



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



entuzjaści
edukacji

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Henryk Szaleniec
Bartosz Kondratek
Filip Kulon
Artur Pokropek
Paulina Skórska
Karolina Świst
Tymoteusz Wołodźko
Mateusz Żółtak

PORÓWNYWALNE WYNIKI EGZAMINACYJNE

Praca zbiorowa

Autorzy:

Henryk Szaleniec

Bartosz Kondratek

Filip Kulon

Artur Pokropek

Paulina Skórska

Karolina Świst

Tymoteusz Wołodźko

Mateusz Żółtak

Recenzenci:

prof. dr hab. Bolesław Niemierko

prof. dr hab. Maciej Karwowski

Wydawca:

Instytut Badań Edukacyjnych

ul. Górczewska 8

01-180 Warszawa

tel. (22) 241 71 00; www.ibe.edu.pl

© Copyright by: Instytut Badań Edukacyjnych, Warszawa 2015

ISBN: 978-83-65115-25-6

Wzór cytowania:

Szaleniec, H., Kondratek, B., Kulon, F., Pokropek, A., Skórska, P., Świst, K., Wołodźko, T. i Żółtak, M. (2015). *Porównywalne wyniki egzaminacyjne*. Warszawa: Instytut Badań Edukacyjnych.

Publikacja współfinansowana przez Unię Europejską ze środków Europejskiego Funduszu Społecznego w ramach projektu: Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego

Spis treści

Spis treści	5
Abstrakt.....	11
Abstract.....	12
1. Wstęp.....	13
2. Cel badań i etapy studium badawczego.....	15
2.1. Etapy studium zrównującego.....	16
3. Charakterystyka egzaminów włączonych do studium badawczego	19
3.1. Sprawdzian	20
3.2. Egzamin gimnazjalny.....	26
3.3. Egzamin maturalny.....	31
3.3.1. Język polski	32
3.3.2. Matematyka	35
3.3.3. Język angielski.....	37
3.3.4. Podsumowanie	39
4. Metodologia	42
4.1. Teoretyczne podstawy zrównywania.....	42
4.2. Dobór próby w badaniach zrównujących.....	45
4.2.1. Elementy schematu badawczego definiującego dobór próby	45
4.2.2. Populacja i operat losowania.....	46
4.2.3. Schemat losowania.....	46
4.3. Narzędzia badawcze	49
4.4. Dobór zadań do zrównania.....	50
4.4.1. Wybór zadań do narzędzi pomiaru w badaniach zrównujących	50
4.4.2. Wybór zadań do analiz zrównujących	58
4.5. Plan nierównoważnych grup z testem kotwiczącym i zrównywanie z wykorzystaniem IRT ..	59
4.6. Implementacja metody łącznej kalibracji modelu IRT do zrównania egzaminów gimnazjalnych i sprawdzianu	63

4.7.	Zrównywanie wyników obserwowanych metodą ekwicytylową	63
4.8.	Generowanie wyników obserwowanych na wspólnej skali w oparciu o IRT	64
4.9.	Generowanie wartości potencjalnych/możliwych (PV)	65
4.10.	Założenia skalowania wertykalnego	67
5.	Realizacja badań terenowych	70
5.1.	Rekrutacja szkół	70
5.2.	Badanie pilotażowe	72
5.3.	Szkolenie koordynatorów i ankieterów	73
5.4.	Realizacja badań w szkołach i ocenianie	73
5.5.	Progi realizacji	74
5.6.	Kontrola badań terenowych	75
5.7.	Podwójne ocenianie losowej próby prac	75
5.7.1.	Sprawdzian	77
5.7.2.	Egzamin gimnazjalny	79
5.7.3.	Egzamin maturalny	81
6.	Wyniki zrównywania	83
6.1.	Sprawdzian	83
6.1.1.	Zmiany trudności egzaminu w latach 2002–2013	83
6.1.2.	Wyniki w latach 2002–2013 na skali zmiennej ukrytej	90
6.2.	Egzamin gimnazjalny – część matematyczno-przyrodnicza	92
6.2.1.	Zmiany trudności egzaminu w latach 2002–2013	92
6.2.2.	Wyniki w latach 2002–2013 na skali zmiennej ukrytej	95
6.3.	Egzamin gimnazjalny – część humanistyczna	97
6.3.1.	Zmiany trudności egzaminu w latach 2002–2013	97
6.3.2.	Wyniki w latach 2002–2013 na skali zmiennej ukrytej	99
6.4.	Egzamin maturalny z matematyki na poziomie podstawowym	101
6.4.1.	Zmiany trudności egzaminu w latach 2010–2013	101
6.4.2.	Wyniki w latach 2010–2013 na skali zmiennej ukrytej	104
6.5.	Egzamin maturalny z języka polskiego na poziomie podstawowym	105

6.5.1.	Specyficzne problemy związane ze zrównywaniem egzaminu maturalnego z języka polskiego	105
6.5.2.	Zmiany trudności egzaminu w latach 2010–2013	109
6.5.3.	Wyniki w latach 2010–2013 na skali zmiennej ukrytej	111
6.6.	Egzamin maturalny z języka angielskiego na poziomie podstawowym	112
6.6.1.	Zmiany trudności egzaminu w latach 2010–2013	112
6.6.2.	Wyniki w latach 2010–2013 na skali zmiennej ukrytej	114
6.7.	Zrównanie ekwicytylowe.....	115
6.8.	Skalowanie wertykalne	120
6.8.1.	Język polski	125
6.8.2.	Matematyka	126
6.9.	Badania ankietowe	128
6.9.1.	Motywacja testowa uczniów	128
6.9.2.	Badania ankietowe nauczycieli.....	134
6.9.3.	Podsumowanie	140
7.	Analizy z użyciem porównywalnych wyników egzaminacyjnych.....	142
7.1.	Ogólne wyniki egzaminów w poszczególnych latach	142
7.1.1.	Sprawdzian	143
7.1.2.	Egzamin gimnazjalny	144
7.1.3.	Egzamin maturalny	146
7.2.	Regiony kraju.....	148
7.2.1.	Sprawdzian	149
7.2.2.	Egzamin gimnazjalny	150
7.2.3.	Egzamin maturalny	151
7.3.	Płeć uczniów	154
7.3.1.	Sprawdzian	154
7.3.2.	Egzamin gimnazjalny	155
7.3.3.	Egzamin maturalny	157
7.4.	Diagnoza dysleksji	159
7.4.1.	Sprawdzian	159

7.4.2.	Egzamin gimnazjalny	161
7.4.3.	Egzamin maturalny	162
7.5.	Typ szkoły	164
7.5.1.	Sprawdzian	165
7.5.2.	Egzamin gimnazjalny	167
7.5.3.	Egzamin maturalny	169
7.6.	Lokalizacja szkoły	172
7.6.1.	Sprawdzian	173
7.6.2.	Egzamin gimnazjalny	173
7.7.	Podsumowanie	175
8.	Prezentacja porównywalnych wyników egzaminacyjnych	177
8.1.	Budowa serwisu.....	177
8.2.	Prezentacja wyników	179
8.2.1.	Skala prezentacji wyników.....	179
8.2.2.	Wykres liniowy	180
8.2.3.	Wykres skrzynkowy	180
8.2.4.	Tabela danych	181
8.3.	Przykłady	181
8.3.1.	Przykład 1.	182
8.3.2.	Przykład 2.	183
8.3.3.	Przykład 3.	184
9.	Baza danych IBE.....	186
9.1.	Dane zawarte w bazie danych.....	186
9.1.1.	Wyniki uczniów i informacje o uczniach	187
9.1.2.	Baza szkół	187
9.1.3.	Bank zadań.....	188
9.1.4.	Oszacowania umiejętności uczniów	188
9.1.5.	Zagregowane wartości wskaźników PWE i EWD.....	188
9.2.	Dostęp do danych znajdujących się w bazie	189
9.2.1.	Serwisy WWW	189

9.2.2.	API HTTP	189
9.2.3.	Pakiet ZPD dla R	190
9.2.4.	Bezpośrednie wykonywanie zapytań SQL na bazie.....	191
10.	Rekomendacje – możliwe scenariusze rozwoju zrównywania wyników polskich egzaminów.....	192
10.1.	Rozwiązania stosowane na świecie	193
10.2.	Warianty docelowo rekomendowane dla polskiego systemu egzaminów zewnętrznych ...	195
10.2.1.	Wariant A – włączenie do egzaminu niejawnych zadań kotwiczących	195
10.2.2.	Wariant B – jednoczesna standaryzacja zadań z wielu edycji egzaminu	197
10.2.3.	Wariant C – dodatkowe badania na reprezentatywnej próbie.....	199
10.3.	Podsumowanie	201
	Bibliografia	203
	Spis ilustracji.....	212
	Spis tabel	215
	Aneksy	218
Aneks 1.	Tabele przeliczeniowe	218
1.1.	Sprawdzian	218
1.2.	Egzamin gimnazjalny w części humanistycznej	219
1.3.	Egzamin gimnazjalny w części matematyczno-przyrodniczej.....	220
1.4.	Matura z j. angielskiego	222
1.5.	Matura z matematyki	223
1.6.	Matura z j. polskiego.....	224
Aneks 2.	Plany zrównania wyników egzaminów	225
2.1.	Sprawdzian 2012-2014.....	225
2.2.	Gimnazjum 2011–2014	226
2.3.	Matura z języka angielskiego 2014	227
2.4.	Matura z matematyki 2013-2014	228
2.5.	Matura z polskiego 2014	229
Aneks 3.	Zadania wchodzące w skład zeszytów testowych.....	230
3.1.	Sprawdzian	230

3.2.	Egzamin gimnazjalny – część humanistyczna	236
3.3.	Egzamin gimnazjalny – część matematyczno-przyrodnicza	248
3.4.	Matura z języka angielskiego	258
3.5.	Matura z języka matematyki	261
3.6.	Matura z języka polskiego	266
Aneks 4.	Kwestionariusze ankiet.....	270
4.1.	Skala Motywacji Testowej – ankieta dla ucznia	270
4.2.	Skala Opinii Uczniów (polskie tłumaczenie Student Opinion Scale, SOS; Sundre, 2007) – ankieta dla ucznia	275
4.3.	Ankiety dla nauczyciela (szkoła podstawowa)	276
4.4.	Ankiety dla nauczyciela (gimnazjum)	279
4.5.	Przykładowa ankieta nauczyciela (szkoła ponadgimnazjalna).....	282
Aneks 5.	Psychometryczne własności zadań egzaminacyjnych	284
Aneks 6.	Zgodność kodowania zadań otwartych	292
6.1.	Egzamin gimnazjalny – część humanistyczna 2011	292
6.2.	Egzamin gimnazjalny – część matematyczno-przyrodnicza 2011	296
6.3.	Matura z matematyki 2013	297
6.4.	Matura z matematyki 2014	299
6.5.	Matura z języka angielskiego 2014	300
6.6.	Matura z języka polskiego 2014	304

Abstrakt

Niniejsza monografia zawiera podsumowanie badań zrównujących wyniki egzaminacyjne przeprowadzonych w latach 2011-2014. Badania zrównujące objęły kolejno egzamin gimnazjalny, sprawdzian na zakończenie szkoły podstawowej i wybrane przedmioty (język polski, język angielski, matematyka) na poziomie maturalnym. Porównywalne wyniki egzaminacyjne (PWE) przedstawiono w skali standardowej o średniej 100 i odchyleniu standardowym 15, zakotwiczonej do roku bazowego 2012.

W poszczególnych rozdziałach monografii przedstawiono charakterystykę zrównywanych egzaminów, oraz główne założenia metodologiczne przeprowadzonego studium. W części poświęconej wynikom zamieszczono PWE dla sprawdzianu oraz egzaminu gimnazjalnego w latach 2002-2013 oraz egzaminu maturalnego w latach 2010-2013. W rozdziale tym znajdują się również rezultaty analiz uzupełniających – wyniki zrównania ekwicyntylowego oraz skalowania wertykalnego. W kolejnych częściach monografii zamieszczono analizy z wykorzystaniem porównywalnych wyników oraz zmiennych kontekstowych dla wszystkich zrównywanych egzaminów. Przedstawiono również wyniki badań ankietowych przeprowadzonych wśród uczniów i nauczycieli biorących udział w studium. Zaprezentowano także opis serwisu <http://pwe.ibe.edu.pl> oraz bazy danych IBE pozwalających na bezpośredni dostęp do PWE i dostarczających narzędzi umożliwiających samodzielne wykonanie analiz. Niniejsza publikacja zawiera też rekomendacje obejmujące potencjalne scenariusze rozwoju zrównywania w Polsce. Warianty te różnią się nakładem kosztów, czasu oraz ingerencji w system egzaminów zewnętrznych. Dołączone do raportu aneksy pozwalają na szczegółowe prześledzenie sposobu realizacji i analiz przeprowadzonych w studium zrównującym.

Abstract

This monograph presents a summary of equating studies of examination results from the years 2011-2014. The equating studies encompass the following: exam at the end of lower secondary school (gimnazjum), exam at the end of primary school and chosen subjects (Polish, English, mathematics) at the matriculation level (matura). Comparable examination results (CER) are presented on a standard scale with a mean of 100 and standard deviation of 15, anchored to the base year 2012.

The respective chapters of the monograph present the characteristics of the equated exams and the main methodological assumptions of the study. The part with the results presents CER for the exam at the end of primary school and the exam at the end of secondary school in 2002-2013 and the matura exam in 2010-2013. The results of supplementary analyses – equipercentile equating and vertical scaling – are also included in this chapter. The monograph also includes analyses using comparable results and context variables for all equated exams. The results of questionnaire studies conducted among students and teachers taking part in the study are also presented. The website <http://pwe.ibe.edu.pl> and IBE data bases offering direct access to CER and providing tools enabling individual analyses are also given. The publication also presents recommendations encompassing potential scenarios for the development of equating in Poland. These options vary with respect to costs, time and intervention in the system of external exams. Annexes included in the report enable the reader to follow the implementation and analyses conducted in the equating study.

1. Wstęp

Polski system egzaminów zewnętrznych pod koniec ubiegłego wieku w momencie przygotowania do wdrożenia był nowatorski i dobrze dopasowany do potrzeb przy wprowadzaniu tak dużej zmiany. Jednocześnie uwzględniał panujące w kraju uwarunkowania społeczne, związane z egzaminacyjną tradycją szkolną, i finansowe. W przygotowaniu wdrożenia systemu egzaminacyjnego w Polsce brali udział eksperci z wiodących w Europie instytucji prowadzących egzaminy, jak CITO¹ (*Central Institute for Test Development*) i AEB² (*Associated Examining Board*), co dostarczyło wiedzy o trendach w tej dziedzinie oraz użytecznych rozwiązaniach stosowanych w innych systemach egzaminacyjnych. Z biegiem lat polski system egzaminacyjny stawał się jednak coraz bardziej odległy od założeń, które były wypracowane na etapie jego projektowania. Budowaniu wieloletniej strategii doskonalenia egzaminów nie sprzyjały szczególnie dwa zjawiska. Pierwsze to upolitycznienie egzaminów. Spowodowało ono między innymi przeniesienie wdrożenia zewnętrznej matury z roku 2002 na rok 2005 przy jednoczesnej rezygnacji z obowiązkowej matury z matematyki (MEN, 2001c) oraz wprowadzenie tzw. amnestii maturalnej w 2006 roku (MEN, 2006). Jako drugie niepożądane zjawisko można wymienić słabą pozycję Centralnej Komisji Egzaminacyjnej (CKE), między innymi z powodu dużej rotacji na stanowisku dyrektora (10 zmian w ciągu 13 lat).

Jednym z problemów związanym z funkcjonowaniem egzaminów zewnętrznych w Polsce, którego rozwiązanie wymaga podejścia strategicznego, jest brak wbudowanego w system zrównywania wyników egzaminacyjnych. W rezultacie porównywalność wyników z roku na rok i z sesji na sesję nie jest bezpośrednio zapewniona. Oznacza to, że na podstawie komunikowanych po egzaminie wyników nie można określić, czy ten sam egzamin, zdawany w różnych latach, był tak samo trudny. System egzaminacyjny nie jest też w stanie dostarczyć informacji, czy uczniowie z jednego rocznika posiadają wyższy poziom umiejętności, czy niższy niż ich rówieśnicy z innego rocznika. Do chwili obecnej nie wprowadzono żadnego systemowego rozwiązania umożliwiającego coroczne zrównywanie wyników egzaminów, chociaż pierwsza próba pilotażowego zrównywania wyników sprawdzianu w szóstej klasie szkoły podstawowej miała miejsce w CKE w latach 2003–2005. W ramach tego pilotażu w celu zrównania *post factum* sprawdzianu 2002, a następnie 2004 i 2005 do roku 2003 (który został przyjęty jako rok bazowy), Bolesław Niemierko zastosował metodę ekwicytylową (Niemierko, 2004; 2007). Badania zrównujące odbyły się na próbie celowej szkół w klasach piątych, czyli o rok niższych. Wyniki zrównane wykazały spadek pomiędzy rokiem 2002 i rokiem 2003 (przyjętym za bazowy) i wzrost w kolejnych latach – 2004 i 2005. Różnice okazały się statystycznie istotne ($p < 0,05$). Niemierko sugerował, że zaobserwowany wzrost zrównanych wyników w latach 2004 i 2005 mógł być spowodowany pozytywnym efektem zwrotnym wpływu systemu egzaminacyjnego na program i realizację kształcenia, a także wzrostem obycia egzaminacyjnego szóstoklasistów (*test-taking skill*) (Niemierko, 2004; 2007). Podobne rezultaty w przypadku sprawdzianu uzyskał Henryk Szaleniec (Szaleniec, 2005), stosując dla tego samego zbioru danych (z lat 2003-2005) jednoparametryczny model logistyczny i oprogramowanie OPLM (*One Parametr Logistic Model*). Oszacowana różnica osiągnięć uczniów pomiędzy rokiem 2003 i 2004 wynosiła 7,7% odchylenia standardowego, a pomiędzy rokiem 2004 i 2005 – 5,6%. Nieco inne wyniki otrzymał Marcin Smolik (Smolik, 2007), stosując do analiz model Rascha i program Winsteps. Na podstawie analiz przeprowadzonych przez Smolika pomiędzy

¹ Krajowy Instytut Pomiaru Edukacyjnego założony przez rząd holenderski w 1968 roku. W 1999 roku został sprywatyzowany. Obecnie ma także oddziały w USA, Niemczech i Turcji.

² Dziś AQA (Assessment and Qualifications Alliance).

rokiem 2003 i 2004 różnica w wynikach zrównanych sprawdzianu jest bardzo mała i nieistotna statystycznie. Natomiast wzrost średnich zrównanych wyników pomiędzy rokiem 2003 i 2005 wynosi 12% odchylenia standardowego i jest statystycznie istotny ($p < 0,01$). Jak już wspomniano wcześniej, te działania miały charakter pilotażowy i obejmowały tylko jeden egzamin. Ponadto pilotaż dotyczył wąskiego przedziału czasowego – obejmował wyłącznie trzy roczniki.

Warto w tym miejscu zauważyć, że na świecie zarówno funkcjonujące od dziesięcioleci systemy testowania, jak i niektóre z nowo powstałych systemów wprowadzają już na etapie konstrukcji testów mechanizmy, które pozwalają na zrównywanie wyników pomiędzy latami. Można wśród nich wymienić np. amerykański SAT (*Scholastic Assessment Test*) i ACT (*American College Test*), PET (*Psychometric Entrance Test*) w Izraelu, SweSAT (*Swedish Scholastic Assessment Test*), testy funkcjonujące w ramach programu ewaluacyjnego w prowincji Ontario w Kanadzie – EQAO (*Education Quality and Accountability Office*), NAPLAN (*National Assessment Program – Literacy and Numeracy*) w Australii czy węgierski test *National Assessment of Basic Competencies*. Szerzej na temat zrównywania w innych krajach pisze Artur Pokropek (Pokropek, 2011b). Zagadnienie to zostało także omówione przez Artura Pokropka i Bartosza Kondratka (Pokropek i Kondratak, 2012).

Dysponowanie porównywalnymi pomiędzy latami wynikami ma znaczenie nie tylko w odniesieniu do całego systemu edukacyjnego ze względu na monitorowanie trendów poziomu osiągnięć na poszczególnych etapach edukacyjnych, ale także na poziomie jednostkowym. Jeśli np. planowane są jakiegokolwiek zmiany w systemie i wyniki z kolejnych lat mają być wykorzystywane do podejmowania decyzji, to kluczowe jest, by wyniki były porównywalne między latami. Na poziomie jednostkowym porównywalność wyników pomiędzy latami ma szczególne znaczenie w przypadku egzaminu maturalnego, kiedy absolwenci z różnych lat (zdający egzaminy o zróżnicowanym poziomie trudności) konkurują o indeks w tej samej szkole wyższej.

Podjęte przez Instytut Badań Edukacyjnych (IBE) czteroetapowe studium (zob. Rozdział 2) miało na celu zrównanie *post factum* archiwalnych wyników zgromadzonych przez system egzaminów zewnętrznych. Studium dotyczyło sprawdzianu w szóstej klasie szkoły podstawowej i egzaminu gimnazjalnego od początku funkcjonowania systemu egzaminacyjnego (od 2002 roku) oraz wyników trzech obowiązkowych egzaminów na poziomie maturalnym (język polski, język angielski i matematyka) od roku 2010, czyli od roku, w którym egzamin maturalny z matematyki stał się obowiązkowy. Ponadto badania były poszukiwaniem rozwiązań metodologicznych, które mogłyby być zastosowane w systemie egzaminacyjnym w trakcie sesji. Przygotowane na podstawie wyników badań rekomendacje wdrożenia są istotną częścią rezultatów osiągniętych w trakcie realizacji studium badawczego.

2. Cel badań i etapy studium badawczego

Badania przeprowadzone w ramach studium zrównującego miały charakter zrównywania *post factum*. Oznacza to, że w trakcie badań dla losowej próby uczniów z danego rocznika (2011, 2012, 2013 lub 2014) zostały zastosowane zeszyty testowe zawierające zadania egzaminacyjne z wcześniejszych lat³. Oprócz głównych celów wymienionych poniżej, kluczowe było poszukiwanie rozwiązań metodologicznych i statystycznych, które mogłyby być rekomendowane do zastosowania w egzaminach zewnętrznych bezpośrednio w trakcie sesji, gdyby zrównywanie było wbudowane w system egzaminacyjny.

Głównymi celami badań zrównujących było:

1. Zrównanie wyników sprawdzianu przeprowadzonego w latach 2002–2013 do roku bazowego.
2. Zrównanie wyników egzaminu gimnazjalnego w części humanistycznej i matematyczno-przyrodniczej przeprowadzonego w latach 2002–2013 do roku bazowego.
3. Zrównanie wyników egzaminu maturalnego z matematyki przeprowadzonego w latach 2010–2013 do roku bazowego.
4. Zrównanie wyników egzaminu maturalnego z języka angielskiego przeprowadzonego w latach 2010–2013 do roku bazowego.
5. Zrównanie wyników egzaminu maturalnego z języka polskiego przeprowadzonego w latach 2010–2013 do roku bazowego.
6. Przedstawienie wyników sprawdzianu i egzaminu gimnazjalnego w części humanistycznej i części matematyczno-przyrodniczej przeprowadzanych w latach 2002–2013 w skali standardowej o średniej 100 i odchyleniu standardowym 15, zakotwiczonej do roku bazowego.
7. Przedstawienie wyników egzaminu maturalnego na poziomie podstawowym z języka polskiego, języka angielskiego i matematyki przeprowadzonych w latach 2010–2013 w skali standardowej o średniej 100 i odchyleniu standardowym 15, zakotwiczonej do roku bazowego.
8. Przedstawienie wyników sprawdzianu i egzaminu gimnazjalnego w części humanistycznej i matematyczno-przyrodniczej przeprowadzanego w latach 2002–2013 w skali standardowej o średniej 100 i odchyleniu standardowym 15, zakotwiczonej do roku bazowego z uwzględnieniem:
 - a) płci ucznia,
 - b) lokalizacji szkoły,

³ Należy pamiętać, że w badaniu zrównującym wykorzystano wyłącznie zadania z arkuszy standardowych (dla uczniów bez niepełnosprawności i uczniów ze specyficznymi trudnościami w uczeniu się), tak więc uzyskane wyniki pozwalają wnioskować wyłącznie o poziomie umiejętności uczniów piszących egzamin z wykorzystaniem standardowych arkuszy egzaminacyjnych.

- c) typu szkoły (szkoły publiczne i niepubliczne),
 - d) diagnozy dysleksji,
 - e) uwarunkowań regionalnych.
9. Przedstawienie wyników egzaminu maturalnego na poziomie podstawowym z języka polskiego, języka angielskiego i matematyki przeprowadzonych w latach 2010–2013 w skali standardowej o średniej 100 i odchyleniu standardowym 15, zakotwiczonej do roku bazowego z uwzględnieniem:
- a) płci ucznia,
 - b) diagnozy dysleksji,
 - c) uwarunkowań regionalnych,
 - d) typu szkoły (liceum ogólnokształcące, liceum profilowane i technikum).
10. Przygotowanie ogólnodostępnego serwisu komunikowania wyników porównywalnych sprawdzianu, egzaminu gimnazjalnego i maturalnego z języka polskiego, języka angielskiego i matematyki.
11. Przygotowanie rekomendacji bieżącego zrównywania wyników.

Jako element uzupełniający badanie zrównujące, przeprowadzono badania ankietowe uczniów i nauczycieli. Celem badania ankietowego uczniów było znalezienie odpowiedzi na pytanie, czy motywacja do rozwiązywania zadań testowych jest związana z wynikami testów niskiej doniosłości (którymi są badania zrównujące, nie mające poważnych konsekwencji dla egzaminowanych). Związek motywacji z wynikami badania zrównującego byłby cechą niewskazaną, świadczącą o ich obciążeniu. Cel badań ankietowych wiązał się z odpowiedzią na trzy poniższe pytania badawcze.

1. Czy deklarowany poziom motywacji testowej uczniów związany jest z wynikami osiągniętymi w testach wykorzystanych w badaniach zrównujących?
2. Ile procent zmienności wyników uzyskanych w badaniach zrównujących wyjaśnia poziom motywacji testowej?
3. Czy oba te efekty różnicują się w zależności od rozwiązywanego testu oraz etapu edukacyjnego?

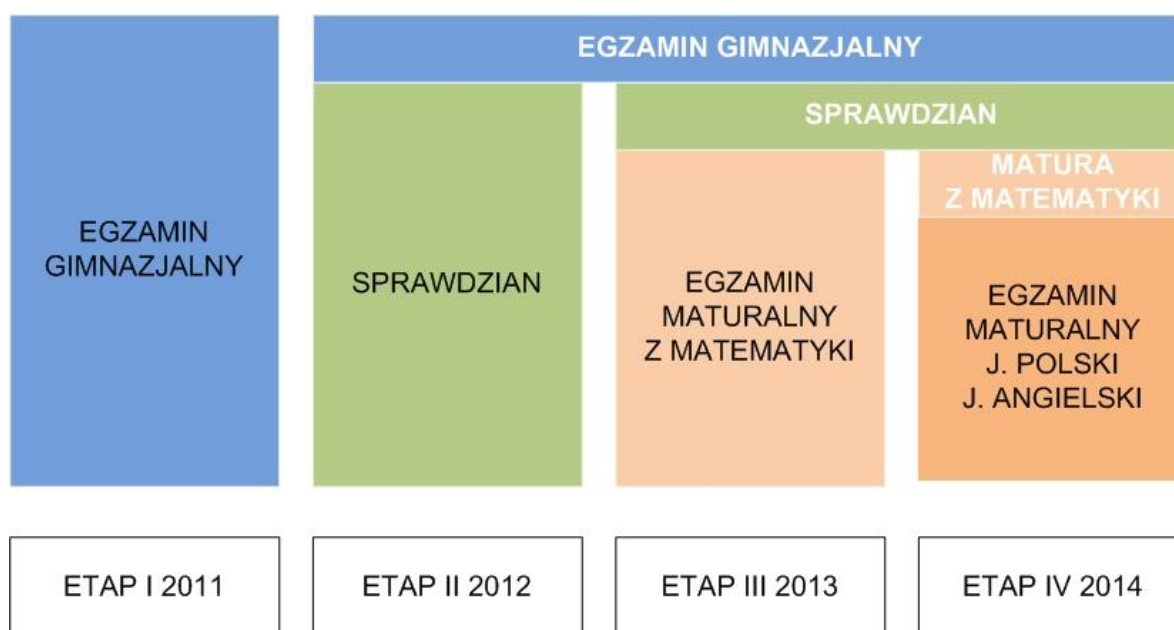
Badania nauczycieli pozwoliły na uzyskanie następujących informacji kontekstowych: ocenianego poziomu przygotowania uczniów do egzaminu zewnętrznego, szacowanego poziomu motywacji uczniów do rozwiązywania testów wykorzystywanych w badaniach zrównujących, potencjalnego wpływu udziału uczniów w badaniu zrównującym na ich wynik w egzaminach zewnętrznych, korzystania z arkuszy z poprzednich egzaminów podczas lekcji oraz opinii na temat porównywalności egzaminów zewnętrznych.

2.1. Etapy studium zrównującego

Badania zrównujące zaplanowane i przeprowadzone zostały przez IBE jako czteroetapowe studium zrównujące (zob. Rysunek 2.1). Pierwszy etap badań przeprowadzono w 2011 roku. W tym etapie zebrano dane pozwalające na zrównanie egzaminu gimnazjalnego dla lat 2002–2010. Drugi etap badań, zrealizowany w 2012 roku, dotyczył głównie sprawdzianu w szóstej klasie szkoły podstawowej i

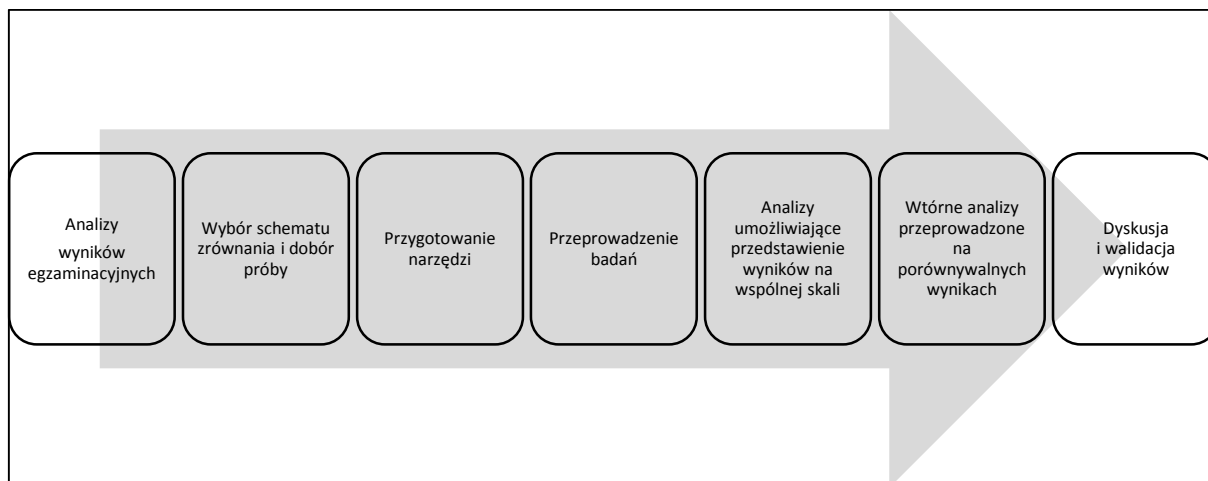
jego podstawowym celem było zrównanie wyników sprawdzianu przeprowadzonego w latach 2002–2011. W tym etapie zebrano również dane pozwalające na kontynuację zrównywania egzaminu gimnazjalnego. Trzeci etap, przeprowadzony w roku 2013, dotyczył głównie egzaminu maturalnego z matematyki, ale zebrano również dane pozwalające na kontynuację zrównania sprawdzianu i egzaminu gimnazjalnego. Czwarty etap, którego realizacja przypadła na 2014 rok, skoncentrowany był na wybranych egzaminach maturalnych – języku polskim oraz angielskim. Dodatkowym elementem czwartego etapu badań było zebranie danych umożliwiających kontynuację wcześniejszych etapów zrównywania.

Jak już wspomniano, badania przeprowadzone w ramach studium zrównującego miały charakter zrównywania *post factum*, czyli losowa próba uczniów z danego rocznika (2011, 2012, 2013 lub 2014) rozwiązywała zadania egzaminacyjne z wcześniejszych lat. Następnie na podstawie informacji o rozwiązanych zadaniach z różnych edycji egzaminu pierwotne wyniki były przekształcane w taki sposób, aby zapewnić porównywalność (zob. Rozdział 4).



Rysunek 2.1. Cztery etapy studium zrównywania wyników przeprowadzonego w IBE w latach 2011–2014

W każdym etapie zrównywania ogólny schemat badania zrównującego był taki sam. Przedstawia go Rysunek 2.2. Podczas kolejnych etapów doskonalona była koncepcja i realizacja analiz statystycznych (zob. Rozdział 4). Uległ też zmianie rok bazowy, do którego odnoszone były zrównane wyniki. Podczas pierwszego etapu za rok odniesienia dla egzaminu gimnazjalnego przyjęto rok 2003. Podczas drugiego etapu badań przyjęto rok 2004 za rok odniesienia dla sprawdzianu. Wybór innego roku bazowego dla sprawdzianu spowodowany był uchybieniami koordynacji oceniania sprawdzianu w 2003 roku w jednej z okręgowych komisji egzaminacyjnych (OKE). Począwszy od trzeciego etapu badań przyjęto za rok bazowy (dla wszystkich egzaminów włączonych do studium badawczego) rok 2012. Zmiana roku bazowego wynikała z dwóch przesłanek. Po pierwsze, rok 2012 jest też pierwszym rokiem, w którym uczniowie gimnazjum kształceni zgodnie z nową podstawą programową, zdawali egzamin gimnazjalny według zmienionej formuły. Po drugie, zrównywanie dla egzaminów maturalnych obejmowało okres rozpoczynający się w 2010 roku, w którym po raz pierwszy matematyka została włączona do kanonu przedmiotów obowiązkowych.



Rysunek 2.2. Schemat procesu badawczego w przeprowadzonych badaniach zrównujących

3. Charakterystyka egzaminów włączonych do studium badawczego

Rozdział ten przybliży ogólny opis egzaminów włączonych do studium badawczego ze szczególnym zwróceniem uwagi na okresy, dla których przeprowadzono zrównanie wyników. W kolejnych podrozdziałach przedstawiono charakterystyki sprawdzianu w szóstej klasie szkoły podstawowej (3.1), egzaminu gimnazjalnego (3.2) i maturalnego z języka polskiego, matematyki i języka angielskiego (3.3). Dla każdego z opisywanych egzaminów przedstawiono grupy umiejętności będące podstawą jego przeprowadzenia z wyszczególnieniem ich wagi w punktacji za rozwiązanie zadań z całego arkusza egzaminacyjnego. Podano także wybrane własności psychometryczne dla tych egzaminów w odniesieniu do populacji, które stanowią uczniowie szkół dla młodzieży, z wyłączeniem szkół specjalnych i przyszpitalnych, piszący egzamin z zastosowaniem arkusza standardowego⁴ dla danego egzaminu. Natomiast parametry psychometryczne poszczególnych zadań egzaminacyjnych zarówno w ujęciu KTT (Klasycznej Teorii Testu) jak i parametry IRT (Item Response Theory) dostępne są za pośrednictwem bazy danych IBE (zob. Rozdział 9 i Aneks 5). Rozdział kończy dyskusja wymagań tożsamego konstruktów w odniesieniu do poszczególnych egzaminów i stabilności ich rzetelności w okresie objętym zrównywaniem wyników.

