekiwaniami nauczyciela). Badania pokazują (Frary, Cross i Weber, 1993), że nauczyciele, wystawiając oceny, uwzględniają różne elementy zachowania ucznia, nie tylko poziom jego umiejętności (Kopciewicz, 2008). Dlatego problemy wychowawcze chłopców i wspomniane kłopoty z utrzymaniem uwagi mogą być systematycznie związane z ich ocenami. Poza tym warto zauważyć, że nauczyciele przydzielają inne typy zadań chłopcom i dziewczętom (Gromkowska-Melosik, 2011), udzielają chłopcom i dziewczętom różnej informacji zwrotnej na temat wykonania zadania (Suchocka, 2011) oraz przejawiają różną gotowość do pracy z chłopcami i dziewczętami (Gromkowska-Melosik,

2011). Dodatkowo podręczniki szkolne charakteryzują się asymetrią relacji płciowych w zależności od nauczanego przedmiotu (Gromkowska-Melosik, 2011). Także funkcja motywacyjna ocen różnicuje się ze względu na płeć (Konarzewski, 1996) – tylko chłopcy o niskim statusie społeczno-ekonomicznym zmieniają swój nakład pracy pod wpływem ocen, jakie uzyskują.

Z kolei wyższe wyniki chłopców na testach szerokiego zasięgu, zwłaszcza z matematyki, tłumaczy się ich **wysokim poziomem pewności siebie** (Preckel, Goetz, Pekrun i Kleine, 2008). Według Steele'a i in. (1997; 1999) kobiety osiągają niższe wyniki na standaryzowanych egzaminach zewnętrznych, gdyż doświadczają **zagrożenia stereotypem**, co wpływa na jakość ich wyników. Badania dowodzą, że stereotypy płciowe na temat wyższych zdolności matematycznych chłopców mogą być wzmacniane przez samych rodziców. Jacobs (1991) wykazał na przykład, że matki, które podzielają pogląd, że chłopcy są bardziej uzdolnieni matematycznie niż dziewczęta, nie doszacowują zdolności matematycznych swoich córek (w stosunku do ich wskaźników osiągnięć). Guiso, Monte, Sapienza i Zingales (2008), wykorzystując dane PISA 2003, analizowali kilka wskaźników równości płci w krajach biorących udział w badaniu, m.in. wskaźnik ***Gender Gap Index*** **(GGI)**. W krajach, które charakteryzują się wysokim wskaźnikiem równości płci, różnica w wynikach testów szerokiego zasięgu między kobietami i mężczyznami jest znacznie mniejsza. Dotyczy to zarówno średnich wyników na testach, jak i poszczególnych przedziałów wyników.

Podsumowując, na wczesnych etapach edukacyjnych wyższe oceny dziewcząt niż chłopców można próbować wyjaśnić większym posłuszeństwem dziewcząt, lepszym zachowaniem w klasie i bardziej pozytywnym stosunkiem do nauki (Zill i West, 2001), a także wspomnianymi cechami osobowościowymi, takimi jak sumienność i samodyscyplina.

Na etapach edukacji, na których uczeń ma możliwość wyboru przedmiotu, tj. głównie na studiach, wskazuje się, że możliwą przyczyną różnic w ocenach między mężczyznami i kobietami jest to, że kobiety wybierają łatwiejsze, łagodniej oceniane kursy (Ramist i in., 1994; Willingham i Cole, 1997). Prawdopodobieństwo uzyskania wyższych ocen z takich przedmiotów jest wyższe. Badania nad tym zjawiskiem dają jednak różne wyniki (Stricker, Rock i Burton, 1993; Duckworth i Seligman, 2006). Nowsze badania (np. Gallagher i Kaufman, 2005) wskazują, że kobiety równie często uczęszczają na zajęcia z matematyki, na których są stawiane wysokie wymagania, i osiągają lepsze oceny niż mężczyźni. Kobiety częściej niż mężczyźni podejmują też na studiach kursy z zakresu nauk ścisłych, takich jak biologia czy chemia (Xie i Shauman 2003). Nawet gdyby efekt wybierania trudniejszych przedmiotów na studiach przez mężczyzn był potwierdzony, zjawisko to nie pozwalałoby na wyjaśnienie różnic międzypłciowych w osiągnięciach wewnątrzszkolnych uczniów na wcześniejszych etapach edukacji, gdy wszyscy uczęszczają na te same przedmioty.

Poza cechami osobowości i predyspozycjami ucznia wskazuje się, że zróżnicowanie płciowe wskaźników osiągnięć ucznia może zależeć od **właściwości narzędzia**, które je mierzy. Loewen, Rosser i Katzman (1988) wskazują, że zadania z części matematycznej SAT mogą być obciążone – odnoszą się w większym stopniu do sytuacji i doświadczeń znanych chłopcom (np. sporty

drużynowe)[1]. Efekt ten może być uwarunkowany również **formatem zadania egzaminacyjnego**. Willingham i Cole (1997) demonstrują, że chłopcy osiągają wyższe wyniki w **zadaniach zamkniętych wielokrotnego wyboru** (*multiple-choice questions*) – najczęściej wykorzystywanym formacie w standaryzowanych testach, natomiast dziewczęta osiągają lepsze wyniki w zadaniach otwartych (*constructed-response questions*), na przykład w wypracowaniach i esejach. Różnice płciowe (*gender gap*) w wynikach standaryzowanych testów szerokiego zasięgu, przejawiające się w faworyzacji chłopców, nasilają się wraz ze wzrostem złożoności i stopnia skomplikowania testu (AAUWEF, 1998).

