

**Filip Kulon**

**Agata Gajewska-Dyszkiewicz**

**Katarzyna Paczuska**

Instytut Badań Edukacyjnych

## **Porównanie wyników Europejskiego Badania Kompetencji Językowych i egzaminu gimnazjalnego z języków obcych**

### **Wprowadzenie**

Od dekady w Europie wyznacznikiem standardów w edukacji językowej jest Europejski System Opisu Kształcenia Językowego (ESOKJ). Jest to obszerne opracowanie Rady Europy z 2001 roku dotyczące parametrów opisu kształcenia językowego, procesów uczenia się i nauczania języków, opisu poziomów biegłości językowej, jej testowania i oceny. Podobnie jak w wielu krajach europejskich, zmiany w kształceniu językowym w Polsce w ostatnich latach zmierzają ku pełniejszemu odniesieniu celów dydaktyki językowej oraz pomiaru jej efektów do zapisów ESOKJ.

Nawiązania do Europejskiego Systemu Opisu Kształcenia Językowego widoczne są już w poprzedniej podstawie programowej dla języków obcych nowożytnych z 2002 roku, która kładzie nacisk na rozwijanie promowanych przez ESOKJ umiejętności komunikacyjnych. Również obowiązujące standardy egzaminu maturalnego z 2003 roku odzwierciedlają wprowadzony przez ESOKJ podział działań językowych. Warto wspomnieć o niesfinalizowanych pracach Centralnej Komisji Egzaminacyjnej z 2004 roku nad pełnym odniesieniem egzaminu maturalnego do ESOKJ i wyrażaniem jego wyniku w skali biegłości językowej ESOKJ (Trzcińska, 2009). Na ESOKJ powołują się także autorzy standardów egzaminu gimnazjalnego obowiązujących w latach 2009-11 (MEN, 2007b).

Nowa podstawa programowa odwołuje się już bezpośrednio do modelu komunikacyjnej kompetencji językowej w ujęciu ESOKJ zarówno w treściach kształcenia, jak i języku ich opisu. Wraz z nią istotnym modyfikacjom uległ także egzamin gimnazjalny w formule obowiązującej od 2012 roku, a od 2015 roku zmieni się również egzamin maturalny. Opracowując egzaminy, CKE opiera się w dużej mierze na deskryptorach biegłości językowej Europejskiego Systemu Opisu Kształcenia Językowego (Smolik i Trzcińska, 2011).

W świetle prac podejmowanych w kierunku zwiększenia jakości nauczania języków w systemie oświaty m.in. poprzez wprowadzenie założeń ESOKJ jako ram określających cele i efekty edukacji językowej oraz zapewnienia większej zbieżności egzaminów zewnętrznych ze skalami biegłości językowej tego dokumentu Rady Europy, warto skorzystać z możliwości, jaką dają wyniki Europejskiego Badania Kompetencji Językowych (European Survey on Language Competences, ESLC) do skonfrontowania rezultatów wprowadzanych zmian.

ESLC to jak do tej pory jedyny, poza egzaminem gimnazjalnym, pomiar zewnętrzny dający informację o poziomie umiejętności językowych całej populacji polskich gimnazjalistów. Oparte w całości na Europejskim Systemie Opisu Kształcenia Językowego badanie daje zatem unikalną sposobność wypowiedzenia się w odniesieniu do jego kryteriów o osiągnięciach w zakresie nauki języków obcych polskich uczniów realizujących założenia poprzedniej podstawy programowej pośrednio nawiązującej już do kompetencji komunikacyjnej w ujęciu ESOKJ. Porównanie wyników egzaminu gimnazjalnego oraz testów ESLC jest także okazją do refleksji nad skutecznością pomiaru, jakim był egzamin gimnazjalny w formule obowiązującej do roku szkolnego 2010/11. Powtórzenie procesu zestawiania wyników nowego egzaminu gimnazjalnego w pełni nawiązującego do ESOKJ z badaniem ESLC po kolejnym cyklu badania w 2016 roku pozwoli wykazać celowość wprowadzenia obowiązującej od 2012 roku formuły egzaminu gimnazjalnego, a także określić kierunek zmian w umiejętnościach uczniów objętych w 2009 roku nową podstawą programową. Prezentowane w tym artykule analizy dotyczące rzetelności pomiaru ESLC 2011 i egzaminu gimnazjalnego z 2011 roku oraz korelacji pomiędzy nimi, stanowić mogą pierwszy krok i przyczynek do dalszych prac, np. we wspomnianych obszarach.

### **Charakterystyka badania ESLC i egzaminu gimnazjalnego**

Realizowane na zlecenie Komisji Europejskiej badanie ESLC ma na celu dostarczenie rzetelnych i porównywalnych danych o poziomie znajomości języków obcych uczniów szkół ponadpodstawowych w Europie. ESLC stosuje standardy losowania próby, przygotowania narzędzi pomiarowych, ich pilotażu i standaryzacji oraz zbierania i analizy danych analogiczne do innych międzynarodowych badań edukacyjnych, takich jak PISA, TIMSS czy PIRLS. Za opracowanie, kalibrację testów językowych oraz analizę ich wyników odpowiadają czołowe europejskie instytucje w dziedzinie psychometrii i pomiaru językowego, m.in. Cito (Centraal instituut voor toetsontwikkeling, Holandia), CIEP (Centre international d'études pédagogiques, Francja), University of Cambridge ESOL Examinations (Wielka Brytania) czy Goethe-Institut (Niemcy).

Założenia pomiarowe kompetencji językowych w ESLC oparte są na socjokognitywnym modelu uczenia się i posługiwania się językiem obcym zawartym w ESOKJ, a wyniki testów wyrażone w poziomach biegłości językowej skal ESOKJ (Komisja Europejska, 2012a, 2012b). Realizowane przez eksperckie instytucje rygorystyczne procedury odnoszenia zadań testowych do deskryptorów biegłości językowej na poziomie specyfikacji i opracowywania zadań testowych, po ich pilotażu oraz ponownie przed ostateczną analizą wyników (*standard setting*) mają zapewnić wysoką trafność pomiaru w odniesieniu do Europejskiego Systemu Opisu Kształcenia Językowego.

W pierwszym cyklu ESLC w 2011 roku badanie objęło pomiar kompetencji językowych uczniów z 14 krajów europejskich w zakresie dwóch najczęściej nauczanych w danym kraju języków obcych spośród angielskiego, francuskiego, hiszpańskiego, niemieckiego i włoskiego. W Polsce był to język angielski i niemiecki. Testy podzielone były na trzy części: rozumienie ze słuchu,

rozumienie tekstu pisanego oraz wypowiedź pisemna. Docelową populacją badania byli uczniowie ostatniego roku szkolnictwa ponadpodstawowego (ISCED 2). W Polsce badaną populacją byli zatem uczniowie trzeciej klasy gimnazjum. Z populacji docelowej wyłączeni byli uczniowie szkół specjalnych oraz szkół dla dorosłych, uczniowie niepełnosprawni fizycznie lub umysłowo (w stopniu uniemożliwiającym uczestnictwo w badaniu), uczniowie nieposługujący się językiem narodowym w stopniu umożliwiającym zrozumienie pytań kwestionariuszowych oraz uczniowie o orzeczonej dysleksji. Dyslektyków dopuszczano jednak do udziału w badaniu, więc ostatecznie znaleźli się oni w badanej populacji. W badaniu losowano oddzielną, reprezentatywną próbę uczniów dla każdego z dwóch języków. Zastosowano dwuetapowy dobór próby: na poziomie szkół (minimum 71 szkół na język) oraz na poziomie uczniów (średnio 25 na szkołę). Łącznie w badaniu wzięło udział 3324 gimnazjalistów (angielski: 1764, niemiecki: 1560), co stanowiło 85% próby.