Instytucjonalne wprowadzenie zewnętrznych egzaminów w Polsce w pierwszym etapie poprzedziły regionalnie badania kompetencji uczniów na zakończenie szkoły podstawowej prowadzone w latach 1997–1999 (Szmigiel i Szaleniec, 1999, Ankudewicz-Blukacz i Małecki, 1999), których wyniki uznawane były przy rekrutacji do szkół ponadpodstawowych. Kolejnym krokiem milowym były działania w kierunku przekształcenia egzaminu dojrzałości przeprowadzanego i ocenianego w szkole według zasad stanowiących odrębnie przez każde kuratorium oświaty w zewnętrzną maturę. W 1994 roku został powołany krajowy program „Nowa Matura” mający na celu przygotowywanie szkół do zmiany egzaminów maturalnych. Uczestnikami programu Nowa Matura byli nauczyciele szkół ponadpodstawowych, pracownicy kuratoriów oświaty i ośrodków doskonalenia oraz nauczyciele akademicy. W latach 1997–1999 program „Nowa Matura” został wzmocniony przez program SMART (komponent 02) (Małecki, 1999) finansowany z funduszy PHARE, który objął maturalne przedmioty przyrodnicze: chemię, fizykę i geografię. W ramach SMART powstał też komponent, którego celem było przygotowanie wdrożenia sprawdzianu i egzaminu gimnazjalnego. Uczestnikami SMART-u byli, podobnie jak programu Nowa Matura, nauczyciele szkół, pracownicy kuratoriów oświaty i ośrodków doskonalenia oraz nauczyciele akademicy. Program „Nowa Matura” w pierwszej fazie obejmował cały kraj w podziale na sześć, a potem osiem Regionalnych Komisji Egzaminów Szkolnych (RKES). Te osiem Regionalnych Komisji Egzaminów Szkolnych dało początek Okręgowym Komisjom Egzaminacyjnym. W 1998 roku Sejm RP zmieniając ustawę o systemie oświaty powołał od 1 stycznia 1999 roku Centralną Komisję Egzaminacyjną (CKE) i osiem okręgowych komisji egzaminacyjnych (OKE) oraz wprowadził sprawdzian i egzaminy zewnętrzne począwszy od 2002 roku (Kancelaria Sejmu, 1998).

Obecnie w szkołach ogólnokształcących przeprowadzane są egzaminy na zakończenie drugiego, trzeciego i czwartego etapu edukacji. Na zakończenie drugiego etapu edukacji przeprowadzany jest w szóstej klasie sprawdzian, który jest obowiązkowy i ma charakter diagnostyczny. W ostatniej klasie

⁴ Arkusz egzaminacyjny dla uczniów bez niepełnosprawności i uczniów ze specyficznymi trudnościami w uczeniu się.

gimnazjum uczniowie przystępują do obowiązkowego egzaminu, który nie ma progu zaliczenia, ale jego wyniki są wykorzystywane przy rekrutacji do szkoły ponadgimnazjalnej. Absolwenci szkół ponadgimnazjalnych przystępują do egzaminu maturalnego, który nie jest obowiązkowy, ale jego wyniki są podstawą rekrutacji na studia. Okręgowe komisje egzaminacyjne przeprowadzają także egzaminy eksternistyczne (szkoła podstawowa, gimnazjum, liceum ogólnokształcące, zasadnicza szkoła zawodowa) i egzaminy potwierdzające kwalifikacje zawodowe. Egzaminy te nie były włączone do studium badawczego i nie będziemy się do nich odnosić w tym opracowaniu.

Za organizację i przeprowadzenie sprawdzianu, egzaminu gimnazjalnego i maturalnego w poszczególnych szkołach na terenie całego kraju są odpowiedzialni przewodniczący szkolnych zespołów egzaminacyjnych (PSZE), którymi najczęściej są dyrektorzy lub wicedyrektorzy tych placówek. PSZE powołuje dyrektor okręgowej komisji egzaminacyjnej. OKE przeprowadza także szkolenie PSZE dotyczące procedur obowiązujących w jednolity sposób w całym kraju podczas sesji egzaminacyjnej. Najczęściej są to jednodniowe spotkania lub sesje z wykorzystaniem e-konferencji. Dystrybucja narzędzi egzaminacyjnych prowadzona jest centralnie, zgodnie ze złożonymi przez szkoły zamówieniami na odpowiednie wersje arkuszy egzaminacyjnych. Za ocenianie prac egzaminacyjnych, przetwarzanie i komunikowanie wyników, w tym wydawanie zaświadczeń o wynikach, odpowiedzialne są okręgowe komisje egzaminacyjne, których działania koordynuje Centralna Komisja Egzaminacyjna. W celu zapewnienia przestrzegania etycznych zachowań uczniów i nauczycieli w trakcie egzaminów w 2003 roku wprowadzono możliwość nadzoru egzaminu przez ekspertów zewnętrznych powoływanych przez dyrektora OKE i obserwatorów, którymi mogli być nauczyciele z innej szkoły. W kolejnych latach do szkolnych zespołów nadzorujących egzamin zostali obowiązkowo włączeni nauczyciele z innej szkoły niż ta, w której odbywa się egzamin. Począwszy od 2004 roku stosuje się też wersje równoległe testu standardowego, które zostały wprowadzone w celu utrudnienia prób odpisywania. W latach 2007–2009 w egzaminie gimnazjalnym i egzaminie z języków obcych na maturze stosowane były trzy ściśle równoległe wersje testów egzaminacyjnych. Na egzaminie maturalnym z matematyki dwie wersje testów stosowane są od 2010 roku. W wymienionych egzaminach do chwili obecnej stosowane są dwie wersje ściśle równoległe. Arkusze w wersji A i w wersji B różnią się jedynie kolejnością odpowiedzi w zadaniach zamkniętych.

3.1. Sprawdzian

Sprawdzian w szóstej klasie szkoły podstawowej zaplanowany był jako interdyscyplinarny test diagnostyczny obowiązkowy dla wszystkich uczniów. Ze sprawdzianu zwolnieni są tylko uczniowie z upośledzeniem umysłowym w stopniu umiarkowanym lub głębokim. Przystąpienie do sprawdzianu warunkuje uzyskanie świadectwa ukończenia szkoły podstawowej, chociaż jego wyniki przedstawiane są na odrębnym zaświadczeniu wydawanym przez okręgową komisję egzaminacyjną i nie mają znaczenia dla promocji. Celem sprawdzianu jest diagnoza opanowania umiejętności niezbędnych na wyższym etapie kształcenia (gimnazjum), co podkreśla jego funkcję diagnostyczną. Zasadniczo wyniki sprawdzianu nie służą do rekrutacji do gimnazjum, które jest szkołą obowiązkową i rejonową. Jednak dla wielu uczniów w dużych miastach ubiegających się o przyjęcie do szkoły spoza rejonu wyniki sprawdzianu stały się, obok wyników oceniania szkolnego, kryterium rekrutacyjnym.

Po raz pierwszy sprawdzian przeprowadzono w 2002 roku. Zasady i tryb przeprowadzania sprawdzianu do 2014 roku określało rozporządzenie Ministra Edukacji Narodowej z 2001 roku (MEN, 2001a). Począwszy od sesji wiosennej 2015 roku sprawdzian odbywa się w nowej formule i składa się z dwóch części: pierwszej obejmującej zagadnienia z zakresu języka polskiego i matematyki oraz części drugiej – z zakresu języka obcego nowożytnego.

Do 2014 roku na sprawdzianie badany był poziom opanowania umiejętności określonych w *Standardach wymagań egzaminacyjnych będących podstawą przeprowadzania sprawdzianu w ostatnim roku nauki w szkole podstawowej* (MEN, 1999b). Ponieważ standardy wymagań egzaminacyjnych zostały opracowane na podstawie obowiązującej w tych latach podstawy programowej, to zadania zawarte w zestawach egzaminacyjnych sprawdzianu nie wykraczały poza treści zawarte w *Podstawie programowej kształcenia ogólnego* (MEN, 1999a).

Standardy wymagań egzaminacyjnych obejmowały umiejętności z następujących obszarów:

- czytanie,
- pisanie,
- rozumowanie,
- korzystanie z informacji,
- wykorzystanie wiedzy w praktyce.

Sprawdzian miał charakter ponadprzedmiotowy, co oznaczało, że np. czytanie obejmowało nie tylko umiejętności odczytywania tekstów literackich, ale także wykresów, map itp., a zadania były tak konstruowane, aby dawały możliwość sprawdzania umiejętności kształconych w obrębie różnych przedmiotów. Twórcy koncepcji sprawdzianu zakładali, że odejście od ściśle przedmiotowego podziału materiału przeniesie akcent w procesie dydaktycznym z materiału nauczania na kształcenie ogólnych umiejętności uczniów na różnych treściach i w rozmaity sposób. Pierwotnie zakładano, że elementem integrującym zadania sprawdzianu będzie motyw przewodni. Od tej zasady odstąpiono dopiero w 2010 roku. Tabela 3.1 pokazuje, jakie motywy przewodnie zastosowano w arkuszach standardowych⁵ w latach 2002-2009.

Tabela 3.1. Motyw przewodni sprawdzianu w latach 2002–2009

Rok sprawdzianu	Tytuł arkusza standardowego
2002	Pory roku
2003	Przed telewizorem
2004	Chleb
2005	W wodzie
2006	Pszczółki i miody
2007	W szkole
2008	Jasne jak słońce
2009	O zwierzętach

Za rozwiązanie zadań na sprawdzianie do 2014 roku uczeń włącznie mógł uzyskać maksymalnie 40 punktów. Zestaw zadań zawierał 20 zadań zamkniętych wielokrotnego wyboru – uczeń wybierał jedną poprawną odpowiedź z czterech możliwości. Zadania otwarte to zadania krótkiej i rozszerzonej

⁵ Arkusze egzaminacyjne dla uczniów bez niepełnosprawności i uczniów ze specyficznymi trudnościami w uczeniu się.

odpowiedzi. Umiejętność pisania była sprawdzana poprzez dłuższą wypowiedź, np. opowiadanie na zadany temat. Sprawdzian trwał 60 minut.

Przydział punktów za poszczególne grupy umiejętności pokazuje Tabela 3.2. Początkowo 12 punktów w teście (30% możliwych do uzyskania punktów) przeznaczono na pisanie i tylko 2 punkty (5% możliwych do uzyskania punktów) na sprawdzenie umiejętności korzystania z informacji. W 2005 roku zmieniono te proporcje: na zadania z pisania i czytania przeznaczano po 10 punktów (po 25% możliwych do uzyskania punktów), a 4 punkty na zadania sprawdzające umiejętność korzystania z informacji. Za zadania sprawdzające umiejętności rozumowania i wykorzystania wiedzy w praktyce uczeń mógł uzyskać 8 punktów (20% możliwych).

Tabela 3.2. Przydział punktów dla poszczególnych standardów wymagań egzaminacyjnych na sprawdzianie

Umiejętności	Lata 2002–2004		Lata 2005–2014	
	maksymalna liczba punktów	udział procentowy	maksymalna liczba punktów	udział procentowy
Czytanie	10	25%	10	25%
Pisanie	12	30%	10	25%
Rozumowanie	8	20%	8	20%
Korzystanie z informacji	2	5%	4	10%
Wykorzystywanie wiedzy w praktyce	8	20%	8	20%

Jak już wcześniej wspomniano, do 2014 roku podstawę do przeprowadzenia sprawdzianu stanowiły standardy wymagań egzaminacyjnych, które nawiązywały do *Podstawy programowej kształcenia ogólnego dla pierwszego i drugiego etapu edukacyjnego* (MEN, 2008a). Były one równocześnie układem odniesienia do zapewnienia trafności sprawdzianu. W latach 2002–2010 przygotowanie arkuszy egzaminacyjnych (w tym propozycji zadań, kluczy do zadań zamkniętych, kryteriów oceny zadań otwartych, modeli odpowiedzi/schematów punktowania, recenzji) odbywało się w okręgowych komisjach egzaminacyjnych według jednolitych zasad i procedur tworzenia arkuszy sprawdzianu ustalonych przez Centralną Komisję Egzaminacyjną. W okręgowych komisjach egzaminacyjnych za przygotowanie arkusza począwszy od planu i kartoteki sprawdzianu, poprzez przygotowanie autorskich zadań, próbne zastosowanie wiązek zadań, recenzje nauczycielskie i akademickie aż do składu gotowego do przekazania do CKE arkusza egzaminacyjnego odpowiedzialny był koordynator sprawdzianu. Koordynatorem najczęściej był kierownik pracowni sprawdzianu w danej okręgowej komisji egzaminacyjnej.

Finalny arkusz egzaminacyjny rekomendowany przez recenzentów i koordynatora sprawdzianu był zatwierdzany przez dyrektora OKE. Zatwierdzenie arkusza oznaczało przyjęcie odpowiedzialności przez dyrektora OKE za jego poprawność merytoryczną, konstrukcyjną i edytorską. Centralna Komisja Egzaminacyjna podejmowała decyzję o wyborze arkuszy na poszczególne sesje egzaminacyjne i poddawała je kolejnym recenzjom oraz dalszemu doskonaleniu przed skierowaniem do druku. W wyniku tych działań ostateczną odpowiedzialność za jakość arkusza egzaminacyjnego przejmowała CKE.

W latach od 2010 do 2013 roku arkusze przygotowywane były przez ogólnopolski zespół złożony z przedstawicieli poszczególnych okręgowych komisji, którego prace koordynował dyrektor OKE we Wrocławiu.

Oprócz arkusza standardowego S-1 na każdą edycję sprawdzianu przygotowywane są dostosowane arkusze egzaminacyjne dla uczniów o specjalnych potrzebach edukacyjnych: dla uczniów słabo widzących – arkusze S-4 i S-5 ze zwiększoną czcionką, odpowiednio 16 i 24 pkt., arkusze S-6 dla uczniów niewidomych, przygotowywane w piśmie Braille'a, arkusze S-7 dla uczniów słabo słyszących i niesłyszących. Odrębny zestaw zadań jest przygotowywany dla uczniów z upośledzeniem w stopniu lekkim (arkusz S-8). Uczniowie o specjalnych potrzebach edukacyjnych mogą mieć wydłużony czas pracy o 30 minut. Dla uczniów o specjalnych potrzebach edukacyjnych dostosowano także standardy wymagań egzaminacyjnych. Każdego roku sprawdzian z wykorzystaniem arkuszy dostosowanych pisało około 2 procent zdających.

Uczniowie ze specyficznymi trudnościami w uczeniu się (z dysleksją rozwojową) powinni spełniać określone w standardach (do 2014 roku, a w podstawie programowej od 2015 roku), jednolite dla wszystkich uczniów wymagania. Dlatego podczas sprawdzianu i innych egzaminów otrzymują arkusze o takiej samej treści, jak uczniowie bez dysleksji. Jednak ze względu na trudności w pisaniu tej grupy uczniów zmodyfikowano ogólne kryteria oceniania, w szczególności w zakresie stosowania zasad ortografii i interpunkcji. Z założenia dostosowanie kryteriów oceniania ma charakter kompensacyjny i powinno dać szansę tym uczniom na osiągnięcie porównywalnych wyników z uczniami bez dysleksji posiadających taki sam poziom umiejętności. Jak już wspomniano wcześniej uczniowie z diagnozą dysleksji rozwojowej mogą skorzystać z wydłużonego czasu pracy o 30 minut.

Tabela 3.3 podaje wybrane własności psychometryczne⁶ sprawdzianów z lat 2002–2014. Zamieszczony w tabelach Tabela 3.3, Tabela 3.8, Tabela 3.9, Tabela 3.11 i Tabela 3.13 opis własności psychometrycznych egzaminów włączonych do studium zrównującego odnosi się do populacji, które stanowią uczniowie szkół dla młodzieży umożliwiających zdawanie egzaminu maturalnego, z wyłączeniem szkół specjalnych i przyszpitalnych, piszący egzamin z zastosowaniem arkusza standardowego dla danego egzaminu.

⁶ Sposoby uzyskania informacji o obliczonych w trakcie badań parametrach psychometrycznych zadań egzaminacyjnych, upublicznianych za pośrednictwem bazy danych IBE, opisuje szczegółowo Aneks 5.

Tabela 3.3. Wybrane własności psychometryczne sprawdzianów z lat 2002–2014

Rok egzaminu	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Wynik średni w skali utworzonej z sumy punktów	29,49	28,61	25,55	29,50	25,32	26,57	25,77	22,64	24,56	25,27	22,75	24,03	25,83
Wynik średni wyrażony w % maks. l. punktów	74	72	64	74	63	66	64	57	61	63	57	60	65
Mediana	31	30	26	31	26	28	27	23	25	26	23	24	27
Modalna	35	33	25	36	33	33	30	24	30	30	23	32	30
Odchylenie standardowe	6,83	6,73	7,83	7,43	8,56	7,82	7,52	7,63	8,03	7,50	7,63	8,38	8,00
Wariancja	46,66	45,32	61,24	55,16	73,33	61,19	56,62	58,18	64,51	56,32	58,23	70,23	64,07
Skośność rozkładu	-0,66	-0,62	-0,17	-0,61	-0,24	-0,55	-0,49	-0,14	-0,16	-0,29	-0,10	0,01	-0,44
Zmienność sd/wynik _{sr}	0,23	0,24	0,31	0,25	0,34	0,29	0,29	0,34	0,33	0,30	0,34	0,35	0,31
Kwartyle	Q ₁	25	24	20	25	19	21	21	17	18	20	17	20
	Q ₂	31	30	26	31	26	28	27	23	25	26	23	27
	Q ₃	35	34	32	35	32	33	32	28	31	31	29	32
Rit (średnia korelacja zadania z testem)	0,44	0,42	0,45	0,42	0,46	0,44	0,43	0,42	0,44	0,43	0,44	0,46	0,45
Liczba zadań	25	25	25	26	25	26	25	25	25	26	26	26	26
Alfa Cronbacha	0,77	0,77	0,79	0,78	0,79	0,82	0,79	0,77	0,79	0,80	0,81	0,83	0,81
Alfa Feldt-Raju	0,84	0,82	0,85	0,85	0,86	0,86	0,85	0,83	0,86	0,84	0,84	0,86	0,86
Współczynnik rzetelności liczony w oparciu o oszacowania IRT	0,85	0,85	0,89	0,84	0,88	0,87	0,86	0,87	0,87	0,87	0,86	0,87	—

Różnice w średnim wyniku surowym ze sprawdzianu (średnia liczba punktów za wszystkie zadania w teście), sięgające nawet wartości jednego odchylenia standardowego, są spowodowane głównie różnicami w trudności arkuszy egzaminacyjnych w poszczególnych latach (zob. podrozdział 6.7). Przez wszystkie lata w analizowanym okresie rzetelność szacowana za pomocą alfa Cronbacha⁷ zmieniała się od 0,77 (lata 2002, 2003, 2009) do 0,83 (rok 2013). Średnia wartość alfa Cronbacha w ciągu 13 lat wynosi 0,79 przy odchyleniu standardowym 0,02. We wszystkich latach rzetelność określoną na podstawie alfa Cronbacha można uznać za przeciętną, co oznacza, że spójność testu jest do zaakceptowania. Warto jednak pamiętać, że alfa Cronbacha jest wskaźnikiem dla dolnej granicy rzetelności, który dostarcza niedoszacowanej wartości, szczególnie dla testów z zadaniami punktowanymi na skali o różnej długości. Lepszego oszacowania dostarcza współczynnik alfa Feldt-Raju, który uwzględnia wagi poszczególnych zadań. Tabela 3.3 podaje dla porównania wartości obydwu wskaźników. Dla alfa Feldt-Raju średnia wartość wskaźnika dla sprawdzianu wynosi 0,85 przy odchyleniu standardowym 0,01.

Dla wszystkich egzaminów włączonych do badań podano także współczynnik rzetelności liczony z wykorzystaniem teorii odpowiedzi na pozycje testowe (*Item Response Theory*, IRT). Korzystając z modelowania IRT uzyskujemy informację o warunkowym błędzie pomiaru, tj. o wielkości wariancji błędu pomiaru w zależności od poziomu umiejętności uczniów. Natomiast klasyczny współczynnik rzetelności jest pojedynczą dla całego testu miarą, w której wariancja błędu pomiaru jest uwzględniana jako średnia wartość dla całej populacji. Możliwe jest jednak wykorzystanie informacji o warunkowej wariancji błędów IRT, do obliczenia wskaźników analogicznych do klasycznych miar rzetelności dla testu (Adams, 2005; Wu, 2005). Jednym ze sposobów, aby osiągnąć ten cel, jest zestawienie wariancji oszacowań umiejętności EAP z wariancją umiejętności oszacowaną w modelu. W prezentowanych badaniach niektóre z pojedynczych kryteriów oceny przed dopasowaniem modelu IRT zostały zsumowane, tak aby uniknąć problemu z nadmierną korelacją wiązek kryteriów oceny pochodzących z tego samego zadania lub odnoszących się np. do tego samego tekstu źródłowego. Takie sumowanie zostało poprzedzone przeprowadzeniem analizy skupień na kryteriach egzaminacyjnych oraz przesłankami treściowymi. Zdarzyło się w przypadku niektórych egzaminów, że pojedyncze zadania egzaminacyjne zostały wykluczone z modelu IRT ze względu na niezadowalające dopasowanie. Takie wykluczane zadania były jednocześnie zadaniami o słabych własnościach psychometrycznych, w szczególności bardzo niskiej (bądź nawet ujemnej) korelacji z resztą testu. Jest to istotne przy zestawieniu z klasycznymi miarami rzetelności, dla których takich zadań nie wykluczano, a ich uwzględnienie może zaniżać oszacowane wartości współczynnika rzetelności. Oszacowany w ten sposób współczynnik rzetelności dla sprawdzianu w latach, dla których przeprowadzono zrównanie wyników jest wyższy (z wyjątkiem roku 2005) od alfa Cronbacha i alfa Feldt-Raju. Średni współczynnik rzetelności liczony z wykorzystaniem IRT wynosi 0,86 przy odchyleniu standardowym 0,01.

W 2015 roku sprawdzian po raz pierwszy pisali uczniowie, dla których proces dydaktyczny bazował już na nowej, zmienionej podstawie programowej (MEN, 2008a). Standardy wymagań egzaminacyjnych przestały obowiązywać, a arkusze do sprawdzianu przygotowywane były zgodnie z wymaganiami opisanymi w podstawie programowej dla pierwszego i drugiego etapu edukacyjnego. Zmieniła się też formuła sprawdzianu. Od sesji kwietniowej 2015 roku sprawdzian składa się z dwóch części. Część 1. obejmuje zadania z języka polskiego oraz z matematyki (dwa rozdzielne testy w jednym zeszycie testowym) i trwa 80 minut. Zadania z języka polskiego i matematyki mogą być oparte na tekstach lub informacjach z zakresu historii i przyrody. Część 2. to zadania z języka obcego nowożytnego, na których rozwiązanie uczeń ma 45 minut. W drugiej części uczeń może wybrać język (angielski, francuski,

⁷ Obliczana dla sprawdzianu i pozostałych egzaminów w przypadku zadań rozszerzonej odpowiedzi z uwzględnieniem podziału tych zadań na części odpowiadające kryteriom oceniania.

hiszpański, niemiecki, rosyjski lub włoski), w zależności od tego, którego uczył się w szkole podstawowej jako obowiązkowego. Liczbę zadań o poszczególnych formatach i czas trwania dla poszczególnych części ilustruje Tabela 3.4. Wyniki będą komunikowane w procentach wyniku maksymalnego dla obydwu części oraz dodatkowo w podziale na język polski i matematykę w części pierwszej.

Tabela 3.4. Liczba zadań określonego formatu i czas trwania dla poszczególnych części sprawdzianu od 2015 roku

	Część 1		Część 2
	Język polski	Matematyka	Język obcy nowożytny
Liczba zadań zamkniętych	8–12	8–12	35–45
Liczba zadań otwartych	2–4	2–4	
Czas	80 minut		45 minut

3.2. Egzamin gimnazjalny

Egzamin gimnazjalny jest obowiązkowo zdawany na zakończenie nauki w gimnazjum. Do sesji 2009 roku składał się z dwóch części – humanistycznej i matematyczno-przyrodniczej. W 2009 roku został rozszerzony o język obcy nowożytny. Za każdą z części uczeń mógł otrzymać maksymalnie 50 punktów. Każdą z części egzaminu uczniowie zdawali innego dnia, przy czym w pierwszym dniu była to część humanistyczna, w dniu drugim – matematyczno-przyrodnicza i w dniu trzecim – język obcy nowożytny. Egzamin zwykle odbywa się w kwietniu, zaś jego wyniki ogłaszane w czerwcu są wykorzystywane w procesie rekrutacji do szkół ponadgimnazjalnych.

Zawartość merytoryczna egzaminu nie zmieniała się w latach 2002–2011. Istotne zmiany zostały wprowadzone w 2012 roku, kiedy do egzaminu przystąpili uczniowie edukowani zgodnie z nową podstawą programową (MEN, 2008a). Do roku 2011 włącznie w części humanistycznej sprawdzane były umiejętności w obszarach (CKE, 2010a):

- czytania i odbioru tekstów kultury,
- tworzenia własnego tekstu.

Tabela 3.5. Przydział punktów dla poszczególnych standardów wymagań egzaminacyjnych dla egzaminu gimnazjalnego w części humanistycznej

Umiejętności	Lata 2002–2011	
	maksymalna liczba punktów	udział procentowy
Czytanie i odbiór tekstów kultury	25	50%
Tworzenie własnego tekstu	25	50%

Tabela 3.5 wskazuje na przydział punktów do poszczególnych standardów wymagań dla części humanistycznej. Arkusze egzaminacyjne zawierały 20 zadań zamkniętych i około 10 zadań otwartych krótkiej i rozszerzonej odpowiedzi. Ostatnie z zadań, dotyczące umiejętności tworzenia własnego

tekstu, najczęściej wymagało przygotowania rozprawki i można było za nie otrzymać maksymalnie kilkanaście punktów.

W części matematyczno-przyrodniczej (przydział punktów do poszczególnych standardów obrazuje Tabela 3.6) sprawdzane były umiejętności w następujących obszarach:

- stosowanie terminów, pojęć i procedur z zakresu przedmiotów matematyczno-przyrodniczych niezbędnych w praktyce życiowej i dalszym kształceniu,
- wyszukiwanie i stosowanie informacji,
- wskazywanie i opisywanie faktów, związków i zależności, w szczególności przyczynowo-skutkowych, funkcjonalnych, przestrzennych i czasowych.
- Stosowanie zintegrowanej wiedzy i umiejętności do rozwiązywania problemów.

Tabela 3.6. Przydział punktów dla poszczególnych standardów wymagań egzaminacyjnych dla egzaminu gimnazjalnego w części matematyczno-przyrodniczej

Umiejętności	Lata 2002–2011	
	maksymalna liczba punktów	udział procentowy
Stosowanie terminów, pojęć i procedur...	15	30%
Wyszukiwanie i stosowanie informacji	12	24%
Wskazywanie i opisywanie faktów, związków i zależności...	15	30%
Stosowanie zintegrowanej wiedzy...	8	16%

Każda z części zawierała zadania zamknięte (punktowane 0–1) i otwarte. Wśród zadań otwartych znajdowały się zadania krótkiej wypowiedzi i zadania rozszerzonej wypowiedzi.

W obowiązującej od 2009 roku części dotyczącej języka obcego nowożytnego sprawdzane są umiejętności w obszarach:

- odbioru tekstu słuchanego,
- odbioru tekstu pisanego,
- redagowania językowego.

Począwszy od roku 2012 zmieniła się formuła egzaminu gimnazjalnego (CKE, 2010a). Konstrukcja, na którym bazuje egzamin, od 2012 roku wyznaczany jest bezpośrednio przez wymagania ogólne i szczegółowe zdefiniowane w podstawie programowej (MEN, 2008a), a nie jak uprzednio przez standardy wymagań egzaminacyjnych. Zmianie uległa też struktura egzaminu gimnazjalnego oraz rozwiązania dotyczące budowy arkuszy egzaminacyjnych. Na egzaminie gimnazjalnym sprawdzane są umiejętności określone w podstawie programowej dla III etapu edukacyjnego, ale mogą też występować zadania odnoszące się do drugiego etapu.

Nowy egzamin gimnazjalny obejmuje w części humanistycznej:

1. Test z języka polskiego – pisany przez 90 minut.
2. Test z historii i wiedzy o społeczeństwie – pisany przez 60 minut.

Test z historii i WOS zbudowany jest tylko z zadań zamkniętych, natomiast test z języka polskiego oprócz zadań zamkniętych zawiera także zadania otwarte, wśród których jest dłuższa wypowiedź pisemna.

W części matematyczno-przyrodniczej uczeń rozwiązuje:

1. Test z matematyki – 90 minut.
2. Test z przedmiotów przyrodniczych – 60 minut.

Test w zakresie przyrodniczym zawiera zadania zamknięte sprawdzające umiejętności szczegółowe i ogólne określone w podstawie programowej dla takich przedmiotów jak biologia, chemia, fizyka, geografia. Test z matematyki zbudowany jest zarówno z zadań zamkniętych, jak i otwartych.

W części egzaminu z języka obcego nowożytnego gimnazjalista zdaje jeden z siedmiu języków: angielski, francuski, hiszpański, niemiecki, rosyjski, włoski lub ukraiński (od roku szkolnego 2013/2014). Do egzaminu gimnazjalista może wybrać tylko ten język, którego uczył się w gimnazjum jako przedmiotu obowiązkowego. Arkusze egzaminacyjne do języka obcego nowożytnego obejmują:

1. Test z języka obcego na poziomie podstawowym, zdawany przez każdego gimnazjalistę, który złożony jest z zadań zamkniętych i pisany przez 60 minut.
2. Test z języka obcego na poziomie rozszerzonym złożony z zadań zamkniętych i otwartych – pisany przez 60 minut.

Test z języka obcego na poziomie rozszerzonym zdają obowiązkowo uczniowie, którzy w gimnazjum kontynuowali naukę języka obcego rozpoczętą w szkole podstawowej. Mogą do niej przystąpić opcjonalnie także pozostali gimnazjaliści. Uczniowie na obu poziomach trzeciej części egzaminu rozwiązują zadania z tego samego języka.

Wybrane własności psychometryczne egzaminu gimnazjalnego prezentuje Tabela 3.7 (część humanistyczna) oraz Tabela 3.8 (część matematyczno-przyrodnicza).

Tabela 3.7. Wybrane własności psychometryczne egzaminu gimnazjalnego w części humanistycznej z lat 2002–2014

Rok egzaminu	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Wynik średni w skali utworzonej z sumy punktów	30,21	31,91	27,10	33,28	31,50	31,62	30,94	31,83	30,51	25,49	41,13	39,05	41,29
Wynik średni wyrażony w % maks. l. punktów	60	64	54	67	63	63	62	64	61	51	63	60	64
Mediana	31	33	27	35	32	33	32	33	31	25	42	39	42
Modalna	34	38	29	39	35	37	38	36	34	27	46	39	46
Odchylenie standardowe	8,79	8,81	9,17	8,63	8,31	9,68	9,74	8,59	8,27	9,23	11,51	10,74	9,93
Wariancja	77,32	77,70	84,17	74,47	69,11	93,68	94,96	73,76	68,33	85,14	132,38	115,33	98,69
Skośność rozkładu	-0,27	-0,37	0,03	-0,60	-0,18	-0,43	-0,33	-0,41	-0,18	0,16	-0,23	0,01	-0,21
Zmienność $sd/wynik_{\bar{s}r}$	0,29	0,28	0,34	0,26	0,26	0,31	0,31	0,27	0,27	0,36	0,28	0,27	0,24
Kwartyle	Q ₁	24	26	20	28	26	25	24	26	25	19	33	35
	Q ₂	31	33	27	35	32	33	32	33	31	25	42	42
	Q ₃	37	39	34	40	38	39	39	38	37	32	50	49
Rit (średnia korelacja zadania z testem)	0,34	0,37	0,40	0,44	0,38	0,46	0,40	0,37	0,40	0,41	0,41	0,38	0,37
Liczba zadań	29	31	31	33	26	30	32	29	29	29	46	46	46
Alfa Cronbacha	0,62	0,68	0,75	0,80	0,70	0,78	0,70	0,72	0,76	0,79	0,88	0,85	0,84
Alfa Feldt-Raju	0,83	0,85	0,86	0,88	0,84	0,89	0,86	0,86	0,84	0,86	0,90	0,88	0,87
Współczynnik rzetelności liczony w oparciu o oszacowania IRT	0,90	0,89	0,89	0,88	0,87	0,90	0,90	0,89	0,87	0,88	0,92	0,91	–

Tabela 3.8. Wybrane własności psychometryczne egzaminu gimnazjalnego w części matematyczno-przyrodniczej z lat 2002–2014

Rok egzaminu	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Wynik średni w skali utworzonej z sumy punktów	28,2	25,84	24,58	24,35	24,00	25,41	27,24	26,18	24,05	23,79	27,43	30,72	28,10
Wynik średni wyrażony w % maks. l. punktów	56	52	49	49	48	51	54	52	48	48	49	54	50
Mediana	28	25	23	23	23	24	26	25	23	23	25	29	27
Modalna	25	17	18	19	19	18	21	19	19	18	20	24	20
Odchylenie standardowe	8,87	10,87	11,02	10,12	10,27	10,20	10,61	10,97	9,59	9,33	10,80	10,90	10,94
Wariancja	78,76	118,23	121,34	102,38	105,41	104,06	112,53	120,39	91,95	86,96	116,65	118,76	119,78
Skośność rozkładu	0,07	0,23	0,43	0,40	0,29	0,42	0,35	0,32	0,33	0,25	0,67	0,47	0,30
Zmienność $sd/wynik_{\text{sr}}$	0,31	0,42	0,45	0,42	0,43	0,40	0,39	0,42	0,40	0,39	0,39	0,35	0,39
Kwartyle	Q ₁	22	17	16	16	16	17	19	17	17	16	19	19
	Q ₂	28	25	23	23	23	24	26	25	23	23	25	27
	Q ₃	35	35	33	32	31	33	36	35	31	30	35	36
Rit (średnia korelacja zadania z testem)	0,41	0,44	0,45	0,45	0,44	0,44	0,44	0,46	0,39	0,43	0,39	0,41	0,42
Liczba zadań	36	34	34	35	34	34	33	36	36	36	47	47	47
Alfa Cronbacha	0,86	0,87	0,88	0,88	0,88	0,88	0,86	0,90	0,87	0,88	0,88	0,89	0,90
Aalfa Feldt-Raju	0,88	0,90	0,91	0,90	0,91	0,90	0,90	0,92	0,89	0,89	0,90	0,90	0,90
Współczynnik rzetelności liczony w oparciu o oszacowania IRT	0,89	0,90	0,91	0,91	0,92	0,91	0,91	0,92	0,92	0,90	0,90	0,91	–

Dla egzaminu gimnazjalnego w części humanistycznej średnia wartość alfa Cronbacha w ciągu 13 lat wynosi 0,76 przy odchyleniu standardowym 0,07. W pierwszych dwóch latach funkcjonowania egzaminu gimnazjalnego rzetelność była niska. Alfa Cronbacha dla tych lat wynosiła odpowiednio 0,62 i 0,68. W pozostałych latach rzetelność określona na podstawie alfa Cronbacha można uznać za przeciętną, co oznacza, że spójność testu jest do zaakceptowania. Najwyższą rzetelność osiągnięto w trzech ostatnich latach, w których odbywał się egzamin już w nowej formule z rozdzieleniem części z języka polskiego od historii i WOS. Wyższa rzetelność szacowana wskaźnikiem Alfa Cronbacha w latach 2012–2014 (odpowiednio 0,88, 0,85 i 0,84) uzyskana została prawdopodobnie poprzez wydłużenie testu (składającego się z dwóch części) o 50 procent zadań (z 30 do 45) i poprzez zmniejszenie liczby zadań otwartych. Do roku 2012 część humanistyczna zawierała 20 zadań zamkniętych i około 10 otwartych, za które uczeń uzyskiwał 30 procent punktacji. Po zmianie formuły, dwusekcyjny egzamin gimnazjalny w części humanistycznej zawiera tylko dwa zadania otwarte (jedno krótkiej i jedno rozszerzonej odpowiedzi), które obejmują 18 procent całkowitej punktacji za egzamin.