Część różnic w wynikach egzaminacyjnych próbowano także przypisać **stylowi odpowiadania na pozycje testowe** – na przykład tendencji do zgadywania (powiązanej z cechami osobowości, np. wyższą u mężczyzn skłonnością do ryzyka) bądź omijania odpowiedzi w pytaniach wielokrotnego wyboru. W badaniu przeprowadzonym przez Ben-Shakhara i Sinai (1991) okazało się, że chłopcy mają mniejszą (niż dziewczynki) **skłonność do omijania zadań testowych**, nawet w pytaniach z przedmiotów, w których dziewczynki zwykle przewyższają chłopców wynikami. Zgadywanie poprawnej odpowiedzi i tendencja do nieomijania zadań testowych wiążą się z większym prawdopodobieństwem udzielenia poprawnej odpowiedzi i mogą tłumaczyć wyższe wyniki chłopców w testach szerokiego zasięgu.

Cele opracowania

Celem niniejszego opracowania jest określenie, czy i w jakim stopniu wskaźniki osiągnięć ucznia są uwarunkowane jego płcią. Na realizację tego zadania złoży się odpowiedź na następujące pytania badawcze:

- Jaki jest efekt wpływu płci na szkolne oceny semestralne?
- Jak silny jest ten efekt w przypadku poszczególnych przedmiotów: czy istnieją zauważalne różnice między przedmiotami humanistycznymi i matematyczno-przyrodniczymi?
- Czy efekt płci charakterystyczny dla ocen szkolnych jest również widoczny na poziomie ich średniej?
- W jaki sposób płeć wpływa na wynik egzaminu gimnazjalnego? Czy zauważalne są różnice między przedmiotami humanistycznymi (historia, język polski) i ścisłymi (matematyka, przyroda)?

Metodologia

Opis wykorzystywanych danych

W niniejszej analizie wykorzystano połączone wyniki egzaminu gimnazjalnego z Centralnej Komisji Egzaminacyjnej oraz dane dotyczące wystawionych za pierwszy semestr ocen szkolnych, pochodzących z drugiego i trzeciego etapu badań zrównujących przeprowadzonych przez Zespół Analiz Osiągnięć Uczniów z Instytutu Badań Edukacyjnych. Próba w tych badaniach była dobierana w sposób losowy, operatem była lista szkół udostępniona przez CKE w 2011 i 2012

[1] W badaniach tych sprawdzano różne funkcjonowanie zadań w grupie chłopców i dziewcząt, jednak nie kontrolowano poziomu osiągnięć uczniów.

roku. Przeprowadzono losowanie, które miało charakter warstwowy (proporcjonalne do liczebności szkoły na poziomie klas/klasy oraz wielopoziomowe) (ZAOU, 2013, s. 53). Poniżej przedstawiono także wielkości prób w przypadku poszczególnych testów wraz z podziałem liczebności prób uczniów na płeć.

Tabela 1. Wielkości prób w badaniach ZAOU dla poszczególnych części egzaminu gimnazjalnego w latach 2012-2013 z podziałem na płeć

Rodzaj egzaminu	Rok egzaminu	Wielkość próby	Liczba chłopców	Liczba dziewczynek
historia i WOS	2012	1655	854	801
	2013	1701	852	849
matematyka	2012	1645	849	796
	2013	1682	838	844
j. polski	2012	1656	852	804
	2013	1700	848	852
przyroda	2012	1642	848	794
	2013	1679	836	843

Źródło: opracowanie własne.

Wykorzystane metody

W niniejszych analizach wykorzystano dwie metody: hierarchiczne modelowanie liniowe oraz metody szacowania wielkości efektu (*effect size*) wywodzące się z metaanalizy. Pierwszym krokiem było oszacowanie wpływu płci na oceny szkolne[2] i wyniki egzaminu gimnazjalnego z wykorzystaniem **dwupoziomowego (hierarchicznego) modelu regresji**. Płeć została uwzględniona jako predyktor na poziomie indywidualnym. Wykorzystano w tym przypadku model z losową stałą (*random intercept*). Przesłanką wykorzystania modelu dwupoziomowego był fakt, że struktura populacji uczniów jest hierarchiczna: uczniowie grupują się w klasach, a klasy w szkołach (Geiser i Santelices, 2007). Z uwagi na zgrupowanie uczniów w klasach zostaje **złamane założenie o homoskedastyczności** (niezależności) składnika błędu (*random residual error term*). Oznacza to, że wielkość błędu w modelu regresji nie jest stała dla wszystkich predyktorów. W takim wypadku (Hox, 2010) nie zaleca się stosowania do analizy liniowej regresji wielokrotnej – powoduje to błędne wnioskowanie o zależnościach w populacji. Specyfikacja modelu z losową stałą wygląda następująco (Snijders i Bosker, 1999):

$$Y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \mathrm{U}_{0j} + R_{ij} \qquad (1)$$

gdzie:

Y_{ij} – zmienna zależna, wynik egzaminu gimnazjalnego *i-tego* ucznia w *j-tej* klasie;

γ_{00} – średni współczynnik stałej regresji (*mean intercept*);

[2] Oceny w prowadzonych analizach traktowano jako zmienne mierzone na ilościowej skali. Jest to efektem tradycyjnego w edukacji podejścia do ocen, np. obliczania na ich bazie średniej arytmetycznej. Podejście to jest zgodne z przyjmowanym na świecie, np. przywoływanych w tym artykule badaniach nad efektem wpływu płci na oceny szkolne.

$\gamma_{10}x_{ij}$ – średni współczynnik regresji dla zmiennej niezależnej (ocen/ średniej ocen) *i-tego* ucznia w *j-tej* klasie;

U_{0j} – efekt losowy (*random effect*) na poziomie klasy; reprezentuje odchylenie stałej dla *j-tej* klasy od średniej stałej γ_{00};

R_{ij} – efekt losowy (*random effect*) na poziomie ucznia.