Pomiar kompetencji z zakresu języka obcego nowożytnego wprowadzono jako trzecią część egzaminu gimnazjalnego w roku 2009. W niezmienionej formule obowiązywał on do 2012 roku. Zakres testu wyznaczały standardy wymagań egzaminacyjnych i podstawa programowa kształcenia ogólnego. Test językowy, tak jak pozostałe części egzaminu, był w formie pisemnej, obejmował odbiór tekstu słuchanego, odbiór tekstu czytanego oraz reagowanie językowe (umiejętność wyboru i produkcji odpowiedniej i skutecznej reakcji językowej w danej sytuacji komunikacyjnej). Egzamin sprawdzał umiejętności gimnazjalistów w jednym z sześciu języków: angielskim, francuskim, hiszpańskim, niemieckim, rosyjskim lub włoskim. W 2011 roku egzamin na arkuszu standardowym z języka angielskiego zdawało w sumie 331 686 uczniów, a z języka niemieckiego 71 405 (CKE, 2011).

Choć testów językowych ESLC i egzaminu gimnazjalnego nie sposób uznać za tożsame, istnieje szereg podobieństw, które czynią porównanie ich zasadnym. Tabela 1. przedstawia zestawienie charakterystyk obu testów w aspektach istotnych dla prezentowanych poniżej analiz. Należy pamiętać, że badanie ESLC dotyczyło uczniów objętych „starą” podstawą programową i przystępujących do egzaminu gimnazjalnego w nieobowiązującej już formule.

**Tabela 1. Ogólna charakterystyka egzaminu gimnazjalnego z języka obcego oraz badania ESLC**

	Egzamin gimnazjalny 2011	ESLC 2011
Populacja	Uczniowie klas III gimnazjum uczący się języka, w zakresie którego przystępują do egzaminu jako przedmiotu obowiązkowego	Uczniowie klas III gimnazjum uczący się badanego języka obcego co najmniej rok przed testem
Próba	Cała populacja	Próba reprezentatywna (n=3324)
Co jest badane	Rozumienie ze słuchu, rozumienie tekstu pisanego, reagowanie językowe	Rozumienie ze słuchu, rozumienie tekstu pisanego, wypowiedź pisemna
Model	Nieokreślony; przez podstawę programową i standardy egzaminacyjne nawiązujący pośrednio do ESOKJ	Konstrukt pisania, czytania i rozumienia ze słuchu oparty na modelu posługiwania językiem się i nauki języka w ujęciu ESOKJ

	Egzamin gimnazjalny 2011	ESLC 2011
Zestawienie badanych mikroumiejętności w Załączniku 1		
Konstrukcja testu	Każde zadanie sprawdza umiejętności lub wiedzę określone w konkretnym standardzie, cały egzamin pokrywa wszystkie umiejętności i wszystkie standardy	Każdy uczeń przystępuje do dwóch z trzech badanych umiejętności, poszczególne arkusze testowe pokrywają tylko część konstruktów danej umiejętności
Wynik	Raportowany dla każdego ucznia, wyrażony w procentach maksymalnej liczby punktów	Raportowany dla populacji, wyrażony przy użyciu <i>plausible values</i> i ważony (na podstawie skalowania IRT)
Stawka	Państwowy egzamin obowiązkowy dla wszystkich gimnazjalistów	Dobrowolne uczestnictwo, anonimowy, brak informacji zwrotnej o wynikach testu na poziomie ucznia
Czas przeznaczony na rozwiązywanie zadań	90 minut	60-75 minut
Termin pomiaru	Kwiecień 2011	Marzec 2011

### Łączenie danych

W danych egzaminacyjnych z 2011 roku udało się odnaleźć wyniki 3296 uczniów biorących udział w badaniu ESLC, co stanowi 99% wszystkich gimnazjalistów, którzy ostatecznie wzięli udział w badaniu (nie odnaleziono danych dla zaledwie 28 osób). Na egzaminie uczniowie mieli do wyboru kilka języków obcych - oprócz angielskiego i niemieckiego były to: francuski, hiszpański, rosyjski i włoski. Część uczniów przystępowała do testu na egzaminie gimnazjalnym z innego języka niż w badaniu ESLC - 62 osoby zdawały na egzaminie niemiecki, a w ESLC angielski, 518 gimnazjalistów na egzaminie rozwiązywało zadania z angielskiego, a w ESLC z niemieckiego, ponadto 5 osób zdawało francuski, a 2 uczniów rosyjski. Ostatecznie otrzymano połączone wyniki 2709 uczniów przystępujących do testu ESLC i egzaminu gimnazjalnego z tego samego języka, w tym 1683 z 1764 (95%) dla języka angielskiego i 1026 z 1560 (66%) dla niemieckiego.

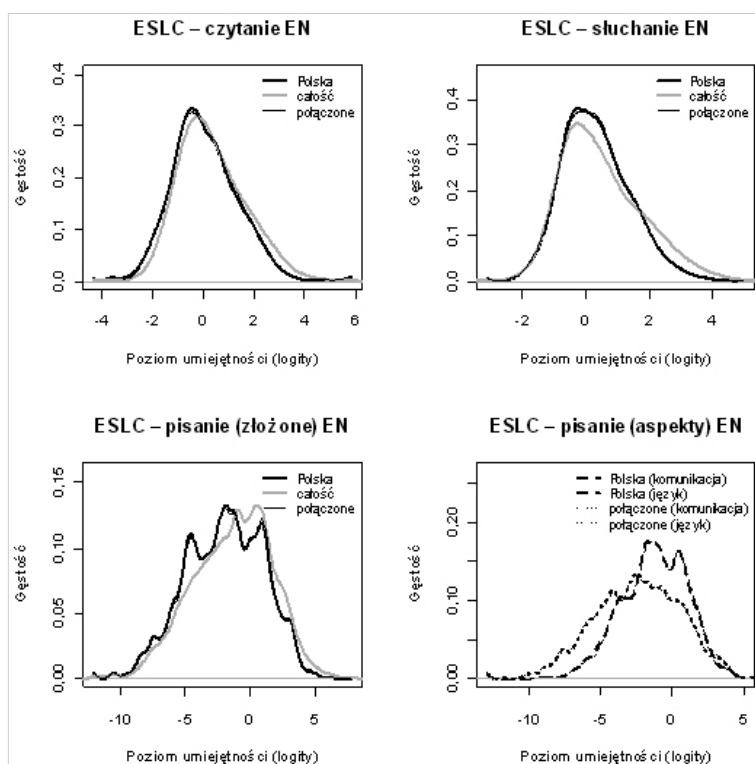
Powyższe podsumowanie wyraźnie pokazuje, że dla języka niemieckiego udało się połączyć zdecydowanie mniej przypadków niż dla języka angielskiego. Niestety na 71 szkół, w których uczniowie w badaniu ESLC przystępowali do testów z języka niemieckiego, wyniki z 13 szkół w ogóle nie zostały połączone. Świadczy to o tym, że opisywane zjawisko najprawdopodobniej nie ma charakteru losowego, co z kolei może ograniczać możliwość uogólnienia wyników porównania na populację. Istotne jest jednak to, czy rozkład wyników uczniów, których wyniki połączono, odbiega od rozkładu wyników dla wszystkich zdających egzamin gimnazjalny. W dalszej części opracowania przedstawiono analizę wyników obydwu testów, z której można wnioskować, że rozkłady umiejętności uczniów, których wyniki połączono, nie odbiegają w dużym stopniu od rozkładu umiejętności dla wszystkich uczniów, zatem niekompletne połączenie danych dla języka niemieckiego nie powinno być przeszkodą do uogólniania wyników na populację.