Średnia wartość alfa Feldt-Raju w ciągu 13 lat wynosi 0,86 przy odchyleniu standardowym 0,01 i w żadnym roku nie wykazuje wartości poniżej 0,80 co pozwala ją uznać za akceptowalną. Średni współczynnik rzetelności liczony z wykorzystaniem IRT w ciągu 12 lat (2002–2013) wynosi 0,91 przy odchyleniu standardowym 0,01.

Dla egzaminu gimnazjalnego w części matematyczno-przyrodniczej średnia wartość alfa Cronbacha w ciągu 13 lat wynosi 0,88 przy odchyleniu standardowym 0,01. W pierwszym roku funkcjonowania egzaminu gimnazjalnego rzetelność była najniższa (0,86). We wszystkich latach określana na podstawie alfa Cronbacha rzetelność mieściła się w górnej granicy przedziału pozwalającego uznać ją za przeciętną, co oznacza, że spójność testu jest do zaakceptowania. W roku 2009 i 2014 rzetelność szacowana w ten sposób była wysoka. Natomiast rzetelność oszacowana za pomocą alfa Feldt-Raju może być uznana za wysoką oprócz pierwszego roku i roku 2010. Średnia wartość alfa Feldt-Raju w ciągu 13 lat wynosi 0,90 przy odchyleniu standardowym 0,01. Podobnie średni współczynnik rzetelności liczony z wykorzystaniem IRT wynosi 0,91 przy odchyleniu standardowym 0,01. Pomimo, że dla części matematyczno-przyrodniczej egzaminu gimnazjalnego w nowej formule obserwujemy podobny przyrost długości arkusza (zwiększenie liczby zadań), jak dla części humanistycznej i zmniejszenie liczby zadań otwartych do trzech (18 procent punktów), to nie obserwujemy wzrostu rzetelności tego egzaminu.

3.3. Egzamin maturalny

Jak już wspomniano na początku rozdziału, wprowadzenie tzw. „nowej matury” przygotowywane było w projekcie o takiej samej nazwie – „Nowa Matura”, poczynsz od 1995 roku. W ramach projektu wypracowano rozwiązania metodologiczne i organizacyjne zewnętrznej matury oraz przygotowywana została kadra dla mających powstać komisji egzaminacyjnych.

Wprowadzany od 2002 roku nowy egzamin maturalny obejmował część wewnętrzną zdawaną i ocenianą w szkole i część zewnętrzną ocenianą przez przygotowanych egzaminatorów wpisanych do rejestru egzaminatorów Centralnej Komisji Egzaminacyjnej. Egzamin wewnętrzny zdawany był ustnie z języka polskiego i języka obcego nowożytnego. Absolwent przystępujący do matury powinien zdawać obowiązkowo egzaminy pisemne z czterech przedmiotów – egzamin z języka polskiego, języka obcego nowożytnego i matematyki oraz jeden egzamin do wyboru spośród listy obejmującej biologię, chemię, fizykę, geografię, historię, drugi język obcy nowożytny, język obcy klasyczny. Można było też zdawać egzamin dodatkowy spośród znajdujących się na liście, ale inny niż już przedmiot wybrany (MEN, 1999c). W pierwotnych założeniach warunkiem zdania matury było uzyskanie co najmniej 40% punktów możliwych do uzyskania z czterech obowiązkowych egzaminów. Warunki szczegółowe dla każdego egzaminu zawierały sylabusy przedmiotowe publikowane przez Centralną Komisję Egzaminacyjną.

Pomimo pełnego przygotowania przez system arkuszy egzaminacyjnych i warunków organizacji egzaminu nastąpiło jego przesunięcie na rok 2004/2005 w wyniku decyzji Ministra Edukacji Narodowej (MEN, 2001). Jednocześnie w tym samym akcie prawnym poszerzono listę przedmiotów do wyboru o filozofię, historię sztuki, historię muzyki i wiedzę o tańcu i ustalono próg zdawalności przy 30% wyniku maksymalnego. Ponieważ decyzja o zmianie warunków egzaminu została podjęta na kilka miesięcy przed sesją, to uczniom pozostawiano wybór, czy w 2002 roku chcą zdawać nową maturę, czy obowiązujący dotychczas egzamin dojrzałości.

Wprowadzony od 2005 roku zewnętrzny egzamin maturalny podlegał kilku mniejszym lub większym zmianom w zakresie regulaminu zdawania, konstrukcji arkuszy egzaminacyjnych i procedur oceniania zadań otwartych rozszerzonej odpowiedzi. Począwszy od 2007 roku maturzyści, którzy nie uzyskali minimum 30% punktów z jednego egzaminu obowiązkowego, mogą przystępować do egzaminu w sesji poprawkowej. Sesja poprawkowa organizowana jest w ostatnim tygodniu sierpnia.

Matura to bardzo ważny egzamin, którego wyniki mają decydujący wpływ na rekrutację na polskich uczelniach. W odróżnieniu od rekrutacji do szkół ponadgimnazjalnych, gdzie o przyjęcie do szkoły wyższego szczebla starają się absolwenci z danego roku, w przypadku matury rywalizują ze sobą kandydaci, którzy zdawali egzamin w różnych latach (sesjach egzaminacyjnych). Aby możliwym było porównywanie wyników egzaminu kandydatów, którzy zdawali egzaminy maturalne w różnych latach (różnych sesjach egzaminacyjnych), niezbędne jest wprowadzenie takich mechanizmów, które pozwolą na ich zrównanie.

Zrównanie wyników egzaminu maturalnego na poziomie podstawowym w sesji głównej oraz poprawkowej z lat 2010–2013 było częścią czteroletniego projektu zrównywania wyników egzaminów zewnętrznych. Rozpoczęło się w III etapie studium zrównującego w 2013 roku (dotyczyło wtedy tylko matematyki) i było kontynuowane w 2014 (etap IV) obejmując trzy egzaminy maturalne: język polski, język angielski i matematykę. Pomimo, że zewnętrzne egzaminy maturalne wprowadzono od 2005⁸ roku, to do studium badawczego włączone zostały wyniki począwszy od roku 2010, ponieważ jest to pierwszy rok, w którym matematyka została włączona do koszyka obowiązkowych przedmiotów zdawanych na maturze.

3.3.1. Język polski

Egzamin maturalny z języka polskiego składa się z dwóch części – pisemnej i ustnej. Część pisemna zdawana jest na poziomie podstawowym lub rozszerzonym i oceniana jest przez egzaminatorów zewnętrznych pracujących w ramach danej okręgowej komisji egzaminacyjnej. Pisemna część egzaminu składa się z dwóch części, z których pierwsza sprawdza rozumienie czytanego tekstu, a druga umiejętność pisanie w odniesieniu do tekstu literackiego zamieszczonego w arkuszu. Część sprawdzająca rozumienie czytanego tekstu składa się z 10–14 zadań zamkniętych lub otwartych typu krótkiej odpowiedzi dotyczących przytoczonego tekstu. W części sprawdzającej umiejętność pisanie własnego tekstu zdający ma do wyboru jeden z dwóch zaprezentowanych tematów. Na poziomie podstawowym za część sprawdzającą rozumienie czytanego tekstu zdający maksymalnie może uzyskać 20 punktów, natomiast w części sprawdzającej umiejętność pisanie 50 punktów.

⁸ Pierwszy zewnętrzny egzamin maturalny z obowiązkową matematyką odbył się w 2002 roku (tylko dla ochotników, ponieważ w sesji egzaminacyjnej 2002 pozostawiono uczniom możliwość wyboru starego egzaminu dojrzałości w szkole lub nowej zewnętrznej matury) w rezultacie przeniesienia terminu wdrożenia nowej matury na 2005 rok.

Prace egzaminacyjne oceniane są według takich samych kryteriów w całym kraju. Do sesji wiosennej 2014 roku włącznie na poziomie podstawowym podczas oceniania pracy pisemnej przyjęte były następujące wagi kryteriów:

- rozwinięcie tematu – 25 punktów (50%),
- kompozycja – 5 punktów (10%),
- styl – 5 punktów (10%),
- język – 12 punktów (24%),
- zapis (ortografia i interpunkcja) – 3 punkty (6%).

Za szczególne walory pracy maturzysty mógł otrzymać dodatkowe 4 punkty. Punkty te przyznaje się za pracę ocenioną poniżej maksymalnej liczby punktów, jednak wyróżniającą się oryginalnością w zakresie rozwinięcia tematu, kompozycji, stylu lub języka. Łączna liczba punktów za wypowiedź pisemną nie może jednak przekroczyć maksymalnej liczby punktów do osiągnięcia za tę część egzaminu.

Od 2010 do 2014 roku włącznie pisemna część egzaminu maturalnego z języka polskiego na poziomie rozszerzonym polegała na napisaniu własnego tekstu w odniesieniu do tekstu literackiego zamieszczonego w arkuszu egzaminacyjnym. Podobnie jak w przypadku egzaminu maturalnego na poziomie podstawowym, zdający miał do wyboru jeden z dwóch tematów. Za pracę zdający mógł otrzymać maksymalnie 40 punktów. Zgodnie z przyjętymi przez CKE kryteriami oceny, zdający mógł otrzymać maksymalnie za:

- rozwinięcie tematu – 26 punktów (65%),
- kompozycję – 2 punkty (5%),
- styl – 2 punkty (5%),
- język – 8 punktów (20%),
- zapis (ortografia i interpunkcja) – 2 punkty (5%).

Podobnie jak w przypadku egzaminu na poziomie podstawowym, zdający mógł otrzymać dodatkowe 4 punkty za pracę wyróżniającą się oryginalnością w zakresie rozwinięcia tematu, kompozycji, stylu lub języka (szczególne walory). Podczas egzaminu pisemnego z języka polskiego zdający na rozwiązanie zadań z arkusza z poziomu podstawowego miał do dyspozycji 170 minut, a z arkusza z poziomu rozszerzonego 180 minut.

Część ustna była obowiązkowa dla uczniów zdających zarówno poziom podstawowy, jak i rozszerzony i była oceniana w szkole przez przedmiotowy zespół egzaminacyjny. Obejmowała ona dwie części: wypowiedź zdającego na wybrany temat oraz rozmowę z zespołem oceniającym, dotyczącą tematu prezentacji oraz bibliografii. Dla części pisemnej i ustnej egzaminu maturalnego z języka polskiego obowiązywał taki sam próg zdania egzaminu, jak w przypadku egzaminu z matematyki i języka obcego, czyli 30% maksymalnej liczby punktów.

Tabela 3.9. Wybrane własności psychometryczne egzaminu maturalnego z języka polskiego na poziomie podstawowym z lat 2010–2014

Rok egzaminu		2010	2011	2012	2013	2014
Wynik średni w skali utworzonej z sumy punktów		41,47	38,72	38,58	39,37	36,19
Wynik średni wyrażony w % maks. l. punktów		59	55	55	56	52
Mediana		42	39	39	39	36
Modalna		42	39	39	40	36
Odchylenie standardowe		11,46	11,19	10,29	10,70	11,27
Wariancja		131,42	125,12	105,94	114,56	127,02
Skośność rozkładu		-0,14	-0,08	-0,12	0,10	0,05
Zmienność $sd/wynik_{sr}$		0,28	0,29	0,27	0,27	0,31
Kwartyle	Q₁	34	31	32	32	28
	Q₂	42	39	39	39	36
	Q₃	50	46	46	47	44
Rit (średnia korelacja zadania z testem)		0,32	0,35	0,34	0,33	0,39
Liczba zadań		15	14	15	15	14
Alfa Cronbacha		0,66	0,70	0,70	0,70	0,72
Alfa Feldt-Raju		0,81	0,83	0,83	0,84	0,83
Współczynnik rzetelności liczony w oparciu o oszacowania IRT		0,87	0,87	0,85	0,86	–

Tabela 3.9 prezentuje wybrane własności psychometryczne egzaminów maturalnych z języka polskiego. Jak można zauważyć, wskaźniki rzetelności dla egzaminu maturalnego na poziomie podstawowym, szacowane za pomocą współczynnika alfa Cronbacha, są bardzo niskie. Jak już wcześniej wspomniano, do roku 2014 włącznie egzamin maturalny z języka polskiego na poziomie podstawowym to 14 lub 15 zadań badających rozumienie czytanego tekstu, za które zdający mógł uzyskać maksymalnie 20 punktów (29 procent całej skali) i jedno zadanie badające umiejętność pisania własnego tekstu punktowane kryterialnie od 0 do 50 punktów (71 procent długości skali). Dlatego też alfa Cronbacha jest wskaźnikiem, który dla języka polskiego, nawet w przypadku liczenia z uwzględnieniem podziału zadań rozszerzonej odpowiedzi zgodnie z wyróżnionymi kryteriami oceniania (co zwiększa liczbę zadań), powoduje mocne niedoszacowanie rzetelności tego egzaminu. Lepsze oszacowanie uzyskujemy za pomocą alfa Feldt-Raju uwzględniając wagi poszczególnych zadań. Średnia wartość wskaźnika alfa Feldt-Raju dla egzaminu maturalnego z języka polskiego w latach 2010–2014 wynosi 0,83 przy odchyleniu standardowym 0,01. Podobną wartość uzyskujemy dla współczynnika rzetelności liczonego w oparciu o oszacowania IRT – średnia w latach 2010–2013 to 0,86 przy odchyleniu standardowym 0,01, co pozwala mówić o przeciętnej rzetelności tego egzaminu.

Analizując wyniki egzaminu maturalnego z języka polskiego borykamy się z dwoma problemami. Pierwszy dotyczy braku porównywalności w zakresie trudności egzaminu w zależności od wyboru tematu. Zagadnienie to związane jest z jednej strony z autoselekcją przy wyborze jednego z dwóch

tematów, a z drugiej, ze zróżnicowaniem stosowania kryteriów oceniania dla tematów, szczególnie w kryterium oceny treści, za które zdający maksymalnie może uzyskać 36 procent punktów dla całej części pisemnej egzaminu na poziomie podstawowym. Szerzej na ten temat pisze Artur Pokropek na przykładzie egzaminu z 2010 roku (Pokropek, 2011a). Drugi problem wiąże się z nieporównywalnie wysokim (w porównaniu z egzaminem z matematyki) efektem egzaminatora. Jak wykazały badania prowadzone w IBE, różnica łagodności na poziomie testu pomiędzy przeciętnym a najbardziej surowym egzaminatorem dla języka polskiego wynosi (w zależności od roku i tematu wypowiedzi pisemnej) od 6 do 8,5 punktu procentowego wyniku egzaminu⁹, podczas gdy dla matematyki różnice te przyjmują wartości od 1,7 do 2,5 punktu procentowego wyniku egzaminu (Szaleniec i inni, 2015). Obydwa wymienione problemy nie pozostają bez wpływu na podjęte próby zrównania wyników tego egzaminu w latach 2010–2013. Można się spodziewać, że wprowadzona od 2015 roku zmiana formuły egzaminu maturalnego z języka polskiego częściowo zmniejszy wpływ wskazanych czynników powodując jednocześnie wzrost rzetelności tego egzaminu.

3.3.2. Matematyka

Począwszy od roku 2010 egzamin z matematyki na poziomie podstawowym jest obowiązkowy (MEN 2008b). Zakres egzaminu maturalnego z matematyki (standardy wymagań egzaminacyjnych i opis struktury arkuszy egzaminacyjnych obowiązujących od sesji wiosennej 2010 do 2014 roku) zawiera *Informator o egzaminie maturalnym od 2010 r. – matematyka* (CKE, 2007).

Egzamin maturalny z matematyki zdawanej jako przedmiot obowiązkowy trwa 170 minut i polega na rozwiązaniu zadań zawartych w arkuszu egzaminacyjnym sprawdzających rozumienie pojęć i umiejętność ich zastosowania w życiu codziennym oraz zadań o charakterze problemowym. Zadania egzaminacyjne odnoszą się do zakresu wymagań dla poziomu podstawowego. Egzamin maturalny z matematyki jako przedmiot dodatkowy jest zdawany na poziomie rozszerzonym. Egzamin trwa 180 minut, a zawarte w nim zadania dotyczą głównie rozwiązywania problemów matematycznych.

Do sesji wiosennej 2015, kompetencje, które były przedmiotem pomiaru na egzaminie maturalnym z matematyki, zdefiniowane były poprzez pięć standardów wymagań egzaminacyjnych. Egzamin powinien sprawdzać, czy zdający posiada umiejętności w zakresie (dokładny opis standardów dla poziomu podstawowego i rozszerzonego przedstawia Tabela 3.10):

- wykorzystania i tworzenia informacji (INF),
- wykorzystania i interpretowania reprezentacji (REP),
- modelowania matematycznego (MOD),
- użycia i tworzenia strategii (STR),
- rozumowania i argumentacji (ROZ).

⁹ Wyniku dla całego egzaminu – łącznie zadania krótkiej odpowiedzi i wypracowanie dla j. polskiego oraz zamknięte i otwarte dla matematyki.

Tabela 3.10. Standardy wymagań egzaminacyjnych – matura z matematyki

Standard	Poziom podstawowy	Poziom rozszerzony
INF	Zdający interpretuje tekst matematyczny i formułuje uzyskane wyniki.	Zdający używa języka matematycznego do opisu rozumowania i uzyskanych wyników.
REP	Zdający używa prostych, dobrze znanych obiektów matematycznych.	Zdający rozumie i interpretuje pojęcia matematyczne i operuje obiektami matematycznymi.
MOD	Zdający dobiera model matematyczny do prostej sytuacji.	Zdający buduje model matematyczny danej sytuacji, uwzględniając ograniczenia i zastrzeżenia.
STR	Zdający stosuje strategię, która jasno wynika z treści zadania.	Zdający tworzy strategię rozwiązania problemu.
ROZ	Zdający prowadzi proste rozumowanie, składające się z niewielkiej liczby kroków.	Zdający tworzy łańcuch argumentów i uzasadnia jego poprawność.

Arkusz egzaminacyjny na poziomie podstawowym składa się z trzech grup zadań:

- pierwszej grupy, zawierającej od 20 do 30 zadań zamkniętych wielokrotnego wyboru, punktowanych 0–1,
- drugiej grupy, zawierającej od 5 do 10 zadań otwartych krótkiej odpowiedzi, punktowanych w skali 0–2,
- trzeciej grupy, zawierającej od 3 do 5 zadań otwartych rozszerzonej odpowiedzi, punktowanych w skali 0–4, 0–5 lub 0–6.

Łączna punktacja za rozwiązanie wszystkich zadań na poziomie podstawowym to maksymalnie 50 punktów. Wybrane własności psychometryczne egzaminu na poziomie podstawowym zawiera Tabela 3.11. Jest to egzamin, który cechuje się wysoką rzetelnością niezależnie od metody jej szacowania. Alfa Cronbacha tylko w 2010 i 2011 przyjmuje wartość poniżej 0,9 (0,89). Natomiast zarówno alfa Feldt-Raju (w latach 2010–2014), jak i współczynnik rzetelności liczony w oparciu o oszacowania IRT (w latach 2010–2013) wynoszą powyżej 0,9.

Arkusz egzaminacyjny na poziomie rozszerzonym składa się z kilkunastu zadań otwartych rozszerzonej odpowiedzi, które są punktowane w skali 0–4, albo 0–5, albo 0–6, za które zdający może łącznie uzyskać maksymalnie 50 punktów.

Tabela 3.11. Wybrane własności psychometryczne egzaminu maturalnego z matematyki na poziomie podstawowym z lat 2010–2014

Rok egzaminu	2010	2011	2012	2013	2014
Wynik średni w skali utworzonej z sumy punktów	30,45	25,93	30,81	30,08	24,81
Wynik średni wyrażony w % maks. l. punktów	60,90	51,86	61,62	60,16	49,62
Mediana	31	24	32	30	24
Modalna	30	18	46	48	15
Odchylenie standardowe	11,31	11,82	11,97	12,18	11,95
Wariancja	127,93	139,72	143,34	148,34	142,82
Skośność rozkładu	-0,15	0,49	-0,30	0,02	0,20
Zmienność $sd/wynik_{sr}$	0,37	0,46	0,39	0,40	0,48
Kwartyle	Q ₁	22	17	21	15
	Q ₂	31	24	32	30
	Q ₃	40	35	41	34
Rit (średnia korelacja zadania z testem)	0,49	0,49	0,52	0,51	0,52
Liczba zadań	34	33	34	34	34
Alfa Cronbacha	0,89	0,89	0,90	0,90	0,91
Alfa Feldt-Raju	0,92	0,93	0,93	0,93	0,93
Współczynnik rzetelności liczony w oparciu o oszacowania IRT	0,92	0,93	0,93	0,93	–

3.3.3. Język angielski

Trzecim obowiązkowym przedmiotem na maturze jest język obcy nowożytny. Maturzysta może wybierać jeden z koszyka sześciu języków: język angielski, język niemiecki, język rosyjski, język francuski, język hiszpański, język włoski. Przeszło 80 procent zdających wybiera język angielski i w tym zakresie obserwowany jest trend wzrostowy (zob. Tabela 3.12).

Tabela 3.12. Wybieralność języka angielskiego na egzaminie maturalnym w latach 2010–2014

Rok egzaminu	2010	2011	2012	2013	2014
Procent zdających wybierających język angielski	81,0	83,1	84,1	84,0	87,4

Egzamin z języka angielskiego zdawany jest w części ustnej i pisemnej. W części pisemnej może być zdawany na poziomie podstawowym i rozszerzonym. Od sesji egzaminacyjnej 2012 roku część ustna zdawana jest bez podziału na poziomy. Szczegółowy opis egzaminu zawarty jest w informatorach o egzaminie maturalnym z języka angielskiego (CKE, 2007, 2010b).

Część pisemna egzaminu maturalnego z języka angielskiego na poziomie podstawowym zawiera zadania z następujących obszarów umiejętności:

- rozumienie ze słuchu (30% możliwych do uzyskania punktów),
- rozumienie tekstów pisanych (40% możliwych do uzyskania punktów),
- wypowiedź pisemna (30% możliwych do uzyskania punktów).

Rozumienie ze słuchu sprawdzane jest z wykorzystaniem tekstów odtworzonych z płyty CD o łącznym czasie trwania około 6 minut. Całe nagranie trwa 20 minut i obejmuje 3 teksty (komunikaty instrukcje, rozmowy, wywiady, teksty narracyjne), które są dwukrotnie odczytywane wraz z poleceniami. Zdający odpowiadają na 15 zadań zamkniętych w 3 wiązkach. Rozumienie tekstu czytanego sprawdzane jest za pomocą 3 adaptowanych tekstów (teksty informacyjne, publicystyczne, narracyjne, literackie) załączonych w arkuszu egzaminacyjnym. Tekstom tym towarzyszy 20 zadań zamkniętych zgrupowanych w 3 wiązkach. Zarówno do sprawdzania umiejętności rozumienia ze słuchu jak i rozumienia tekstów pisanych stosowane są zadania typu prawda/fałsz, na dobieranie i wielokrotnego wyboru. Wypowiedź pisemna to 2 teksty użytkowe pisane zgodnie ze wskazówkami dotyczącymi ich treści. Krótszy, punktowany od 0 do 5, to np. notatka, ogłoszenie, zaproszenie, ankieta pocztówka itp. Dłuższy tekst użytkowy, punktowany od 0-10, to np. list prywatny lub formalny. Za cały egzamin, na który przeznaczone jest 120 minut, zdający może maksymalnie uzyskać 50 punktów

Wybrane własności psychometryczne egzaminu maturalnego z języka angielskiego na poziomie podstawowym opisuje Tabela 3.13.

Egzamin maturalny z języka angielskiego na poziomie rozszerzonym zawiera zadania badające umiejętności z następujących obszarów:

- rozumienie ze słuchu (30% możliwych do uzyskania punktów),
- rozumienie tekstów pisanych (18% możliwych do uzyskania punktów),
- rozpoznawanie struktur leksykalno-gramatycznych (6% możliwych do uzyskania punktów),
- stosowanie struktur leksykalno-gramatycznych (10% możliwych do uzyskania punktów),
- wypowiedź pisemna (36% możliwych do uzyskania punktów).

Na poziomie rozszerzonym egzamin składa się z dwóch części. W pierwszej, trwającej 120 minut, zdający rozwiązują zadania otwarte z zakresu stosowania struktur leksykalno-gramatycznych oraz formułują wypowiedź pisemną. W drugiej części, trwającej 70 minut, zdający rozwiązują zadania zamknięte sprawdzające umiejętności rozumienia ze słuchu, rozumienia tekstów pisanych i rozpoznawania struktur leksykalno-gramatycznych.

Tabela 3.13. Wybrane własności psychometryczne egzaminu maturalnego z języka angielskiego na poziomie podstawowym z lat 2010–2014

Rok egzaminu		2010	2011	2012	2013	2014
Wynik średni w skali utworzonej z sumy punktów		32,82	36,84	35,72	34,64	35,24
Wynik średni wyrażony w % maks. l. punktów		66	74	71	69	70
Mediana		34,00	40,00	38,50	37,00	38,50
Modalna		49	50	48	49	49
Odchylenie standardowe		12,11	11,35	11,43	11,46	12,20
Wariancja		146,65	128,83	130,65	131,33	148,80
Skośność rozkładu		-0,29	-0,84	-0,73	-0,62	-0,80
Zmienność sd/wynikśr		0,37	0,31	0,32	0,33	0,35
Kwartyle	Q₁	23,00	28,50	27,00	25,50	25,50
	Q₂	34	40	38,5	37	38,5
	Q₃	44	47	46	45	46
Rit (średnia korelacja zadania z testem)		0,78	0,78	0,75	0,78	0,79
Liczba zadań		35	35	35	35	35
Alfa Cronbacha		0,90	0,90	0,89	0,90	0,90
Alfa Feldt-Raju		0,92	0,93	0,93	0,93	0,93
Współczynnik rzetelności liczony w oparciu o oszacowania IRT		0,94	0,93	0,94	0,93	–

Podobnie jak egzamin maturalny z matematyki, egzamin maturalny z języka angielskiego cechują wysokie współczynniki rzetelności niezależnie od metody jej szacowania.

3.3.4. Podsumowanie

Aby łączenie wyników z kolejnych edycji egzaminu (*linking*) mogło być uznane za zrównywanie (*equating*), konieczne jest spełnienie szeregu założeń, które szerzej omówione są w kolejnym rozdziale (zob. Rozdział 4.1). Podsumowując opis egzaminów włączonych do studium zrównującego zwrócimy uwagę na dwa założenia:

1. wymóg tożsamesego konstruktu (*equal construct requirement*): testy mierzące różne konstrukty nie powinny być zrównywane;
2. wymóg równej rzetelności (*equal reliability requirement*): testy mierzące ten sam konstrukt, ale różniące się rzetelnością, nie powinny być zrównywane.

Sprawdzian w latach 2002-2014 przeprowadzany był na bazie tych samych standardów wymagań egzaminacyjnych. Ponadto arkusze egzaminacyjne sprawdzianu z założenia budowane były tak, że za zadania sprawdzające umiejętności z pięciu grup standardów egzaminacyjnych uczeń mógł otrzymać

w kolejnych latach prawie taką samą liczbę punktów (zob. Tabela 3.2). Pewna zmiana nastąpiła począwszy od 2005 roku. Zmniejszył się procent punktów, które zdający mógł uzyskać za zadania sprawdzające umiejętność pisania (z 30% do 25% na korzyść dla umiejętności korzystania z informacji - zmiana z 5% do 10 %). Od 2010 roku zrezygnowano z motywu przewodniego dla arkuszy egzaminacyjnych (zob. Tabela 3.1), co głównie uwolniło autorów od poszukiwania tekstów do wiązek zadań i używania w treści zadań sformułowań związanych z danym motywem przewodnim. Zmiany te jednak nie dotyczyły samego konstruktów i można uznać, że w przypadku sprawdzianu spełniony był wymóg tożsameskiego konstruktów w latach 2002-2014. Także spełniony był wymóg stałej rzetelności w analizowanym okresie. Rzetelność wyrażona zarówno przez alfa Feldt-Raju, jak i współczynnik rzetelności liczony w oparciu o oszacowania IRT (zob. Tabela 3.3) wykazuje stabilność w czasie przy średniej wartości odpowiednio 0,85 i 0,86 przy odchyleniu standardowym równym 0,01 i rozstępie odpowiednio 0,05 i 0,04.

W przypadku egzaminu gimnazjalnego te same standardy wymagań egzaminacyjnych obowiązywały w latach 2002-2011. Jak już wspomniano wcześniej, począwszy od 2012 roku egzamin gimnazjalny zdawali uczniowie kształceni według nowej podstawy programowej. Egzamin konstruowany jest z bezpośrednim odniesieniem do tej podstawy, gdyż zrezygnowano z odrębnego definiowania takich standardów, a wymagania zarówno ogólne jak i szczegółowe zawarte są bezpośrednio w podstawie programowej. Zmieniła się też formuła egzaminu w wyniku rozdzielenia części humanistycznej na język polski oraz historię z WOS, a część matematyczno-przyrodniczą na matematykę i przedmioty przyrodnicze. Pojawiły się też nowe typy zadań. Zmiana podstawy programowej zapewne ma znaczenie dla procesu kształcenia, o czym mogłyby świadczyć wyniki PISA z 2012 roku (por. Dolata, Jakubowski i Pokropek 2013). Czy te zmiany mają wpływ na zmianę konstruktów, na którym bazuje egzamin gimnazjalny? Elżbieta Tyralska-Wojtyca (Tyralska-Wojtyca, 2013) zwraca uwagę, że np. dla przedmiotów przyrodniczych zapisy wymagań (przytoczone w informatorze o egzaminie gimnazjalnym od 2012 roku) w większości zapisane są językiem obowiązujących wcześniej standardów wymagań egzaminacyjnych. Nie zmieniły się też w zasadniczy sposób proporcje punktacji przydzielanej dla grup zadań sprawdzających umiejętności z zakresu poszczególnych przedmiotów obejmowanych przez egzamin.

Podobnie jak dla sprawdzianu także dla egzaminu gimnazjalnego spełniony jest wymóg stałej rzetelności. Dla części humanistycznej średnia wartość alfa Feldt-Raju w ciągu 13 lat, podobnie jak dla sprawdzianu, wynosi 0,86 przy odchyleniu standardowym 0,01 i rozstępie 0,05. Natomiast średni współczynnik rzetelności liczony z wykorzystaniem IRT w ciągu 12 lat (2002–2013) jest znacznie wyższy i wynosi 0,91 przy odchyleniu standardowym 0,01 i rozstępie 0,04 (zob. Tabela 3.7). W części matematyczno-przyrodniczej średnia wartość alfa Feldt-Raju w ciągu 13 lat wynosi 0,90 przy odchyleniu standardowym 0,01 i rozstępie 0,3, natomiast średni współczynnik rzetelności liczony z wykorzystaniem IRT wynosi 0,91 przy odchyleniu standardowym 0,01 i rozstępie 0,2. (zob. Tabela 3.8)

W egzaminie maturalnym objęty zrównywaniem był stosunkowo krótki okres (4 lata), w którym obowiązywała ta sama podstawa programowa i formuła egzaminu. Pierwsi absolwenci uczący się zgodnie z nową formułą egzaminu przystąpili do matury w 2015 roku. Można więc uznać bez zastrzeżeń, że warunek pierwszy (tożsameskiego konstruktów) jest spełniony. Także dla wszystkich trzech egzaminów spełniony jest także warunek stabilności rzetelności (zob. Tabela 3.9, Tabela 3.11, Tabela 3.13). Dla egzaminu z języka polskiego średnia wartość alfa Feldt-Raju w ciągu 4 lat wynosi 0,81 przy odchyleniu standardowym 0,01 i rozstępie 0,3 natomiast średni współczynnik rzetelności liczony z wykorzystaniem IRT wynosi 0,86 przy odchyleniu standardowym 0,01 i rozstępie 0,2. (zob. Tabela 3.9). W przypadku egzaminu z matematyki i języka angielskiego rzetelność w ciągu analizowanych trzech lat była nie tylko stała, ale i wysoka. Dla matematyki średnia alfa Feldt-Raju i średni współczynnik rzetelności liczony z wykorzystaniem IRT wynosi 0,93 przy odchyleniu standardowym poniżej 0,001 i

rozstępie 0,01, natomiast dla języka angielskiego odpowiednio 0,93 i 0,94 przy takich samych miarach rozrzutu jak dla egzaminu z matematyki.

Ze względu na rzetelność włączone do studium badawczego egzaminy można podzielić na dwie grupy. Sprawdzian, egzamin gimnazjalny w części humanistycznej i maturalny z języka polskiego na poziomie podstawowym charakteryzują się przeciętną rzetelnością, natomiast egzamin gimnazjalny w części matematyczno-przyrodniczej, maturalny na poziomie podstawowym z języka angielskiego i matematyki - wysoką i przy bardzo dobrej stabilności.

4. Metodologia

W rozdziale przedstawione zostaną zagadnienia metodologiczne czteroetapowego studium zrównującego. W pierwszym podrozdziale zostaną przybliżone teoretyczne podstawy zrównywania, które warunkowały zarówno organizację zbierania danych empirycznych, jak i przeprowadzone analizy. W kolejnych podrozdziałach przedstawiono definicję populacji i schemat doboru próby oraz wyjaśniono logikę konstrukcji planów zrównywania wyników egzaminacyjnych. W dalszej kolejności wyjaśniono logikę planu nierównoważnych grup z testem kotwiczącym i zasady zrównywania z wykorzystaniem IRT w tym implementację metody łącznej kalibracji modelu IRT do zrównania wyników sprawdzianu w szóstej klasie szkoły podstawowej, egzaminu gimnazjalnego w części humanistycznej i matematyczno-przyrodniczej oraz trzech obowiązkowych egzaminów maturalnych- język polski, język angielski i matematyka. Rozdział zamyka dyskusja dotycząca sposobów horyzontalnego zrównania wyników egzaminów włączonych do studium zrównującego (zarówno metodą ekwicytylową, jak i w oparciu o metodologię IRT) a także rozważania na temat skalowania wertykalnego.

4.1. Teoretyczne podstawy zrównywania

Zrównywanie testów jest łączeniem testów, które ma na celu umożliwienie zamiennego, równoważnego posługiwania się wynikami z tych testów. Wynika z tego, że jest to procedura dotycząca testów mierzących ten sam konstrukt i tworzonych zgodnie z tymi samymi specyfikacjami (*blueprint*). Konieczność zrównywania testów jest konsekwencją faktu praktycznej niemożności stworzenia dwóch testów, które byłyby ściśle równoległe¹⁰ (Holland, Dorans i Petersen, 2007). Cytując Alinę von Davier (2011, s. 1–2.):

„Zrównywanie jest konieczne tylko z tego względu, że standaryzowany pomiar edukacyjny korzysta z wielu form testu, które różnią się trudnością, mimo że są tworzone zgodnie z tymi samymi specyfikacjami (...). Zrównywanie można postrzegać jako procedurę statystycznej kontroli zmiennej zakłócającej (*confounding variable*), za którą przyjmuje się właśnie formę testu. Gdyby proces tworzenia testu był idealny, nie byłoby potrzeby zrównywania.”¹¹

Aby łączenie wyników testowych (*linking*) mogło być uznane za zrównywanie (*equating*), konieczne jest spełnienie szeregu restrykcyjnych założeń. Założenia te (wymogi) sformułowane w zbliżonej formie można znaleźć u wielu autorów (np.: Lord, 1980; Kolen i Brennan, 2004). Poniżej wymienimy je w formie przytoczonej przez Neila Doransa i Paula Hollanda (2000, s. 282–283):

„(a) Wymóg tożsamego konstruktu (*equal construct requirement*): testy mierzące różne konstrukty nie powinny być zrównywane.

¹⁰ Dwa testy są ściśle równoległe (*strictly parallel*), jeżeli każda badana osoba z populacji w obu testach będzie miała taką samą wariancję błędów pomiaru oraz taki sam wynik prawdziwy. Mniej formalnie, dwa testy ściśle równoległe są w zupełności sobie równoważnymi (*perfectly equivalent, interchangeable*) narzędziami pomiarowymi (Gruijter i van der Kamp, 2005).

¹¹ Tłum. własne.

(b) Wymóg równej rzetelności (*equal reliability requirement*): testy mierzące ten sam konstrukt, ale różniące się rzetelnością, nie powinny być zrównywane.

(c) Wymóg symetrii (*symmetry requirement*): funkcja zrównująca wyniki w teście Y z wynikami w teście X powinna być odwrotnością funkcji zrównującej wyniki w teście X z wynikami w teście Y .