Aby ocenić efekt wpływu płci na wielkość wskaźników osiągnięć zewnątrz- i wewnątrzszkolnych, skorzystano z procedur wykorzystywanych w ramach metaanalizy[3]. Oszacowano wielkość efektu płci, używając **statystyki wielkości efektu** (*effect size*) **g Hedgesa**. Wskaźnik ten należy do szerszego zbioru statystyk opartych na wskaźniku d Cohena, wykorzystujących wystandaryzowaną różnicę pomiędzy średnimi w obydwu porównywanych grupach. Wskaźniki te przyjmują poniższą formę, różnią się natomiast sposobem wyliczenia odchylenia standardowego.

$$\delta = \frac{\mu_1 - \mu_2}{\sigma} \qquad (2)$$

gdzie:

μ_1 – to średnia analizowanej zmiennej (oceny uczniów/ wyniki standaryzowanych testów szerokiego zasięgu) w jednej grupie (kobiety);

μ_2 – to średnia analizowanej zmiennej (oceny uczniów/ wyniki standaryzowanych testów szerokiego zasięgu) w drugiej grupie (mężczyźni);

σ – to łączne odchylenie standardowe wyników w obu porównywanych grupach, obliczane jako $\sigma + \sqrt{\sigma_{g1}^2 + \sigma_{g2}^2 / 2}$, (gdzie σ_{g1} i σ_{g2} to odpowiednio odchylenie standardowe zmiennej w grupie pierwszej i drugiej).

Korekta wprowadzona przez Hedgesa (1981; Hedges i Olkin, 1985) pozwala na redukcję obciążenia ze względu na niewielkie próby (n<20). Wzór na wielkość efektu g Hedgesa wygląda następująco:

$$\vartheta = (1 - \frac{3}{4N-9})\,\delta \qquad (3)$$

gdzie: N jest wielkością próby.

Rice (2009) zaproponował następujące kryteria oceny wielkości efektu m.in. dla g Hedgesa. Mały efekt przyjmuje wartość równą lub mniejszą niż 0,2; średni efekt wynoszący około 0,5, natomiast duży efekt – równą lub większą niż 0,8.

Każda wielkość efektu posiada odpowiadającą jej wagę, odzwierciedlającą precyzję pomiaru efektu w poszczególnych próbach. Najczęstszym rodzajem wag stosowanych w metaanalizie jest odwrotność kwadratowego błędu standardowego w danej próbie (Lipsey i Wilson, 2001). Dzięki użyciu wag można obliczyć średnią wielkość efektu. W praktyce jest to średnia ważona obliczana jako stosunek zsumowanych iloczynów wielkości efektu oraz odpowiadających im wag (*inverse variance weight*) odniesionych do sumy tych wag. Poprawna interpretacja

[3] Wykorzystane podejście nie jest metaanalizą, gdyż nie korzystamy z istniejących, opublikowanych artykułów naukowych, których wyniki podsumowujemy, ale ze statystycznych danych wtórnych. Używane procedury liczenia wielkości efektu są jednak jednym z elementów procedury wykorzystywanej w metaanalizie.

przeciętnych wielkości efektów wymaga także przeprowadzenia analizy homogeniczności (Hedges i Olkin, 1985). Pozwala ona odpowiedzieć na pytanie, czy wszystkie statystyki wielkości efektu, które uśredniamy, estymują ten sam efekt w populacji. W tym przypadku skorzystano ze statystyki I^2 (Higgins i Thompson, 2004; Higgins i in., 2003). Statystyka I^2 wskazuje na procent wariancji efektów wynikający z ich heterogeniczności. Ponieważ w przypadku analizowanych efektów test statystyki I^2 jest istotny i wskazuje na około 90% wariancję efektów spowodowaną niejednorodnością efektów, ich wagi zostały skorygowane zgodnie z logiką modelu efektów losowych (*random effects model*) zamiast modelu efektów stałych (*fixed effects model*) (van den Noortgate i Onghena, 2003)[4].

Wyniki

Efekt luki płci dla ocen szkolnych (i ich średniej)

Współczynniki regresji dla modeli hierarchicznych, w których płeć jest jedynym predyktorem uzyskanej oceny zostały przedstawione w tabeli 2. Gwiazdką zaznaczono współczynniki istotne statystycznie. Można zauważyć, że wszystkie współczynniki są istotne statystycznie i ujemne, co oznacza, że bycie mężczyzną obniża prognozowaną ocenę w porównaniu z byciem kobietą. Efekt ten jest wysoki w przypadku języka polskiego (-0,63 w roku 2012 i -0,71 w roku 2013), niski w przypadku fizyki, matematyki i historii w 2012 roku. W przypadku języka polskiego (w 2012 i 2013 roku) i biologii (w 2013 roku) płeć wpływa na zmianę oceny semestralnej o ponad połowę. Niezależnie od wielkości uzyskanego efektu warto zauważyć, że w przypadku wszystkich przedmiotów współczynniki regresji dla ocen semestralnych są ujemne, wskazując na lepsze oceny dziewczynek. Można też zauważyć, że wszystkie efekty są większe w roku 2013 niż w roku 2012.