Warto w tym miejscu poświęcić jeszcze uwagę na sygnalizowaną wcześniej kwestię uczniów z orzeczoną dysleksją. Jak wspomniano w opisie ESLC, w badaniu dopuszczano ich udział. W przypadku egzaminu gimnazjalnego arkusz standardowy przeznaczony jest dla uczniów bez dysfunkcji oraz dla uczniów ze specyficznymi trudnościami w uczeniu się (np. dysleksją). Możemy uznać, że obydwie populacje mają w gruncie rzeczy takie same charakterystyki. Z danych egzaminacyjnych wynika, że dyslektycy stanowią około 7% populacji dla zdających język niemiecki na egzaminie i około 9% dla zdających angielski (arkusz standardowy). Wśród uczniów, którzy wzięli udział w badaniu ESLC odsetek dyslektyków jest dwukrotnie niższy i wynosi po zważeniu około 3% dla języka niemieckiego i około 4% dla angielskiego (mowa tu o wszystkich uczniach, którzy wzięli udział w badaniu, dla których udało się uzyskać dane z egzaminu niezależnie od języka zdawanego na egzaminie). Wyraźnie zatem widać, że ostatecznie część dyslektyków wzięła udział w badaniu. Dla danych, których ostatecznie użyto w analizie, czyli połączonych tak, aby język obu testów był ten sam, odsetek dyslektyków dla angielskiego spada do 3%. Mniejszy udział uczniów z orzeczoną dysleksją w badaniu (i połączonych danych) nie stanowi jednak wystarczającego powodu, aby wykluczać ich z analizy. Stracilibyśmy w ten sposób możliwość wypowiedzania się o całości wyników uzyskanych za pomocą arkusza standardowego z egzaminu gimnazjalnego.

### **Analiza wyników testów**

Rysunki 1-4 ilustrują rozkład gęstości prawdopodobieństwa dla wyników z obu omawianych testów dla obydwu języków. Poszczególne umiejętności przedstawiono osobno, w przypadku egzaminu zaprezentowano również wynik sumaryczny, a dla ESLC zaprezentowano obydwa aspekty pisania (komunikacja i język) na jednym wykresie.

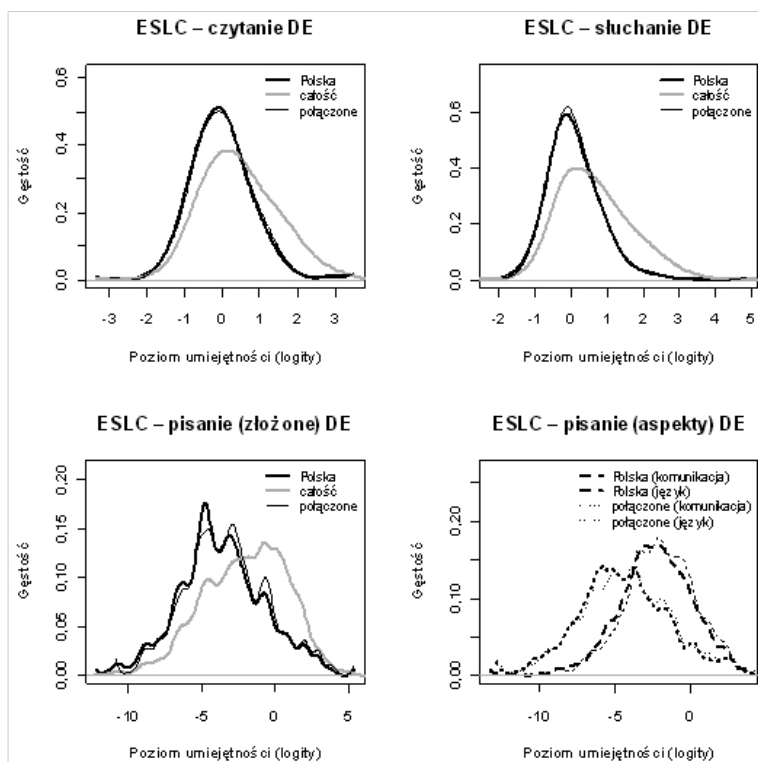
W przypadku ESLC na wykresach widoczne są trzy krzywe - dla wyników wszystkich krajów, w których testowano dany język, dla Polski oraz wyniki uczniów, dla których połączono dane z egzaminu gimnazjalnego (wszystkie zostały zważone za pomocą wag dostarczonych wraz ze zbiorem danych ESLC). Wyjątek stanowią wykresy, gdzie zestawiono obydwa aspekty pisania (komunikację i język) - dla czytelności nie umieszczano na nich wyników dla wszystkich krajów. W ocenie wypowiedzi pisemnej zastosowano dwa kryteria: komunikacji określonych treści i poprawności językowej. Dla każdego kryterium osobno obliczono wyniki, a także stworzono ogólny wskaźnik umiejętności pisania (zwany dalej złożonym).



Rysunek 1. Wykresy gęstości prawdopodobieństwa poszczególnych umiejętności ESLC z języka angielskiego (EN) dla wszystkich krajów, Polski oraz wyników połączonych

Wyniki dla wszystkich krajów, w których badano dany język, umieszczone zostały w celu odniesienia wyników Polski do ogólnych wyników danego testu językowego. Najbardziej interesujące jest porównanie wyników całej populacji uczniów gimnazjów z wynikami próby uczniów z połączonymi wynikami (uogólnionymi na populację). Jak widać na rysunkach 1. i 2. rozkład zważonych wyników uczniów, których wyniki połączono z egzaminem, nie odbiega znacząco od rozkładu dla całej Polski.

Oprócz wizualnego porównania rozkładów wyników całej populacji gimnazjalistów i wyników połączonych warto przyrzeć się parametrom tych rozkładów: średniej i odchyleniu standardowemu. W tabeli 2. umieszczono zestawienie tych parametrów dla ESLC. Wyraźnie widać, że nie różnią się one znacząco od siebie nawet dla języka niemieckiego, gdzie połączonych zostało mniej danych.

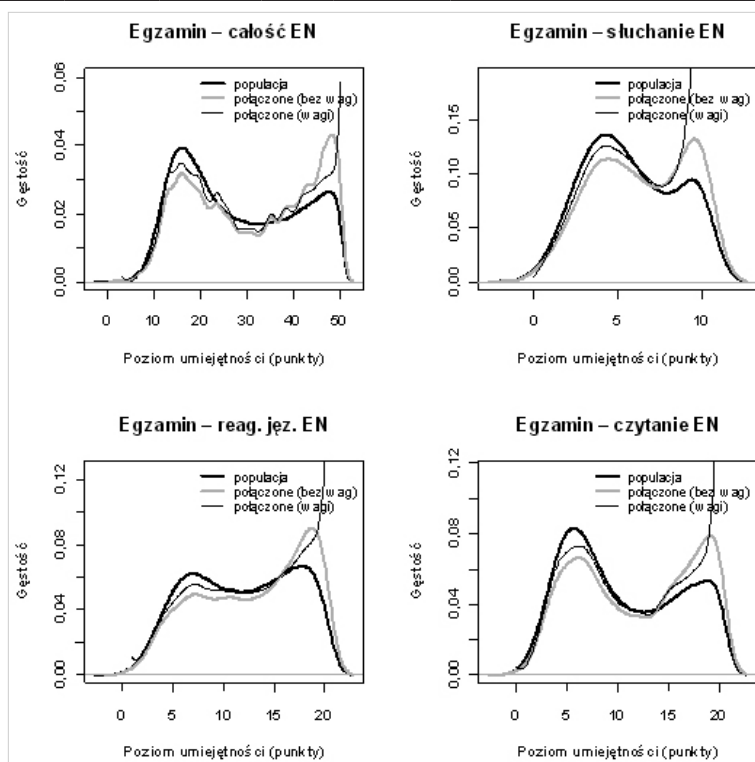


Rysunek 2. Wykresy gęstości prawdopodobieństwa poszczególnych umiejętności ESCL z języka niemieckiego (DE) dla wszystkich krajów, Polski oraz wyników połączonych

Tabela 2. Porównanie średnich i odchyłeń standardowych poszczególnych umiejętności z badania ESCL dla populacji i połączonych danych (ważonych) wraz z błędami standardowymi