(d) Wymóg równoważności (*equity requirement*): nie powinno mieć żadnego znaczenia dla osoby rozwiązującej test, którą z wersji testu rozwiązuje, gdy testy są zrównane.

(e) Wymóg niezmienniczości względem populacji (*population invariance requirement*): wybór (sub)populacji użytej do obliczenia funkcji zrównującej wyniki w testach X oraz Y nie powinien mieć znaczenia, tj. funkcja zrównująca używana do łączenia wyników w testach X oraz Y powinna być *niezmiennicza względem populacji*.

Dla bliższego wytłumaczenia tych pięciu wymogów można się odwołać do komentarza zawartego w artykule *Equating test scores* (Holland i in., 2007). Wymogi tożsamesgo konstruktu (a) oraz równej rzetelności (b) oznaczają, że zrównywane testy powinny być skonstruowane tak, aby były zgodne pod względem treści oraz statystycznych własności. Wymóg symetrii (c) wyklucza możliwość zastosowania metod regresji do zrównywania testów. Wymóg równoważności (d) poniekąd tłumaczy konieczność wymogu tożsamesgo konstruktu (a) – jeżeli testy mierzyłyby różne konstrukty, to osoby mające je rozwiązywać preferowałyby wybór tego testu, w którym, w ich mniemaniu, miałyby szansę uzyskać lepszy wynik. Wymóg niezmienniczości względem populacji (e) można wykorzystać do uzasadnienia wymogów tożsamesgo konstruktu (a) oraz równej rzetelności (b). Jeżeli testy byłyby tworzone zgodnie z różnymi wzorcowymi specyfikacjami (*blueprint*), to funkcja zrównująca wyniki mogłaby się różnić w zależności od wyboru subpopulacji. Przykładowo, łącząc wyniki testu badającego umiejętność rozumowania na podstawie „materiału” niewerbalnego z wynikami testu badającego umiejętność rozumowania na podstawie „materiału” werbalnego, zapewne uzyskano by różne funkcje łączące w zależności od płci. Paul Holland, Neil Dorans i Nancy Petersen (2007) podają również odwołania do krytycznej dyskusji na temat omawianych wymogów dla zrównywania, niemniej zgadzają się, że tworzą one ogólną i intuicyjną „teorię” zrównywania testów.

Pewnego dopowiedzenia wymaga wymóg równoważności (d), gdyż formalnie pojawia się on w dwóch, istotnie różniących się, wersjach (Kolen i Brennan, 2004):

$$1. \quad \forall_{\tau} \mathbb{P}(\text{eq}_Y(X) \leq y | \tau) = \mathbb{P}(Y \leq y | \tau), \quad (4.1)$$

$$2. \quad \forall_{\tau} \mathbb{E}(\text{eq}_Y(X) | \tau) = \mathbb{E}(Y | \tau), \quad (4.2)$$

gdzie $\text{eq}_Y: X \rightarrow Y$ jest funkcją zrównującą test X z Y . Pierwsza wersja równoważności (*equity*) (4.1) stanowi, że dla każdego wyniku prawdziwego τ warunkowy, względem τ , rozkład wyników otrzymywanych w teście Y jest taki sam, jak w zrównanym do niego teście X . Natomiast, druga wersja równoważności (4.2) osłabia wymóg warunkowej równości dwóch rozkładów do warunkowej równości jedynie względem pierwszego momentu zwykłego (czyli wartości średniej) tych rozkładów. W szczególności wersja (4.2) nie wymaga równości między warunkowymi wariancjami, czyli nie wymaga równości warunkowego błędu pomiaru. Pierwsza wersja (4.1), sformułowana przez Frederica Lorda (1980), jest bardzo restrykcyjnym wymogiem, który u Michaela Kolena i Roberta Brennana (2004), spotyka się ze słusznym komentarzem, iż „korzystanie z równoważności Lorda jako kryterium oznacza, że zrównywanie albo nie jest możliwe, albo nie jest potrzebne” (por. van der Linden, 2011, Lord, 1980).

Większość metod zrównywania wyników stawia sobie wprost za cel spełnienie słabszej formy równoważności (*weak equity*, *first-order equity*).

Pogłębioną refleksję w kwestii problemu równoważności w „silnym” sformułowaniu Lorda (4.1) możemy znaleźć u Wima van der Lindena (2011), który zwracając uwagę na lokalny charakter zrównania proponuje zrównywanie oparte na konstrukcji lokalnych funkcji zrównujących. Pojawia się tu ścisła zależność pomiędzy wymogiem równoważności (d), a wymogiem niezmienniczości względem populacji (e). Mimo iż zależność jest taka, że (e) implikuje (d), van der Linden (2011) sugeruje, że przybliżanie się do spełnienia wymogu niezmienniczości względem populacji również przybliża spełnienie wymogu równoważności. Ujęcie van der Lindena również wskazuje, że istotą problemu jest fakt, że pomiar edukacyjny jest obarczony błędem, co umyka w nielokalnych modelach zrównywania wyników. Zignorowanie tego faktu przy stosowaniu pojedynczej funkcji zrównującej $eq_Y(x)$ prowadzi do lokalnego obciążenia. Niestety, wydaje się, że lokalne funkcje zrównujące wyniki obserwowane nie mogłyby zostać praktycznie wykorzystane do raportowania zależności między zrównywanymi testami – np. przy zastosowaniu IRT, oznaczałoby to różne przekształcenia na test Y dla osób o tym samym wyniku w teście X , jeżeli różniłyby się ich oszacowania θ . Niemniej, koncepcja lokalnych funkcji zrównujących i związek pomiędzy wymogiem równoważności, a niezmienniczością względem populacji dostarczają ważnych narzędzi do empirycznej weryfikacji spełnienia założenia o równoważności.

Jinghua Liu i Michael Walker (2007, s.115) dokonując przeglądu wymogów stawianych dla procedury zrównywania testów przez wymienionych wcześniej autorów, tj. Lorda (1980), Doransa i Hollanda (2000) oraz Kolena i Brennana (2004), zdecydowali się na wyszczególnienie dodatkowych trzech punktów na podstawie pracy tych ostatnich:

„Wymóg takich samych inferencji (*the same inferences*): testy powinny mieć wspólne cele pomiarowe i powinny być zaprojektowane do wyciągania takich samych typów wniosków.

Wymóg takiej samej populacji docelowej (*the same target population*): testy powinny mieć taką samą populację docelową.

Wymóg takich samych charakterystyk/warunków pomiarowych (*the same measurement characteristics/conditions*): testy powinny mieć taką samą specyfikację, być przeprowadzane w takich samych warunkach oraz być równoważne pod względem psychometrycznych właściwości.”

Zauważalna jest pewna redundantność zbioru wszystkich, już ośmiu, wymienionych wymogów niezbędnych do przeprowadzenia zrównywania wyników testowych. Jednak wydaje się, że sformułowanie wszystkich *explicite* daje jaśniejszy obraz tego, czym „w teorii” zrównywanie wyników ma być. Natomiast w praktyce, niektóre z wymogów mogą być trudne do weryfikacji (np. wymóg równoważności) lub mniej istotne. W kwestii wagi poszczególnych wymogów toczy się dyskusja, którą w skrócie omawiają Holland i in. (2007). Natomiast w kwestii praktycznej weryfikacji wymienionych wymogów, warto odwołać się do pracy Liu i Walkera (2007), którzy zastosowali interesujący zestaw kryteriów zrównywalności (*equatability*) testu SAT w wersji funkcjonującej do 2004 roku z nową wersją, która weszła do użycia w 2005 roku. Znamienne jest, że zadanie zrównywania wyników zostało podjęte w obliczu znacznej zmiany w zakresie wzorcowych specyfikacji testu (*test blueprint*). Przy konserwatywnym traktowaniu wszystkich wymogów stawianych przed zrównywaniem, mogłoby zostać uznane za argument dyskwalifikujący możliwość dokonania zrównania. Zaproponowane kryteria zrównywalności były następujące:

1. Podobieństwo konstruktów (*construct similarity*) – weryfikowane przez stopień podobieństwa treści jak i statystyczne właściwości testu.
2. Empiryczna relacja pomiędzy nowym i starym testem – weryfikowana poprzez współczynnik korelacji między dwoma testami w odniesieniu do współczynnika rzetelności każdego z testów (wyznaczającego górną granicę dla takiej korelacji).
3. Precyzja pomiaru – weryfikowana zarówno poprzez współczynnik rzetelności, jak i poprzez lokalne miary błędu pomiaru umiejętności.
4. Niezmienniczość w podgrupach (*subgroup invariance*) – weryfikowana poprzez zbadanie relacji między średnimi wynikami w zależności od istotnych zmiennych grupujących oraz poprzez analizę postaci funkcji łączącej wyniki w zależności od istotnych zmiennych grupujących.

4.2. Dobór próby w badaniach zrównujących

4.2.1. Elementy schematu badawczego definiującego dobór próby

Podstawowym sposobem gromadzenia danych koniecznych do zrównania wyników egzaminacyjnych było przeprowadzenie badań z wykorzystaniem zeszytów testowych. Celowi temu podporządkowane były wszystkie inne czynności badawcze. W ramach zrównywania wyników wykorzystane zostały unikalne zeszyty testowe w dwóch wersjach A i B (wynikających z rotacji kolejności odpowiedzi w zadaniach zamkniętych). Liczba zeszytów zastosowana na każdym etapie studium zrównującego określona została w zależności od liczby sesji egzaminacyjnych objętych zrównywaniem dla danego egzaminu (zgodnie z przyjętymi planami zrównania).

W trakcie badania każdy uczeń rozwiązywał zadania z jednego zeszytu testowego. Taki schemat badania implikował schemat doboru próby. Jako że każdy zeszyt testowy musiał być rozwiązany przez losowo dobranych uczniów, pozwala to myśleć o próbie dla każdego badania, jako o specyficznej liczbie (określonej dla danego egzaminu i etapu badań) oddzielnych losowych „podprób” (jedna „podpróba” to uczniowie rozwiązujący zadania z jednego zeszytu testowego w wersji A i B). Trzeba jednak pamiętać, że w czasie procedury losowania dana liczba „podprób” musiała zostać wylosowana jednocześnie, co wynika z przyjętego złożonego schematu doboru próby i skończonej liczby placówek w populacji szkół. Na przykład w badaniach dla egzaminu gimnazjalnego w 2011 roku zastosowano 11 zeszytów testowych (w wersji A i B). Dla każdej z „podprób” wylosowanych zostało 40 szkół, co łącznie daje 440 szkół. Tabela 4.1 przedstawia liczbę wylosowanych szkół dla każdego z etapów badań.

Tabela 4.1. Liczebność prób w poszczególnych latach badania w podziale na typy szkół oraz egzaminów

Rok	Typ szkoły (egzamin)	Liczebność szkół w próbie	Liczba podprób
2011	Gimnazjum (egzamin gimnazjalny)	440	11
2012	Szkoła podstawowa (sprawdzian po VI klasie szkoły podstawowej)	440	11
	Gimnazjum (egzamin gimnazjalny)	80	2
2013	Szkoła ponadgimnazjalna (matura z matematyki)	200	7
	Szkoła podstawowa (sprawdzian po VI klasie szkoły podstawowej)	80	2
	Gimnazjum (egzamin gimnazjalny)	80	2
	Szkoła ponadgimnazjalna (matura z języka polskiego)	300	10
2014	Szkoła ponadgimnazjalna (matura z języka angielskiego)	360	12
	Szkoła ponadgimnazjalna (matura z matematyki)	120	4
	Szkoła podstawowa (sprawdzian po VI klasie szkoły podstawowej)	80	2
	Gimnazjum (egzamin gimnazjalny)	80	2

4.2.2. Populacja i operat losowania

We wszystkich etapach studium zrównującego populacja docelowa została określona, jako: uczniowie szkół dla młodzieży, z wyłączeniem szkół specjalnych i przyszpitalnych, piszący egzamin z zastosowaniem arkusza standardowego dla danego egzaminu. W przypadku egzaminu maturalnego dodatkowo ograniczono populację do uczniów liceów ogólnokształcących, liceów profilowanych i techników. Operatami pierwotnymi w kolejnych losowaniach były listy szkół udostępnione przez CKE, sporządzone na podstawie zamówień na arkusze egzaminacyjne.

Aby zapewnić minimalny błąd pomiaru dla narzuconych ograniczeń kosztowych, postanowiono ograniczyć populację docelową do uczniów uczących się w szkołach nie mniejszych niż dziesięcioosobowe (na poziomie klasy przystępującej do egzaminu). Tę populację nazywamy rzeczywistą populacją losowania, której odpowiada dostosowany do niej operat wtórny. Decyzja ta została podjęta między innymi po uwzględnieniu wyników symulacji niezbędnych do oszacowania błędów próbkowania. Przyjęte ograniczenie wyłącza z operatu nie więcej niż 5% uczniów w każdym etapie badań.

4.2.3. Schemat losowania

W każdym z etapów badania wykorzystano taki sam schemat losowania. Miało ono charakter losowania warstwowego, proporcjonalnego do liczebności szkoły na poziomie klas/klasy oraz wielopoziomowego. Każda z prób, definiowana przez jeden zeszyt testowy, została podzielona na cztery warstwy, według kategorii wielkości miejscowości (miasto do 20 tys.; miasto od 20 tys. do 100 tys.; miasto powyżej 100 tys.; wieś). Dokonano alokacji zbliżonej do procentowego udziału uczniów w danej warstwie. Dodatkowo każda z warstw została podzielona na $n/2$ warstw drugiego rzędu (gdzie n to liczba wylosowanych szkół w danej warstwie) utworzonych ze względu na wyniki egzaminacyjne szkoły. Z każdej tak powstałej warstwy wylosowane zostały dwie szkoły. Szkoły losowane były proporcjonalnie do liczby uczniów. Użyta została klasyczna metoda kumulatywna. W każdej wylosowanej szkole do udziału w badaniu poproszony został jeden oddział wylosowany w sposób prosty. Jeżeli w danej szkole był tylko jeden oddział ostatniego rocznika, automatycznie wchodził on do badania. Wraz z próbą

główną została wylosowana próba rezerwowa. Każda szkoła miała przypisane, wylosowane dodatkowo szkoły rezerwowe, które w razie odmowy udziału w badaniu zastępowały wylosowaną szkołę. Liczba zeszytów testowych i liczba zadań warunkowała liczebność próby, która różniła się między etapami badania, jak i typami szkół. Główne założenie mówiło o tym, że każde z zadań z zeszytu zrównującego powinno być rozwiązane około 800 razy. Dla przykładu w głównych badaniach zrównujących egzamin gimnazjalny przeprowadzanych w 2011 roku wzięło udział 10398 uczniów z 440 szkół rozwiązując 22 zeszyty testowe (każdy uczeń rozwiązywał zadania z dwóch zeszytów – jeden z części matematyczno-przyrodniczej i jeden z części humanistycznej). Zastosowany plan zrównywania wyników egzaminów włączonych do studium

Teoretyczne omówienie planu zrównania i implementację metody łącznej kalibracji modelu IRT do zrównania egzaminu gimnazjalnego i sprawdzianu przedstawiono w podrozdziale 4.6. W tym podrozdziale przedstawimy na przykładzie badań zrównujących dla sprawdzianu w szóstej klasie szkoły podstawowej, w jaki sposób zostało zaplanowane i zrealizowane zbieranie danych koniecznych do zrównania *post factum*. Analogiczne plany dla zrównania wszystkich typów egzaminów zawiera Aneks 2. Główne badania zrównujące wyniki sprawdzianu miały miejsce w 2012 roku na próbie losowej 440 szkół. W kolejnych latach – 2013 i 2014 do badań wylosowano po 80 szkół w każdym z tych lat (por. Tabela 4.1).

Rysunek 4.1 przedstawia schematycznie plan badawczy zastosowany do zebrania danych wykorzystanych do zrównywania wyników sprawdzianu na zakończenie szóstej klasy szkoły podstawowej we wszystkich trzech etapach studium zrównującego (II etap 2002-2011, III etap – 2011-2012 i IV etap 2012-2013). Zrównywanie wyników sprawdzianu przeprowadzone w latach 2012-2014 obejmowało 14 różnych populacji uczniów: $P_{02}, P_{03}, \dots, P_{14}$. Każda populacja w sesji egzaminacyjnej rozwiązywała zadania z odpowiedniej edycji sprawdzianu dla których arkusze egzaminacyjne dla poszczególnych lat oznaczono odpowiednio: $T_{02}, T_{03}, \dots, T_{14}$.

Wyniki sprawdzianu z poszczególnych lat to niepołączone dane, archiwizowane przez system egzaminacyjny, które zostały udostępnione do badań przez Centralną Komisję Egzaminacyjną. W pierwszym roku zrównywania wyników sprawdzianu (II etap studium) wszyscy uczniowie z populacji piszącej sprawdzian w 2012 roku - P_{12} rozwiązywali zadania sprawdzianu T_{12} , ale ponadto, z tej populacji wylosowano 11 losowo równoważnych prób uczniów: $S_{12}^1, S_{12}^2, \dots, S_{12}^{11}$, którzy wzięli udział w dodatkowej sesji zrównującej. Podobnie podczas badań zrównujących dla sprawdzianu w 2013 roku wszyscy uczniowie z tego rocznika rozwiązywali sprawdzian T_{13} . Z populacji P_{13} zostały wylosowane próby S_{13}^1 oraz S_{13}^2 , z których uczniowie rozwiązywali zadania z zeszytów testowych zastosowanych w sesji zrównującej złożonych z wybranych zadań z arkuszy egzaminacyjny T_{11} i T_{12} lub T_{04} i T_{12} oraz dodatkowych zadań kotwiczących. Podobnie w IV etapie badań w 2014 roku wszyscy uczniowie z populacji z 2014 roku P_{14} rozwiązywali zadania sprawdzianu T_{14} i losowa próba S_{14} rozwiązywała zadania z zeszytu testowego zawierającego odpowiednio zadania T_{12} i T_{13} oraz dodatkowe zadania kotwiczące. Każda z prób w sesji zrównującej rozwiązywała zadania z zeszytu testowego składającego się z dwóch podzbiorów zadań kotwiczących - wybranych z poprzednich egzaminów (T) oraz podzbioru dodatkowych zadań kotwiczących (do skalowania wertykalnego z egzaminem gimnazjalnym i wynikami sprawdzianu w kolejnych latach).

Przykładowo, próba S_{12}^5 rozwiązywała zadania z zeszytu testowego z podzbiórami kotwiczących zadań T_{06} oraz T_{07} (podzbiory zadań z arkuszy egzaminacyjnych z lat 2006 oraz 2007) oraz z podzbiorem dodatkowych zadań kotwiczących C^2 i C^3 .

Źródło	Populacja	Próba	N	Zbiory zadań egzaminacyjnych														Zbiory dodatkowych zadań kotwiczących			
				T_{02}	T_{03}	T_{04}	T_{05}	T_{06}	T_{07}	T_{08}	T_{09}	T_{10}	T_{11}	T_{12}	T_{13}	T_{14}	C^1	C^2	C^3	C^4	
Dane egzaminacyjne	$\mathcal{P}_{02}$		$5,4 \cdot 10^5$	■																	
	$\mathcal{P}_{03}$		$5,3 \cdot 10^5$		■																
	$\mathcal{P}_{04}$		$5,2 \cdot 10^5$			■															
	$\mathcal{P}_{05}$		$4,9 \cdot 10^5$				■														
	$\mathcal{P}_{06}$		$4,6 \cdot 10^5$					■													
	$\mathcal{P}_{07}$		$4,4 \cdot 10^5$						■												
	$\mathcal{P}_{08}$		$4,1 \cdot 10^5$							■											
	$\mathcal{P}_{09}$		$4,0 \cdot 10^5$								■										
	$\mathcal{P}_{10}$		$3,8 \cdot 10^5$									■									
	$\mathcal{P}_{11}$		$3,7 \cdot 10^5$										■								
	$\mathcal{P}_{12}$		$3,6 \cdot 10^5$	■	■	■	■	■	■	■	■	■	■	■				■	■	■	■
	$\mathcal{P}_{13}$		$3,5 \cdot 10^5$			■								■	■			■	■	■	■
	$\mathcal{P}_{14}$		$3,4 \cdot 10^5$												■	■		■	■		■
	Dane z sesji zrównujących	$\mathcal{P}_{12}$	S_{12}^1	800	■	■									■			■	■		
$\mathcal{P}_{12}$		S_{12}^2	800		■	■								■				■		■	
$\mathcal{P}_{12}$		S_{12}^3	800			■	■							■			■			■	
$\mathcal{P}_{12}$		S_{12}^4	800				■	■						■			■		■		
$\mathcal{P}_{12}$		S_{12}^5	800					■	■					■				■	■		
$\mathcal{P}_{12}$		S_{12}^6	800						■	■				■				■		■	
$\mathcal{P}_{12}$		S_{12}^7	800							■	■			■			■			■	
$\mathcal{P}_{12}$		S_{12}^8	800								■	■		■			■		■		
$\mathcal{P}_{12}$		S_{12}^9	800									■	■	■				■	■		
$\mathcal{P}_{12}$		S_{12}^{10}	800		■								■	■				■		■	
$\mathcal{P}_{12}$		S_{12}^{11}	800						■				■	■			■			■	
$\mathcal{P}_{13}$		S_{13}^1	800										■	■	■		■		■	■	
$\mathcal{P}_{13}$		S_{13}^2	800			■								■	■		■	■		■	
$\mathcal{P}_{14}$		S_{14}	1600											■	■	■	■	■		■	

Legenda:

- ☐ uczniowie nie rozwiązują żadnego zadania
- ☒ uczniowie rozwiązują wszystkie zadania
- ☒ uczniowie rozwiązują wszystkie zadania na sesji egzaminacyjnej
- ☒ część uczniów rozwiązuje wszystkie zadania na sesji badawczej
- ☒ uczniowie rozwiązują podzbiór zadań
- ☒ część uczniów rozwiązuje podzbiór zadań na sesji badawczej

Rysunek 4.1. Schemat planu badawczego – sprawdzian w VI klasie szkoły podstawowej

4.3. Narzędzia badawcze

W badaniach zastosowane zostały trzy typy narzędzi badawczych:

1. Testy kotwiczące (zeszyty testowe) przeznaczone do rozwiązywania przez uczniów włączonych do próby badawczej:
 - a) w pierwszym etapie studium zrównującego dla egzaminu gimnazjalnego,
 - b) w drugim etapie dla sprawdzianu i dla egzaminu gimnazjalnego,
 - c) w trzecim etapie dla egzaminu maturalnego z matematyki na poziomie podstawowym oraz dla sprawdzianu i egzaminu gimnazjalnego,
 - d) w czwartym etapie dla egzaminu maturalnego na poziomie podstawowym z języka polskiego, języka angielskiego i matematyki oraz dla sprawdzianu i egzaminu gimnazjalnego.
2. Ankieta ucznia przeznaczona do zastosowania po każdej części badań.
3. Ankieta adresowana do nauczycieli uczących w badanej klasie przedmiotów, których zakres obejmuje treści sprawdzane na sprawdzianie i na egzaminie gimnazjalnym.

Celem badania ankietowego uczniów było znalezienie odpowiedzi na pytanie, czy motywacja do rozwiązywania zadań testowych wpływa na uzyskiwane wyniki w testach niskiej doniosłości (jakimi są badania zrównujące, nie mające poważnych konsekwencji dla egzaminowanych). Wpływ motywacji na wyniki badania zrównującego byłby cechą niewskazaną, świadczącą o ich obciążeniu. Cel badań ankietowych wiąże się z odpowiedzią na trzy poniższe pytania badawcze.

1. Jak deklarowany poziom motywacji testowej uczniów wpływa na wyniki osiągnięte w testach wykorzystanych w badaniach zrównujących?
2. Ile procent zmienności wyników uzyskanych w badaniach zrównujących wyjaśnia poziom motywacji testowej?
3. Czy oba te efekty różnicują się w zależności od rozwiązywanego testu oraz od etapu edukacyjnego?

W roku 2014 zastąpiono stosowaną we wcześniejszych latach Skalę Motywacji Testowej (Węziak-Białowolska, 2011) polskim tłumaczeniem Skali Opinii Uczniów (Sundre, 2007). Własności psychometryczne obydwu narzędzi zostały przetestowane w pilotażu badań zrównujących w 2014 roku – Skala Opinii Uczniów wykazała lepsze dopasowanie do danych. O specyfice narzędzi badawczych zastosowanych w badaniu ankietowych oraz o jego wynikach Czytelnik może przeczytać w podrozdziale 6.9.

4.4. Dobór zadań do zrównania

4.4.1. Wybór zadań do narzędzi pomiaru w badaniach zrównujących

Celem tego podrozdziału jest przybliżenie Czytelnikowi sposobu przygotowania podstawowych narzędzi – zeszytów testowych stosowanych w badaniach zrównujących. Zwrócono uwagę zarówno na zastosowane procedury, jak i przedstawiono identyfikatory poszczególnych zadań, które składały się na zeszyty testowe we wszystkich czterech etapach badań. Dzięki temu zainteresowany Czytelnik, może odtworzyć pełen schemat badań zrównujących.

We wszystkich etapach studium zrównywania wyników i dla wszystkich egzaminów włączonych do badań zostały zastosowane takie same zasady tworzenia zeszytów testowych. Zastosowano powiązanie wspólnymi zadaniami zeszyty testowe dla różnych poziomów egzaminacyjnych (sprawdzian, egzamin gimnazjalny, egzamin maturalny), co stanowić może punkt wyjścia do skalowania wertykalnego wyników egzaminacyjnych.

Budując narzędzia będące podstawą zrównania wyników egzaminacyjnych do zeszytów testowych dobierano zadania charakteryzujące się pożądanymi własnościami psychometrycznymi i reprezentujące zawartość treściową arkuszy egzaminacyjnych w taki sposób, aby możliwe było kotwiczenie zgodne z przyjętym schematem badania.

4.4.1.1. Test kotwiczący jako podstawowe narzędzie badawcze w zrównaniu

Podstawowymi narzędziami pomiaru w badaniach zrównujących były testy kotwiczące (zeszyty testowe), rozwiązywane przez uczniów szkół podstawowych, gimnazjalnych i ponadgimnazjalnych, włączonych do próby badawczej (Szaleniec i in., 2013). Test kotwiczący to test, który składa się z zadań „nowych” (nieznanych uczniom) oraz zadań „starych”, które były stosowane w testach referencyjnych. Wszystkie zadania składające się na test kotwiczący określa się mianem „zadań wspólnych”.

Testy kotwiczące wykorzystuje się w celu zrównania wyników uzyskiwanych za pomocą testów zbudowanych według tych samych założeń odnośnie zawartości oraz własności psychometrycznych. Przygotowanie testów kotwiczących opierało się na założeniu, że ich budowa (skład) ma minimalizować błędy zrównania wynikające z różnicy w poziomie umiejętności uczniów rozwiązujących test referencyjny i test kotwiczący wykorzystywany do zrównywania. W realizowanych badaniach zrównujących problem ten odnosił się do potencjalnie występujących różnic w poziomie kompetencji uczniów rozwiązujących zadania arkusza egzaminacyjnego na egzaminie w danym roku i poziomu umiejętności uczniów rozwiązujących testy kotwiczące w trakcie studium zrównującego, które odbywało się na około miesiąc przed odpowiadającymi sesjami egzaminacyjnymi. Występowanie tego typu błędów jest powszechnie znane (por. Liu, Moses i Low, 2009).

Kluczowym etapem w budowanie testu kotwiczącego jest ustalenie liczby zadań wspólnych (Węziak, 2007). Badacze (por. Wright i Stone, 1979; Wright i Masters, 1982) uważają, że ich liczba powinna wynosić od 5 do 15. Dla testów dłuższych, liczących 60 pozycji liczba zadań wspólnych w teście kotwiczącym powinna mieścić się w przedziale 10–20 (Afrassa i Keeves; 1999). William Angoff (1971) proponuje, by liczba zadań wspólnych wynosiła 20 lub 20% zadań całego testu (w zależności od tego, która liczba jest większa). Autor podkreśla, że jeśli liczba zadań kotwiczących wynosi 15–25, to dodanie kolejnych zadań nie podnosi znacząco precyzji zrównywania.

W praktyce pula potencjalnych zadań wspólnych, stanowiących podstawę zrównania powinna być większa od założonej docelowo liczby zadań wspólnych. W analizie danych często okazuje się, że część zadań wykazuje niedostateczne dopasowanie do założonego modelu. Zadania takie są usuwane z ostatecznej kalibracji, dlatego powinna istnieć pula możliwych zadań wspólnych, mogących je zastąpić.

Jeśli chodzi o własności zadań kotwiczących, to powinny one charakteryzować się poziomem trudności zbliżonym do przeciętnego poziomu umiejętności¹² (ze względu na najmniejsze błędy standardowe oszacowania). Zadania kotwiczące nie powinny mieć skrajnego poziomu trudności (zadania bardzo łatwe i bardzo trudne). Zakres parametrów trudności zadań wspólnych powinien mieścić się w przedziale od 1,5 do 2 logitów względem średniego poziomu umiejętności uczniów rozwiązujących dany test. Rozkład oszacowań trudności powinien być bardziej zbliżony do rozkładu jednostajnego niż normalnego. Zawartość merytoryczna zadań wspólnych ma istotne znaczenie dla dokładności zrównania zwłaszcza wtedy, gdy grupa uczniów rozwiązujących test zrównywany znacząco różni się poziomem umiejętności od grupy referencyjnej (Hu, Rogers i Vukmirovic, 2008). W opisywanych badaniach, sesje zrównujące odbywały się na około miesiąc przed właściwym egzaminem, dlatego też można przyjąć, że poziom umiejętności uczniów rozwiązujących testy zrównujące był porównywalny z poziomem umiejętności uczniów rozwiązujących zadania z arkuszy egzaminacyjnych.

4.4.1.2. Procedura przygotowania zeszytów testowych w badaniach zrównujących

Każdy zeszyt zrównujący wykorzystany w badaniach składał się z następujących części:

- zadania z egzaminu z roku n (część A),
- zadania z egzaminu z roku m (część B),
- zadania nowe, nieznane uczniom (część C).

Arkusze testowe w kolejnych edycjach badań zrównujących przygotowywane były przez specjalistów okręgowych komisji egzaminacyjnych i Centralnej Komisji Egzaminacyjnej, mających doświadczenie w zakresie budowania zadań i arkuszy egzaminacyjnych. Autorzy arkuszy otrzymali zestaw procedur i wytycznych do tworzenia zeszytów testowych do badań zrównujących ze schematem rozdziału zadań do zeszytów kotwiczących. Przed zastosowaniem przedstawiony schemat zrównania został skonsultowany z ekspertem zewnętrznym – Antonem Béguinem (CITO, Arnhem). Oprócz tego autorzy zeszytów testowych do badań otrzymali parametry zadań z poszczególnych arkuszy egzaminacyjnych (obliczone zgodnie z założeniami zarówno KTT, jak i IRT). Pozwoliły im one wybrać do zeszytów testowych wykorzystanych w badaniach zrównujących zadania o odpowiednich własnościach psychometrycznych. Celem autorów było przygotowanie planów zeszytów testowych według następujących zasad¹³:

1. Każdy zeszyt powinien posiadać dwie wersje różniące się kolejnością odpowiedzi w zadaniach zamkniętych wielokrotnego wyboru.

¹² Opierając się na założeniach modeli IRT, wyrażenie poziomu trudności i umiejętności w logitach pozwala na bezpośrednie porównanie obu parametrów.

¹³ W przypadku konfliktu zasad 2–6, należało je stosować w kolejności zgodnej z numeracją.

- a) Podczas budowania zeszytów testowych najpierw budowana jest wersja podstawowa A.
 - b) Jeżeli na egzaminie występowały równoległe wersje, to układ odpowiedzi w wersji B zeszytu testowego powinien być identyczny jak w arkuszach egzaminacyjnych wersji B.
 - c) Dla egzaminów, w których stosowana była jedna wersja autorzy powinni utworzyć wersję B zeszytu testowego przez rotację kolejności odpowiedzi.
2. Wkład zadań „starych” (znanych z arkuszy egzaminacyjnych, tj. w części A i B zeszytów testowych) w wynik za cały zeszyt testowy wykorzystywany w zrównaniu powinien wynosić 80% (po 40% dla części A i B). Wkład „nowych” zadań, przygotowanych specjalnie na potrzeby badań zrównujących, powinien wynosić 20%. Liczba zadań w każdej części (A, B, C) powinna być proporcjonalna do wskazanego wkładu procentowego punktów.
3. Zeszyty testowe przygotowywane do zrównania nie powinny zawierać zadań, które charakteryzują się niewystarczającymi własnościami psychometrycznymi, tj. zadań, których:
- a) moc różnicująca jest ujemna lub bliska 0 (nie przekraczająca progu 0,2),
 - b) łatwość bliska jest 0 lub 1,
 - c) tylko jeden dystraktor jest atrakcyjny (wskazywany przez uczniów jako poprawna odpowiedź w karcie odpowiedzi).
4. Zawartość merytoryczna zadań w części A i B powinna być reprezentatywną próbką zawartości merytorycznej całego arkusza referencyjnego, z którego pochodzą zadania z części odpowiednio A i B.
5. W ramach części A i B zróżnicowanie parametrów łatwości/trudności zadań wybranych do arkusza testowego powinno być najbardziej zbliżone do zróżnicowania parametrów łatwości/trudności zadań całego arkusza referencyjnego, z którego wybierane są zadania.
6. Suma punktów możliwych do zdobycia w przypadku każdego zeszytu testowego powinna w miarę możliwości być taka sama (dla danego typu egzaminu wskazano autorom liczbę punktów). Maksymalne dopuszczalne odchylenie wynosi 3 punkty powyżej lub poniżej liczby punktów wskazanej dla danego typu egzaminu.

Plany zeszytów testowych przygotowane przez autorów zostały zweryfikowane przez zespół badawczy. Następnym krokiem było przygotowanie zeszytów testowych z wybranych zgodnie z wymienionymi powyżej kryteriami zadań. Jednocześnie zostały sprawdzone zarówno identyczność tych zadań z ich oryginałami zastosowanymi na egzaminie, jaki i zachowanie standardów edycyjnych stosowanych w oryginalnych arkuszach referencyjnych. Kolejnym krokiem był pilotaż zeszytów testowych przeznaczonych do zrównywania. Na tej podstawie przeprowadzono analizę danych, która stała się podstawą korekty i przygotowania finalnej wersji zeszytów testowych. Ostateczne wersje zeszytów testowych zostały wydrukowane łącznie z odpowiadającymi im kartami odpowiedzi.

4.4.1.3. Struktura zeszytów testowych

Zrównanie sprawdzianu

Wyniki sprawdzianu podlegały zrównaniu w trzech edycjach badań zrównujących, w latach 2012, 2013 oraz 2014. We wszystkich edycjach zrównania sprawdzianu wykorzystano łącznie 154 zadania „stare”, pochodzące z arkuszy egzaminacyjnych sprawdzianu z lat 2002–2013 oraz 14 zadań „nowych”, nieznanych uczniom szkół podstawowych. W badaniu 2012 wykorzystano 11 zeszytów testowych, w 2013 roku dwa zeszyty testowe, a w 2014 roku jeden zeszyt testowy. Liczba zadań „nowych” wahała się w zeszytach testowych od pięciu do ośmiu. Tabela 4.2 wskazuje na strukturę zeszytów testowych w poszczególnych edycjach zrównywania sprawdzianu, z uwzględnieniem liczby zadań „starych” (z arkuszy egzaminacyjnych sprawdzianu 2002–2013) oraz „nowych”. Aneks 3.1 zawiera dokładny opis zeszytów testowych, wskazujący na konkretne zadania budujące test.