Tabela 2. Współczynniki nachylenia w dwupoziomowym modelu regresji płci na semestralne oceny szkolne w roku 2012 i 2013

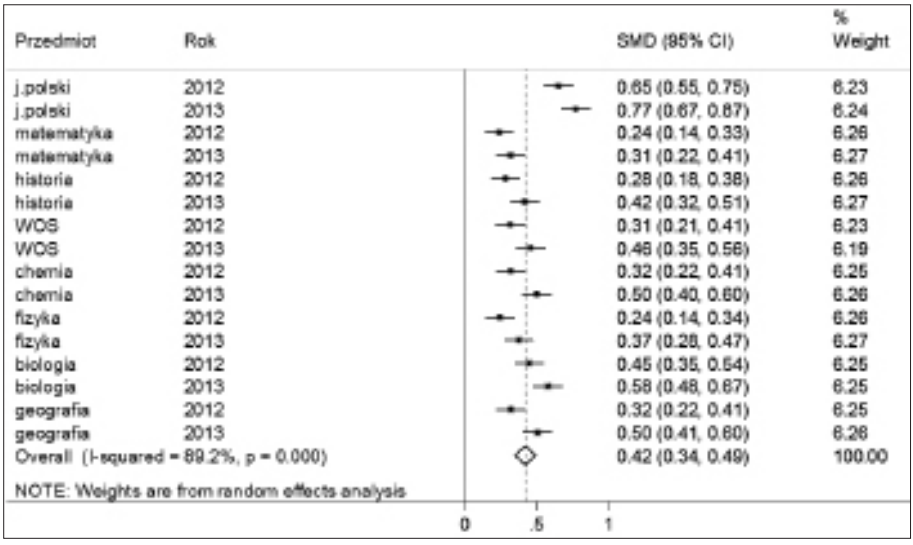
Ocena	Współczynnik regresji dla płci (0 = K, 1 = M)	
	estymator punktowy	95% przedział ufności
j. polski 2012	-0,63*	(-0,72; -0,54)
j. polski 2013	-0,71*	(-0,8; -0,63)
matematyka 2012	-0,23*	(-0,34; -0,13)
matematyka 2013	-0,31*	(-0,4; -0,21)
historia 2012	-0,25*	(-0,35; -0,15)
historia 2013	-0,43*	(-0,52; -0,34)
WOS 2012	-0,32*	(-0,42; -0,22)
WOS 2013	-0,5*	(-0,6; -0,4)
chemia 2012	-0,33*	(-0,44; -0,23)
chemia 2013	-0,5*	(-0,6; -0,41)
fizyka 2012	-0,22*	(-0,31; -0,12)

[4] Dalsza diagnostyka modelu może iść też w kierunku przeprowadzenia tzw. meta-regresji (Thompson i Higgins, 2002) oraz przeprowadzenia kilku metaanaliz w podgrupach efektów. Analiza taka wykracza jednak poza potrzeby i możliwości niniejszego opracowania.

Ocena	Współczynnik regresji dla płci (0 = K, 1 = M)	
	estymator punktowy	95% przedział ufności
fizyka 2013	-0,38*	(-0,47; -0,3)
biologia 2012	-0,45*	(-0,55; -0,35)
biologia 2013	-0,59*	(-0,68; -0,5)
geografia 2012	-0,3*	(-0,39; -0,2)
geografia 2013	-0,47*	(-0,56; -0,38)

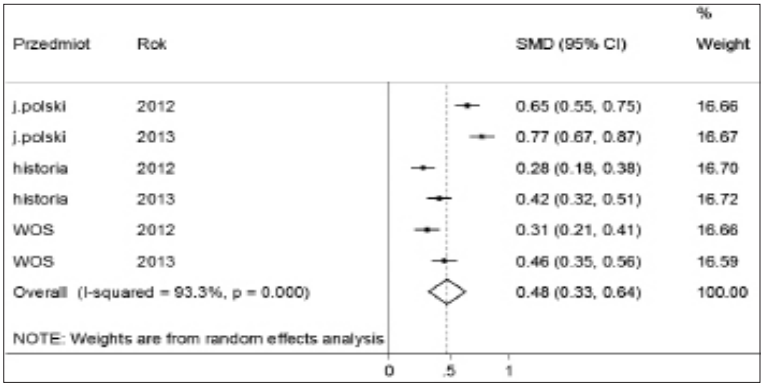
Źródło: opracowanie własne.

Wyniki uzyskane dzięki wykorzystaniu hierarchicznego modelu liniowego zostały potwierdzone poprzez porównanie wielkości efektu (*effect size*) g Hedgesa. Poniżej przedstawiono wykres drzewkowy (*forest plot)* prezentujący statystyki wielkości efektu (i ich przedziały ufności, a także odpowiadające im wagi) dla poszczególnych ocen w poszczególnych latach (wykres 1). Można zauważyć, że w przypadku wszystkich analizowanych przedmiotów dziewczynki otrzymują wyższe oceny niż chłopcy. Efekt wyższej oceny szkolnej u dziewcząt niż u chłopców jest największy w przypadku języka polskiego (0,65 i 0,77 w roku 2012 i 2013). Najniższe wielkości efektu płci charakteryzują oceny z matematyki (0,24 w 2012 roku i 0,31 w 2013 roku) oraz fizyki (0,24 w 2012 roku i 0,37 w 2013 roku). Można też zauważyć, że wielkości efektów w 2013 roku są systematycznie wyższe niż w roku 2012 (różnica pomiędzy wielkościami efektów w przypadku ocen z matematyki jest stosunkowo najniższa). Średni efekt płci dla ocen szkolnych w przypadku wszystkich przedmiotów wynosi 0,42 (z przedziałem ufności: 0,34; 0,49).