Umiejętność	Język	Populacja	Średnia	Odchylenie standardowe	Błąd standardowy średniej	Błąd standardowy odchylenia standardowego
Czytanie	język angielski	populacja	0,08	1,26	0,10	0,07
		wyniki połączone	0,08	1,25	0,10	0,07
	język niemiecki	populacja	-0,01	0,83	0,04	0,04
		wyniki połączone	0,05	0,84	0,05	0,06
Słuchanie	język angielski	populacja	0,30	1,04	0,08	0,06
		wyniki połączone	0,34	1,05	0,07	0,06
	język niemiecki	populacja	0,08	0,76	0,05	0,04
		wyniki połączone	0,12	0,70	0,04	0,04

Umiejętność	Język	Populacja	Średnia	Odchylenie standardowe	Błąd standardowy średniej	Błąd standardowy odchylenia standardowego
Pisanie (komunikacja)	język angielski	populacja	-1,18	2,15	0,19	0,16
		wyniki połączone	-1,20	2,21	0,18	0,13
	język niemiecki	populacja	-2,29	2,41	0,20	0,23
		wyniki połączone	-1,96	2,32	0,19	0,15
Pisanie (język)	język angielski	populacja	-2,29	2,98	0,25	0,29
		wyniki połączone	-2,36	2,98	0,25	0,26
	język niemiecki	populacja	-4,35	3,30	0,34	0,24
		wyniki połączone	-4,11	3,46	0,34	0,25
Pisanie (złożone)	język angielski	populacja	-1,88	2,87	0,24	0,22
		wyniki połączone	-1,92	2,91	0,24	0,19
	język niemiecki	populacja	-3,66	2,92	0,28	0,19
		wyniki połączone	-3,33	2,98	0,28	0,19

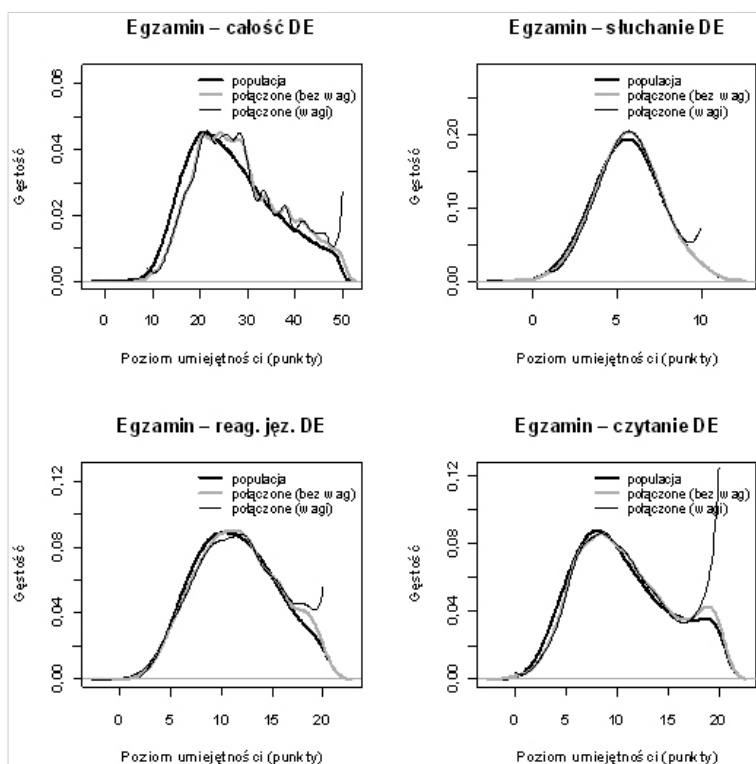


Rysunek 3. Wykresy gęstości prawdopodobieństwa wyniku sumarycznego i poszczególnych umiejętności egzaminu gimnazjalnego z języka angielskiego (EN) dla populacji oraz nieważonych i ważonych wyników połączonych



Dla wyników z egzaminu gimnazjalnego na każdym z wykresów również przedstawiono trzy krzywe - dla całości wyników, czyli całej populacji gimnazjalistów zdających dany język oraz nieważone i ważne wyniki z połączonych danych. Na wykresach można zauważyć, że uczniowie, których wyniki połączono, a więc biorący udział w badaniu ESLC, otrzymali nieco lepsze wyniki niż cała populacja. Być może jest to związane z tym, że uczniowie o niższym poziomie umiejętności rezygnowali z udziału w badaniu, aby nie obciążać się dodatkowym testowaniem. W przypadku języka angielskiego zważenie wyników niweluje nieco te rozbieżności i przybliża rozkład ważonych wyników do rozkładu populacji. Jeśli chodzi o język niemiecki, to zważenie nie „poprawia” zbyt rozkładu wyników, ale z drugiej strony, rozkłady dla populacji i połączonych danych są do siebie bardziej zbliżone niż dla języka angielskiego.

Ponownie warto zwrócić uwagę na średnie i odchylenia standardowe dla poszczególnych umiejętności i wyniku sumarycznego oraz porównać wyniki populacji z ważonymi wynikami uczniów, dla których połączono dane. Tabela 3. zawiera analogiczne zestawienie jak w przypadku tabeli 2., lecz dla danych z egzaminu gimnazjalnego.



Rysunek 4. Wykresy gęstości prawdopodobieństwa wyniku sumarycznego i poszczególnych umiejętności egzaminu gimnazjalnego z języka niemieckiego (DE) dla populacji oraz nieważonych i ważonych wyników połączonych

Zebrane wartości potwierdzają oczywiście to, co na pierwszy rzut było widoczne na wykresach - wyniki uczniów z próby są nieco wyższe niż wyniki całej populacji. Uwzględniając błędy standardowe, obserwowana różnica średnich wyniku sumarycznego wynosi dla języka niemieckiego zaledwie 1 punkt, a dla języka angielskiego średnie wszystkich wyników z egzaminu mieszczą się w przedziałach ufności dla średnich z ważonych wyników uczniów, których dane połączono. Można domniemywać, że różnica dla języka niemieckiego wynika z niekompletnego połączenia danych (wyniki uczniów, których danych nie połączono, z czytania i pisania ESLC są nieco niższe niż tych, których dane połączono - jest bardzo możliwe, że na egzaminie z języka niemieckiego zależność ta również miałyby miejsce).

**Tabela 3. Porównanie średnich i odchyłeń standardowych wyniku sumarycznego i poszczególnych umiejętności z egzaminu gimnazjalnego dla populacji i połączonych danych (ważonych) wraz z błędami standardowymi**

Umiejętność	Język	Populacja	Średnia	Odchylenie standardowe	Błąd standardowy średniej	Błąd standardowy odchylenia standardowego
Wynik sumaryczny	język angielski	populacja	28,28	12,50	-	-
		wyniki połączone	29,64	12,73	0,85	0,24
	język niemiecki	populacja	26,75	9,19	-	-
		wyniki połączone	28,68	9,31	0,53	0,25
Słuchanie	język angielski	populacja	5,81	2,68	-	-
		wyniki połączone	6,08	2,73	0,17	0,05
	język niemiecki	populacja	5,65	1,84	-	-
		wyniki połączone	5,80	1,83	0,09	0,06
Reagowanie językowe	język angielski	populacja	11,89	5,13	-	-
		wyniki połączone	12,33	5,23	0,34	0,09
	język niemiecki	populacja	11,44	4,01	-	-
		wyniki połączone	11,83	4,12	0,24	0,09
Czytanie	język angielski	populacja	10,58	5,60	-	-
		wyniki połączone	11,23	5,68	0,35	0,10
	język niemiecki	populacja	10,54	4,67	-	-
		wyniki połączone	11,05	4,72	0,25	0,10

### Trafność i rzetelność testów

Bez wątplenia dwoma bardzo istotnymi parametrami wszelkich testów, w tym testów osiągnięć, jest ich rzetelność i trafność. Raporty Centralnej Komisji Egzaminacyjnej publikowane corocznie po przeprowadzeniu egzaminów zawierają oszacowanie rzetelności użytych arkuszy egzaminacyjnych w postaci współczynnika alfa Cronbacha. Trafność jest zagadnieniem dużo bardziej złożonym niż rzetelność i nie jest możliwe jej potwierdzenie tylko za pomocą

procedur statystycznych. Weryfikacja trafności teoretycznej arkusza nie jest zadaniem łatwym i wymaga systematycznych prac związanych głównie z badaniem jakości zadań i struktury arkusza. Trafność treściową egzaminu możemy uznać za spełnioną na podstawie dokumentów omawianych we wstępie. Jest ona zagwarantowana poprzez procedury tworzenia narzędzi zgodnych z odpowiednimi rozporządzeniami zawierającymi standardy wymagań egzaminacyjnych, weryfikowane przez Centralną Komisję Egzaminacyjną. Potwierdzenia trafności kryterialnej egzaminu gimnazjalnego z języka obcego w zasadzie na dzień dzisiejszy nie mamy i właśnie badanie ESLC daje nam szansę na przeprowadzenie analizy w tej materii.