Tabela 4.2. Liczba zadań i ich źródła w zeszytach testowych wykorzystanych w kolejnych edycjach badań zrównujących (sprawdzian)

Rok egzaminu (sprawdzian)	Kod arkusza	Edycje zrównania i zeszyty														
		2012											2013		2014	
		1	2	3	4	5	6	7	8	9	10	11	1	2	1	
2002	S-A1-021	12	–	–	–	–	–	–	–	–	–	–	–	–	-	
2003	S-A1-032	11	11	–	–	–	–	–	–	–	11	–	–	–	-	
2004	S-A1-04	–	11	11	–	–	–	–	–	–	–	–	–	10	-	
2005	S-A1-052	–	–	11	11	–	–	–	–	–	–	–	–	–	-	
2006	S-A1-062	–	–	–	15	15	–	–	–	–	–	–	–	–	-	
2007	S-1-072	–	–	–	–	12	12	–	–	–	–	–	–	–	-	
2008	S-1-082	–	–	–	–	–	11	11	–	–	–	–	–	–	-	
2009	S-1-092	–	–	–	–	–	–	16	16	–	–	–	–	–	-	
2010	S-1-102	–	–	–	–	–	–	–	15	15	–	–	–	–	-	
2011	S-1-112	–	–	–	–	–	–	–	–	12	12	12	12	–	-	
2012	S-1-122	–	–	–	–	–	–	–	–	–	–	–	13	12	13	
2013	S-1-132	–	–	–	–	–	–	–	–	–	–	–	–	–	14	
Zadania „nowe”		7	8	7	6	7	8	7	6	7	8	7	7	5	7	

Zrównanie egzaminu gimnazjalnego

Wyniki egzaminu gimnazjalnego podlegały zrównaniu w czterech edycjach badań zrównujących, w latach 2011–2014. We wszystkich edycjach zrównania wyników egzaminu gimnazjalnego w części humanistycznej wykorzystano łącznie 177 zadań „starych”, pochodzących z arkuszy egzaminu gimnazjalnego w części humanistycznej lat 2002–2013¹⁴ oraz 73 zadania „nowe”, nieznane uczniom gimnazjów. W badaniu 2011 wykorzystano 11 zeszytów testowych, a w latach 2013 i 2014 po dwa zeszyty testowe. Liczba zadań „nowych” wahała się w zeszytach testowych od 6 do 31. Tabela 4.3 wskazuje na strukturę zeszytów testowych w poszczególnych edycjach zrównywania egzaminu

¹⁴ Przy czym od 2012 roku były to osobne arkusze dla egzaminu z języka polskiego oraz historii i wiedzy o społeczeństwie.

gimnazjalnego z przedmiotów humanistycznych. Aneks 3.2 zawiera dokładny opis zeszytów testowych, wskazujący na konkretne zadania budujące test.

Tabela 4.3. Liczba zadań i ich źródła w zeszytach testowych wykorzystanych w kolejnych edycjach badań zrównujących (egzamin gimnazjalny – część humanistyczna)

Rok egzaminu (gimnazjalny)	Kod arkusza	Edycje zrównania i zeszyty																
		2011											2012		2013		2014	
		część humanistyczna											H	JP	H	JP	H	JP
		1	2	3	4	5	6	7	8	9	10	11	1	1	1	1	1	1
2002	GH-A1-021	11	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
2003	GH-A1-031	12	12	–	–	–	–	–	–	12	–	12	–	–	–	–	–	–
2004	GH-A1-042	–	13	13	–	–	–	–	–	–	–	–	–	–	–	–	–	–
2005	GH-A1-052	–	–	13	13	–	–	–	–	–	–	–	–	–	2	–	–	–
2006	GH-A1-062	–	–	–	10	10	–	–	–	10	10	–	–	–	–	–	–	–
2007	GH-A1-072	–	–	–	–	10	10	–	–	–	–	–	–	–	–	–	–	–
2008	GH-A1-082	–	–	–	–	–	14	14	–	–	–	–	6	–	–	–	–	–
2009	GH-A1-092	–	–	–	–	–	–	11	11	–	–	–	5	–	3	–	–	–
2010	GH-A1-102	–	–	–	–	–	–	–	12	–	12	12	–	–	–	–	–	–
2011	GH-A1-112	–	–	–	–	–	–	–	–	–	–	–	5	8	4	8	–	–
2012	GH-HA1-122	–	–	–	–	–	–	–	–	–	–	–	–	–	16	12	13	12
2013	GH-PA1-132	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	14	13
Zadania „nowe”		9	15	16	17	17	17	16	16	9	7	7	9	31	9	18	6	7

H – egzamin z historii; JP – egzamin z języka polskiego

W przypadku zrównania wyników egzaminu gimnazjalnego w części matematyczno-przyrodniczej wykorzystano łącznie 170 zadań „starych”, pochodzących z arkuszy egzaminu gimnazjalnego w części matematyczno-przyrodniczej lat 2002-2013¹⁵ oraz 60 zadań „nowych”, nieznanymi uczniom gimnazjów. Podobnie, jak w przypadku zrównania wyników egzaminu w części humanistycznej wykorzystano 11 zeszytów testowych w edycji zrównania 2011 oraz po dwa zeszyty testowe w pozostałych edycjach zrównania. Liczba zadań „nowych” we wszystkich arkuszach waha się od 4 do 17. Strukturę zeszytów testowych w poszczególnych edycjach zrównywania egzaminu gimnazjalnego z przedmiotów matematyczno-przyrodniczych podsumowuje Tabela 4.4. Aneks 3.3 zawiera opis zeszytów testowych, wskazujący na konkretne zadania budujące test.

¹⁵ Przy czym od 2012 roku były to osobne arkusze dla matematyki oraz przedmiotów przyrodniczych.

Tabela 4.4. Liczba zadań i ich źródła w zeszytach testowych wykorzystanych w kolejnych edycjach badań zrównujących (egzamin gimnazjalny – część matematyczno-przyrodnicza)

Rok egzaminu (gimnazjalny)	Kod arkusza	Edycje zrównania i zeszyty																
		2011											2012		2013		2014	
		część matematyczno-przyrodnicza											M	P	M	P	M	P
		1	2	3	4	5	6	7	8	9	10	11	1	1	1	1	1	1
2002	GM-A1-021	12	–	–	–	–	–	–	–	–	12	–	–	–	–	–	–	–
2003	GM-A1-031	13	13	–	–	–	–	–	–	13	–	13	–	–	–	–	–	–
2004	GM-A1-042	–	12	12	–	–	–	–	–	–	–	–	–	–	–	–	–	–
2005	GM-A1-052	–	–	12	12	–	–	–	–	–	–	–	–	–	–	–	–	–
2006	GM-A1-062	–	–	–	14	14	–	–	–	14	–	–	–	–	–	–	–	–
2007	GM-A1-072	–	–	–	–	13	13	–	–	–	–	–	–	–	–	–	–	–
2008	GM-A1-082	–	–	–	–	–	13	13	–	–	–	–	–	–	–	–	–	–
2009	GM-A1-092	–	–	–	–	–	–	13	13	–	–	–	–	–	–	–	–	–
2010	GM-A1-102	–	–	–	–	–	–	–	12	–	12	12	–	–	–	–	–	–
2011	GM-A1-112	–	–	–	–	–	–	–	–	–	–	–	7	8	7	7	–	–
2012	GM-PA1-122	–	–	–	–	–	–	–	–	–	–	–	–	–	11	10	11	10
2013	GM-MA1-132	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	11	9
Zadania „nowe”		11	12	14	13	12	12	13	14	13	12	14	15	17	14	16	4	6

M – egzamin z matematyki; P – egzamin z przyrody

Zrównanie egzaminu maturalnego

Badania zrównujące wyniki matury z matematyki prowadzono w latach 2013–2014, a matury z języka polskiego i języka angielskiego w 2014 roku. Arkusze testowe wykorzystywane w zrównaniu wyników matury składały się z czterech typów zadań:

1. Pochodzących z egzaminu gimnazjalnego (zob. podrozdział 4.10). Umieszczenie zadań pochodzących z egzaminu gimnazjalnego w arkuszach testowych miało na celu podjęcie próby zrównania pionowego wyników.
2. Zadań pochodzących z arkuszy egzaminacyjnych matury w latach 2010–2013.
3. Zadań pochodzących z arkuszy egzaminacyjnych matury poprawkowej z lat 2010–2013.
4. „Nowych”, nieznanymi wcześniej uczniom zadań, w liczbie od siedmiu do dziewięciu na zeszyt testowy.

Tabela 4.5, Tabela 4.6 i Tabela 4.7 wskazują na liczbę i źródło poszczególnych typów zadań zastosowanych w zeszytach testowych w badaniach mających na celu zrównanie wyników matury, kolejno z języka polskiego, matematyki oraz języka angielskiego. Aneksy 3.4, 3.5 i 3.6 zawierają dokładny opis składu kolejnych zeszytów testowych, wskazujący na konkretne zadania budujące testy.

Tabela 4.5. Liczba zadań i ich źródła w zeszytach testowych wykorzystanych w kolejnych edycjach badań zrównujących (matura – język polski)

Typ egzaminu	Rok egzaminu	Kod arkusza	Edycje zrównania i zeszyty									
			2014									
			1	2	3	4	5	6	7	8	9	10
gimnazjalny	2012	GH-PA1-122	5	5	5	5	5	5	5	5	5	5
matura	2010	MPO-P1A1P-102	11	–	–	–	1	10	–	–	–	2
matura	2011	MPO-P1A1P-112	–	11	1	–	–	2	10	–	–	–
matura	2012	MPO-P1A1P-122	1	–	11	–	–	–	2	10	–	–
matura	2013	MPO-P1A1P-132	–	1	–	10	–	–	–	2	9	–
matura poprawkowa	2013	MPO-P1A1P-134	–	–	1	–	11	–	–	–	2	10

Tabela 4.6. Liczba zadań i ich źródła w zeszytach testowych wykorzystanych w kolejnych edycjach badań zrównujących (matura – matematyka)

Typ egzaminu	Rok egzaminu	Kod arkusza	Edycje zrównania i zeszyty											
			2013							2014				
			1	2	3	4	5	6	7	1	2	3	4	
gimnazjalny	2011	GM-A1-112	3	–	–	–	3	–	4	–	–	–	–	
gimnazjalny	2012	GM-MA1-122	–	–	–	–	–	–	–	3	3	3	3	
gimnazjalny	2013	GM-MA1-132	–	–	–	–	–	–	–	6	6	6	6	
matura	2010	MMA-P1A1P-102	11	13	–	–	–	12	–	–	–	–	–	
matura	2011	MMA-P1A1P-112	11	1	13	14	–	–	–	–	–	–	–	
matura	2012	MMA-P1A1P-122	–	–	11	–	11	12	–	10	–	–	10	
matura	2013	MMA-P1A1P-132	–	–	–	–	–	–	–	–	11	11	11	
matura poprawkowa	2010	MMA-P1A1P-104	–	11	10	–	–	–	11	–	–	–	–	
matura poprawkowa	2011	MMA-P1A1P-114	–	9	–	9	9	–	–	–	–	–	–	
matura poprawkowa	2012	MMA-P1A1P-124	–	–	–	10	–	10	11	11	11	–	–	
matura poprawkowa	2013	MMA-P1A1P-134	–	–	–	–	–	–	–	–	–	10	–	
Zadania „nowe”			8	–	–	–	9	–	7	–	–	–	–	

Tabela 4.7. Liczba zadań i ich źródła w zeszytach testowych wykorzystanych w kolejnych edycjach badań zrównujących (matura – język angielski)

Typ egzaminu	Rok egzaminu	Kod arkusza	Edycje zrównania i zeszyty											
			2014											
			1	2	3	4	5	6	7	8	9	10	11	12
gimnazjalny	2012	GJA-A1-20122	2	2	2	2	2	2	2	2	2	2	2	2
gimnazjalny	2013	GJA-A1-20132	1	1	1	1	1	1	1	1	1	1	1	1
matura	2010	MJA-P1A1P-102	8	–	–	8	8	–	–	–	8	–	–	–
matura	2011	MJA-P1A1P-112	13	13	–	–	–	13	–	–	–	13	–	–
matura	2012	MJA-P1A1P-122	–	13	13	–	–	–	13	–	–	–	13	–
matura	2013	MJA-P1A1P-132	–	–	9	9	–	–	–	9	–	–	–	9
matura poprawkowa	2010	MJA-P1A1P-104	–	–	–	–	7	–	–	–	–	–	–	7
matura poprawkowa	2011	MJA-P1A1P-114	–	–	–	–	–	3	–	–	3	–	–	–
matura poprawkowa	2012	MJA-P1A1P-124	–	–	–	–	–	–	3	–	–	3	–	–
matura poprawkowa	2013	MJA-P1A1P-134	–	–	–	–	–	–	–	3	–	–	3	–

Skalowanie wertykalne

W badaniach zrównujących podjęto pierwszą w Polsce próbę skalowania wertykalnego wyników egzaminacyjnych (por. podrozdział 6.8). W tym celu w zeszytach testowych znalazły się zadania, które miały umożliwić takie zrównanie. Łącznie we wszystkich zeszytach badawczych wykorzystano:

- osiem zadań dla zeszytów testowych zrównujących wyniki sprawdzianu i egzaminu gimnazjalnego z przedmiotów humanistycznych, pochodzących z arkuszy niepublikowanych, a więc zawierających zadania „nowe”, nieznane wcześniej uczniom¹⁶,
- sześć zadań dla zeszytów testowych zrównujących wyniki sprawdzianu i egzaminu gimnazjalnego z przedmiotów matematyczno-przyrodniczych, pochodzących z arkuszy niepublikowanych, a więc zawierających „zadania nowe”, nieznane wcześniej uczniom¹⁷,
- pięć zadań dla zeszytów testowych zrównujących wyniki egzaminu gimnazjalnego z przedmiotów humanistycznych i matury z języka polskiego. Spis tych zadań zawiera Aneks 3.6.
- 24 zadania dla zeszytów testowych zrównujących wyniki egzaminu gimnazjalnego z przedmiotów matematyczno-przyrodniczych i matury z matematyki: z czego 13 zadań pochodziło z arkuszy egzaminu gimnazjalnego (spis zadań zawiera Aneks 3.5) oraz 11 zadań

¹⁶ Były to następujące zadania: 1) z arkusza ht1: 25(1506), 29(1507); 2) z arkusza ht2: 26(1515), 27(1510), 30(1511), 33(1516), 35(1504), 37(1505).

¹⁷ Były to następujące zadania: 1) z arkusza mc1: 29(1508), 30(1503), 36(1514); 2) z arkusza mc2: 26(1502), 40(1501); 3) z arkusza mcs: 23(1509).

pochodziło z arkuszy niepublikowanych, a więc zawierających zadania „nowe”, nieznanie wcześniej uczniom¹⁸.

- trzy zadania dla zeszytów testowych zrównujących wyniki egzaminu gimnazjalnego z języka angielskiego i matury z języka angielskiego. Spis tych zadań zawiera Aneks 3.4.

4.4.2. Wybór zadań do analiz zrównujących

Przed przystąpieniem do zrównania wyników egzaminacyjnych w kolejnych edycjach badania prowadzono kilkuetapową analizę danych. Jej celem było wybranie zbioru zadań o odpowiednich właściwościach do oszacowania funkcji zrównującej wyniki egzaminów ze sobą.

Pierwszy etap analiz polegał na określeniu właściwości psychometrycznych wszystkich zadań arkuszy egzaminacyjnych z lat 2002–2013 za pomocą procedur rozwiniętych w ramach Klasycznej Teorii Testów (KTT), jak i *Item Response Theory* (IRT). Na tej podstawie wybrano pulę zadań egzaminacyjnych, które weszły w skład zeszytów testowych wykorzystanych w badaniach zrównujących (o czym była mowa powyżej).

W kolejnym etapie, już po przeprowadzeniu badań zrównujących, przeanalizowano właściwości psychometryczne zarówno zadań „starych”, pochodzących z arkuszy egzaminacyjnych, jak i „nowych” zadań nieegzaminacyjnych.

Przyjęto założenie, że zadanie może stać się podstawą analiz, jeśli zachowuje minimalnie dobre właściwości psychometryczne, zarówno w prawdziwym egzaminie, jak i badaniu zrównującym. Zadania: o bardzo niskiej dyskryminacji (korelacja zadania z resztą testu: poniżej 0,2¹⁹), bardzo łatwe (łatwość powyżej 95%), i niedopasowane do modelu odpowiedzi na pozycje testowe (IRT) zostały usunięte z analiz. W przypadku sprawdzianu przyjęto mniej restrykcyjne kryteria, ograniczając się do identyfikacji zadań niedopasowanych do modelu IRT.

Wykorzystano metody symulacyjne polegające na usuwaniu kolejnych zadań z puli zrównującej i przeprowadzeniu zrównywania. W toku analiz przeprowadzono wielokrotne zrównywanie liczbą $n-1$ zadań, gdzie n oznacza liczbę zadań zrównujących. Innymi słowy w każdym zrównaniu pomijano jedno zadanie²⁰. Metoda ta pozwoliła na oszacowanie wrażliwości wybranej metody zrównywania na dobór zadań, a także dostarczyła informacji o błędzie zrównywania. Gdy rozrzut wyników po zrównywaniu dla określonego roku przy różnym doborze zadań jest stosunkowo wysoki, należy stwierdzić, iż błąd zrównywania jest wysoki, a tym samym precyzja oszacowania wyniku stosunkowo mała. Analiza pozwala również zidentyfikować przypadki odstające w sytuacji, gdy po usunięciu danego zadania następuje istotna zmiana wyników zrównywania.

¹⁸ Były to następujące zadania: 1) z arkusza mc1: 17(1583), 29(1508), 30(1503), 31(1586); 2) z arkusza mc2: 26(1502), 30(1588), 40(1501); 3) z arkusza mcs: 18(1589), 21(1590), 23(1509), 25(1594).

¹⁹ Kryterium to nie było używane, jeżeli zadania charakteryzowało się bardzo dobrym poziomem dopasowania do modelu dwuparametrycznego.

²⁰ W analizie te włączono także zadania wytypowane jako problematyczne ze względu na niewystarczające właściwości psychometryczne.

Ostatnim etapem uzupełniającym była analiza zróżnicowanego funkcjonowania zadań (*differential item functioning*, DIF). Jej celem było określenie czy wybrane do zeszytów badawczych zadania funkcjonują psychometrycznie tak samo w egzaminie i badaniu zrównującym. Miało to służyć odpowiedzi na pytanie, czy w związku z tym, że badania opierały się o testy niskiej doniosłości (*low-stakes tests*) mogą dostarczyć nieobciążone dane będące podstawą oszacowania zrównanych wyników egzaminacyjnych. Zadania funkcjonujące odmiennie w obu sytuacjach testowych zidentyfikowano na podstawie wielkości efektu P-DIF (Kondratek i Grudniewska, 2013). Jako zadania potencjalnie różnie funkcjonujące w grupie badawczej i egzaminacyjnej uznano takie, dla których wartość bezwzględna średniej wielkości efektu przekracza próg 0,067²¹. Wyniki analiz wskazały, że jeśli wystąpiły zadania różnie funkcjonujące w grupie uczniów w badaniach zrównujących i w grupie uczniów piszących egzamin, to efekty DIF znoszą się. Oznacza to, że część zadań sklasyfikowanych jako różnie funkcjonujące, faworyzuje grupę uczniów biorących udział w badaniach (jest dla niej łatwiejsze), a część faworyzuje grupę uczniów piszących egzamin. Wynika to z relatywnej natury DIF, co oznacza, że należy analizować potencjalne obciążenie całego testu, a nie pojedynczego zadania. Nawet jeśli w teście znajduje się zadanie wykazujące DIF, to na poziomie całego testu efekty te powinny się zerować. Tak też było w tym przypadku, gdyż liczba zadań faworyzujących grupę badawczą i egzaminacyjną jest porównywalna. Analizy ujawniły, że nie występuje także efekt polegający na zwiększeniu się liczby zadań łatwiejszych dla grupy badawczej, pochodzących z egzaminów w latach poprzedzających badanie (w porównaniu do lat wcześniejszych). Można więc założyć, że w trakcie badań zrównujących nie ujawnił się efekt większej znajomości u badanych zadań występujących w arkuszach egzaminacyjnych z lat poprzedzających badania w porównaniu do tych, występujących we wcześniejszych egzaminach.

Na podstawie powyżej opisanych analiz część zadań została wykluczona z analiz zrównujących. Informacje o tym, które zadania zostały wykluczone z analiz oraz powody wykluczenia zostały opisane przez Henryka Szaleńca i współpracowników (Szaleniec i in., 2013).

4.5. Plan nierównoważnych grup z testem kotwiczącym i zrównywanie z wykorzystaniem IRT

Klasycznie wyróżniane są następujące plany zbierania danych umożliwiające przedstawianie na wspólnej skali wyników z różnych narzędzi pomiarowych:

- plan grup równoważnych (*equivalent groups*, EG), w którym uczniowie z dwóch różnych prób rozwiązują dwa testy, przy czym zakłada się, że populacje, z jakich wywodzą się obie grupy, charakteryzują się tym samym poziomem umiejętności,

²¹ Zróżnicowane funkcjonowanie zadań oceniano w siedmiu różnych modelach, ostatecznie wybrano trzy z nich: a) dla danych opierających się na zadaniach egzaminacyjnych, „starych”; b) dla danych opierających się zarówno na zadaniach egzaminacyjnych, „starych”, jak i „nowych”, niejawnych, bez wykorzystania modelu 3PLM; c) dla danych opierających się zarówno na zadaniach egzaminacyjnych, „starych”, jak i „nowych”, niejawnych, z wykorzystaniem modelu 3PLM. Ostateczne oszacowania wielkości efektu opierają się na uśrednionych statystykach P-DIF ze wskazanych trzech typów modeli. Wielkości efektu DIF wskazującą na konieczność analizy treściowej pod kątem występowania stronniczości ustalono arbitralnie na wartość $2/3 \cdot 0,1 \approx 0,067$. Wartość taka została dobrana przy uwzględnieniu rozkładu obserwowanych efektów DIF (uwzględniała wszystkie przypadki odstające) oraz ze względu na przesłankę rzędu efektu, który miałby potencjalnie istotne praktyczne znaczenie."

- plan pojedynczej grupy (*single group*, SG), gdzie ta sama próba rozwiązuje oba testy,
- plan zrównoważony (*counterbalanced design*, CB), to w gruncie rzeczy dwa plany SG, gdzie dwie grupy uczniów rozwiązują oba testy w różnej kolejności,
- plan nierównoważnych grup z testem kotwiczącym (*nonequivalent groups with anchor test*, NEAT), w tym planie biorą udział dwie grupy, pierwsza rozwiązuje test zrównywany oraz tzw. test kotwiczący, a druga ten sam test kotwiczący i test, do którego następuje zrównanie.

Spośród wymienionych planów jedynie plan NEAT jest modelowym schematem zbierania danych, w którym grupy uczniów rozwiązujących różne testy różnią się poziomem umiejętności. W planie tym mamy dwie różne populacje uczniów $\mathcal{P}$ oraz $\mathcal{Q}$, z których uczniowie rozwiązują dwa różne testy, odpowiednio X oraz Y . Plan ten uwzględnia również trzeci, dodatkowy podzbiór zadań – test A noszący nazwę kotwicy, który jest rozwiązywany przez uczniów z obu populacji. Bez zastosowania dodatkowego testu A rozwiązywanego zarówno przez próbę uczniów z populacji $\mathcal{P}$ jak i $\mathcal{Q}$ porównanie wyników różnych grup uczniów jakie uzyskują w różnych testach X oraz Y nie byłoby możliwe – nie sposób byłoby rozróżnić trudności testów X oraz Y od różnic w poziomie umiejętności między uczniami z $\mathcal{P}$ oraz $\mathcal{Q}$. Schematycznie plan NEAT przedstawia Tabela 4.8.

Tabela 4.8. Schemat planu NEAT

Populacja	Próba	X	Y	A
$\mathcal{P}$	S_1	✓		✓
$\mathcal{Q}$	S_2		✓	✓

Zadanie ustalenia porównywalności wyników z różnych testów mierzących tą samą umiejętność stanowi problem zrównywania wyników testowych (czy ogólniej – linkowania/łączenia wyników testowych). Teoretyczne i metodologiczne szczegóły dotyczące problemu zrównywania testów można znaleźć u wielu autorów, na przykład u: Lorda (1980), Kolena i Brennana (2004), Doransa i Hollanda (2000) czy Hollanda i współpracowników (2007), a także w raporcie ze zrównywania wyników egzaminu gimnazjalnego 2002–2010 (Szaleniec i in., bdw). Tu ograniczymy się jedynie do ogólnej typologii metod zrównywania wyników ze szczególnym naciskiem na metody wykorzystujące modelowanie IRT.

Na najogólniejszym poziomie, podziału metod zrównywania wyników testowych można dokonać w zależności od tego czy:

- zrównywanie odbywa się na skali wyników obserwowanych, czy wyników prawdziwych,
- zrównywanie odbywa się z bezpośrednim odwołaniem do modelu pomiarowego, czy nie.

Większość technik wykorzystywanych do zrównywania wyników testowych zalicza się do kategorii zrównywania wyników obserwowanych (*observed score equating*), gdzie przez wynik obserwowany rozumie się klasyczny sumaryczny wynik w teście. Nacisk na przeprowadzanie zrównywania na poziomie wyników obserwowanych jest konsekwencją tego, że w przeważającej większości takie właśnie wyniki są wykorzystywane do raportowania rezultatów testowania. Zrównywanie wyników obserwowanych może zostać przeprowadzone bez konieczności odwoływania się w modelu statystycznym do sparametryzowanego mechanizmu leżącego u podstaw obserwowanych wyników, ale także z wykorzystaniem takiego modelu, tj. z wykorzystaniem IRT (*IRT observed score equating*).

W obrębie podejścia opartego na modelach IRT, pojawia się możliwość dokonania zrównywania wyników prawdziwych (*IRT true score equating*). Przez wynik prawdziwy danego ucznia rozumie się tu wartość oczekiwaną wyniku obserwowanego tego ucznia. Aby zrównanie zostało przeprowadzone na skali wyników prawdziwych klasycznej teorii testu, konieczne jest oszacowanie parametrów modelu pomiarowego leżącego u podstaw obserwowanych odpowiedzi. Zrównywania na skali wyników prawdziwych nie można zatem przeprowadzić „ateoretycznie”, jak w przypadku wyników obserwowanych. Omawiane zależności między metodami zrównywania wyników schematycznie opisuje Tabela 4.9.

Tabela 4.9. Schematyczny podział metod zrównywania wyników testowych

	Zrównywanie wyników obserwowanych	Zrównywanie wyników prawdziwych
Metody niezależne od modelu pomiarowego	zrównywanie ekwicyntylowe bez IRT, liniowe, metoda <i>circle-arc</i>	
Metody oparte na modelu pomiarowym	Zrównywanie wyników obserwowanych w oparciu o IRT	Zrównywanie wyników prawdziwych w oparciu o IRT

Zrównywanie testów w oparciu o model pomiarowy IRT nakłada na zrównywane testy założenie, że zastosowany model IRT poprawnie opisuje udzielane przez uczniów odpowiedzi na zadania zrównywanych testów. Wielowymiarowy model IRT (MIRT) stanowi, że dla każdego ucznia j prawdopodobieństwo udzielenia konkretnego wektora odpowiedzi $\mathbf{u}_j$ da się sparametryzować w następujący sposób (por. Reckase, 2009):

$$\mathbb{P}(\mathbf{U}_j = \mathbf{u}_j | \boldsymbol{\theta}_j) = \prod_i \mathbb{P}(U_{i,j} = u_{i,j} | \boldsymbol{\theta}_j), \quad (4.3)$$

$$\mathbb{P}(U_{i,j} = u_{i,j} | \boldsymbol{\theta}_j) = f_i(u_{i,j} | \boldsymbol{\beta}_i, \boldsymbol{\theta}_j), \quad (4.4)$$

gdzie $\boldsymbol{\theta}_j = [\theta_j^{(1)}, \theta_j^{(2)}, \dots, \theta_j^{(m)}]$ jest wektorem parametrów określających pozycję j -tego ucznia na m -wymiarowej zmiennej ukrytej $\boldsymbol{\theta}$ (umiejętność, *ability*), a $\boldsymbol{\beta}_i$ jest wektorem parametrów określających kształt f_i – funkcji charakterystycznej i -tego zadania (*item characteristic function*, ICC). IRT zatem wprost postuluje rozdzielanie parametrów określających właściwości ucznia od parametrów określających właściwości zadań/testu w celu opisanego mechanizmu udzielania odpowiedzi na zadania. Rozdzielenie parametrów zadań od parametrów uczniów stanowi klucz do zrozumienia popularności metod IRT przy zrównywaniu wyników.

Założenie (4.3) jest wielowymiarową wersją założenia o lokalnej homogeniczności pomiarów (*local homogeneity*; Ellis i van der Woldenberg, 1993) i oznacza, że cała informacja o współzmienności zadań zawarta jest w $\boldsymbol{\theta}$ – ustaliwszy wartość parametru umiejętności ucznia $\boldsymbol{\theta}$, odpowiedzi na zadania stają się zdarzeniami niezależnymi. Założenie (4.4) natomiast określa parametryczną postać zależności pomiędzy wartościami $\boldsymbol{\theta}$, a prawdopodobieństwem określonych odpowiedzi w zadaniu $U_{i,j}$. Od spełnienia obu tych bardzo silnych założeń zależy jakość wniosków opartych na zastosowaniach IRT, w szczególności jakość zrównywania testów wykorzystującego IRT. Z tego względu (por. Livingston, 2004; von Davier i inni, 2004) nadal często preferowane są, omówione wcześniej, nieczyniące tak daleko idących założeń „ateoretyczne” metody zrównywania wyników. Opis metod weryfikacji spełnienia

założeń IRT w kontekście zrównywania oraz analizy odporności (*robustness*) metod zrównywania IRT na niespełnienie założeń (4.3) i (4.4) można znaleźć w pracach Antona Béguina (2000) oraz Ceesa Glasa i Antona Béguina (2011).

Model IRT w (4.3) i (4.4) został określony dla ogólnego, wielowymiarowego przypadku umiejętności θ . Dla uproszczenia, dalszy przegląd metod zrównywania w oparciu o modele IRT został przeprowadzony dla przypadku jednowymiarowego $\theta = \theta$. Informacje o zrównywaniu z wykorzystaniem modeli MIRT można znaleźć u Marka Reckase (2009), Antona Béguina (2000), Yuana Li i Roberta Lissitza (2000) oraz Lihua Yao i Keitha Boughtona (2009). W dalszej części przyjmujemy, zatem, założenie, że zrównywane w planie NEAT testy X , Y oraz kotwica A są jednowymiarowe²² i mierzą ten sam pojedynczy konstrukt (von Davier i von Davier, 2011).

Zmienna umiejętności θ jest w modelu IRT zmienną losową, tzn. umiejętność ucznia θ_j pojawiająca się w równaniach (4.3) i (4.4) nie jest parametrem modelu podlegającym bezpośredniej estymacji, a jedynie losowym efektem zmiennej θ o określonym rozkładzie. To właśnie parametry rozkładu zmiennej umiejętności, oznaczmy go ψ , są estymowane dla modelu IRT. Jest to bardzo ważne w kontekście zrównywania – modelowanie rozkładu umiejętności oddzielnie dla każdej populacji jest istotą całej procedury. W związku z czym model IRT przedstawimy również w formie bezwarunkowej, tj. scałkowanej po rozkładzie umiejętności w populacji $\mathcal{P}$:

$$P(U = \mathbf{u}|\mathcal{P}) = \int f(\mathbf{u}, \theta, \boldsymbol{\beta}) \psi_{\mathcal{P}}(\theta) d\theta, \quad (4.5)$$

Modele IRT poprzez uwzględnienie wprost parametrów rozkładu populacji oraz bezproblemowe radzenie sobie z niekompletnymi schematami zbierania danych stanowią bardzo dobre narzędzie do rozwiązywania problemu zrównywania. W obrębie IRT wypracowano wiele metod do umieszczania na wspólnej skali rozkładów θ dla populacji $\mathcal{P}$ oraz $\mathcal{Q}$ oraz parametrów $\boldsymbol{\beta}$ dla testów X , Y oraz A . Wyróżnić można następujące metody (Kolen i Brennan, 2004):

1. Łączna kalibracja (*concurrent calibration*) wszystkich trzech testów.
2. Oddzielna kalibracja (*separate calibration*) par testów (X , A) oraz (Y , A), po której stosuje się sprowadzające do wspólnej skali przekształcenia oparte na:
 - a) liniowej funkcji parametrów kotwicy – metody średnia/średnia lub średnia/sigma (*mean/mean, mean/sigma*),
 - b) krzywych charakterystycznych kotwicy – metoda Stockinga-Lorda lub Haebary.
3. Metoda ustalonych parametrów (*fixed parameters method*) dla kotwicy A .
4. Metoda przekształcania umiejętności (*proficiency transformation*).

W zrównywaniu wyników sprawdzianu, egzaminu gimnazjalnego oraz wybranych przedmiotów maturalnych omawianym w niniejszej monografii wybrano metodę kalibracji łącznej. Uczyniono tak ze względu na skomplikowany schemat zbierania danych (zob. Aneks 2) oraz ze względu na to, że parametry rozkładu umiejętności w różnych populacjach w tej metodzie są szacowane w sposób

²² Dyskusję o ścisłej zależności pomiędzy założeniem (4.3), a pojęciem „wymiarowości” testu można znaleźć w przełomowej monografii Frederica Lorda i Melvina Novicka (1968).

bezpośredni. Łączna kalibracja dla planu NEAT polega na dopasowaniu do wszystkich danych zebranych z prób S_1 oraz S_2 modelu IRT w pojedynczej procedurze estymacji. Strukturalny brak odpowiedzi na zadania testu Y w próbie S_1 oraz brak odpowiedzi na zadania testu X w próbie S_2 nie stanowi problemu dla maksymalizujących funkcję wiarygodności metod estymacji wykorzystywanych do szacowania parametrów modeli IRT. W celu uzyskania nieobciążonych parametrów w planie NEAT, konieczne jest bezpośrednie uwzględnienie w modelu IRT odrębnych parametrów dla rozkładów θ dla populacji $\mathcal{P}$ oraz $\mathcal{Q}$ i oszacowanie ich z prób S_1 oraz S_2 .

4.6. Implementacja metody łącznej kalibracji modelu IRT do zrównania egzaminów gimnazjalnych i sprawdzianu

W powyższej części rozdziału opisano metodę łącznej kalibracji w zastosowaniu do planu NEAT. Logika postępowania w metodzie łącznej kalibracji przy bardziej skomplikowanym planie zrównywania, w skład którego wchodzi więcej populacji oraz więcej różnych testów, jest analogiczna – w pojedynczym kroku estymujemy parametry zadań z wszystkich testów oraz parametry rozkładu umiejętności dla każdej populacji. Dla planu zbierania danych zastosowanego przy zrównywaniu polskich egzaminów zewnętrznych (zob. Aneks 2) rozmiary macierzy danych przerastały możliwości obliczeniowe sprzętu oraz oprogramowania dostępnego na początkowych etapach projektu. Dla przezwyciężenia tego technicznego problemu pierwsze próby zrównywania przeprowadzano korzystając z wielokrotnie dokonywanej estymacji parametrów modelu IRT na losowych podpróbach z pełnego zbioru dostępnych wyników egzaminacyjnych. Wspomniane procedury próbkowania danych, opisane szczegółowo w raportach podsumowujących wcześniejsze etapy projektu (Szaleniec i in., bdw; 2013), wiązały się ze znaczną czasochłonnością i nastroczały wielu kłopotów przy estymacji błędów szacowania.

Ostatecznie dopasowanie modelu IRT przeprowadzono korzystając z opracowanego w Instytucie Badań Edukacyjnych oprogramowania, działającego w ramach Stata i umożliwiającego szacowanie parametrów wielogrupowego modelu IRT dla danych o tak dużych rozmiarach jak te, z którymi przyszło zmierzyć się w projekcie zrównującym wyniki egzaminacyjne. Zastosowany program stanowił adaptację powszechnie wykorzystywanego do szacowania modeli IRT algorytmu EM (Dempster, Laird i Rubin, 1977), w którym zmienna losowa nieznanego poziomu umiejętności ucznia jest traktowana jako brak danych.