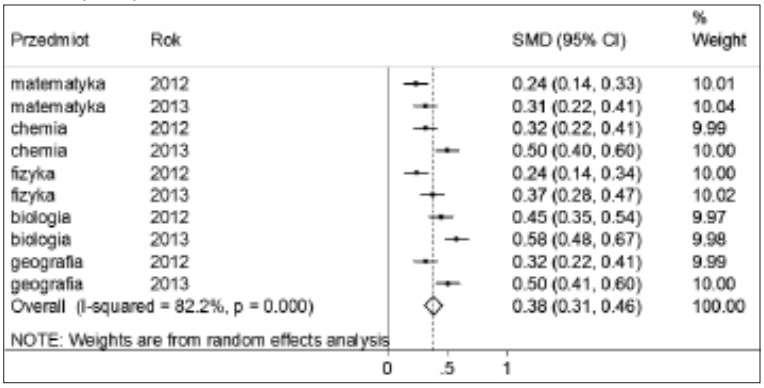


Wykres 1. Wielkości efektu (*effect size*) płci dla ocen semestralnych z poszczególnych przedmiotów szkolnych

Poniższe wykresy przedstawiają wielkość efektu w rozbiciu na przedmioty humanistyczne (wykres 2) i matematyczno-przyrodnicze (wykres 3). Średni efekt przewagi dziewczynek w wysokości ocen szkolnych z przedmiotów humanistycznych wynosi 0,48 (0,33; 0,64) i jest wyższy od analogicznego efektu w przypadku przedmiotów matematyczno-przyrodniczych, który wynosi 0,38 (0,31; 0,46).



Wykres 2. Wielkości efektu (*effect size*) płci dla ocen semestralnych z przedmiotów humanistycznych



Wykres 3. Wielkości efektu (*effect size*) płci dla ocen semestralnych z przedmiotów matematyczno-przyrodniczych

Podobne wyniki (do tych uzyskanych dla ocen szkolnych z poszczególnych przedmiotów) uzyskano dla średniej ocen szkolnych. Współczynniki regresji dla modeli, w których płeć jest jedynym predyktorem uzyskanej średniej ocen (w podziale na rok badania) zostały przedstawione w tabeli 3. Gwiazdką zaznaczono współczynniki istotne statystycznie. Można zauważyć, że w roku 2012 średnia ocen chłopców jest o około jedną trzecią niższa niż średnia ocen dziewcząt. W roku 2013 średnia ocen chłopców jest o około pół jednostki niższa niż średnia ocen dziewcząt. Prawidłowości te replikują wielkości efektów g Hedgesa dla płci, które wskazują na znacząco wyższą średnią ocen wśród kobiet.

Tabela 3. Współczynniki regresji płci na średnią ocen szkolnych i wielkość efektu płci g Hedgesa

Rok badania	Współczynnik regresji dla płci (0 = K, 1 = M)		Wielkość efektu – g Hedgesa
	estymator punktowy	95% przedział ufności	
2012	-0,33*	(-0,42; -0,25)	0,41
2013	-0,49*	(-0,56; -0,41)	0,6

Źródło: opracowanie własne.

Efekt luki płci dla wyników standaryzowanych testów szerokiego zasięgu (egzamin gimnazjalny)

W celu określenia wielkości wpływu płci na wyniki egzaminu gimnazjalnego wykorzystano model dwupoziomowej regresji, w którym jedynym predyktorem wyniku egzaminacyjnego uczniów była ich płeć. Współczynniki regresji dla poszczególnych testów (oraz ich 95% przedziały ufności) przedstawia tabela 4. Gwiazdką zaznaczono współczynniki istotne statystycznie.

Istotne statystycznie efekty występują w przypadku egzaminów z matematyki w 2012 roku, historii w 2012 oraz obydwu egzaminów z języka polskiego. W przypadku egzaminu gimnazjalnego z języka polskiego wynik egzaminacyjny chłopców (w porównaniu z wynikiem dziewczynek) spada przeciętnie o około 2,25 punktu (w roku 2012) i 2,5 punktu (w roku 2013), przy maksymalnej możliwej do osiągnięcie liczbie punktów równej 32. w przypadku dwóch pozostałych egzaminów przejście z grupy kobiet do grupy mężczyzn powoduje przyrost wyniku egzaminacyjnego o 0,8 punktu (historia) i 0,55 punktu (matematyka)[5].

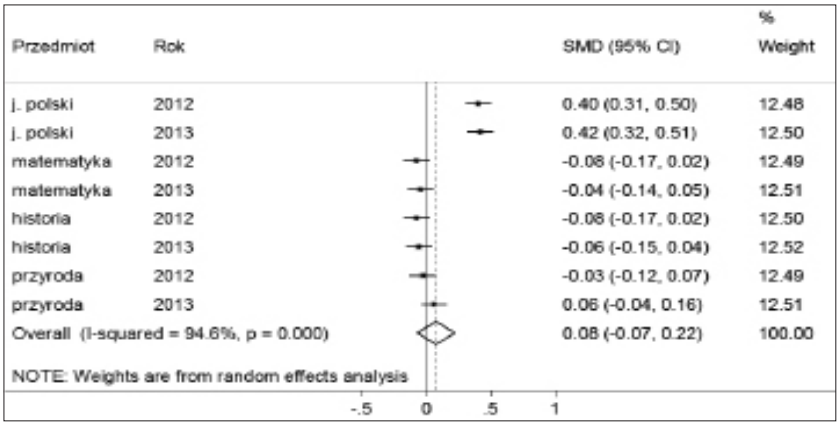
Tabela 4. Współczynniki nachylenia w dwupoziomowym modelu regresji płci na wyniki egzaminu gimnazjalnego w 2012 i 2013 roku

Egzamin	Współczynnik regresji dla płci (0 = K, 1 = M)		Maksymalna liczba punktów do uzyskania z danego egzaminu
	estymator punktowy	95% przedział ufności	
matematyka 2012	0,55*	(0,01; 1,08)	29
matematyka 2013	0,29	(-0,26; 0,85)	30
przyroda 2012	0,37	(-0,03; 0,78)	26
przyroda 2013	-0,29	(-0,69; 0,12)	28
j. polski 2012	-2,23*	(-2,75; -1,7)	32
j. polski 2013	-2,46*	(-2,98; -1,93)	32
historia 2012	0,8*	(0,24; 1,35)	33
historia 2013	0,4	(-0,07; 0,88)	33

Źródło: opracowanie własne.