Podstawowym sposobem sprawdzenia trafności kryterialnej testu jest skorelowanie jego wyników z wynikami innego testu mierzącego ten sam konstrukt. Jeśli przyjmiemy, że badanie ESLC i egzamin gimnazjalny z języka obcego mierzą te same umiejętności, to możliwe jest takie porównanie. Opisane na wstępie relacje omawianych testów do ESOKJ pozwalają na przyjęcie takiego założenia. Kolejnym warunkiem porównywania obydwu testów jest ich w miarę duża rzetelność - nie ma sensu opierać analiz i wniosków na nierzetelnych pomiarach.

**Tabela 4. Współczynniki alfa Cronbacha dla egzaminu gimnazjalnego**

		Cały egzamin	Słuchanie	Reagowanie językowe	Czytanie
W rozbiciu na działania	język angielski	0,95	0,76	0,88	0,90
	język niemiecki	0,89	0,44	0,77	0,83
	liczba pozycji	50	10	20	20
Dla zadań	język angielski	0,93	0,69	0,85	0,86
	język niemiecki	0,85	0,26	0,71	0,78
	liczba pozycji	13	3	5	5

W tabeli 4. umieszczono zestawienia współczynników alfa Cronbacha dla testów gimnazjalnych z obydwu języków. Rzetelność egzaminu z języka niemieckiego jest niższa niż egzaminu z języka angielskiego, niezależnie od tego, czy bierzemy pod uwagę współczynnik alfa liczony dla skali rozbitej na poszczególne działania, czy też składającej się z punktacji za zadania (po zsumowaniu odpowiednich działań). Ponieważ współczynnik alfa zależy od liczby pozycji skali, porównywanie poszczególnych umiejętności między sobą jest utrudnione (słuchanie ma mniej pozycji niż pozostałe umiejętności). Również z tego powodu rzetelność poszczególnych umiejętności jest niższa niż całego egzaminu. Można jednak porównać współczynniki pomiędzy językami. O ile czytanie i reagowanie językowe wykazują podobne różnice co całość egzaminu, o tyle rzetelność słuchania z języka niemieckiego jest zdecydowanie niższa niż z języka angielskiego. Różnicę tę bardzo wyraźnie widać w przypadku współczynników obliczanych na zsumowanych działaniach. Porównywanie poszczególnych umiejętności pomiędzy ESLC i egzaminem gimnazjalnym staje się zatem utrudnione ze względu na niską rzetelność słuchania. Jedynie czytanie jest umiejętnością, którą można bez przeszkód porównywać pomiędzy obydwojema testami.

Podsumowując, rzetelność całego egzaminu z języka angielskiego jest bardzo dobra, a z języka niemieckiego dobra. Zarówno czytanie, jak i reagowanie językowe dla obydwu języków oraz słuchanie dla języka angielskiego mają akceptowalną lub dobrą rzetelność. Rzetelność słuchania dla języka niemieckiego jest niższa niż 0,5 i jest poniżej akceptowalnego poziomu, zatem porównania tej skali należy dokonywać z dużą ostrożnością.

Jesteśmy zmuszeni do uczynienia założenia, że trafność badania ESLC jest wysoka (czy też akceptowalna dla naszych potrzeb) bez potwierdzenia w postaci konkretnych wartości jakichkolwiek współczynników. Raporty ESLC nie zawierają bowiem żadnego „twardego” dowodu na to, że deklaracje o wysokiej trafności i rzetelności zostały spełnione. W pewnym sensie czynimy takie założenia na podstawie tego, że było to międzynarodowe badanie przygotowywane przez eksperckie instytucje z różnych krajów, co jest niezbyt mocnym i niemerytorycznym argumentem. Potwierdzenia trafności omawianego badania można by dokonać poprzez porównanie jego wyników z wynikami innych testów badających te same umiejętności. Wcześniej jednak przyjęliśmy, że to właśnie na podstawie ESLC będziemy weryfikować trafność egzaminu.

Istotny z punktu widzenia dalszych analiz jest fakt, że wszystkie obliczenia na zbiorach ESLC oraz na połączonych danych były wykonywane z użyciem dostarczonych wraz ze zbiorem danych ESLC wag probabilistycznych i replikacyjnych. Użycie wag replikacyjnych ma na celu oszacowanie nieobarczonych błędów standardowych estymatorów. W przypadku użycia *plausible values* (PV), z którymi mamy do czynienia w badaniu ESLC<sup>1</sup>, estymator zmiennej ukrytej reprezentowanej przez zestaw PV obliczany jest jako średnia estymatorów obliczonych dla każdej PV, a w obliczaniu błędu standardowego uwzględniana jest wariancja estymatorów z poszczególnych PV (OECD, 2009). Takie postępowanie jest uprawnione w przypadku, gdy rozkład estymatora jest normalny. W naszym przypadku mamy do czynienia z korelacją, a jako że korelacja nie ma charakteru normalnego, powinniśmy najpierw poddać współczynniki transformacji Fishera, obliczyć średnią i ponownie przekształcić tak otrzymaną wartość (ma to również znaczenie przy obliczaniu błędów standardowych). Jednak wariancja obliczanych współczynników korelacji jest na tyle mała, że różnice pomiędzy średnią „surowych” współczynników a średnią obliczoną na podstawie ich wartości standaryzowanych nie wpływają na interpretację. Z tego powodu, mając na celu uproszczenie obliczeń, zdecydowano się na pominięcie etapu transformowania współczynników korelacji do ich formy standaryzowanej. W celu uniknięcia ewentualnego niedoszacowania błędów standardowych współczynników korelacji ilekroć sprawdzano 95% przedziały ufności dla współczynników, były one obliczane jako podwojony błąd standardowy (zamiast współczynnika 1,96) - w związku z tym podawane przedziały w niektórych przypadkach mogą być szersze niż w rzeczywistości.

<sup>1</sup> Celem badania ESLC jest oszacowanie umiejętności językowych całej badanej populacji i nie przewidywano dostarczania informacji o wynikach poszczególnym uczniom. W związku z tym każdemu uczniowi jako jego wyniki przypisano zestaw pięciu *plausible values* dla każdej umiejętności. Zagadnienie *plausible values* opisuje w swoim artykule Wu (Wu, 2005).

W przypadku rzetelności ESLC możemy pośrednio oszacować rzetelność skal przygotowanych w procesie badawczym. Jako jedną z miar rzetelności w wypadku modeli IRT podaje się tzw. rzetelność EAP/PV (Adams, 2005). Jest to stosunek wariancji EAP (*expected a posteriori*) do wariancji PV, gdzie wariancja EAP jest wariancją oszacowania punktowego, a wariancja PV jest całkowitą wariancją zmiennej ukrytej. Niestety w bazach danych z badania ESLC nie zawarto oszacowania EAP (dla każdej umiejętności mamy tylko zestaw pięciu PV), więc nie jest możliwe obliczenia rzetelności EAP/PV. Analogicznym podejściem do opisanej wyżej rzetelności EAP/PV jest obliczenie korelacji pomiędzy PV - zamiast wariancji EAP mamy tutaj do czynienia z kowariancją poszczególnych PV. Jednym z klasycznych sposobów oszacowania rzetelności jest obliczenie korelacji pomiędzy dwoma pomiarami tym samym narzędziem przeprowadzonymi na tych samych badanych obiektach (test-retest). Możemy w naszym wypadku wysnuć również analogię do tego podejścia. Jako kolejne pomiary możemy potraktować kolejne *plausible values* dla danej umiejętności. Losowanie kolejnych wartości z rozkładu *a posteriori* daje nam oszacowania cechy ukrytej - jeśli korelacja pomiędzy tymi oszacowaniami jest duża, możemy uznać, iż pomiar cechy ukrytej był rzetelny.