Jedną z cech algorytmu EM jest to, że nie dostarcza on w sposób bezpośredni oszacowań błędów standardowych dla estymowanych za jego pomocą parametrów. Spośród wielu dostępnych metod szacowania błędów standardowych dla algorytmu EM, wybrano metodę polegającą na numerycznym różniczkowaniu przekształcenia EM, zgodnie z rekomendacjami, jakie można znaleźć w artykule Mortazy Jamshidiana i Roberta Jennricha (2000).

4.7. Zrównywanie wyników obserwowanych metodą ekwicytylową

Opiszemy tu metodę zrównywania wyników obserwowanych jaką jest zrównywanie ekwicytylowe. Metoda ta została użyta do zrównania wyników sprawdzianu, egzaminu gimnazjalnego i egzaminu maturalnego, w celu weryfikacji wyników uzyskanych przy użyciu zrównania opartego o IRT (patrz rozdział 6.7). Idea zrównywania ekwicytylowego opiera się na obserwacji, że wszystkie wartości punktowe x i y testów X i Y są ekwiwalentne jeżeli występują z takim samym prawdopodobieństwem (Davies, Holland i Thayer, 2004), co możemy zapisać w sposób następujący:

$$F_X(x) = u = F_Y(y), \quad (4.6)$$

gdzie F_X oraz F_Y to dystrybuanty X i Y , a u przyjmuje wartości w zakresie $[0, 1]$. Opierając się na tym fakcie, możemy zdefiniować funkcję zrównania ekwicyntylowego dwóch ciągłych i ściśle rosnących dystrybuant F_X oraz F_Y następująco:

$$Y = F_Y^{-1} [F_X(x)] \quad (4.7)$$

czyli złożenie $F_Y^{-1} \circ F_X$ przekształca zmienną losową X w zmienną losową Y . Niestety dystrybuanty F_X oraz F_Y dla wyników obserwowanych w testach X i Y , ze względu na dyskretność wyników, są funkcjami skokowymi i równanie (4.7) nie może zostać bezpośrednio zastosowane. Wszystkie ekwicyntylowe metody zrównywania wyników obserwowanych wymagają uwzględnienia jakiejś formy odpowiedniego (skutkującego różnowartościowością) wygładzania (*smoothing*) dystrybuant, co zostało omówione w dalszej części tekstu. Dwoma najpopularniejszymi metodami wygładzania skokowych dystrybuant są (a) lokalna interpolacja liniowa lub (b) wygładzanie za pomocą estymatora jądrowego (*kernel smoothing*). Poglębiony przegląd pierwszego podejścia można znaleźć u Kolena i Brennana (2004), a drugiego u Aliny von Davier, Paula Hollanda i Dorothy Thayer (2004). Ekwicyntylowa funkcja zrównująca X i Y przedstawiona w równaniu (4.8), przyjmuje wtedy kształt:

$$\text{eq}_Y^{(\text{equi})}(x) = \tilde{F}_Y^{-1}[\tilde{F}_X(x)], \quad (4.8)$$

gdzie $\tilde{F}_X$ oznacza uciągloną dystrybuantę X , a $\tilde{F}_Y^{-1}$ jest odwrotną funkcją uciąglonej dystrybuanty Y . Jak w przypadku pozostałych metod zrównywania, funkcja zrównująca jest odwracalna:

$$\text{eq}_X^{(\text{equi})}(y) = \tilde{F}_X^{-1}[\tilde{F}_Y(y)]. \quad (4.9)$$

Metoda ta sprowadza się więc do przypisania wartościom punktowym testu X takich wartości punktowych testu Y , które występują z takim samym prawdopodobieństwem, jak wartości testu X . Szerzej implikacje wynikające z tej definicji omawiają Kolen i Brennan (2004) oraz von Davier i współpracownicy (2004).

4.8. Generowanie wyników obserwowanych na wspólnej skali w oparciu o IRT

Zrównywanie wyników obserwowanych w oparciu o model IRT (por. Tabela 4.9) wymaga estymacji nieobserwowanych dystrybuant obserwowanych wyników $F_{X|Q}$ lub $F_{X|P}$ lub obu tych dystrybuant w oparciu o parametry modelu IRT wyrażone na wspólnej dla populacji P i Q skali. Ustalając uwagę na $F_{X|Q}$, oznacza to konieczność scałkowania po rozkładzie $\psi_Q(\theta)$ warunkowego prawdopodobieństwa uzyskania każdego z wyników:

$$p_{x|Q} = \int_{\theta} \mathbb{P}(X = x|\theta) \psi_Q(\theta) d\theta. \quad (4.10)$$

Warunkowe prawdopodobieństwa $\mathbb{P}(X = x|\theta)$ są kombinacją warunkowych prawdopodobieństw zaobserwowania wektorów odpowiedzi sumujących się do x . Oszacowanie $F_{X|Q}$ stanowi zatem skomplikowany problem kombinatoryczny połączony z całkowaniem numerycznym. Rekursywny algorytm obliczający szukane prawdopodobieństwa jest podany przez Kolena i Brennana (2004). Cees Glas i Anton Béguin (1996) wskazują również na możliwość oszacowania szukanego $F_{X|Q}$ poprzez

przeprowadzenia stosownego eksperymentu Monte Carlo bazującego na oszacowanym i zrównanym modelu IRT.

W przeprowadzonym badaniu zaadaptowano symulacyjną strategię generowania wyrażonych na wspólnej skali wyników obserwowanych roku bazowego. Do wygenerowania wyników obserwowanych dla danego rocznika na skali z roku bazowego (dla wszystkich egzaminów przyjęto rok 2012) wykorzystano 15 kompletów wartości potencjalnych/możliwych (*plausible values*, PV) wygenerowanych zgodnie z oszacowaną dla tego rocznika średnią i odchyleniem standardowym rozkładu umiejętności θ oraz przy uwzględnieniu parametrów zadań dla roku bazowego, a także dodatkowych zmiennych warunkujących. Generowanie PV, wraz z warunkowaniem, opisano szczegółowo poniżej, w podrozdziale 4.9.

Tak wygenerowane wyniki obserwowane w teście z danego egzaminu z roku 2012 dla każdego objętego badaniem rocznika uczniów pozwalają na wnioskowanie o tym, jak wyglądałby rozkład wyników z egzaminu z roku 2012, gdyby podchodzili do niego uczniowie z innych roczników. Wyrażony dla wszystkich roczników na skali wyników obserwowanych egzaminu z 2012 roku rozkład wyników został w badaniu wykorzystany na dwa sposoby. Po pierwsze, możliwe było przedstawienie średnich wyników uczniów ze wszystkich roczników piszących dany egzamin na wspólnej skali egzaminu z roku 2012. Jest to informacja uzupełniająca prezentację długookresowych trendów w poziomie umiejętności raportowanych na skali zmiennej ukrytej sprowadzonej do skali (100; 15). Takie zestawienie pozwala na oszacowanie stopnia, w jakim obserwowane wahania średniego wyniku egzaminów między latami jest związane ze zmianami trudności egzaminów, a na ile ze zmianami w poziomie umiejętności. Jest to informacja szczególnie istotna w kontekście oceny jakości egzaminów, gdyż zakładamy iż powinny one być tak konstruowane, aby ich trudność w kolejnych latach była porównywalna. Po drugie, możliwe było stworzenie tablic przeliczeniowych (zawiera je Aneks 1) pozwalających obliczyć, jaki wynik uzyskalby uczeń w teście z roku bazowego na podstawie surowego wyniku uzyskanego przez niego w egzaminie, który faktycznie zdawał.

4.9. Generowanie wartości potencjalnych/możliwych (PV)

W wyniku dopasowania do danych modelu IRT otrzymujemy jedynie dwa pierwsze momenty rozkładu umiejętności. Dla zwiększenia precyzji odwzorowania kształtu rozkładu θ przy generowaniu wyników obserwowanych obserwacje z rozkładu θ generowano z wykorzystaniem tak zwanych wartości potencjalnych/możliwych (*plausible values*, PV). PV stanowią realizacje z rozkładu *a posteriori* parametru umiejętności ucznia o wektorze odpowiedzi $\mathbf{u}$ (Wu, 2005):

$$\mathbb{P}(\theta | \mathbf{U} = \mathbf{u}) = \frac{\mathbb{P}(\mathbf{U} = \mathbf{u} | \theta, \boldsymbol{\beta}) \psi_0(\theta)}{\int \mathbb{P}(\mathbf{U} = \mathbf{u} | \theta, \boldsymbol{\beta}) \psi_0(\theta) d\theta} \quad , \quad (4.11)$$

gdzie $\psi_0(\theta)$ jest rozkładem *a priori* umiejętności, a $\mathbb{P}(\mathbf{U} = \mathbf{u} | \theta, \boldsymbol{\beta})$ klasyczną funkcją wiarygodności zależną od parametru umiejętności oraz parametrów zadań (por. równania (4.3) i (4.4)).

Uzyskanie PV zgodnie z równaniem (4.12) wymaga również zastosowania zaawansowanych numerycznych rozwiązań opartych na metodologii Monte Carlo i łańcuchów Markowa (*Markov Chain Monte Carlo*, MCMC; Robert i Casella, 2004). W badaniu łańcuchy Markowa służące do wygenerowania PV stworzono w oparciu o algorytm Metropolis-Hastingsa z symetryczną funkcją generującą kandydatów na kolejne punkty w łańcuchu Markowa. Algorytm ten składa się z następujących kroków (por. Patz i Junker, 1999; de la Torre, 2009):

1. Wylosuj punkt kandydujący θ^* zgodnie z generującym rozkładem $N(\cdot | \theta_t, SE_{\theta_0}^2)$,
2. zaakceptuj kandydata ($\theta_{t+1} = \theta^*$) z prawdopodobieństwem $\min \left\{ 1, \frac{\mathbb{P}(\mathbf{U} = \mathbf{u} | \theta^*, \boldsymbol{\beta}) \psi_0(\theta^*)}{\mathbb{P}(\mathbf{U} = \mathbf{u} | \theta_t, \boldsymbol{\beta}) \psi_0(\theta_t)} \right\}$, a w przeciwnym razie pozostaw łańcuch w miejscu ($\theta_{t+1} = \theta_t$).

Przy czym spełnione są następujące warunki:

- a) wartość startowa łańcucha θ_0 jest punktowym oszacowaniem EAP dostarczonym przez program użyty do dopasowania modelu IRT,
- b) za stałą wartość odchylenia standardowego funkcji generującej kandydatów SE_{θ_0} przyjęto błąd standardowy oszacowania θ_0 również raportowany przez program pierwotnie estymujący parametry IRT, a kształt funkcji jest normalny (symetryczność),
- c) rozkład a priori parametru umiejętności $\psi_0(\cdot)$ jest rozkładem standardowym normalnym.

Ze względu na dobór wartości startowej łańcucha (θ_0) i odchylenia standardowego funkcji generujących kandydatów (SE_{θ_0}) bardzo zbliżony do faktycznego rozkładu *a posteriori* (4.12), łańcuchy MCMC od samego początku znajdowały się w centralnym rejonie swojego docelowego rozkładu stacjonarnego. Rozkład *a posteriori* (4.12) wykorzystany do generowania PV uzyskiwano z 500 iteracji łańcucha po uprzednim odrzuceniu 200 pierwszych iteracji łańcucha (tzw. *burn-in*).

Opisana powyżej metoda, w ramach której generowane są PV została uzupełniona o uwzględnienie szeregu zmiennych kontekstowych, poprzez wielopoziomową regresję względem zmiennej ukrytej:

$$\mathbb{P}(\theta | \mathbf{W} = \mathbf{w}, \mathbf{U} = \mathbf{u}) = \frac{\mathbb{P}(\mathbf{U} = \mathbf{u} | \theta, \boldsymbol{\beta}) \psi_0(\theta, \boldsymbol{\gamma}, \mathbf{w})}{\int \mathbb{P}(\mathbf{U} = \mathbf{u} | \theta, \boldsymbol{\beta}) \psi_0(\theta, \boldsymbol{\gamma}, \mathbf{w}) d\theta} \quad (4.12)$$

Różnica między równaniami (4.12) oraz (4.11) jest taka, że pośrednicząca w wyjaśnianiu obserwowanych współzależności między odpowiedziami na zadania $\mathbf{U} = \mathbf{u}$ zmienna umiejętności ucznia θ , sama jest dodatkowo wyjaśniana poprzez dodatkowe zmienne kontekstowe $\mathbf{W}$, przyjmujące wartości $\mathbf{w}$, zgodnie z modelem wielopoziomowej regresji liniowej o parametrach $\boldsymbol{\gamma}$. Wykorzystanie dodatkowych warunkujących zmiennych wyjaśniających przy generowaniu PV skutkuje dwoma pozytywnymi rezultatami w kontekście wykorzystania uzyskanych PV w dalszych analizach: po pierwsze, ze względu na wykorzystanie dodatkowej informacji niepewność pomiaru jest niższa (zwiększona rzetelność), po drugie, późniejsze analizy z wykorzystaniem PV i tych zmiennych kontekstowych dostarczą wyników nieobciążonych w stronę zera (por. Wu, 2005).

Jedną ze zmiennych warunkujących był wskaźnik rozdzielający uczniów włączonych do definicji ogólnopolskiej populacji podlegającej zrównywaniu wyników od uczniów z niej wykluczonych. Wykluczenie z populacji zrównywanej dotyczyło między innymi uczniów szkół dla dorosłych lub uczniów szkół przyszpitalnych. Do oszacowania parametrów średniej oraz odchylenia standardowego w każdym z lat egzaminów wykorzystano jedynie obserwacje uczniów włączonych do definicji populacji, jednak dla uczniów wykluczonych wykorzystano oszacowane parametry modelu IRT do wygenerowania PV, co umożliwia przeprowadzanie ewentualnych analiz z uwzględnieniem również tych grup uczniów.

W przypadku egzaminów maturalnych do warunkowania używano zmiennej oznaczającej typ szkoły: liceum ogólnokształcące, liceum profilowane, technikum, liceum uzupełniające oraz technikum uzupełniające. Dwa ostatnie typy wiążą się z wykluczeniem z definicji populacji podlegającej zrównywaniu. Przy warunkowaniu zmienną oznaczającą typ szkoły przekodowano ją na zestaw liniowo niezależnych wskaźników (zmiennych przyjmujących wartości 0–1).

Dla wszystkich egzaminów obserwujemy istotną korelację między wynikami tych samych szkół na przestrzeni lat – jedne szkoły uzyskują systematycznie wyniki wysokie, inne szkoły uzyskują wyniki systematycznie niskie itp. W celu uwzględnienia autoregresyjnej natury średnich wyników szkół do warunkowania włączono zmienne zawierające średni wynik, jaki uczniowie z danej szkoły uzyskali w roku poprzedzającym warunkowany bieżący rok oraz dwa lata przed warunkowanym bieżącym rokiem. Zmienne te uwzględniano w przypadku lat, dla których istniały wyniki z lat wcześniejszych (tj. nie wykorzystywano w pierwszym roczniku żadnej zmiennej, a w drugim roczniku wykorzystywano jedynie zmienną z wynikami z poprzedniego roku). Przypadki braków zastąpiono średnim wynikiem wszystkich szkół z danego rocznika.

Regresja warunkująca została przeprowadzona w modelu trójpoziomowym, gdzie drugi poziom odpowiadał skupieniu uczniów w szkole, a trzeci poziom odpowiadał skupieniu szkół w danej gminie. Włączenie zmiennych identyfikujących szkołę i gminę pozwala na uwzględnienie w generowanych PV informacji o istotnym nielosowym podobieństwie wyników uczniów uczęszczających do tej samej szkoły oraz wyników szkół pochodzących z tej samej gminy. Bez uwzględnienia tych zmiennych w modelu warunkującym wszelkie analizy na PV mające na celu oszacowanie wariancji międzyszkolnej, czy wariancji międzygminnej wyników uczniowskich skutkowałyby oszacowaniami wariancji obciążonymi w stronę zera.

Oprócz powyższych zmiennych podczas warunkowania dla każdego typu egzaminu używano zmiennych oznaczających płeć ucznia i posiadanie przez niego (lub nie) orzeczenia o dysleksji rozwojowej. Tabela 4.10 zestawia wszystkie zmienne kontekstowe wykorzystane do warunkowania w rozbiciu na zrównywane egzaminy.

Tabela 4.10. Zmienne wykorzystane przy warunkowaniu podczas generowania PV w rozbiciu na egzaminy

Zmienna warunkująca	Typ zmiennej	Egzamin		
		sprawdzian	gimnazjum	matura
płeć ucznia	wskaźnik	uwzględniono	uwzględniono	uwzględniono
posiadanie przez ucznia orzeczenia o dysleksji rozwojowej	wskaźnik	uwzględniono	uwzględniono	uwzględniono
przynależność do populacji	wskaźnik	uwzględniono	uwzględniono	uwzględniono
rodzaj szkoły ponadgimnazjalnej*	wskaźnik	nie uwzględniono	nie uwzględniono	uwzględniono
średni wynik szkoły w roku poprzedzającym	ciągła	uwzględniono	uwzględniono	uwzględniono
średni wynik szkoły dwa lata wcześniej	ciągła	uwzględniono	uwzględniono	uwzględniono
identyfikator szkoły	losowa	uwzględniono	uwzględniono	uwzględniono
identyfikator gminy	losowa	uwzględniono	uwzględniono	uwzględniono

* Liceum ogólnokształcące, liceum profilowane, technikum, liceum uzupełniające, technikum uzupełniające

4.10. Założenia skalowania wertykalnego

W polskim systemie edukacji sposób oszacowania średniego postępu uczniów pomiędzy kolejnymi etapami kształcenia jest zagadnieniem stosunkowo mało zgłębnionym. Procedura skalowania

wertykalnego pozwala na ocenę przyrostu umiejętności ucznia bez konieczności przeprowadzenia skomplikowanego logistycznie badania podłużnego.

Skalowanie wertykalne (*vertical linking, vertical scaling, cross-grade scaling*) jest szeregiem technik, wykorzystywanych w celu skonstruowania skali pionowej (*vertical scale, developmental score scale*), pozwalającej na monitorowanie przyrostu umiejętności uczniów poprzez ich kolejne etapy nauki (Tong i Kolen, 2007; Carlson, 2011). Podczas zbierania danych w niniejszym badaniu wykorzystano schemat oparty na wspólnych pozycjach testowych (*common-items design*). Schemat ten polega na tym, że grupa uczniów z każdego etapu nauczania dostaje do rozwiązania pewne bloki (zbiory) zadań – niektóre z nich zawierają zadania ściśle odpowiadające poziomowi umiejętności tych uczniów, natomiast w skład innych wchodzi zadania kotwiczące (*anchor items*). Zadania kotwiczące są rozwiązywane przez uczniów z sąsiadujących etapów nauczania (np. sąsiadujących roczników), a następnie zostają wykorzystane jako podstawa skalowania (Kolen i Brennan, 2004; Carlson, 2011). Oznacza to, że zadania z wcześniejszych etapów edukacyjnych zostały wykorzystane jako kotwiczące w późniejszych etapach edukacyjnych. Więcej o samej procedurze skalowania wertykalnego można przeczytać w publikacji Karoliny Świsł, Pauliny Skórskiej i Artura Pokropka (2015).

W drugiej (2012)²³, trzeciej (2013) oraz czwartej (2014) edycji badań zrównujących przeprowadzono kotwiczenie pomiędzy etapami kolejnych standaryzowanych egzaminów zewnętrznych – na koniec szkoły podstawowej, na koniec gimnazjum oraz szkoły kończącej się maturą. W roku 2012 przeprowadzono badania zrównujące w szkole podstawowej i w gimnazjum (w związku z tym możemy ocenić przyrost umiejętności humanistycznych i matematycznych uczniów pomiędzy szkołą podstawową a gimnazjum). W roku 2013 przeprowadzono test umiejętności matematycznych w szkole podstawowej, gimnazjum i na etapie szkoły ponadgimnazjalnej oraz test umiejętności humanistycznych w szkole podstawowej oraz gimnazjum. W roku 2014 przeprowadzono badania umiejętności humanistycznych i matematycznych dla trzech etapów edukacyjnych (szkoły podstawowej, gimnazjum i szkół kończących się maturą). Rok 2014 umożliwia więc zbadanie przyrostu umiejętności uczniów przez te etapy edukacyjne.

Procedura skalowania wertykalnego jest uznawana przez wielu badaczy za kontrowersyjną. Obciążenia oszacowań mogą spowodować takie zjawiska jak: wielowymiarowość skali związana z przesunięciem się mierzonego konstruktu (umiejętności) (*construct shift*), dryft parametrów skali (*item parameter drift*) oraz niestabilność charakterystyk zadań kotwiczących (zob. Camilli, Yamamoto i Wang, 1993; Li i Lissitz, 2012). Szczególnie doniosłym problemem wydaje się być wielowymiarowość skali – umiejętności mierzone na poszczególnych etapach edukacyjnych mogą się jakościowo (treściowo) różnić. Konstrukcja wspólnej skali dla różnych umiejętności może więc prowadzić do trudności interpretacyjnych.

W związku z tym, że skalowanie wertykalne było zagadnieniem pobocznym podczas badań zrównujących, decyzje dotyczące konstrukcji testów wykorzystywanych w badaniach, liczby oraz wyboru zadań kotwiczących były podporządkowane celom zrównywania horyzontalnego. W celu uzyskania jak największej trafności badań zrównujących, zeszyty testowe nie mogły być zbyt długie, zadania z wiązek nie powinny zostać rozdzielone, a uczniowie nie mogli rozwiązywać zadań zbyt trudnych lub zbyt łatwych (ze względu na potencjalne obniżenie motywacji). Dlatego też liczba zadań kotwiczących dla skalowania pomiędzy etapami edukacyjnymi jest niewielka. Niesie to za sobą konsekwencje w postaci wyboru specyficznego modelu do skalowania wertykalnego, odbiegającego od klasycznych propozycji opisywanych w literaturze – modelu podwójnego czynnika (*bi-factor model*). Z

²³ W pierwszej edycji (2011) badano wyłącznie uczniów gimnazjum.

szerszym opisem tego modelu oraz z wynikami skalowania wertykalnego Czytelnik może zapoznać się w podrozdziale 6.8.

5. Realizacja badań terenowych

Każdy z czterech etapów badania przebiegał według takich samych procedur, niezależnie od typu szkoły i egzaminu. Ze względu na różne terminy egzaminów dla poszczególnych typów szkół konieczne było odpowiednie dostosowanie ram czasowych niektórych działań. Użyte procedury miały zapewnić jak najwyższe standardy przeprowadzania badań, a także podobieństwo badań do egzaminów, co miało zapewnić wysoką motywację uczniów do rozwiązywania zeszytów testowych. Dzięki doświadczeniom zebranych w każdym roku przeprowadzania badań możliwe było usprawnienie niektórych procedur, jak rekrutacja szkół czy szkolenie ankieterów. Poniżej opisano najważniejsze etapy realizacji badania.

5.1. Rekrutacja szkół

Zgodnie z założeniami projektu rekrutacja szkół była podzielona na trzy etapy:

1. kontakt listowny,
2. kontakt telefoniczny,
3. kontakt osobisty (ankieter).

Listy informujące o badaniu były wysyłane do dyrektorów szkół z próby głównej na początku roku szkolnego (w formie papierowej). Wczesne poinformowanie szkół o możliwości wzięcia udziału w badaniu pozwalało na uwzględnienie czasu potrzebnego na realizację badań w harmonogramie zajęć. Jednocześnie takie same materiały były wysyłane do szkół pocztą elektroniczną. Podczas kontaktu telefonicznego, który następował w dalszej kolejności dyrektorzy szkół mogli uzyskać szersze informacje o badaniu oraz wyrazić swoją zgodę na przeprowadzenie badania lub jej odmówić. W niektórych przypadkach szkoły nie spełniały kryteriów wzięcia udziału w badaniu, najczęściej były to szkoły dla dorosłych. Niestety operat losowania nie zawierał informacji o tym czy szkoła jest szkołą dla dorosłych lub szkołą specjalną. Tylko w przypadku, gdy nazwa szkoły zawierała taką informację można ją było wykluczyć z operatu.

W przypadku odmów następował kontakt ze szkołą z listy zapasowej. Najczęstszą przyczyną odmów były:

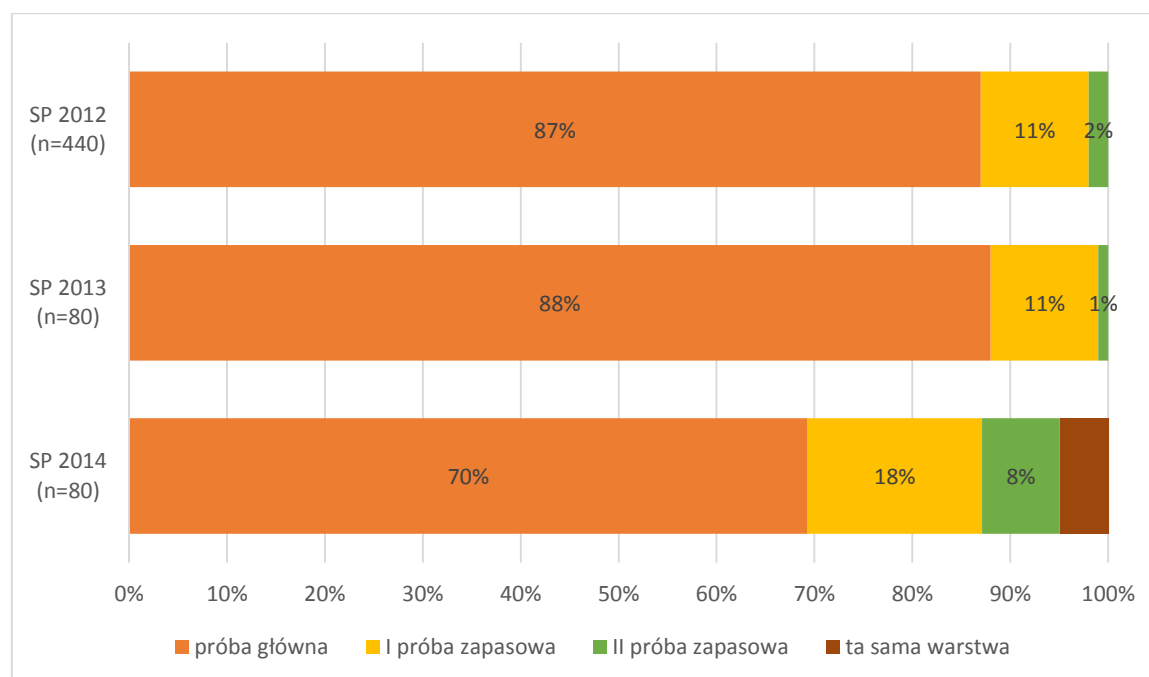
- brak czasu na przeprowadzenie badania – szkoły brały udział w innych badaniach lub egzaminach próbnych,
- zbyt późna prośba o udział w badaniu – dyrektorzy wskazywali, że decyzje o organizacji roku szkolnego podejmują na przełomie sierpnia i września,
- objęcie badaniem tylko jednego oddziału klasowego, co wymaga organizacji innego testu próbnego dla pozostałych oddziałów.

Koncepcja badania przewidywała jego realizację w jednym oddziale z każdej ze szkół, jednakże w niektórych przypadkach w wyniku rekrutacji do badania włączano dodatkowe klasy. Dotyczyło to z reguły szkół, gdzie w ostatnim roczniku były dwa oddziały klasowe i dyrektor szkoły uzależniał swoją zgodę na udział w badaniu od objęcia ich obydwu badaniem. Oprócz dodatkowych oddziałów badania przeprowadzono także w kilku dodatkowych szkołach, z uwagi na zbyt późną decyzję dyrektora o

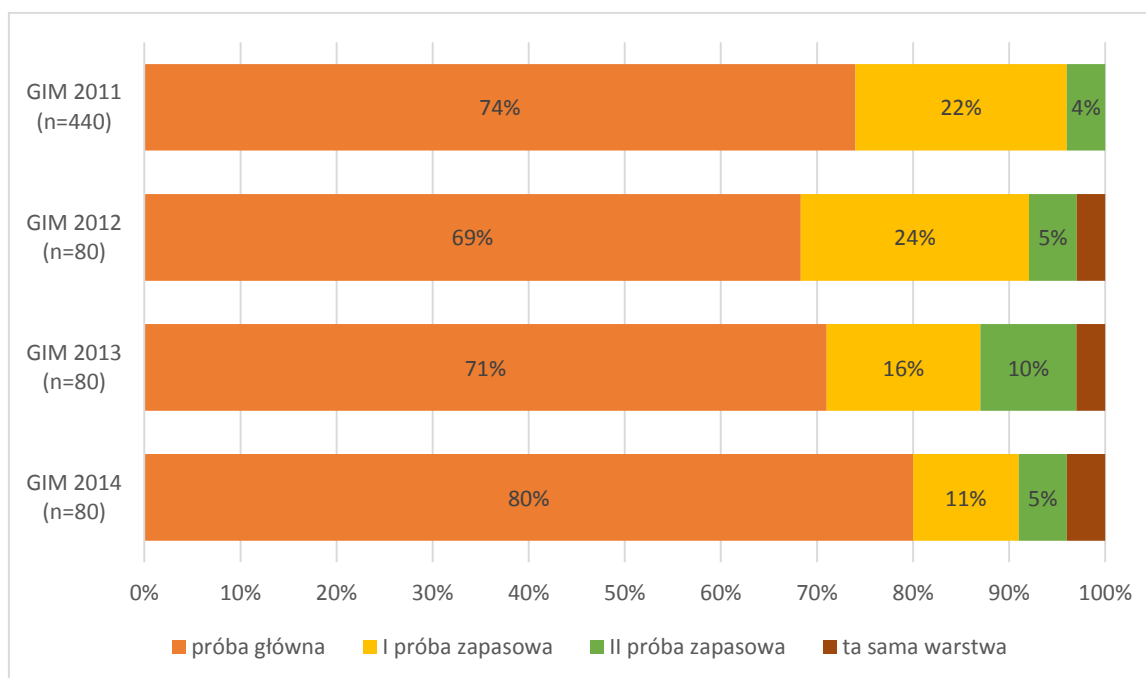
udziale w badaniu w momencie, gdy była już zrekrutowana szkoła zastępcza. Szkoły te były traktowane jako nadmiarowe i nie zastępowały innych szkół z próby.

W zależności od roku badania i typu szkoły liczba rekrutowanych szkół była różna. Największą pulę szkół w danym roku stanowiły te, dla których w danym roku następowało pełne zrównanie wszystkich wcześniejszych edycji egzaminów danego typu, co wynikało z założonego schematu zrównania (por. Rysunek 2.1). W roku 2011 były to gimnazja, w 2012 szkoły podstawowe, w 2013 szkoły ponadgimnazjalne (matura z matematyki), a w 2014 ponownie szkoły ponadgimnazjalne (matura z polskiego i języka angielskiego). Ponadto w latach 2012–2014 przeprowadzano bieżące zrównanie dodatkowej edycji egzaminu dla egzaminów zrównanych w poprzednich latach. Zestawienie liczebności prób szkół dla poszczególnych lat i typów egzaminów zawiera Tabela 4.1.

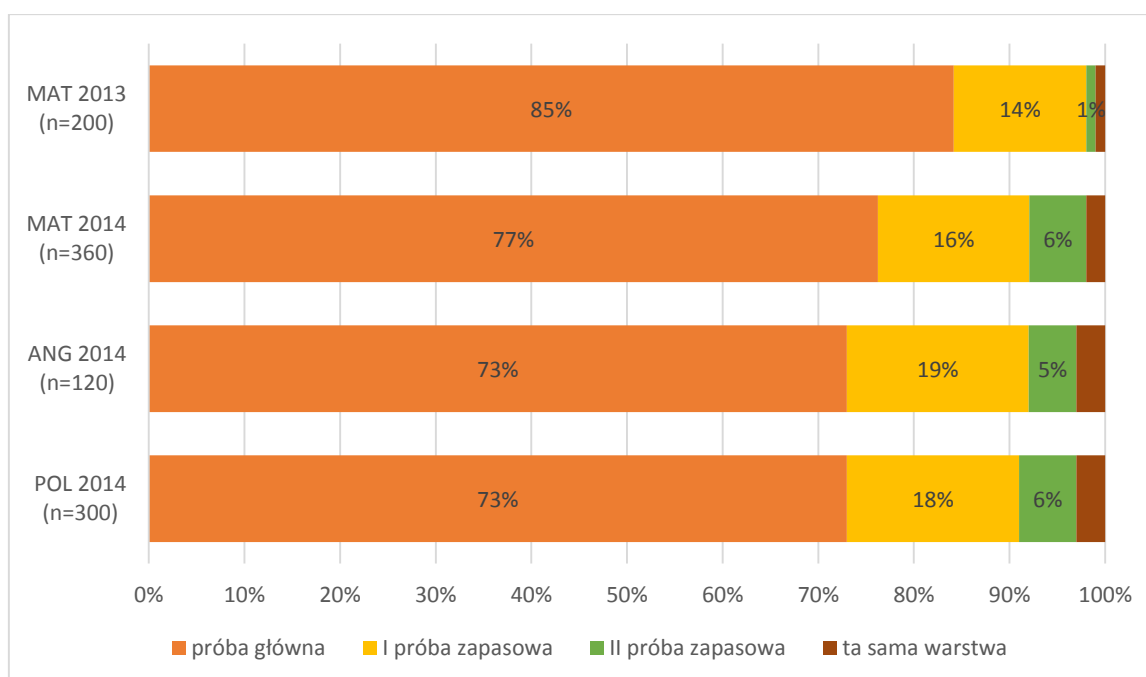
We wszystkich etapach badania udało się zrekrutować wymaganą liczbę szkół zgodnie z założonym schematem. W 2014 roku wymagało to dodatkowego losowania szkół podstawowych i gimnazjalnych z uwagi na wyczerpanie możliwości zastąpienia szkół z pierwotnie losowanych prób. Odsetek szkół z próby głównej, które wzięły udział w badaniu, wahał się w zależności od roku badania i typu szkoły. W przypadku szkół podstawowych w latach 2012 i 2013 było to odpowiednio 87% i 88%, natomiast w roku 2014 69%. W przypadku gimnazjów odpowiednio dla lat 2011-2014 wartości te wynosiły: 74%, 69%, 71% i 80%. W 2013 roku z próby głównej szkół ponadgimnazjalnych w badaniu wzięło udział 85%, a w roku 2014 od 73% do 77% w zależności od przedmiotu. Zdecydowaną większość przypadków, gdy szkoła nie wzięła udziału w badaniu stanowiły odmowy dyrektorów (podjęte samodzielnie lub po konsultacjach z radą pedagogiczną lub rodzicami). Najwięcej przypadków niespełniających kryteriów udziału w badaniu było pośród szkół ponadgimnazjalnych, w 2013 roku było to 3%, a w 2014 od 7% do 8% w zależności od przedmiotu. Dla szkół podstawowych i gimnazjów wartość ta wahała się od 0% do 5%. Pochodzenie szkół w ostatecznej próbie przedstawiają dla szkół podstawowych, gimnazjów i szkół ponadgimnazjalnych odpowiednio: Rysunek 5.1, Rysunek 5.2 i Rysunek 5.3.



Rysunek 5.1. Pochodzenie szkół podstawowych w ostatecznej próbie



Rysunek 5.2. Pochodzenie gimnazjów w ostatecznej próbie



Rysunek 5.3. Pochodzenie szkół ponadgimnazjalnych w ostatecznej próbie

5.2. Badanie pilotażowe

W celu przetestowania procedur badania oraz narzędzi w postaci testów umiejętności szkolnych i kwestionariuszy ankiet na dwa miesiące przed badaniem głównym każdorazowo przeprowadzono badanie pilotażowe. W pilotażu nie był wymagany losowy dobór szkół, w związku z czym stosowano dobór celowy. Liczba szkół biorących w pilotażu w danym roku podyktowana była liczbą zeszytów testowych używanych w każdym roku – istotne było, aby każdy z zeszytów został przetestowany

przynajmniej w jednej szkole. Pilotaż pozwalał na wykrycie błędów w konstrukcji narzędzi, a także pozwalał na weryfikację instrukcji przeprowadzania badania dla ankieterów. Po przeprowadzeniu pilotażu zbierano informacje zarówno od uczniów, jak i od ankieterów i dokonywano odpowiednich korekt w narzędziach i instrukcjach.