[5] Model uwzględniający jedynie płeć jako predyktor wyjaśnia nikły procent wariancji wyników egzaminacyjnych (około 4% w przypadku języka polskiego, w przypadku innych egzaminów procent ten jest bliski 0).

Na poniższym wykresie przedstawiono wielkości efektów w przypadku poszczególnych części egzaminu gimnazjalnego. W przypadku standaryzowanych egzaminów zewnętrznych znaczący efekt na korzyść dziewcząt występuje wyłącznie w przypadku egzaminów z języka polskiego (odpowiednio 0,40 w 2012 roku i 0,42 w 2013 roku). W przypadku pozostałych egzaminów efekt płci jest równy zeru.



Wykres 4. Wielkości efektu (*effect size*) płci dla wyników poszczególnych części egzaminu gimnazjalnego w 2012 i 2013 roku

Dyskusja wyników

Zgodnie z występującymi w literaturze i w badaniach tendencjami dziewczęta w Polsce uzyskują wyższe oceny w przypadku wszystkich analizowanych w tym artykule przedmiotów gimnazjalnych. Efekt ten występuje zarówno w przypadku przedmiotów humanistycznych (najwyższy w przypadku języka polskiego), jak i przedmiotów ścisłych (najwyższy w przypadku chemii, biologii i geografii). Zgodnie z klasyfikacją Rice'a (2009) efekty dla języka polskiego, chemii, biologii i geografii mieszczą się w przedziale pomiędzy średnim a dużym efektem. Warto też zauważyć, że efekt płci jest systematycznie wyższy w roku 2013, co wydaje się być zagadnieniem wartym dalszych analiz. Uzyskane wyniki dla ocen z przedmiotów szkolnych w gimnazjum potwierdza analiza z wykorzystaniem średniej ocen szkolnych – porównując grupę chłopców z grupą dziewcząt, można zauważyć, że średnia ocen szkolnych jest wyższa u dziewcząt. Zgodnie z przytaczaną klasyfikacją Rice'a (2009) wielkość tych efektów jest średnia. W dalszych badaniach należałoby kontrolować potencjalny efekt podwyższania ocen przez nauczycieli, gdyż jak wykazał Konarzewski (1996) nauczyciele są skłonni podwyższać oceny dziewczynkom o niższym statusie społeczno-ekonomicznym i nie są skłonni podwyższać ocen chłopcom, mimo wzrostu ich pracowitości z semestru na semestr. Dodatkowo warto kontrolować w badaniach nad efektem płci status społeczno-ekonomiczny, gdyż może on wchodzić w interakcję z płcią ucznia (Konarzewski, 1996).

W przypadku standaryzowanych egzaminów zewnętrznych efekt (na korzyść dziewcząt) jest w stopniu znaczącym różny od zera wyłącznie dla egzaminu z języka polskiego w 2012 i 2013 roku. Zgodnie z klasyfikacją Rice'a (2009) dla

obydwu tych egzaminów wystąpił średni efekt płci. W przypadku innych egzaminów efekt różnic pomiędzy płciami praktycznie nie różni się od 0. Różnice międzypłciowe w wynikach z egzaminu z języka polskiego są zgodne z wynikami badań PISA (OECD, 2009, s. 16-17). Różnica w wynikach z czytania wśród piętnastolatków zwiększyła się pomiędzy rokiem 2000 a 2006 z 32 do 38 punktów (odchylenie standardowe w badaniach PISA wynosi 100). Jest to spowodowane pogorszeniem się w sposób istotny statystycznie (o 10 punktów) wyników chłopców. Raport OECD (2009) wskazuje, że wynik ten można powiązać z inną zmienną – indeksem zaangażowania i zainteresowania czytaniem. Wielkość ta jest wystandaryzowana (średnia – 0 i odchylenie standardowe – 1). W roku 2000, wartość ta dla chłopców wynosiła -0,24, natomiast dla dziewcząt 0,26 (czyli pół odchylenia standardowego różnicy). We wszystkich krajach chłopcy wykazywali mniejsze zainteresowanie czytaniem. Choć autorzy raportu zastrzegają, że nie da się udowodnić związku przyczynowego pomiędzy tą wielkością a wynikiem uzyskanym w badaniu PISA, różnica jednego odchylenia standardowego w indeksie zainteresowania czytaniem wiąże się ze zmianą wyniku z czytania o 27,9 punktu (OECD, 2009, s. 17). Rozstrzygnięcie, co jest przyczyną występowania zjawiska niższych wyników chłopców na egzaminach z języka polskiego – zróżnicowanie cech osobowościowych w grupie chłopców i dziewcząt (Duckworth i Seligman, 2006; Kling, Noftle i Robins, 2013), uzyskiwanie wyższych wyników przez dziewczęta w zadaniach otwartych (np. DeMars, 2000), które to przeważają w egzaminie z języka polskiego, czy inny styl nauczania języka polskiego w stosunku do dziewcząt i chłopców – wymaga dalszych badań.