W tabeli 5. przedstawiono średnie współczynników korelacji pomiędzy poszczególnymi PV dla wszystkich umiejętności. Dla danych ze wszystkich krajów wszystkie współczynniki mają wartości powyżej 0,8, co wskazuje na dużą zgodność pomiędzy poszczególnymi oszacowaniami. Jeśli ograniczymy dane tylko do Polski, to możemy zauważyć, że współczynniki są nieco niższe, głównie dla słuchania i czytania dla języka niemieckiego, choć dla angielskiego również widoczny jest niewielki spadek. W połączeniu z niższą rzetelnością testu gimnazjalnego z języka niemieckiego może to sugerować pewne problemy w oszacowaniu umiejętności językowych polskich gimnazjalistów, dotyczących języka niemieckiego. Nie powinno to jednak powodować zagrożenia dla poprawności przeprowadzanych analiz - nadal wszystkie współczynniki mają wartości powyżej 0,7. Rzetelność pomiaru żadnej z umiejętności mierzonej w badaniu ESLC nie odbiega znacząco od pozostałych. Błędy standardowe przedstawionych korelacji były minimalne, wszystkie wynosiły poniżej 0,01.

**Tabela 5. Współczynniki korelacji Pearsona pomiędzy *plausible values* dla poszczególnych umiejętności z badania ESLC dla wszystkich krajów i dla Polski**

		Słuchanie	Czytanie	Pisanie (komunikacja)	Pisanie (język)	Pisanie (złożone)
Wszystkie kraje	język angielski	0,90	0,88	0,87	0,89	0,93
	język niemiecki	0,85	0,83	0,86	0,87	0,92
Polska	język angielski	0,86	0,87	0,87	0,87	0,93
	język niemiecki	0,77	0,74	0,83	0,85	0,90

Dla porównania w tabeli 6. zostały podane korelacje pomiędzy *plausible values* obliczonymi dla egzaminu gimnazjalnego. Łatwo zauważyć, że ich wartości są zbliżone do współczynników alfa Cronbacha zaprezentowanych w tabeli 4., a także są one na podobnym poziomie co wartości dla ESLC. Możemy zatem uznać, że w przypadku omawianych testów mamy do czynienia z dobrą rzetelnością.

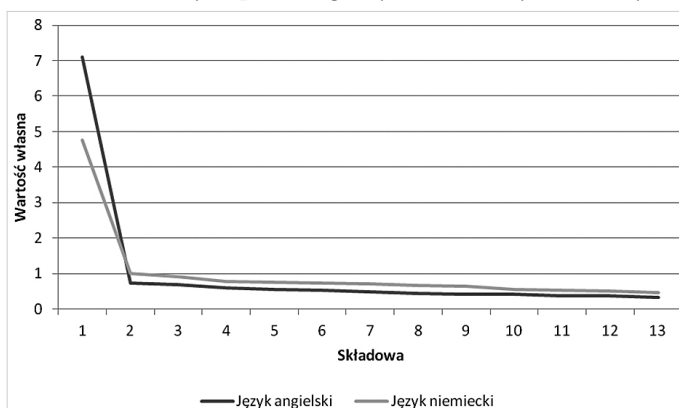
Podczas skalowania wyników z egzaminu z języka niemieckiego zsumowaniu uległy działania zadania 3, a działanie 1\_2 zostało usunięte ze względu na ujemną dyskryminację. Obydwa zadania wchodziły w skład słuchania. W przypadku skalowania tej umiejętności dodatkowo usunięte zostało działanie 1\_3. W przypadku języka angielskiego nie było potrzeby usuwania żadnych zadań. Dla każdego języka wygenerowano po pięć PV dla każdej umiejętności oraz wyniku sumarycznego. Spojrzenie na średnią korelacji pomiędzy *plausible values* dla całego egzaminu i poszczególnych umiejętności zawarte w tabeli 6. nie prowadzi do innych wniosków niż wysnute powyżej na podstawie współczynnika alfa Cronbacha. Słuchanie znów odstaje od pozostałych umiejętności, a w przypadku języka niemieckiego średnia współczynników korelacji Pearsona dla pięciu PV wynosi mniej niż 0,5 i potwierdza niską rzetelność tej podskali.

Tabela 6. Współczynniki korelacji Pearsona pomiędzy *plausible values* dla wyniku sumarycznego i poszczególnych umiejętności z egzaminu gimnazjalnego

	Cały egzamin	Słuchanie	Reagowanie językowe	Czytanie
Język angielski	0,94	0,77	0,87	0,88
Język niemiecki	0,92	0,49	0,84	0,85

### Jednowymiarowość egzaminu gimnazjalnego

Pomimo wyróżnienia w egzaminie gimnazjalnym z omawianych języków obcych trzech umiejętności analiza czynnikowa metodą głównych składowych nie wskazuje na wielowymiarowość analizowanych danych. Przy analizowaniu zadań (zsumowanych działań) i przyjęciu jako kryterium wyodrębnienia składowych kryterium Kaisera, tylko w przypadku języka niemieckiego otrzymamy dwie składowe. Nie poddają się one jednak interpretacji w odniesieniu do założonej teoretycznej struktury analizowanego arkusza. Gdyby jako kryterium przyjąć procent wariancji wyjaśnianej przez poszczególne składowe, to dla obydwu języków należałoby wyodrębnić tylko jedną składową. Rysunek 5. przedstawia wykres wartości własnych poszczególnych składowych dla obydwu języków.



Rysunek 5. Wykres wartości własnych z analizy głównych składowych dla języka angielskiego i niemieckiego

Analiza przeprowadzona na poszczególnych działaniach nie prowadzi do odmiennych wniosków. W przypadku kryterium Kaisera otrzymalibyśmy odpowiednio sześć składowych dla języka angielskiego i jednaście dla niemieckiego, a otrzymane składowe również nie dają się zinterpretować w odniesieniu do teoretycznej struktury arkuszy. Kryterium wyjaśnianej wariancji pokrywa się z wersją analizy przeprowadzonej na zadaniach.

Analiza czynnikowa przeprowadzana była bez usuwania poszczególnych zmiennych, ponieważ miała na celu zweryfikowanie teoretycznej, założonej przez twórców arkusze egzaminacyjne, struktury. Narzucanie rozwiązań zawierających trzy składowe, a więc tyle, ile umiejętności wyodrębnionych w egzaminie, również nie pozwalało na interpretację. We wszystkich przypadkach testowano także rozwiązania rotowane, lecz zarówno rotacje ortogonalne, jak i nieortogonalne nie wpływały na interpretację otrzymanych wyników.

### Zależność pomiędzy wynikami ESLC i egzaminu gimnazjalnego

Po zweryfikowaniu rzetelności testu ESLC i egzaminu gimnazjalnego oraz potwierdzeniu na poziomie teoretycznym możliwości porównywania ich wyników możemy obliczyć, jak korelują między sobą poszczególne umiejętności obydwu testów oraz wynik sumaryczny z egzaminu gimnazjalnego. Porównywanie między sobą umiejętności jest ograniczone zarówno ze względów teoretycznych opisanych na wstępie, jak też z powodu dość niskiej rzetelności słuchania z egzaminu (zwłaszcza z języka niemieckiego). Nie bez wpływu pozostaje też fakt, że w analizie czynnikowej nie potwierdzono istnienia trzech składowych dla danych z egzaminu. W przypadku wyników ESLC nie mamy do czynienia z sumarycznym wynikiem, a jedynie z wynikami w rozbiciu na umiejętności. Nie możemy zatem porównać „ogólnych” wyników obydwu testów językowych. O ile zestawienie ze sobą słuchania i czytania z obydwu testów wydaje się zasadne, o tyle porównywanie reagowania językowego z pisaniami może nie być uzasadnione. Kluczową informacją będą zatem współczynniki korelacji pomiędzy poszczególnymi umiejętnościami mierzonymi w badaniu i wynikiem sumarycznym z egzaminu. W tabeli 7. zamieszczono korelacje pomiędzy PV dla ESLC i surowymi wynikami egzaminu.