5.3. Szkolenie koordynatorów i ankieterów

Przed przystąpieniem do realizacji badania głównego w terenie konieczne było zorganizowanie szkolenia koordynatorów oraz ankieterów. W pierwszym roku badania zastosowano dwustopniową strukturę szkolenia. Członkowie zespołu badawczego przeszkolili koordynatorów terenowych, którzy z kolei szkolili podległych sobie ankieterów. Niestety takie rozwiązanie spowodowało pewne różnice w interpretacji niektórych zapisów instrukcji, co prowadziło do konieczności przesyłania ankieterom pisemnych wyjaśnień.

W celu zachowania spójności przekazywanych informacji i zebrania pytań od wszystkich ankieterów zaangażowanych w badanie w następnych latach (2012, 2013, 2014) zdecydowano się na organizację centralnego szkolenia. Ze względu na dużą liczbę uczestników szkolenie prowadzono w dwóch równoległych grupach. Nad jego kształtem czuwali przedstawiciele zespołu badawczego. Dzięki centralnemu i jednostopniowemu szkoleniu udało się odpowiedzieć na wszelkie wątpliwości ankieterów oraz wprowadzić odpowiednie zmiany i wyjaśnienia w instrukcji przeprowadzania badania i jego dokumentacji.

5.4. Realizacja badań w szkołach i ocenianie

Realizacja badania głównego była w każdym przypadku dostosowana do harmonogramu egzaminów w danym roku. Termin realizacji badania wyznaczano na około miesiąc przed egzaminem, każdorazowo dane zbierane były przez około dwa tygodnie. Termin taki pozwalał na przeprowadzenie badania, ocenę prac uczniów oraz dostarczenie szkołom wyników jeszcze przed właściwym egzaminem. Jednocześnie bliskość egzaminu w czasie powodowała, że uczniowie mogli go potraktować jako sprawdzenie swoich umiejętności, co stanowiło czynnik podwyższający motywację do rozwiązywania zeszytów testowych. Termin taki gwarantował również, że zdecydowana większość zagadnień obecnych na egzaminie została przedstawiona uczniom przez nauczycieli odpowiednich przedmiotów.

Ocenianie zadań otwartych ze wszystkich rodzajów zeszytów odbywało się w weekendy, w ośrodkach oceniania zlokalizowanych w szkołach w Krakowie. Nad całością oceniania czuwali koordynatorzy. Procedury oceniania były tożsame z procedurami stosowanymi w trakcie sesji egzaminacyjnej, a egzaminatorami były osoby posiadające uprawnienia do oceniania prac w sesjach egzaminacyjnych. W przededniu oceniania odbywało się szkolenie przewodniczących zespołów oceniających (PZE). Pierwszego dnia oceniania przewodniczący zespołów na podstawie wcześniejszego szkolenia przygotowali egzaminatorów do procesu oceniania.

Egzaminatorzy swoje oceny kodowali na kartach oceniania dołączonych do prac uczniów. Po zakończeniu oceniania karty były oddzielane od prac i skanowane (razem z odpowiedziami na zadania zamknięte zakodowanymi przez uczniów). Po ręcznym uzupełnieniu danych, w przypadku błędów odczytu skanera, wyniki uczniów były umieszczane w bazie danych. Podczas oceniania zastosowano procedurę podwójnego oceniania minimum 10% prac z każdej szkoły. Została ona, wraz z jej wynikami, opisana w podrozdziale 5.7.

5.5. Progi realizacji

Założony próg realizacji badania w szkole wynosił 85% w przypadku szkół podstawowych i gimnazjów, a 80% w przypadku szkół ponadgimnazjalnych. Część uczniów mogła nie brać udziału w badaniu, przeważnie były to osoby piszące inny arkusz egzaminacyjny niż standardowy dla uczniów bez dysfunkcji i z dysleksją rozwojową. W indywidualnych przypadkach, np. długotrwała nieobecność czy niepełnosprawność utrudniająca udział w badaniu, również podejmowano decyzję o nie włączaniu ucznia do badania. Jeśli podczas głównego terminu badania w szkole nie uzyskiwano wymaganego progu, przeprowadzano w szkole badanie uzupełniające. Z uwagi na to, że w gimnazjum uczniowie rozwiązywali zeszyty testowe podczas dwóch dni (a od 2012 roku dwa zeszyty w ciągu dnia z przerwą) dla każdej części testu progi realizacji obliczano niezależnie. Oznacza to, że nie było wymagane to, aby uczeń rozwiązał wszystkie zeszyty.

W 2011 roku, pomimo badań uzupełniających, nie we wszystkich badanych szkołach udało się osiągnąć wymagany próg. Sytuacja ta dotyczyła 10% szkół i była spowodowana zbyt krótkim okresem przewidzianym na przeprowadzenie badań uzupełniających. W następnych latach badania główne odbywały się w krótszym czasie, co dawało dłuższy okres na realizację ewentualnych badań uzupełniających. W efekcie we wszystkich latach i typach szkół osiągnięto wymagane progi realizacji. Szczegółowe zestawienie stóp realizacji badania w szkole w podziale na lata i typy egzaminu zawiera Tabela 5.1.

Tabela 5.1. Stopy realizacji badania w podziale na lata i typy szkół (egzaminu)

Rok	Typ szkoły (egzamin)	Część egzaminu	Średnia stopa realizacji	Procent szkół, w którym zbadano wszystkich uczniów z wylosowanej klasy
2011	Gimnazjum (egzamin gimnazjalny)	Część humanistyczna	92%	23%
		Część matematyczno-przyrodnicza	92%	21%
2012	Szkoła podstawowa (sprawdzian po VI klasie szkoły podstawowej)	–	96%	48%
	Gimnazjum (egzamin gimnazjalny)	Język polski	96%	46%
		Historia i WOS	96%	45%
		Matematyka	95%	41%
		Przyroda	95%	40%
2013	Szkoła ponadgimnazjalna (matura z matematyki)	Matematyka	93%	28%
	Szkoła podstawowa (sprawdzian po VI klasie szkoły podstawowej)	–	95%	39%

	Gimnazjum (egzamin gimnazjalny)	Język polski	95%	36%
		Historia i WOS	95%	38%
		Matematyka	94%	30%
		Przyroda	94%	29%
2014	Szkoła ponadgimnazjalna (matura z języka polskiego)	–	91%	16%
	Szkoła ponadgimnazjalna (matura z języka angielskiego)	–	91%	18%
	Szkoła ponadgimnazjalna (matura z matematyki)	–	93%	28%
	Szkoła podstawowa (sprawdzian po VI klasie szkoły podstawowej)	–	96%	43%
	Gimnazjum (egzamin gimnazjalny)	Język polski	94%	24%
		Historia i WOS	94%	26%
		Matematyka	94%	26%
		Przyroda	93%	24%

5.6. Kontrola badań terenowych

W losowo wybranych szkołach pracownicy terenowi IBE przeprowadzili kontrolę zgodności badań z ustalonymi procedurami. W trakcie kontroli nie odnotowano rażącego odstępstwa od przyjętych procedur. Zaobserwowano jedynie niewielkie niedopatrzania nie mające wpływu na poprawność zebranego materiału badawczego (głównie związane z nieutrzymywaniem dyscypliny wśród uczniów czy rozdaniem wersji zeszytów w odmienny sposób niż przewidziany). W niektórych przypadkach uczniowie pisali testy w salach nie odpowiadających wytycznym, lecz było to spowodowane skromnymi warunkami lokalowymi poszczególnych szkół. Warto podkreślić, że w tych szkołach właściwy egzamin odbywa się w takich samych warunkach.

5.7. Podwójne ocenianie losowej próby prac

Podczas studium zrównującego 10% prac oceniono podwójnie, podobnie jak podczas egzaminu zewnętrznego. Etapy procedury podwójnego oceniania odzwierciedlały organizację zgodną ze stosowaną podczas sesji egzaminacyjnych, która składa się z trzech wymienionych poniżej etapów:

1. spotkania koordynatora i przewodniczących zespołów egzaminatorów w dniu poprzedzającym ocenianie,
2. szkolenia egzaminatorów do oceniania zadań zawartych w poszczególnych zeszytach testowych,
3. oceniania prac.

Do oceny przez drugiego egzaminatora została wylosowana próba prac, która została skopiowana przed rozpoczęciem oceniania. W przypadku różnych ocen któregośkolwiek z zadań otwartych ostateczną decyzję podejmował przewodniczący zespołu egzaminatorów. Przyznana przez niego punktacja była ostateczna.

Dla każdego zadania otwartego oszacowany został wskaźnik zgodności kappa. Współczynnik kappa pozwala ocenić stopień zgodności dwukrotnych pomiarów tej samej zmiennej. Współczynnik obliczany jest za pomocą poniższego wzoru (Cohen, 1960):

$$\kappa = \frac{\Pr(o) - \Pr(e)}{1 - \Pr(e)}$$

gdzie:

$\Pr(o)$ – obserwowana zgodność kodowania,

$\Pr(e)$ – oczekiwana zgodność kodowania, czyli zgodność, która powstałaby w przypadku losowego rozłożenia wartości skali punktowania danego zadania.

Kappa może przyjmować wartości od -1 do +1. Wartość 0 oznacza zgodność na takim poziomie jaki powstałby dla losowego rozkładu wartości na całej skali punktowania zadania. Szeroko przyjętą interpretację wskaźnika zgodności kappa przedstawia Tabela 5.2.

Tabela 5.2. Interpretacja wskaźnika zgodności kappa (Bland i Altman, 1999)

Poziom zgodności	Wartość kappa	Interpretacja poziomu zgodności
1	poniżej 0	brak zgodności
2	0,00–0,20	niewielki
3	0,21–0,40	dostateczny
4	0,41–0,60	średni
5	0,61–0,80	znaczny
6	0,81–1,00	idealny

Poniżej zaprezentowano oszacowanie zgodności kodowania dla każdego z analizowanych etapów edukacyjnych w poszczególnych latach przeprowadzenia studium zrównującego. W tabelach podano obserwowaną i oczekiwaną zgodność kodowania oraz obliczoną na tej podstawie wartość współczynnika zgodności kappa, wraz z błędem standardowym oraz wartością statystyki testowej z pozwalającej stwierdzić czy współczynnik kappa jest statystycznie różny od zera. Oprócz tego podano numery zadań pozwalające je odnaleźć w Bazie danych IBE (zob. Rozdział 9). Powtarzające się numery oznaczają, że dane zadanie było wykorzystywane w więcej niż jednym zeszycie testowym.

Przeprowadzone analizy pozwalają na stwierdzenie, że wyszczególnione zadania otwarte cechują się w większości bardzo wysoką zgodnością kodowania. Wykorzystanie ich w badaniach zrównujących nie powinno więc w żaden sposób zniekształcać uzyskanych wyników.

5.7.1. Sprawdzian

5.7.1.1. Rok 2012

Procedurze podwójnego oceniania w 2012 roku w przypadku sprawdzianu poddano 1106 prac z 29 zadaniami otwartymi. Tabela 5.3 przedstawia uzyskane wyniki analizy zgodności oceniania zadań otwartych. W przypadku większości zadań osiągnięto wysoką zgodność kodowania: w sześciu zadaniach współczynnik zgodności kappa jest wyższy niż 0,8, co oznacza idealną zgodność, tylko w przypadku pięciu zadań współczynnik kappa jest niższy niż 0,8, jednak wyższy niż 0,6, co można interpretować jako znaczną zgodność.

Tabela 5.3. Zgodność kodowania zadań otwartych – zrównywanie sprawdzianu 2012

Nr zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd std.	z
		obserwowana	oczekiwana			
1	1111	97,98%	59,43%	0,95	0,10	9,45
2	1112	90,72%	24,68%	0,88	0,05	16,64
3	1144	91,46%	27,26%	0,88	0,03	28,49
4	1501	94,86%	35,57%	0,92	0,03	28,92
5	1505	82,78%	51,27%	0,65	0,06	10,70
6	1507	97,45%	58,32%	0,94	0,05	17,64
7	1514	100,00%	47,01%	1,00	0,05	18,51
8	1264	98,79%	46,26%	0,98	0,06	16,65
9	1265	100,00%	53,75%	1,00	0,08	13,38
10	1299	94,53%	25,30%	0,93	0,03	28,70
11	1303	99,66%	37,03%	1,00	0,04	23,62
12	1253	98,59%	52,53%	0,97	0,12	8,18
13	1336	92,86%	22,12%	0,91	0,04	24,34
14	1340	76,02%	36,48%	0,62	0,05	13,10
15	1277	92,52%	63,96%	0,79	0,06	13,71
16	1369	98,56%	35,12%	0,98	0,05	19,70
17	1370	96,15%	32,80%	0,94	0,05	20,82
18	1325	97,83%	45,32%	0,96	0,08	11,56
19	1501	94,86%	35,57%	0,92	0,03	28,92
20	1402	97,74%	42,56%	0,96	0,06	16,29
21	1507	97,45%	58,32%	0,94	0,05	17,64
22	1436	96,34%	36,39%	0,94	0,04	21,69

23	1437	99,25%	55,35%	0,98	0,05	20,71
24	1444	89,16%	32,95%	0,84	0,04	22,49
25	1144	91,46%	27,26%	0,88	0,03	28,49
26	1160	87,72%	41,54%	0,79	0,05	16,63
27	1193	96,83%	40,69%	0,95	0,06	17,37
28	1195	94,27%	27,37%	0,92	0,04	21,81
29	1231	81,55%	28,68%	0,74	0,04	17,08

5.7.1.2. Rok 2013

W 2013 roku w części dotyczącej zrównywania sprawdzianu procedurze podwójnego oceniania poddano 195 prac z 11 zadaniami otwartymi. Tabela 5.4 przedstawia uzyskane wyniki analizy zgodności kodowania. W przypadku większości zadań osiągnięto wysoką zgodność kodowania: w dziewięciu zadaniach współczynnik zgodności kappa jest wyższy niż 0,8, co oznacza idealną zgodność. Dla dwóch zadań osiągnięto zgodność kodowania rzędu 0,64 i 0,75, jednak zgodnie z regułami interpretacji, które prezentuje Tabela 5.2, można interpretować ją jako znaczną.

Tabela 5.4. Zgodność kodowania zadań otwartych – zrównywanie sprawdzianu 2013

Nr zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd std.	z
		obserwowana	oczekiwana			
1	1507	94,37%	63,54%	0,85	0,12	7,15
2	1501	93,15%	34,85%	0,89	0,08	10,75
3	1436	100,00%	38,88%	1,00	0,08	12,34
4	1437	96,15%	59,70%	0,90	0,09	10,22
5	1444	74,12%	28,10%	0,64	0,06	10,06
6	1466	97,10%	28,73%	0,96	0,07	13,29
7	1466	96,67%	28,11%	0,95	0,06	15,85
8	1515	97,50%	33,75%	0,96	0,08	12,13
9	1179	97,89%	42,20%	0,96	0,08	12,02
10	1193	85,42%	40,52%	0,75	0,07	11,33
11	1195	92,31%	21,99%	0,90	0,06	15,84

5.7.1.3. Rok 2014

Podczas zrównywania sprawdzianu w 2014 roku procedurze podwójnego oceniania poddano 190 prac z siedmioma zadaniami otwartymi. Tabela 5.5 podaje uzyskane wyniki analiz zgodności kodowania. W przypadku pięciu zadań współczynnik kappa jest wyższy od 0,8, co oznacza idealną zgodność kodowania. Pozostałe zadania cechują się zgodnością kodowania w przedziale od 0,68 do 0,73, czyli ich zgodność kodowania jest znaczna.

Tabela 5.5. Zgodność kodowania zadań otwartych – zrównywanie sprawdzianu 2014

Nr zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd std.	z
		obserwowana	oczekiwana			
1	4015	96,84%	82,40%	0,82	0,06	14,47
2	4016	82,63%	45,79%	0,68	0,06	10,68
3	1505	84,74%	43,20%	0,73	0,06	12,36
4	1466	94,74%	23,33%	0,93	0,04	24,65
5	1501	94,74%	26,25%	0,93	0,04	21,81
6	4011	98,95%	37,57%	0,98	0,05	18,57
7	4013	98,95%	34,08%	0,98	0,05	19,77

5.7.2. Egzamin gimnazjalny

5.7.2.1. Rok 2011

Podczas zrównywania egzaminu gimnazjalnego w 2011 roku, procedurze podwójnego oceniania poddano zadania z obydwu części egzaminu. W części humanistycznej ponownej ocenie poddano 960 prac uczniów, zawierających w sumie 183 zadania otwarte znajdujące się w 11 zeszytach testowych. W części matematyczno-przyrodniczej oceniano w sposób podwójny 950 prac uczniów, zawierających w sumie 58 zadań wchodzących w skład 11 zeszytów testowych. Ze względu na rozmiary tabel, zaprezentowano je w Aneksach 6.1 oraz 6.2. Większość zadań w części humanistycznej posiada idealną, bądź znaczną zgodność kodowania – jedynie w przypadku trzech z nich zgodność kodowania można określić jako średnią. Analizując zadania otwarte w części matematyczno-przyrodniczej możemy dojść do podobnego wniosku. Zgodność kodowania jest idealna, bądź znaczna dla wszystkich analizowanych zadań.

5.7.2.2. Rok 2012

Na etapie zrównywania egzaminu gimnazjalnego w 2012 roku procedurze podwójnego oceniania poddano 293 prace uczniów, zawierające 11 zadań otwartych, w tym trzy zadania z zakresu przedmiotów matematyczno-przyrodniczych i osiem zadań humanistycznych. Tabela 5.6 przedstawia wyniki analizy zgodności oceniania tych zadań otwartych. We wszystkich zadaniach wartość współczynnika kappa jest wyższa niż 0,8, co oznacza bardzo wysoką zgodność kodowania. W przypadku pięciu zadań osiągnięto 100% zgodność kodowania.

Tabela 5.6. Zgodność kodowania zadań otwartych – zrównywanie egzaminu gimnazjalnego 2012

Zeszyt testowy	Nr zadania	Numer kryterium/zadania w bazie IBE	Zgodność		kappa	Błąd std.	z
			obserwowana	oczekiwana			
M	1	1595	100,00%	43,65%	1,00	0,06	16,61
	2	1596	91,72%	28,45%	0,88	0,05	19,73
	3	1501	100,00%	39,48%	1,00	0,06	16,30
JP	1	1544	99,51%	53,63%	0,99	0,07	14,13
	2	1545	93,65%	52,45%	0,87	0,07	11,99
	3	1546	96,25%	42,94%	0,93	0,06	14,93
	4	1547	97,87%	73,16%	0,92	0,07	12,66
	5	1548	98,99%	50,73%	0,98	0,07	13,79
	6	1505	96,48%	57,26%	0,92	0,07	12,97
	7	1627	97,96%	42,85%	0,96	0,06	15,47
	8	1557	99,39%	50,47%	0,99	0,08	12,69
	9	1561	100,00%	50,04%	1,00	0,08	13,38
	10	1562	100,00%	57,40%	1,00	0,07	13,49
	11	1563	97,42%	51,99%	0,95	0,08	11,78

M – matematyka, JP – j. polski

5.7.2.3. Rok 2013

Podczas zrównywania egzaminu gimnazjalnego w 2013 roku procedurze podwójnego oceniania poddano 324 prace z 11 zadaniami otwartymi (trzema z matematyki, ośmioma z języka polskiego). Tabela 5.7 zawiera uzyskane wyniki analiz zgodności oceniania. Można zauważyć, że w przypadku dziewięciu zadań zgodność kodowania jest idealna (przyjmuje wartości większe niż 0,8), natomiast w przypadku 2 zadań jest ona znaczna (0,73 i 0,76).

Tabela 5.7. Zgodność kodowania zadań otwartych – zrównywanie egzaminu gimnazjalnego 2013

Zeszyt testowy	Nr zadania	Nr kryterium/zadania w bazie IBE	Zgodność		Kappa	Błąd std.	Z
			obserwowana	oczekiwana			
M	1	1595	98,10%	41,78%	0,97	0,05	20,69
	2	1596	97,82%	24,69%	0,97	0,04	27,08
	3	1501	97,50%	38,97%	0,96	0,05	20,44
JP	4	1548	98,37%	52,39%	0,97	0,06	16,90
	5	516	97,96%	38,12%	0,97	0,05	20,40
	6	1505	89,53%	60,93%	0,73	0,06	12,67
	7	1557	99,62%	50,01%	0,99	0,06	16,06
	8	1561	98,05%	50,15%	0,96	0,06	15,41
	9	1562	100,00%	56,54%	1,00	0,06	16,79
	10	1563	94,14%	51,27%	0,88	0,07	13,18
	11	546	84,44%	34,76%	0,76	0,04	17,58

M – matematyka, JP – j. polski

5.7.2.4. Rok 2014

W 2014 roku podczas zrównywania egzaminu gimnazjalnego procedurze podwójnego oceniania poddano 365 prac z siedmioma zadaniami otwartymi (trzema z matematyki, czterema z języka polskiego). Tabela 5.8 zawiera wyniki analizy zgodności kodowania tych zadań. Można zauważyć, że w przypadku trzech zadań współczynnik kappa jest na poziomie wyższym lub równym 0,95; natomiast w przypadku czterech zadań na poziomie wyższym lub równym 0,8. Zgodnie z regułami interpretacji wielkości współczynników kappa, można przyjąć, że dla wszystkich analizowanych zadań zgodność oceniania jest idealna.

Tabela 5.8. Zgodność kodowania zadań otwartych – zrównywanie egzaminu gimnazjalnego 2014

Zeszyt testowy	Numer zadania	Numer kryterium/zadania w bazie IBE	Zgodność		Kappa	Błąd std.	Z
			Obserwowana	Oczekiwana			
M	1	1595	96,99%	37,57%	0,95	0,04	26,45
	2	1501	98,36%	35,79%	0,97	0,03	28,87
	3	4150	91,78%	35,77%	0,87	0,03	26,08
JP	4	1505	93,42%	39,77%	0,89	0,04	22,40
	5	546	89,32%	31,50%	0,84	0,03	27,68
	6	1548	98,90%	36,29%	0,98	0,04	26,64
	7	4172	86,58%	32,14%	0,80	0,03	25,76

M – matematyka, JP – j. polski

5.7.3. Egzamin maturalny

5.7.3.1. Rok 2013

Podczas oceniania zadań dotyczących zrównywania egzaminu maturalnego z matematyki w 2013 roku podwójnie ocenionych zostało 535 prac (66 zadań otwartych). Ze względu na rozmiary tabeli, zawiera

ją Aneks 6.3 wraz z numerami zadań pozwalającymi je odnaleźć w Centralnej Bazie IBE. W przypadku powtarzania się identyfikatorów, oznacza to, że dane zadanie wykorzystano więcej niż w jednym zeszycie testowym. W większości przypadków zgodność kodowania można uznać za znaczną lub idealną. Jednakże dla siedmiu zadań zgodność kodowania można zaklasyfikować jako średnią lub dostateczną, natomiast dla 11 zadań jako niewielką lub jako brak zgodności. Należy także zauważyć, że w części przypadków zgodność kodowania jest różna dla tych samych zadań znajdujących się w różnych zeszytach.

5.7.3.2. Rok 2014

W 2014 roku podczas zrównywania wyników egzaminu maturalnego procedurze podwójnego oceniania poddano 776 prac z języka polskiego (216 zadań otwartych), 306 z matematyki (44 zadania otwarte) i 876 z języka angielskiego (136 zadań otwartych). Ze względu na rozmiary tabel zawierających wyniki analizy zgodności oceniania, zawierają je Aneksy 6.4, 6.5 i 6.6. Analiza współczynników kappa pozwala wnioskować o bardzo wysokiej zgodności kodowania w przypadku wszystkich przedmiotów. Współczynniki kappa przyjmują wartości od 0,87 do 1 dla języka polskiego, 0,94 do 1 dla matematyki oraz 0,86 do 1 dla języka angielskiego.

Prezentowane wyniki z języka polskiego mogą się różnić od tych przedstawionych w publikacji autorstwa Henryka Szaleńca i współpracowników (w druku) dotyczącej efektu egzaminatora. Należy jednak pamiętać, że cel leżący u podłoża przeprowadzenia tamtych badań był diametralnie inny. W przytoczonej monografii chodziło o analizę fenomenu efektu egzaminatora (jako procesu psychologicznego) oraz skali występowania tego zjawiska. Procedura przeprowadzenia tamtych badań różniła się zdecydowanie od opisywanego podwójnego kodowania: zadania otwarte były oceniane przez większą liczbę egzaminatorów (8) oraz bez możliwości uzgadniania ocen pomiędzy oceniającymi oraz przewodniczącym zespołu egzaminatorów. Dlatego też zgodność podwójnego oceniania jest zdecydowanie wyższa niż sugerowałyby wyniki analiz związanych z efektem egzaminatora.

6. Wyniki zrównywania

W tym rozdziale przedstawione zostaną najważniejsze wyniki badań zrównujących. Przeprowadzone analizy umożliwiły rozdzielenie efektu zmian związanych z trudnością różnych edycji tego samego egzaminu od efektu zmian w poziomie umiejętności uczniów między latami. Dzięki temu niezależnie można zanalizować wkład każdego z tych czynników. To z kolei pozwala, z jednej strony, na krytyczną analizę jakości egzaminów (gdyż oczekujemy utrzymywania jak najbardziej zbliżonej trudności egzaminu między latami), a z drugiej strony, na przeprowadzenie bardzo istotnej z punktu widzenia polityki oświatowej analizy wieloletnich trendów w zakresie poziomu umiejętności uczniów.

W pierwszych sześciu podrozdziałach zostaną opisane wyniki zrównania dla kolejnych egzaminów: sprawdzianu (6.1), egzaminu gimnazjalnego z części matematyczno-przyrodniczej (6.2), egzaminu gimnazjalnego z części humanistycznej (6.3) oraz trzech egzaminów maturalnych z poziomu podstawowego – matury z matematyki (6.4), z języka polskiego (6.5) i z języka angielskiego (6.6). Opis każdego z sześciu egzaminów zostanie przeprowadzony w analogiczny sposób. Najpierw, wykorzystując możliwość przedstawienia wyników uczniów w każdym z badanych roczników na skali wyników surowych egzaminu z roku 2012 zostanie zanalizowany problem zmian poziomu trudności egzaminów między latami. Następnie, dla wszystkich lat objętych badaniem wyniki uczniów zostaną przedstawione na wspólnej skali zmiennej ukrytej wraz z informacją o błędach standardowych oszacowań, co pozwoli na analizę problemu długookresowych zmian umiejętności. Jedną z zalet wykorzystania zmiennej ukrytej do analizy zmian w poziomie umiejętności jest fakt, że jej rozkład dla każdego rocznika jest bardzo zbliżony do normalnego i korzystając jedynie ze średniej oraz odchylenia standardowego jesteśmy w stanie przekazać praktycznie całą istotną informację o różnicach między porównywanymi populacjami uczniów.

W podrozdziale 6.7 przedstawione zostaną wyniki analizy zrównującej przeprowadzonej na danych zebranych podczas badań w ramach projektu, ale z wykorzystaniem klasycznego, nieodwołującego się do modelu IRT, zrównywania ekwicytylowego. Zasadniczym celem tej analizy jest weryfikacja wyników uzyskanych w podejściu opartym o IRT. Na koniec, w podrozdziale 6.8, zebrano wyniki zrównywania wertykalnego, którego celem było oszacowanie wielkości przyrostu umiejętności uczniów pomiędzy kolejnymi szczeblami edukacji. Przeprowadzenie zrównywania wertykalnego było możliwe dzięki uwzględnieniu wspólnych zadań dodatkowych w badaniach zrównujących przeprowadzonych dla różnych egzaminów.

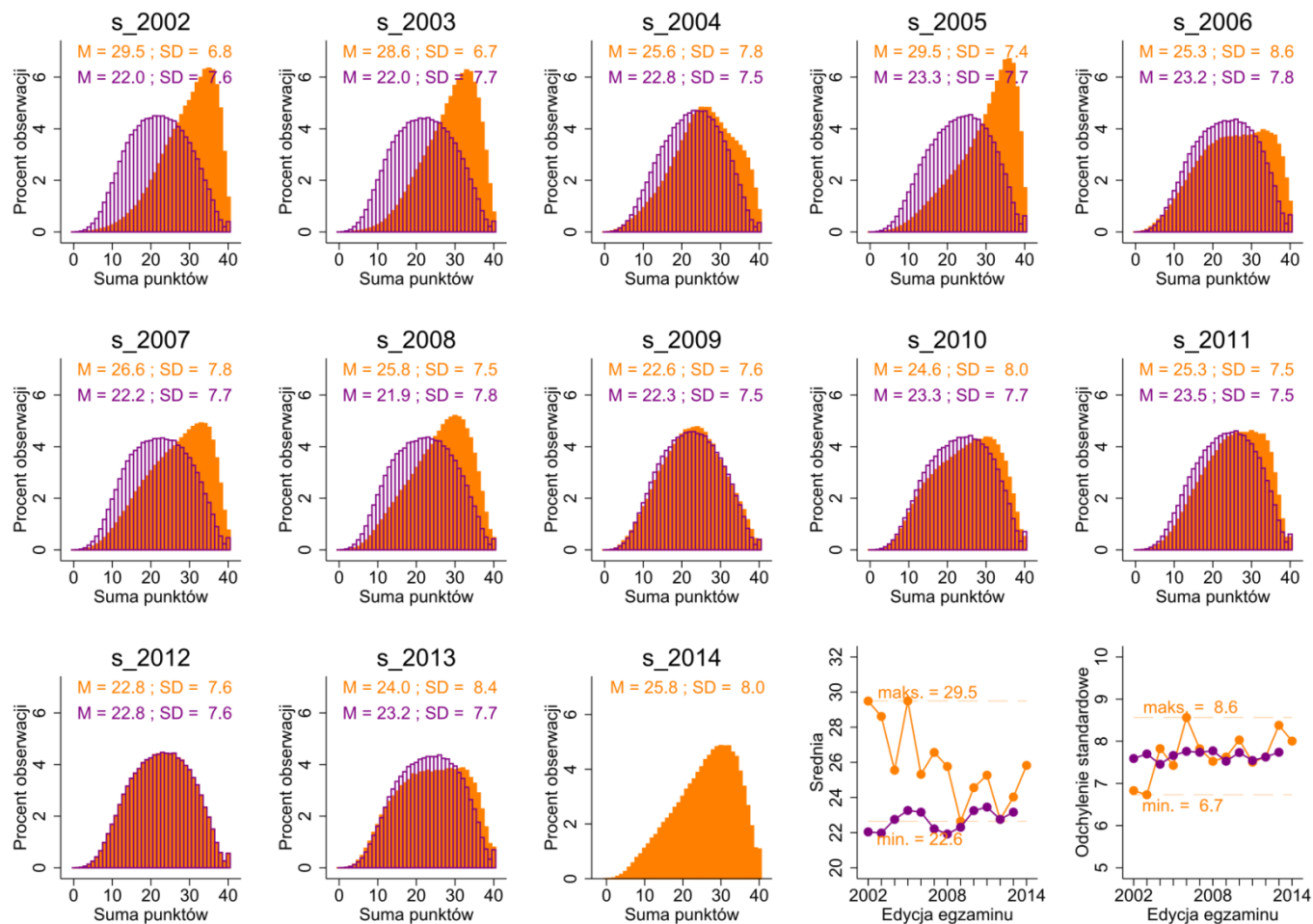
6.1. Sprawdzian

6.1.1. Zmiany trudności egzaminu w latach 2002–2013

W wyniku zastosowanej procedury zrównywania dysponujemy zakotwiczonym na wspólnej skali rozkładem poziomu umiejętności dla każdego roku oraz parametrami zadań opisującymi prawdopodobieństwo udzielenia określonej odpowiedzi na zadania egzaminacyjne w zależności od poziomu umiejętności uczniów. Informacje te pozwalają na oszacowanie, jak wyglądałby rozkład „zwykłej” sumarycznej liczby punktów (zwanej dalej wynikami obserwowanymi) z dowolnego egzaminu w latach 2002–2013, gdyby był rozwiązywany przez populację uczniów z dowolnej kohorty w latach 2002–2013. W szczególności, możliwe jest przedstawienie wyników obserwowanych wszystkich edycji sprawdzianów na skali wyników obserwowanych sprawdzianu z roku 2012, który został wybrany jako referencyjny.

Rysunek 6.1 przedstawia rozkłady rzeczywistych oraz wyrażonych na skali z roku 2012 wyników ze sprawdzianu w latach 2002–2013 wraz z informacjami o średniej (M) oraz odchyleniu standardowym (SD) dla wersji oryginalnej i zrównanej. Na wykresie przedstawiono również wyniki sprawdzianu z roku 2014, którego nie objęła procedura zrównywania ex-post przeprowadzonego w ramach badań. Rysunek 6.1 przedstawia zmiany średniej oraz odchylenia standardowego oryginalnych oraz zrównanych wyników obserwowanych na dwóch ostatnich wykresach. Pierwszą narzucającą się obserwacją jest to, że zmienność kształtu rozkładu, jak i najważniejszych jego parametrów pomiędzy niezrównanymi obserwowanymi wynikami na przestrzeni lat jest o wiele większa niż pomiędzy wynikami zrównanymi. Jest to sygnałem, że występujące między latami fluktuacje rozkładu wyników egzaminów są przede wszystkim konsekwencją zmian w trudności poszczególnych arkuszy, a w mniejszym stopniu zmian w poziomie umiejętności uczniów.

Patrząc na wartości średnich obserwowanych wyników sprawdzianu przed zrównaniem zauważamy silny trend spadkowy między rokiem 2002, a 2013. Średnie niezrównane wyniki w pierwszych dwóch edycjach wynosiły odpowiednio 29,5 oraz 28,5 punktu, gdy w ostatnich trzech objętych zrównaniem latach (2011–2013) notujemy wartości odpowiednio: 25,3, 22,8 oraz 24 punkty. Oznacza to spadek sięgający w skrajnym przypadku prawie 7 punktów na 40 możliwych do uzyskania w teście. Po zrównaniu okazuje się natomiast, że wyniki wyrażone na wspólnej skali 2012 dla trzech ostatnich lat są tak naprawdę wyższe niż dla pierwszych edycji sprawdzianu, a jednocześnie różnice między wszystkimi analizowanymi edycjami sprawdzianu nie wykraczają co do absolutnej wartości poza 1,6 punktu.



Rysunek 6.1. Rozkłady wyników obserwowanych ze sprawdzianu w latach 2002-2014. Pełne słupki odnoszą się do niezrównanych wyników egzaminacyjnych, zrównane wyniki przedstawiono za pomocą słupków pustych

Tak silna fluktuacja trudności egzaminów między latami pokazuje, jak dalece mijalibyśmy się z prawdą przyjmując surowe wyniki sprawdzianu za porównywalną miarę zmian poziomu osiągnięć uczniów na przestrzeni lat. Wynik ten również w negatywnym świetle stawia dotychczasowe rozwiązania obowiązujące w systemie egzaminacyjnym w zakresie konstruowania testu – nie udało się w przypadku sprawdzianu utrzymać stabilnej trudności egzaminu na przestrzeni lat. W tym kontekście zaproponowana procedura przedstawiania rozkładów wyników obserwowanych różnych egzaminów w zestawieniu z wynikami zrównanymi do egzaminu z roku referencyjnego ukazuje się jako interesujące narzędzie do monitorowania jakości egzaminów (spodziewamy się utrzymywania możliwie stałej trudności tego samego egzaminu w różnych latach). Ustaliwszy rok referencyjny można w łatwy sposób ocenić stopień w jakim trudność egzaminu odbiegała od przyjętego układu odniesienia po prostu odejmując średnią zaobserwowaną dla danego egzaminu od średniej na skali wyników obserwowanych roku referencyjnego. Najbardziej trudnością (łatwością) od roku referencyjnego odstają lata 2002, 2003 oraz 2005, gdzie mamy wzrost średniego wyniku uczniów odpowiednio o 7,5, 6,6 oraz 6,2 punktu względem wyniku jaki uczniowie w tych latach uzyskaliby rozwiązując test z roku 2012.