W Polsce nie występuje jednak opisywany wcześniej efekt polegający na niedoszacowaniu wyników egzaminacyjnych w grupie chłopców. Według Guiso i in. (2008) taki efekt może być związany z wyższym wskaźnikiem równości płci (*Gender Gap Index*). Indeks ten składa się z czterech wymiarów: ekonomicznego, politycznego, edukacyjnego i zdrowotnego. Zgodnie z raportem *The Global Gender Gap Report* (2013) Polska w 2013 roku uzyskała 54. miejsce (na 136 analizowanych krajów), natomiast w 2012 roku 53. Analizując wymiar nierówności edukacyjnych, Polska osiągnęła w 2013 roku 37. miejsce (z wynikiem 0,9983, o 0,0015 wyższym niż w 2012 roku). Maksymalna wartość indeksu wynosi 1, tak więc mimo uzyskanego 37. miejsca[6] dostęp do osiągnięć edukacyjnych kobiet i mężczyzn jest wyrównany – stąd też mogą wynikać porównywalne wyniki dziewcząt i chłopców z egzaminu gimnazjalnego.

Podsumowując, prezentowane w tym opracowaniu wyniki są spójne z wynikami publikowanymi w artykułach z zagranicznych czasopism naukowych w zakresie przewagi dziewczynek nad chłopcami w osiąganych ocenach szkolnych i ich średniej. W Polsce nie replikują się natomiast efekty dotyczące wyników chłopców i dziewczynek w standaryzowanych testach szerokiego zasięgu. Chłopcy nie uzyskują znacząco wyższych rezultatów – obie płcie uzyskały na egzaminie gimnazjalnym 2012 i 2013 podobne wyniki. Dla egzaminu z języka polskiego efekt jest odwrotny do spodziewanego (na podstawie literatury naukowej) – wyniki

[6] 25 krajów osiągnęło wartość indeksu równą 1,0, stąd też analiza pozycji Polski wyłącznie na podstawie jej miejsca w rankingu może być niemiarodajna.

dziewczynek przewyższają wyniki chłopców. Niniejsze opracowanie ma charakter eksploracyjny – wskazuje hipotezy mogące tłumaczyć uzyskane rezultaty. Sprawdzenie ich zasadności, a także zbadanie opisywanych zjawisk w innych powszechnych egzaminach zewnętrznych (sprawdzian, matura), powinno się stać przedmiotem dalszych badań nad luką płci we wskaźnikach osiągnięć uczniów.

Bibliografia

1. American Association of University Women Educational Foundation [AAUWEF] (1998). *Gender gaps: Where schools still fail our children*. Washington, DC: Author.
2. Ben-Shakhar, G. i Sinai, Y. (1991). Gender Differences in Multiple-Choice Tests: The Role of Differential Guessing Tendencies. *Journal of Educational Measurement*, *28*(1), 23-35.
3. Cole, N. S. (1997). *The ETS gender study: How males and females perform in educational settings*. Princeton, NJ: Educational Testing Service.
4. DeMars, C.E. (2000). Test Stakes and Item Format Interactions. *Applied Measurement in Education, 13*(1), 55-77.
5. Donnellan, M.B. i Lucas, R.E. (2008). Age differences in the Big Five across the life span: Evidence from two national samples. *Psychology and Aging, 23*, 558-566.
6. Duckworth, A.L. i Seligman, M. E.P. (2006). Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores. *Journal of Educational Psychology, 98*(1), 198-208.
7. Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43*, 95-103.
8. Feingold A. (1994). Gender differences in personality: a meta-analysis. *Psychological Bulletin, 116*(3), 429-456.
9. Frary, R.B., Cross, L.H., i Weber, L.J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. *Educational Measurement: Issues and Practice, 12*(3), 23-30.
10. Gallagher, A.M. i Kaufman, J.C.(2005). (red.). *Gender Differences in Mathematics: an Integrative Psychological Approach*. Cambridge, UK: Cambridge Univ. Press.
11. Geiser, S. i Santelices, M.V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs standardized tests as indicators of four-year college outcomes*. Berkeley, CA: Center for Studies in Higher Education, University of California, Berkeley.
12. Gromkowska-Melosik, A. (2011). *Edukacja i (nie)równość społeczna kobiet. Studium dynamiki dostępu*, Kraków: Oficyna Wydawnicza Impuls.
13. Guiso, L., Monte,F., Sapienza, P. i Zingales, L. (2008). Culture, Gender and Math. *Science, 320*(5880), 1164-1165.
14. Hedges, L.V. (1981). Meta-Analysis. *Journal of Educational Statistics, 4*, 279-296.
15. Hedges, L.V. i Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
16. Hedges, L.V. i Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*, 41-45.
17. Higgins, J. P. T., Thompson, S. G., Deeks, J. J. i Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557-560.
18. Higgins, J. P. T. S. G. i Thompson, J. J. (2004). Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine, 23*, 1663-1682.
19. Hox, J.J. (2010). *Multilevel Analysis. Techniques and Application*. New York: Routledge.