**Tabela 7. Współczynniki korelacji Pearsona pomiędzy *plausible values* dla umiejętności z badania ESLC oraz surowymi wynikami sumarycznymi i dla poszczególnych umiejętności z egzaminu gimnazjalnego**

		Słuchanie	Czytanie	Pisanie (komunikacja)	Pisanie (język)	Pisanie (złożone)
Język angielski	słuchanie	0,63	0,67	0,63	0,63	0,66
	czytanie	0,68	0,72	0,67	0,68	0,71
	reagowanie językowe	0,68	0,72	0,69	0,70	0,73
	egzamin	0,71	0,76	0,72	0,72	0,76
Język niemiecki	słuchanie	0,42	0,42	0,36	0,37	0,39
	czytanie	0,52	0,57	0,64	0,62	0,68
	reagowanie językowe	0,52	0,57	0,63	0,62	0,67
	egzamin	0,58	0,62	0,67	0,66	0,72

Dla języka angielskiego słuchanie z egzaminu lepiej koreluje z czytaniem z badania ESLC niż z analogiczną umiejętnością. Czytanie z egzaminu najwyżej koreluje z czytaniem z ESLC, a reagowanie językowe z pisaniem. Zaobserwowane różnice nie są jednak istotne statystycznie (SE 0,04-0,06) i wahają się w granicach 0,63-0,76 (0,57-0,80 po uwzględnieniu SE).

W przypadku języka niemieckiego ze wszystkimi umiejętnościami z ESLC najniżej koreluje słuchanie z egzaminu, co jest związane ze słabą rzetelnością tej podskali stwierdzoną na podstawie współczynnika alfa i korelacji pomiędzy PV. Korelacja pisania z ESLC i słuchania z egzaminu jest na tyle niska (poniżej 0,40), że istotnie różni się od korelacji pisania z ESLC z pozostałymi umiejętnościami mierzonymi na egzaminie. Ciekawe wydaje się to, że słuchanie mierzone w badaniu koreluje z egzaminem gimnazjalnym (oprócz słuchania) w mniejszym stopniu niż pozostałe umiejętności (część różnic jest istotna statystycznie). Podobnie jak w angielskim, reagowanie językowe najlepiej koreluje z pisaniem.

Najwyższe korelacje można oczywiście zaobserwować dla całości egzaminu gimnazjalnego z poszczególnymi umiejętnościami ESLC. W przypadku języka angielskiego współczynniki dla wszystkich umiejętności wynoszą pomiędzy 0,71 i 0,76 i różnice między nimi nie są istotne statystycznie. Dla niemieckiego są one nieco niższe, przy czym najniższy współczynnik ma słuchanie (0,58), nieco wyższy czytanie (0,62), a najwyższe pisanie (pomiędzy 0,66 i 0,72). Różnica pomiędzy słuchaniem a pisaniem (złożonym) jest istotna statystycznie. Może to budzić pewien niepokój ze względu na fakt, że na egzaminie gimnazjalnym nie była badana umiejętność pisania, a jednak związek pomiędzy wynikiem egzaminu i pisaniem z ESLC jest silniejszy niż ze słuchaniem z ESLC.

Ogólnie wszystkie współczynniki są wyższe dla języka angielskiego niż odpowiadające im współczynniki dla języka niemieckiego. Nie powinno to zbytnio dziwić, biorąc pod uwagę niższą rzetelność obydwu testów z języka niemieckiego. Największe różnice występują w przypadku korelacji słuchania mierzonego na egzaminie ze wszystkimi umiejętnościami badanymi w ESLC - wszystkie wynoszą powyżej 0,20 (i są istotne statystycznie). Różnice w przypadku korelacji dwóch pozostałych umiejętności i wyniku sumarycznego egzaminu ze słuchaniem i pisaniem z ESLC są bardzo podobne i wahają się w granicach 0,13-0,16 (w niektórych przypadkach różnice są istotne statystycznie). Najmniejsze różnice można zaobserwować dla korelacji pisania z ESLC (oprócz wspomnianego słuchania z egzaminu) - wynoszą one 0,03-0,08 i nie są istotne statystycznie.

W tabeli 8. zamieszczono analogiczne współczynniki dla *plausible values* z egzaminu. Nie prowadzą one do odmiennych wniosków niż korelacje wyników surowych i są na nieco niższym poziomie. Jedynie dla słuchania z języka niemieckiego można zauważyć dużą różnicę pomiędzy obydwoma zestawieniami. Prawdopodobnie ma to związek z usunięciem podczas skalowania części pytań, co skutkowało ograniczeniem informacji dostarczonych do budowy tej podskali.



**Tabela 8. Współczynniki korelacji Pearsona pomiędzy *plausible values* dla umiejętności z badania ESLC oraz *plausible values* dla wyniku sumarycznego i poszczególnych umiejętności z egzaminu gimnazjalnego**

		Słuchanie	Czytanie	Pisanie (komunikacja)	Pisanie (język)	Pisanie (złożone)
Język angielski	słuchanie	0,56	0,61	0,56	0,56	0,59
	czytanie	0,64	0,68	0,63	0,64	0,66
	reagowanie językowe	0,66	0,69	0,64	0,65	0,68
	egzamin	0,70	0,75	0,69	0,71	0,74
Język niemiecki	słuchanie	0,26	0,33	0,29	0,27	0,31
	czytanie	0,48	0,53	0,60	0,59	0,64
	reagowanie językowe	0,50	0,54	0,60	0,59	0,64
	egzamin	0,56	0,61	0,66	0,65	0,71

## Podsumowanie

Analiza wyników testu ESLC i egzaminu gimnazjalnego dla całej populacji i dla uczniów, których wyniki połączono, pozwala na stwierdzenie, że rozkłady wyników obydwu tych grup nie odbiegają zbyt od siebie. Jeśli chodzi o wyniki badania ESLC, to parametry (średnia i odchylenie standardowe) rozkładów dla obydwu grup nie różnią się statystycznie istotnie od siebie dla obydwu omawianych języków. W przypadku sumarycznego wyniku z egzaminu gimnazjalnego dla języka niemieckiego występuje istotna różnica średnich wynosząca 1 punkt na korzyść uczniów biorących udział w badaniu ESLC. Pomimo niepełnego połączenia danych dla języka niemieckiego, które prawdopodobnie nie miało losowego charakteru, takie parametry obydwu rozkładów zdają się świadczyć o tym, że łączenie danych nie spowodowało istotnego zaburzenia i przeprowadzanie analiz na połączonych danych oraz uogólnianie ich na populację jest uprawnione.

Ogólna rzetelność egzaminu gimnazjalnego z 2011 roku z języka angielskiego i niemieckiego jest na dobrym poziomie. Poszczególne umiejętności, ze względu na krótszą skalę, mają nieco gorsze oszacowania, co jest bardzo dobrze widoczne w przypadku języka niemieckiego. Najśłabszą rzetelność ma słuchanie z języka niemieckiego.

Przeprowadzona analiza czynnikowa wyników egzaminu gimnazjalnego skłania ku stwierdzeniu, że opisywany egzamin mierzył jednowymiarową strukturę związaną z umiejętnościami językowymi uczniów. Wątpliwy w związku z tym wydaje się podział na poszczególne umiejętności w obrębie wyników egzaminu, co skłania raczej ku używaniu wyniku sumarycznego. O ile porównywanie poszczególnych umiejętności ze sobą jest utrudnione, zarówno ze względów teoretycznych, jak i praktycznych (niskiej rzetelności skali słuchania), o tyle porównanie sumarycznego wyniku egzaminu z poszczególnymi umiejętnościami ESLC wydaje się uprawnione.

Korelacje wyników ESLC i egzaminu obliczone zarówno dla wyników surowych i PV z egzaminu są na dość dobrym poziomie. Są one jednak zbyt niskie, aby wyciągnąć jednoznaczne wnioski na temat trafności egzaminu gimnazjalnego z języka obcego. Gdyby były wyższe, moglibyśmy stwierdzić, że obydwa testy mierzą podobny zakres umiejętności, treściowo odpowiadający ESOKJ. Niestety przy tym poziomie współczynników trudno ocenić, czy konstrukty, które mierzą testy, zbyt znacząco odbiegają od siebie, czy też któryś z testów mierzy coś innego niż powinien. Obserwowane korelacje nie są jednak na tyle niskie, aby powodować niepokój co do trafności egzaminu gimnazjalnego z języka obcego. Do interpretacji współczynników korelacji dla poszczególnych umiejętności badanych na egzaminie gimnazjalnym należy podchodzić z ostrożnością, ze względu na opisane problemy z rzetelnością skal i jednowymiarowość egzaminu.

Następny cykl badania ESLC będzie miał miejsce w momencie, kiedy nowa formuła egzaminu będzie funkcjonowała już kilka lat (o ile znów nie ulegnie zmianie). Kolejne tego typu porównanie pozwoliłoby częściowo rozwiązać opisane wyżej wątpliwości, ponieważ od 2012 roku egzamin bazuje na podstawie programowej bezpośrednio odwołującej się do Europejskiego Systemu Opisu Kształcenia Językowego.

## Załącznik 1

**Tabela A. Zestawienie celów zadań w testach ESLC i standardów egzaminacyjnych dla poszczególnych umiejętności (MEN, 2007a, Komisja Europejska, 2012b)**

ESLC 2011 - Cele zadań	Egzamin gimnazjalny 2009-2011 - Standardy
<b>CZYTANIE</b>	
<ul style="list-style-type: none"> <li>• rozpoznawanie informacji związanych ze znanymi tematami w tekstach prywatnych</li> <li>• wyszukiwanie możliwych do przewidzenia informacji w materiałach informacyjnych, np.: ogłoszeniach, rozkładach jazdy, jadłospisach przy wsparciu materiałem ikonograficznym</li> <li>• zrozumienie oznaczeń, tablic informacyjnych, ogłoszeń</li> <li>• określanie głównej myśli tekstu i szczegółów tekstu</li> <li>• zrozumienie informacji, uczuć i intencji w tekstach prywatnych</li> <li>• wyszukiwanie określonych informacji, szczegółowe rozumienie tekstu, określenie nastawienia i opinii</li> <li>• ogólne i szczegółowe rozumienie tekstu, określenie nastawienia, opinii i intencji nadawcy tekstu; wnioskowanie znaczenia z kontekstu, związku między poszczególnymi częściami tekstu</li> <li>• rozpoznawanie struktury tekstu i związków między poszczególnymi częściami tekstu</li> </ul>	Uczeń: <ul style="list-style-type: none"> <li>• określa główną myśl tekstu</li> <li>• określa główną myśl poszczególnych części tekstu</li> <li>• stwierdza, czy tekst zawiera określone informacje; wyszukuje lub selekcjonuje informacje</li> <li>• określa intencje nadawcy tekstu</li> <li>• określa kontekst sytuacyjny</li> <li>• rozpoznaje związki między poszczególnymi częściami tekstu</li> </ul>
<b>SŁUCHANIE</b>	
<ul style="list-style-type: none"> <li>• rozpoznawanie kluczowych słów/informacji</li> <li>• określenie kontekstu sytuacyjnego i/lub głównej myśli tekstu lub funkcji komunikacyjnej</li> <li>• zrozumienie i interpretacja szczegółów tekstu</li> <li>• zrozumienie i interpretacja najważniejszych treści, nastawienia i opinii nadawcy</li> <li>• określenie i interpretacja głównej myśli tekstu, najważniejszych treści i szczegółów tekstu, określenie nastawienia i opinii nadawcy tekstu</li> </ul>	Uczeń: <ul style="list-style-type: none"> <li>• określa główną myśl tekstu</li> <li>• określa kontekst sytuacyjny</li> <li>• stwierdza, czy tekst zawiera określone informacje; wyszukuje lub selekcjonuje informacje</li> </ul>

WYPOWIEDŹ PISEMNA	REAGOWANIE JĘZYKOWE
<ul style="list-style-type: none"> <li>• wyrażanie ogólnych lub szczegółowych treści przedstawionych w materiale ikonograficznym</li> <li>• wyrażanie ogólnych lub szczegółowych treści w odniesieniu do tekstu lub punktów tekstu</li> <li>• tworzenie wypowiedzi w celu przekazania informacji</li> <li>• tworzenie wypowiedzi o funkcji perswazyjnej</li> </ul>	<p>Uczeń:</p> <ul style="list-style-type: none"> <li>• przetwarza treści tekstu przeczytanego w języku polskim lub treści przedstawione w materiale ikonograficznym i wyraża je w języku obcym</li> <li>• właściwie reaguje językowo w określonych kontekstach sytuacyjnych, w szczególności w celu uzyskania, udzielenia, przekazania lub odmowy udzielenia informacji, rozpoczęcia, podtrzymania i zakończenia rozmowy</li> <li>• rozpoznaje i poprawnie stosuje struktury leksykalno-gramatyczne niezbędne do skutecznej komunikacji</li> </ul>

### Bibliografia:

1. Adams, R. J. (2005), *Reliability as a measurement design effect*. Studies in Educational Evaluation, 31, ss. 162-172.
2. Centralna Komisja Egzaminacyjna (2011), *Osiągnięcia uczniów kończących gimnazjum w roku 2011*, Warszawa: CKE.
3. Komisja Europejska (2012a), *First European Survey on Language Competences: Final Report*, Bruksela: Komisja Europejska.
4. Komisja Europejska (2012b), *First European Survey on Language Competences: Technical Report*, Bruksela: Komisja Europejska.
5. Ministerstwo Edukacji Narodowej (2007a), *Rozporządzenie Ministra Edukacji Narodowej z dnia 28 sierpnia 2007 r. zmieniające rozporządzenie w sprawie standardów wymagań będących podstawą przeprowadzania sprawdzianów i egzaminów* (DZ. U. Nr 157, poz. 1102), Warszawa: MEN.
6. Ministerstwo Edukacji Narodowej (2007b), *Uzasadnienie do projektu rozporządzenia Ministra Edukacji Narodowej zmieniającego rozporządzenie w sprawie standardów wymagań będących podstawą przeprowadzania sprawdzianów i egzaminów*, Warszawa: MEN.
7. OECD (2009), *PISA Data Analysis Manual: SPSS, Second Edition*, PISA, OECD Publishing.
8. Smolik, M., Trzcńska, B. (2011), *Zmiany w systemie egzaminów zewnętrznych a Europejski System Opisu Kształcenia Językowego*. Prezentacja podczas seminarium: *Odnoszenie egzaminów językowych do Europejskiego Systemu Opisu Kształcenia Językowego (ESOKJ)*, Warszawa, 28 kwietnia 2011 [dostęp: 01.07.2012] <<http://beta.eduentuzjasci.pl/images/stories/newsy/01.pdf>>.
9. Trzcńska, B. (2009), *Odnoszenie egzaminów z języków obcych do Europejskiego Systemu Opisu Kształcenia Językowego*. *Języki Obce w Szkole*, 3, ss. 84-89.
10. Wu, M. (2005), *The role of plausible values in large-scale surveys*, Studies In Educational Evaluation, 31, ss. 114-128.