20. Hyde, J.S., Fennema, E., Ryan, M., Frost, L.A. i Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: a meta-analysis. *Psychology of Women Quarterly, 14*(3), 299-342.
21. Hyde, J.S. i Kling, K.C. (2001). Women, motivation, and achievement. *Psychology of Women Quarterly, 25*, 364-378.
22. Jacobs, J.E. (1991). Influence of Gender Stereotypes on Parent and Child Mathematics Attitudes. *Journal of Educational Psychology, 83*(4), 518-527.
23. Kling, K.C., Noftle, E.E. i Robins, R.W. (2013). Why Do Standardized Tests Underpredict Women's Academic Performance? The Role of Conscientiousness. *Social Psychological and Personality Science, 4*(5), 600-606.
24. Konarzewski, K. (1996). Problemy i schematy. Pierwszy rok nauki szkolnej dziecka. Warszawa: "Żak".
25. Kopciewicz L., *Grzeczne dziewczynki, niegrzeczni chłopcy — wytwarzanie różnic rodzajowych w dydaktyczno-wychowawczej pracy szkoły*, [w:] M. Dudzikowa, M. Czerepaniak-Walczak (red.) *Wychowanie. Pojęcia. Procesy. Konteksty. Interdyscyplinarne ujęcie*, tom 4, Gdańsk: GWP, 2008.
26. Leonard, D.K. i Jiang, J. (1999). Gender bias and the college predictions of the SATs: A cry of despair. *Research in Higher Education, 40*(4), 375-407.
27. Linn, R.L. (1973). Fair test use in selection. *Review of Educational Research, 43*, 139-161.
28. Lipsey M.W. i Wilson D. (2001). *Practical Meta-analysis*. London: Sage Publications, Volume 49.
29. Loewen, J.W., Rosser, P. i Katzman, J. (1988). *Gender Bias in SAT Items*. Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA).
30. Maccoby, E.E. i Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
31. Mau, W.C. i Lynn, R. (2001). Gender differences on the Scholastic Aptitude Test, the American College Test and College Grades. *Educational Psychology, 21*, 133-136.
32. Niemierko, B. (2001). *Chłodne oblicze egzaminu zewnętrznego*, „Edukacja" nr 3, 11-22.
33. Noftle, E.E. i Robins, R.W. (2007). Personality Predictors of Academic Outcomes: Big Five Correlates of GPA and SAT Scores. *Journal of Personality and Social Psychology, 93*(1), 116-130.
34. O'Connor, M.C. i Paunonen, S.V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences, 43*, 971-990.
35. OECD (2006). *Education at a glance: OECD Indicators*. Paris: OECD
36. OECD (2009). *Equally prepared for life? How 15-year-old boys and girls perform in school.* Dostępne: http://www.stemequitypipeline.org/_documents/OECD%20%282009%29%20gender%20diff%20similarities%20in%2015-year-old.pdf [dostęp: 23.05.2014].
37. Perkins, R., Kleiner, B., Roey, S., Brown, J. (2004). *The High School Transcript Study: A Decade of Change in Curricula and Achievement, 1990-2000.* Washington, DC: National Center for Education Statistics.
38. Pomerantz, E. M., Altermatt, E. R. i Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *Journal of Educational Psychology, 94*, 396-404.
39. Poropat, A.E. (2009). A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance. *Psychological Bulletin, 135*(2), 322-338.
40. Preckel, F., Goetz, T., Pekrun, R. i Kleine, M. (2008). Gender Differences in Gifted and Average-Ability Students: Comparing Girls' and Boys' Achievement, Self-Concept, Interest, and Motivation in Mathematics. *Gifted Child Quarterly, 52*(2): 146-159.

41. Ramist, L., Lewis, C. i McCamley-Jenkins, L. (1994). *Student group differences in predicting college grades: Sex, language, and ethnic group* (College Board Report Number 93-1; ETS RR-94-27). New York: College Board.
42. Rice, M.J. (2009). Effect Size in Psychiatric Evidence-Based Practice Care. *Journal of the American Psychiatric Nurses Association, 15*, 138-142.
43. Schmitt, D.P., Realo, A., Voracek, M. i Allik, J. (2008). Why Can't a Man Be More Like a Woman? Sex Differences in Big Five Personality Traits Across 55 Cultures. *Journal of Personality and Social Psychology, 94*(1), 168-182.
44. Silverman, I.W. (2003). Gender differences in delay of gratification: a meta-analysis. *Sex Roles, 49*(9-10), 451-463.
45. Snijders, T.A.B i Bosker, R.J. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. London: Sage Publications.
46. Spencer, S.J., Steele C.M. i Quinn, D.M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4-28.
47. Suchocka, A. (2011). *Przemoc symboliczna jako element ukrytego programu kształcenia polskiej szkoły.* Zeszyty Naukowe AMW, 4(187), 293-302.
48. Srivastava, S., John, O.P., Gosling, S.D., Potter, J. (2003). Development of Personality in Early and Middle Adulthood: Set Like Plaster or Persistent Change? *Journal of Personality and Social Psychology, 84*(5), 1041-1053.
49. Steele, C.M. (1997). A threat in the air: how stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613-629.
50. Stricker, L. J., Rock, D. A. i Burton, N. W. (1993). Sex differences in predictions of college grades from Scholastic Aptitude Test scores. *Journal of Educational Psychology, 85*(4), 710-718.
51. Thompson, S.G. i Higgins, J.P.T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21*, 1559-1573.
52. Willingham, W.W. i Cole, N.S. (1997). *Gender and fair assessment*. Mahwah NJ: Lawrence Erlbaum.
53. World Economic Forum (2013). *The Global Gender Gap Report (2013)*. Dostępne: http://www3.weforum.org/docs/WEF_GenderGap_Report_2013.pdf [dostęp: 26.05.2014].
54. van den Noortgate, W. i Onghena, P. (2003). Multilevel meta-analysis: a comparison with traditional meta-analytical procedures. *Educational and psychological measurement, 63*, 765-790.
55. Xie, Y., Shauman, K.A. (2003). *Women in Science: Career Processes and Outcomes*.
56. Cambridge, MA: Harvard University Press.
57. Young, J.W. i Kobrin, J.L. (2001). *Differential Validity, Differential Prediction, and College Admission Testing: A Comprehensive Review and Analysis*. College Entrance Examination Board, New York: College Board Research Report No.2001-6.
58. Zespół Analiz Osiągnięć Uczniów (ZAOU) Instytutu Badań Edukacyjnych (2013). *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 [dostęp: 12.05.2014].
59. Zill, N. i West, J. (2001). *Entering Kindergarten: A Portrait of American Children When They Begin School: Findings from the Condition of Education 2000*. Washington DC: National Center for Education Statistics.