Zasymulowane na podstawie modelu IRT (zob. podrozdział 4.8) rozkłady wyników obserwowanych sprawdzianu z 2012 roku dla wszystkich populacji uczniów w latach 2002–2013 pozwalają na stworzenie tablic przeliczeniowych. Umożliwiają one przyporządkowanie uczniowi uzyskującemu dany wynik na egzaminie przeprowadzonym w dowolnym roku wynik, jaki uzyskałby na sprawdzianie z 2012 roku. Wystarczy w tym celu dokonać zrównania ekwicyntylowego (zob. podrozdział 4.7) wyników obserwowanych testu z roku X z wynikami obserwowanymi w teście 2012 dla uczniów pochodzących do egzaminu w roku X. Tabela 6.1 przedstawia tablicę przeliczeniową dla sprawdzianu, która i zostanie omówiona w dalszej części rozdziału. Natomiast dla kolejnych egzaminów tablice pozwalające na przeliczenie wyników obserwowanych na edycję egzaminu z roku 2012 ze względu na rozmiary zawiera Aneks 1, a ich opis zostanie ograniczony do koniecznego minimum.

Analiza danych ukazanych w tabeli przeliczeniowej (zob. Tabela 6.1) pozwala na zauważenie kolejnych kilku bardzo interesujących zależności między wynikami uczniów z różnych lat w egzaminach, które różnią się łatwością. Dla przykładu, wynik ucznia, który w 2002 lub 2003 roku uzyskał 20 punktów (środek skali), powinien zostać przeliczony na zrównany wynik niższy odpowiednio o 8 lub 7 punktów na skali zrównanej, co odpowiada różnicom rzędu 17,5–20% wyniku sumarycznego. Jednocześnie różnice między średnimi wynikami po zrównaniu między latami 2002–2003 a rokiem 2012 wynoszą w obu przypadkach 0,8 punktu. Podobnie duże rozbieżności dla wybranych lat dotyczą nie tylko uczniów o wynikach surowych równych 20, ale dość szerokiego przedziału punktowego od 16 do 34 punktów. Wniosek ten jest szczególnej wagi, gdyż w tym zakresie punktów znajduje się większość populacji uczniów rozwiązujących egzaminy w latach 2002–2003. Warto odnotowania jest również, że na krańcach skali rozbieżności między uzyskanym a przeliczonym na skalę z 2012 roku wynikiem zaczynają raptownie maleć. Taka nieliniowość przekształcenia zrównującego (widoczna również dla innych zestawień niż 2002 i 2003 z 2012), nakazuje wyciągnąć wniosek, że stosowanie wszelkich poprawek o liniowym charakterze (jak na przykład standaryzacja) nie będzie w stanie rozwiązać problemu nierównoważnych trudności egzaminów między latami.

Tabela 6.1. Tablica przeliczeniowa obserwowanych wyników sprawdzianu na wyniki obserwowane w roku 2012

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012											
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
0	1	0	1	0	1	0	0	1	1	0	0	1
1	1	1	2	1	2	1	1	1	1	1	1	1
2	2	1	3	2	3	2	1	2	2	2	2	2
3	2	2	4	2	4	3	2	3	3	3	3	3
4	3	2	4	3	5	3	3	4	4	4	4	4
5	3	3	5	4	6	4	4	5	5	5	5	5
6	4	4	6	4	7	5	5	6	6	6	6	6
7	4	4	7	5	7	6	5	7	7	6	7	7
8	5	5	8	6	8	6	6	8	8	7	8	8
9	5	5	8	6	9	7	7	9	9	8	9	9
10	6	6	9	7	10	8	8	10	10	9	10	10
11	6	6	10	8	11	9	9	11	11	10	11	11
12	7	7	11	8	12	9	9	12	12	11	12	12

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012											
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
13	8	7	11	9	12	10	10	13	13	12	13	13
14	8	8	12	10	13	11	11	14	13	12	14	14
15	9	9	13	10	14	12	12	15	14	13	15	15
16	9	9	14	11	15	12	12	16	15	14	16	16
17	10	10	14	12	16	13	13	17	16	15	17	17
18	11	11	15	13	17	14	14	18	17	16	18	18
19	11	12	16	13	17	15	15	19	18	17	19	19
20	12	13	17	14	18	15	15	20	19	18	20	19
21	13	13	18	15	19	16	16	21	20	19	21	20
22	14	14	19	15	20	17	17	22	21	20	22	21
23	14	15	20	16	21	18	18	23	21	21	23	22
24	15	16	21	17	22	19	19	24	22	22	24	23
25	16	17	22	18	23	20	20	25	23	23	25	24
26	17	18	23	18	23	21	21	26	24	24	26	25

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012											
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
27	18	19	24	19	24	22	22	27	25	25	27	26
28	19	20	25	20	25	23	23	28	26	26	28	27
29	20	21	26	21	26	24	25	29	27	27	29	27
30	21	22	27	22	27	25	26	30	28	28	30	28
31	22	24	28	23	28	26	27	31	29	29	31	29
32	24	25	29	24	29	27	28	32	30	30	32	30
33	25	27	30	26	30	28	30	32	31	31	33	31
34	26	28	31	27	31	30	31	33	32	32	34	32
35	28	30	32	29	32	31	32	34	34	34	35	33
36	30	31	33	30	33	32	34	35	35	35	36	34
37	31	33	34	32	34	34	35	36	36	36	37	36
38	33	35	35	34	35	35	36	37	37	38	38	37
39	36	37	37	36	37	37	38	38	39	39	39	38
40	38	40	39	39	40	40	40	40	40	40	40	40

6.1.2. Wyniki w latach 2002–2013 na skali zmiennej ukrytej

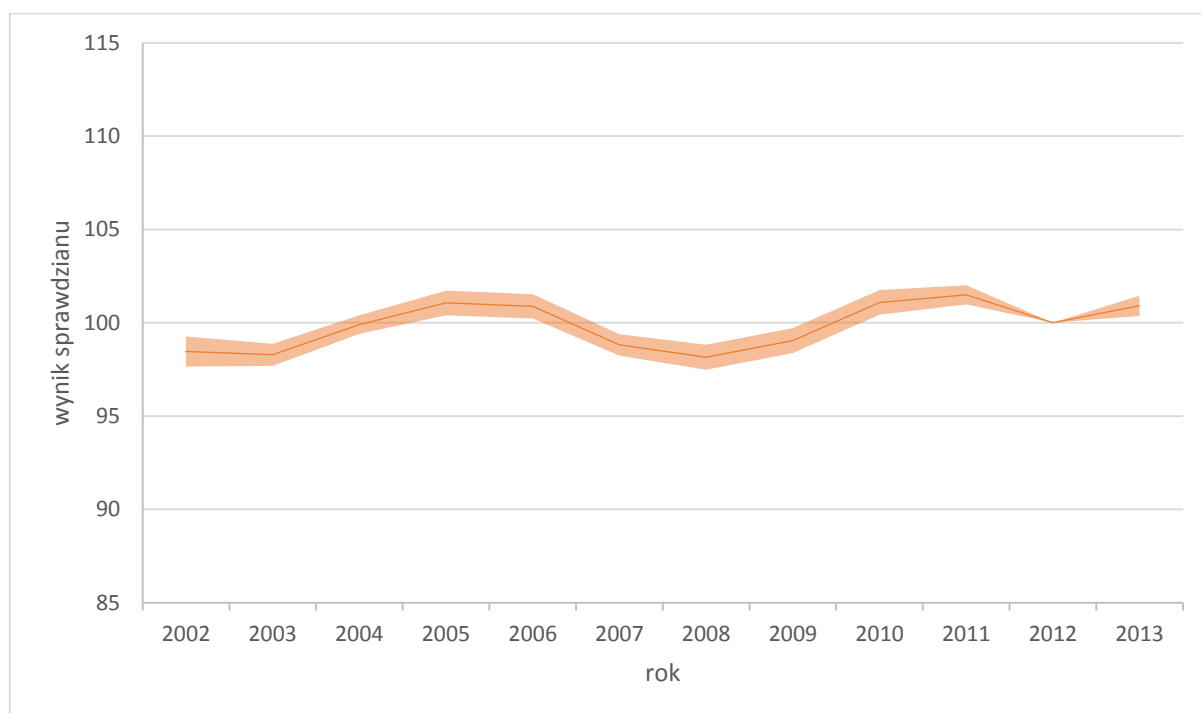
Powyżej skupiliśmy się na analizie zmian trudności egzaminów z różnych lat. W tej części skupimy się na poziomie umiejętności uczniów. Przy pomocy danych egzaminacyjnych postaramy się odpowiedzieć na fundamentalne pytanie z perspektywy szeroko rozumianej polityki oświatowej: czy i jak zmienił się poziom umiejętności uczniów kończących szkołę podstawową w przeciągu lat 2002–2013?

Dopasowanie odpowiedniego modelu IRT do danych egzaminacyjnych połączonych z danymi zebranymi w dodatkowych badaniach zrównujących umożliwiło oszacowanie na wspólnej skali zmiennej ukrytej średniej oraz odchylenia standardowego rozkładu umiejętności uczniów w latach 2002–2013 (metodologię szczegółowo opisano w rozdziale 4). Dla ułatwienia interpretacji wyniki zostały przekształcone liniowo do skali o średniej 100 oraz odchyleniu standardowym 15 – taka skala jest łatwiejsza do prezentacji, ponieważ nie daje ujemnych wyników, jest jedną z najbardziej znanych skal standardowych (np. prezentowane są na niej wyniki IQ), ponadto używana jest już do komunikacji wyników polskich badań np. w EWD oraz OBUT. Tabela 6.2 przedstawia wartości średniej oraz odchylenia standardowego, wraz z wielkością błędu standardowego oraz szerokością 95% przedziału ufności dla przekształconych wyników. Rysunek 6.2 ilustruje dodatkowo średnie z przedziałami ufności.

Tabela 6.2. Średnia oraz odchylenie standardowe zmiennej umiejętności dla sprawdzianu w latach 2002–2013, wyniki zrównane na skali 100; 15 zakotwiczone w roku 2012

Rok egzaminu	Średnia	Błąd standard.	95% przedział ufn.		Odch. stand.	błąd standard.	95% przedział ufn.	
			dolna gr.	górna gr.			dolna gr.	górna gr.
2002	98,5	0,41	97,6	99,3	15,0	0,53	13,9	16,0
2003	98,3	0,30	97,7	98,9	15,3	0,42	14,5	16,1
2004	99,9	0,25	99,4	100,4	14,5	0,34	13,8	15,1
2005	101,1	0,34	100,4	101,7	15,1	0,48	14,2	16,1
2006	100,9	0,33	100,2	101,5	15,4	0,46	14,5	16,3
2007	98,8	0,29	98,2	99,4	15,3	0,39	14,5	16,1
2008	98,2	0,34	97,5	98,8	15,5	0,46	14,6	16,4
2009	99,0	0,34	98,4	99,7	14,7	0,43	13,8	15,5
2010	101,1	0,34	100,4	101,8	15,3	0,45	14,4	16,2
2011	101,5	0,26	101,0	102,0	14,7	0,34	14,1	15,4
2012	100,0	–	–	–	15,0	–	–	–
2013	100,9	0,28	100,4	101,4	15,3	0,39	14,5	16,1

Ponieważ parametry rozkładu umiejętności w 2012 roku zostały ustalone, ich wartość nie wiąże się z jakąkolwiek niepewnością. Tabela 6.2 nie zamieszcza więc dla nich informacji o błędzie pomiaru, ani o szerokości 95% przedziału ufności. Dla wszystkich innych parametrów ta informacja pozwala na odpowiedź na pytanie odnośnie tego czy różnią się one w sposób istotny statystycznie względem wartości w roku referencyjnym, a także między sobą. Dany parametr uznajemy za statystycznie istotnie (na poziomie $\alpha=0,05$) różny od wartości w roku referencyjnym, jeżeli jego 95% przedział ufności jest rozłączny z wartością roku referencyjnego. Analogicznie, dla dwóch niereferencyjnych roczników istotność statystyczną określamy jako rozłączność ich przedziałów ufności.



Rysunek 6.2. Średnie zrównane wyniki sprawdzianu na skali 100;15 wraz z 95% przedziałem ufności

Widzimy, że w latach 2002–2003 oraz 2007–2009 średni poziom umiejętności uczniów mierzony przy pomocy sprawdzianu był statystycznie istotnie niższy niż w roku 2012, natomiast w latach 2005–2006, 2010–2011 oraz 2013 istotnie wyższy niż w roku 2012, a wyniki z roku 2004 nie mogą być uznane za różniące się w sposób statystycznie istotny. Jeżeli przyjrzeć się graficznemu przedstawieniu zmienności średniego poziomu umiejętności na przestrzeni lat (zob. Rysunek 6.2), to widzimy iż mają one charakter sinusoidalny z okresem najwyższych wyników przypadających w okolicach roku 2005 oraz 2011, a najniższych w okolicach 2003 oraz 2008. Jeżeli zestawić kształt zależności między średnimi wynikami na przestrzeni lat, który prezentuje Rysunek 6.2 ze średnimi wynikami przedstawionymi na skali zrównanych do 2012 wyników obserwowanych (przedostatni wykres, który ilustruje Rysunek 6.1), to widzimy między nimi pełną zgodność. Zgodność taka jednak nie dziwi, gdyż jest to ta sama informacja, a jedyna różnica polega na tym, że Rysunek 6.1 przedstawia ją na skali dyskretnych wyników obserwowanych w teście, natomiast Rysunek 6.2 na skali zmiennej ukrytej otrzymanej bezpośrednio w wyniku dopasowania do danych modelu IRT.

Na tle zauważalnie istotnej dynamiki w średnim poziomie umiejętności uczniów ich odchylenie standardowe okazuje się bardzo stabilne (zob. Tabela 6.2). Odchylenia standardowe zmiennej ukrytej

w żadnym roku nie różnią się w sposób istotny statystycznie od roku 2012, nie różnią się również istotnie statystycznie między sobą.

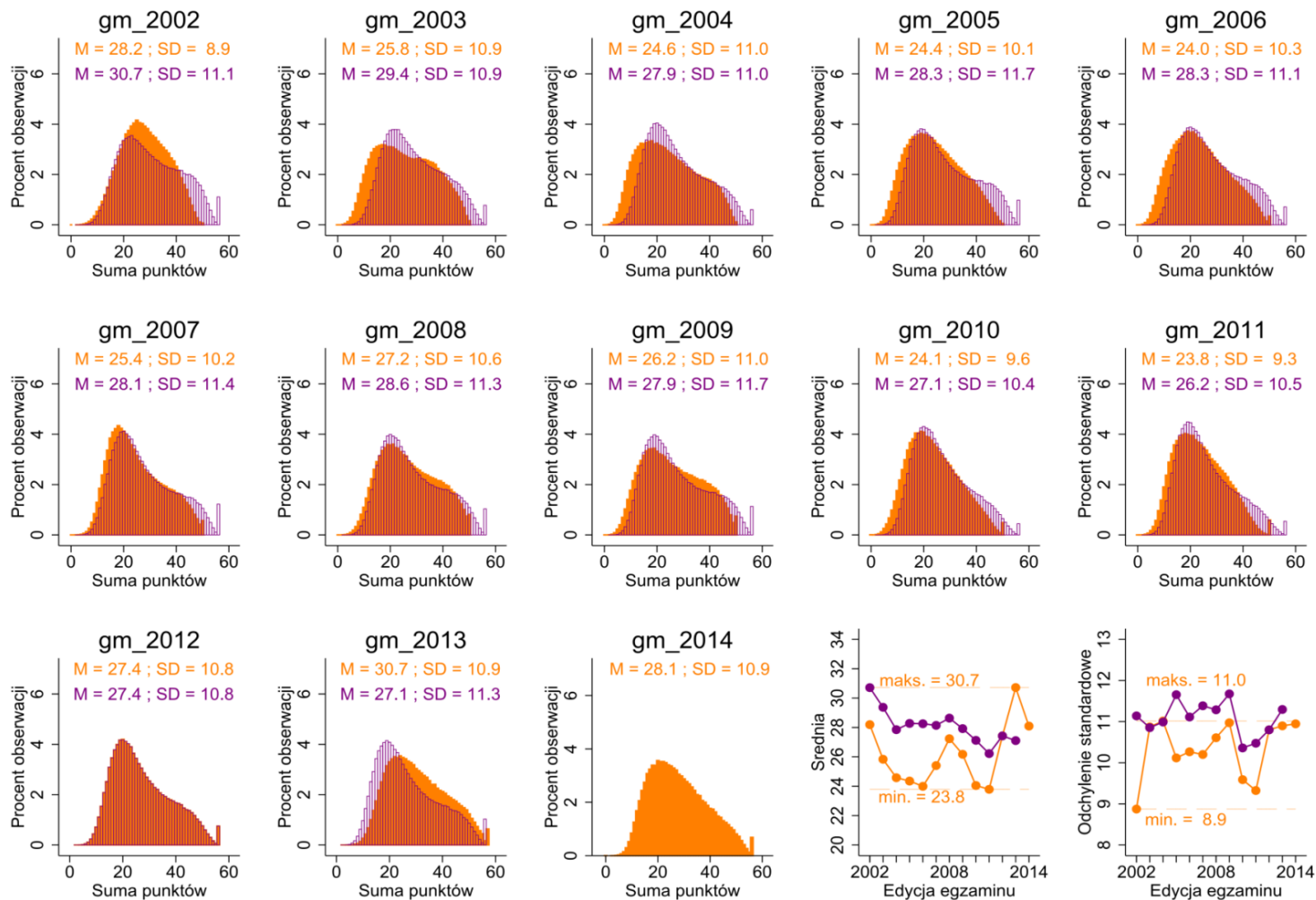
6.2. Egzamin gimnazjalny – część matematyczno-przyrodnicza

6.2.1. Zmiany trudności egzaminu w latach 2002–2013

Rysunek 6.3 w sposób analogiczny do sprawdzianu, zestawia rozkłady rzeczywistych oraz wyrażonych na skali z roku 2012 wyników ze z części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013 wraz z informacjami o średniej (M) oraz odchyleniu standardowym (SD) dla wersji pierwotnej i zrównanej. Przystępując do analizy zmian kształtów rozkładów wyników obserwowanych uzyskanych w danym roku oraz sprowadzonych do wspólnej skali, należy poczynić kilka uwag stwarzających istotny kontekst dla właściwej interpretacji niektórych zjawisk obserwowanych w przypadku egzaminu gimnazjalnego, a w szczególności jego części matematyczno-przyrodniczej.

Pierwszą rzeczą na jaką pragniemy uwrażliwić Czytelnika jest fakt, że w roku 2012 zmianie uległa formuła, w jakiej przeprowadza się egzamin gimnazjalny – dla części matematyczno-przyrodniczej wydzielono oddzielne arkusze matematyczny i z przedmiotów przyrodniczych, a dla części humanistycznej wydzielono oddzielne arkusze z języka polskiego oraz z historii i wiedzy o społeczeństwie. Rozdzielenie arkuszy części matematyczno-przyrodniczej szło w parze ze zwiększeniem łącznej liczby zadań wchodzących w skład tej części egzaminu, a co za tym idzie również maksymalnej liczby możliwych do uzyskania za nią liczby punktów. W latach 2002–2011 za zadania z części matematyczno-przyrodniczej można było uzyskać maksymalnie 50 punktów, natomiast w latach 2012–2014 od 56 do 57 punktów. Opisana zmiana rzutuje na interpretację rozkładów obserwowanych wyników, jaka następuje w kolejnych akapitach, a także znajduje swoje odzwierciedlenie w tablicach przeliczeniowych między wynikami egzaminu gimnazjalnego z danego roku, a egzaminu z roku 2012, które zawiera Aneks 1.

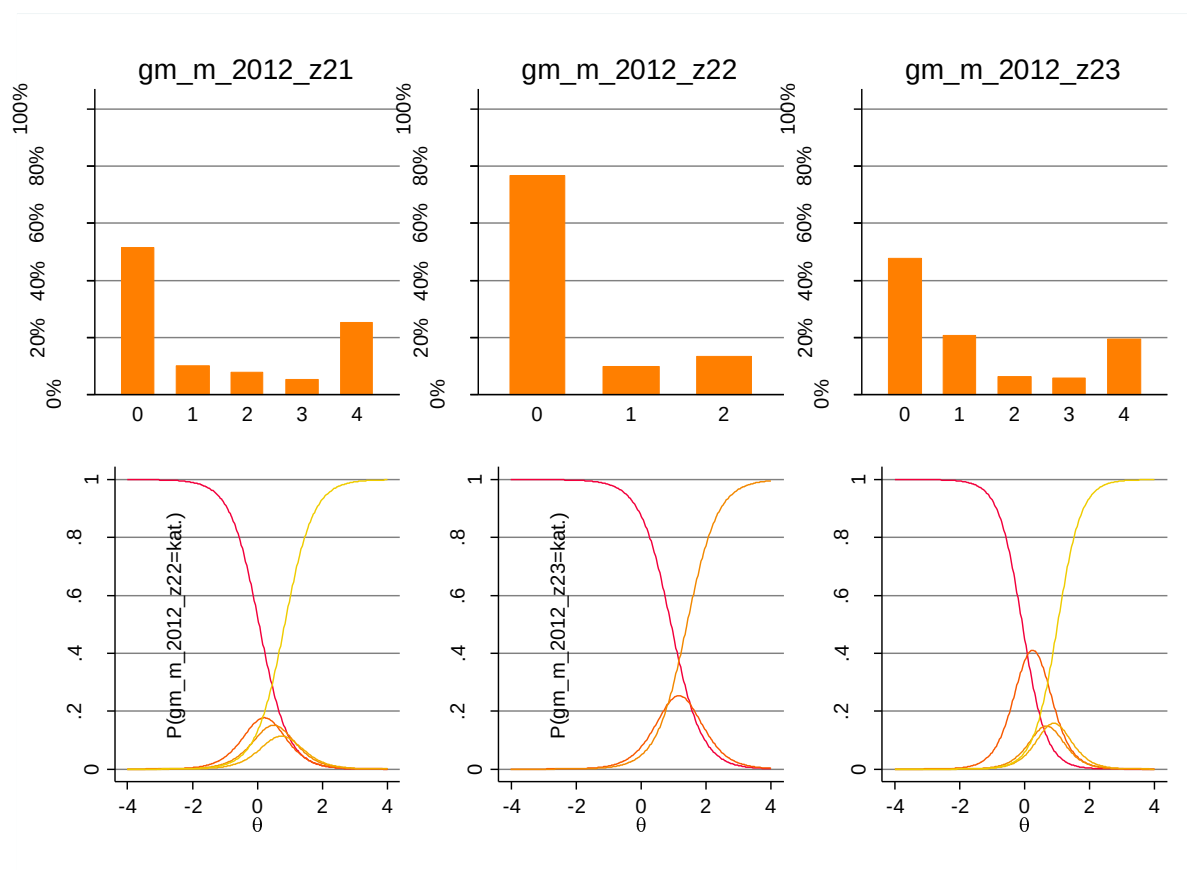
Wydłużenie skali, jakie nastąpiło przy zmianie formuły egzaminu utrudnia nieco interpretację wykresów, które przedstawia Rysunek 6.3 gdyż należy pamiętać, że wyższe średnie w testach w latach 2012–2014 po części są również konsekwencją zwiększenia maksymalnego możliwego do uzyskania w egzaminie wyniku. Aby przeprowadzić analizę zmian trudności egzaminu na przestrzeni lat w sprawdzianie (zob. podrozdział 6.1.1) dokonano omówienia średnich wyników na skali z danego roku w zestawieniu z wynikami zrównanymi do roku referencyjnego. Tutaj, aby przeprowadzić analogiczne rozumowanie należałoby najpierw sprowadzić średnie wyniki do skali procentowej, gdyż surowa średnia przy rozszerzeniu skali traci swoją interpretację porównywalnej w całym badanym okresie miary trudności egzaminu. Czytelnik zainteresowany taką analizą może z łatwością ją przeprowadzić na podstawie dostarczonych przez Rysunek 6.3 danych – w niniejszym rozdziale pozostajemy jednak przy skali najbardziej surowych, punktowych wyników obserwowanych, aby nie pomijać innych interesujących zjawisk, jakie można zaobserwować dla rozkładów egzaminacyjnych na przestrzeni lat. Jeżeli chodzi o analizę wahań trudności egzaminu, to nawet bez sprowadzania wyników do skali procentowej widać, że łatwość egzaminu między latami ulegała dość znacznym zmianom. Przykładowo, w latach 2006–2008 obserwowane średnie surowe wyniki (na skali 0–50 punktów) wzrastały od 24 przez 25,4 do 27,2 punktu, natomiast dla tych samych lat wyniki wyrażone na zrównanej skali wyników egzaminu referencyjnego z 2012 roku (zatem na skali szerszej o 6 punktów) nie wykazywały żadnego trendu, a ich wahania były kilkukrotnie niższe, odpowiednio: 28,3, 28,1 oraz 28,6 punktu. Na podstawie takiej analizy możemy wywnioskować, że wzrost średniego surowego wyniku w przykładowych latach 2006–2008 o ponad 3 punkty jest przede wszystkim związany ze zmianami w łatwości egzaminu w tych latach.



Rysunek 6.3. Rozkłady wyników obserwowanych z części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002-2014. Pełne słupki odnoszą się do niezrównanych wyników egzaminacyjnych, zrównane wyniki przedstawiono za pomocą słupków pustych

Kolejną rzeczą, na jaką należy zwrócić uwagę przy okazji omawiania obserwowanych wyników z części matematyczno-przyrodniczej zrównanych do egzaminu z 2012 roku, jest przedstawiony przez Rysunek 6.3 bardzo specyficzny kształt rozkładu wyników dla tej edycji egzaminu. Rozkład dla części matematyczno-przyrodniczej z roku 2012 jest silnie prawoskośny, średnia (27,4) jest aż o przeszło 7 punktów wyżej niż modalna, przypadająca na wartość 20 punktów. Takie cechy rozkładu wyników z egzaminu z 2012 roku, z mniejszymi lub większymi zmianami, ujawniają się również w innych latach z zakresu 2002–2013 (histogramy z pustymi słupkami ilustrowane przez Rysunek 6.3). Należy zaznaczyć, że taki kształt rozkładu wyników bynajmniej nie jest odzwierciedleniem żadnego istotnego zróżnicowania wewnątrzpopulacyjnego w zakresie umiejętności badanych testem. Przyczyny obserwowanego rozkładu wyników w roku referencyjnym 2012 należy upatrywać w konstrukcji trzech zadań otwartych z matematyki, o numerach w arkuszu egzaminacyjnym 21, 22 oraz 23. Rysunek 6.4 przedstawia rozkłady punktów zdobywanych za te trzy zadania przez uczniów w 2012 roku wraz z krzywymi charakterystycznymi, które informują jakie jest prawdopodobieństwo uzyskania określonej kategorii punktowej w zależności od poziomu umiejętności, θ . Widzimy, że wszystkie te zadania były oceniane w sposób wielokategorialny oraz odznaczały się U-kształtnym rozkładem odpowiedzi. Kategorie pośrednie w tych zadaniach dostarczały zazwyczaj bardzo mało informacji o poziomie umiejętności uczniów. Takie zadania silnie różnicowały uczniów o wyższym poziomie umiejętności (dodatni zakres θ) od uczniów o przeciętnym oraz niskim poziomie umiejętności, jednocześnie umożliwiając uczniom o wyższym poziomie umiejętności zdobycie maksymalnej liczby punktów za zadanie z pominięciem kategorii pośrednich. Doprowadziło to obserwowanego „zajęcia” znacznego zakresu możliwych do zdobycia w teście punktów, po stronie wyników wysokich, przez relatywnie niewielką liczbę uczniów o najwyższym poziomie umiejętności. Taki test pod względem pomiarowym odbiega od optymalnego rozwiązania dla różnicowania umiejętności uczniów w całej populacji podchodzących do egzaminu, gdyż znaczna część kategorii punktowych testu jest wykorzystywana do różnicowania niewielkiej części populacji.

Poczynione uwagi odnośnie specyficznych problemów na jakie napotykamy przy interpretowaniu zrównanych obserwowanych wyników części matematyczno-przyrodniczej egzaminu gimnazjalnego (zmiana skali oraz specyficzne właściwości psychometryczne egzaminu w roku referencyjnym) ilustrują, jakiego typu problemy można napotkać, gdy skala komunikowanych wyników egzaminu opiera się na wynikach surowych. Opisane w dalszej części oparte na modelowaniu IRT zmiany w poziomie umiejętności na skali zmiennej ukrytej są wolne od tych uciążliwości, gdyż rozkład wyników w tej skali jest zbliżony do normalnego i w bardzo niewielkim stopniu zależy od liczby zadań w teście lub właściwości psychometrycznych testu.



Rysunek 6.4. Rozkłady punktów dla zadań 21, 22 oraz 23 w arkuszu matematyczno-przyrodniczym 2012 (górze) wraz z odpowiadającymi tym zadaniom krzywymi charakterystycznymi (dół)

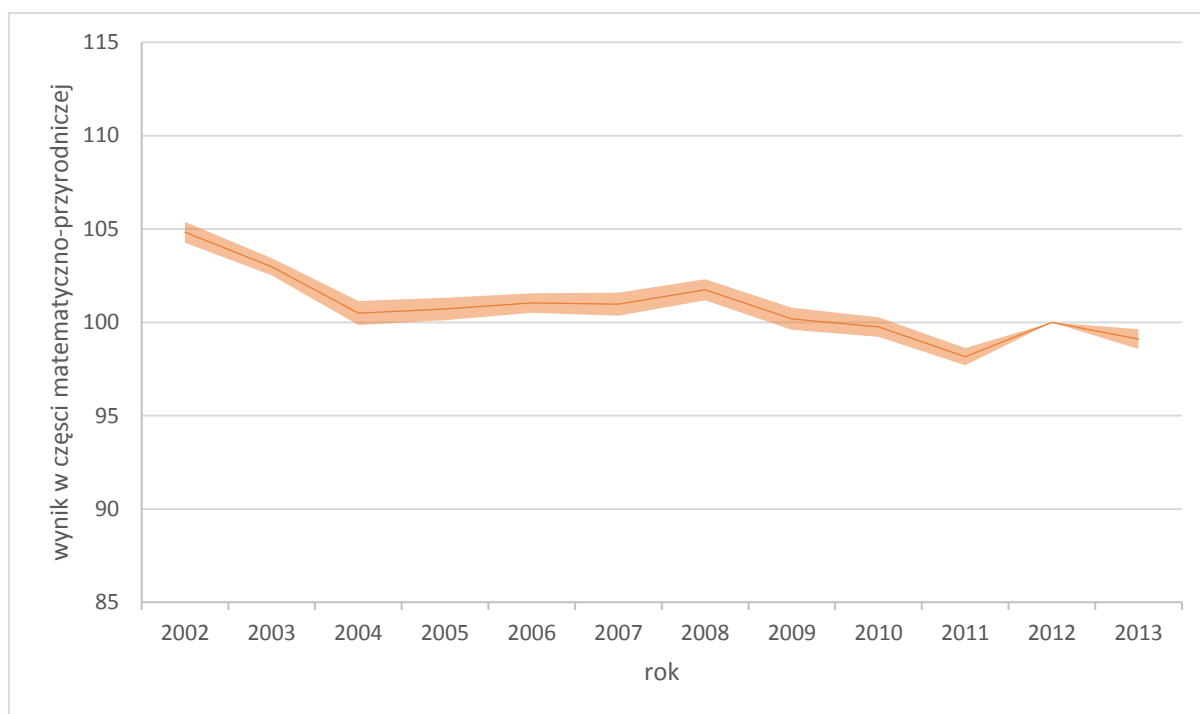
6.2.2. Wyniki w latach 2002–2013 na skali zmiennej ukrytej

Jeżeli przyjrzeć się średnim oraz odchyleniom standardowym wyników obserwowanych zrównanych do skali egzaminu z 2012 roku (dwa ostatnie wykresy – Rysunek 6.3), to oprócz stwierdzenia, że parametry zrównanych wyników zmieniają się na przestrzeni lat znacznie mniej dynamicznie niż ma to miejsce w przypadku wyników surowych, należy zauważyć, iż mimo wszystko zmieniają się w stopniu, który może okazać się istotny z praktycznego oraz statystycznego punktu widzenia. Aby odpowiedzieć w sposób ścisły na pytanie o stabilność rozkładu umiejętności matematyczno-przyrodniczych uczniów kończących gimnazjum w latach 2002–2013 przedstawiono analogiczną jak w przypadku sprawdzianu analizę przedstawiającą zmiany średniej oraz odchylenia standardowego na poziomie zmiennej ukrytej modelowanej za pomocą IRT. Oprócz informacji o istotności statystycznej, taka analiza jest również wolna od trudności interpretacyjnych związanych z opisanymi wcześniej właściwościami skal wyników obserwowanych dla omawianej części egzaminu gimnazjalnego. Tabela 6.3 przedstawia średnie oraz odchylenia standardowe zmiennej ukrytej umiejętności matematyczno-przyrodniczych wraz z informacją o precyzji ich oszacowania, przy czym informacje dla średniej dla ułatwienia naniesiono również na wykres ilustrowany przez Rysunek 6.5.

Tabela 6.3. Średnia oraz odchylenie standardowe zmiennej umiejętności dla części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013, wyniki zrównane na skali 100; 15 zakotwiczone w roku 2012

Rok egzaminu	Średnia	Błąd standard.	95% przedział ufn.		Odch. stand.	Błąd standard.	95% przedział ufn.	
			dolna gr.	górną gr.			dolna gr.	górną gr.
2002	104,8	0,29	104,3	105,4	14,4	0,39	13,6	15,1
2003	103,0	0,24	102,5	103,4	14,2	0,35	13,5	14,8
2004	100,5	0,33	99,9	101,1	15,1	0,45	14,3	16,0
2005	100,7	0,31	100,1	101,3	16,6	0,46	15,7	17,5
2006	101,0	0,27	100,5	101,6	15,3	0,40	14,5	16,1
2007	101,0	0,32	100,4	101,6	15,8	0,48	14,9	16,8
2008	101,7	0,29	101,2	102,3	15,3	0,48	14,4	16,3
2009	100,2	0,30	99,6	100,8	16,8	0,52	15,8	17,8
2010	99,8	0,27	99,2	100,3	14,2	0,39	13,4	14,9
2011	98,2	0,24	97,7	98,6	15,0	0,31	14,4	15,6
2012	100,0	–	–	–	15,0	–	–	–
2013	99,1	0,27	98,6	99,6	16,5	0,37	15,8	17,2

Porównując przedziały ufności dla średniego poziomu umiejętności matematyczno-przyrodniczych (zob. Tabela 6.3 oraz Rysunek 6.5 obserwujemy wyraźny trend spadkowy. Można go podzielić na trzy okresy: lata początkowe 2002–2003, w których średnia wyniosła 104,8 oraz 103 jest istotnie statystycznie wyższa niż we wszystkich pozostałych latach; lata pośrednie 2005–2008, w których wyniki znajdują się w zakresie 100,7–101,7 i są istotnie niższe niż w latach początkowych, ale istotnie wyższe niż w ostatnim analizowanym okresie przypadającym na lata 2011–2013 ze średnią w zakresie 98,2–100. Przy interpretowaniu zaobserwowanych wyników trzeba jednak być bardzo ostrożnym, w szczególności jeżeli chodzi o pierwszy okres, w którym obowiązywał egzamin. Były to pierwsze lata funkcjonowania systemu egzaminów zewnętrznych i nie ma pewności na ile udało się zapewnić prawidłowe przestrzeganie wszystkich procedur przeprowadzania egzaminu (uwaga ta odnosi się również do pierwszych wyników sprawdzianu oraz części humanistycznej egzaminu gimnazjalnego). Dla odchylenia standardowego poziomu umiejętności matematyczno-przyrodniczych obserwujemy w przypadku kilku roczników różnice istotne statystycznie. W szczególności w roku 2003 oraz 2010 odchylenie standardowe umiejętności było niższe niż w roku referencyjnym, natomiast w latach 2005, 2009 oraz 2013 wyższe. Jest to interesujące, gdyż dla sprawdzianu zanotowano bardzo silną stabilność parametru rozproszenia umiejętności uczniów.

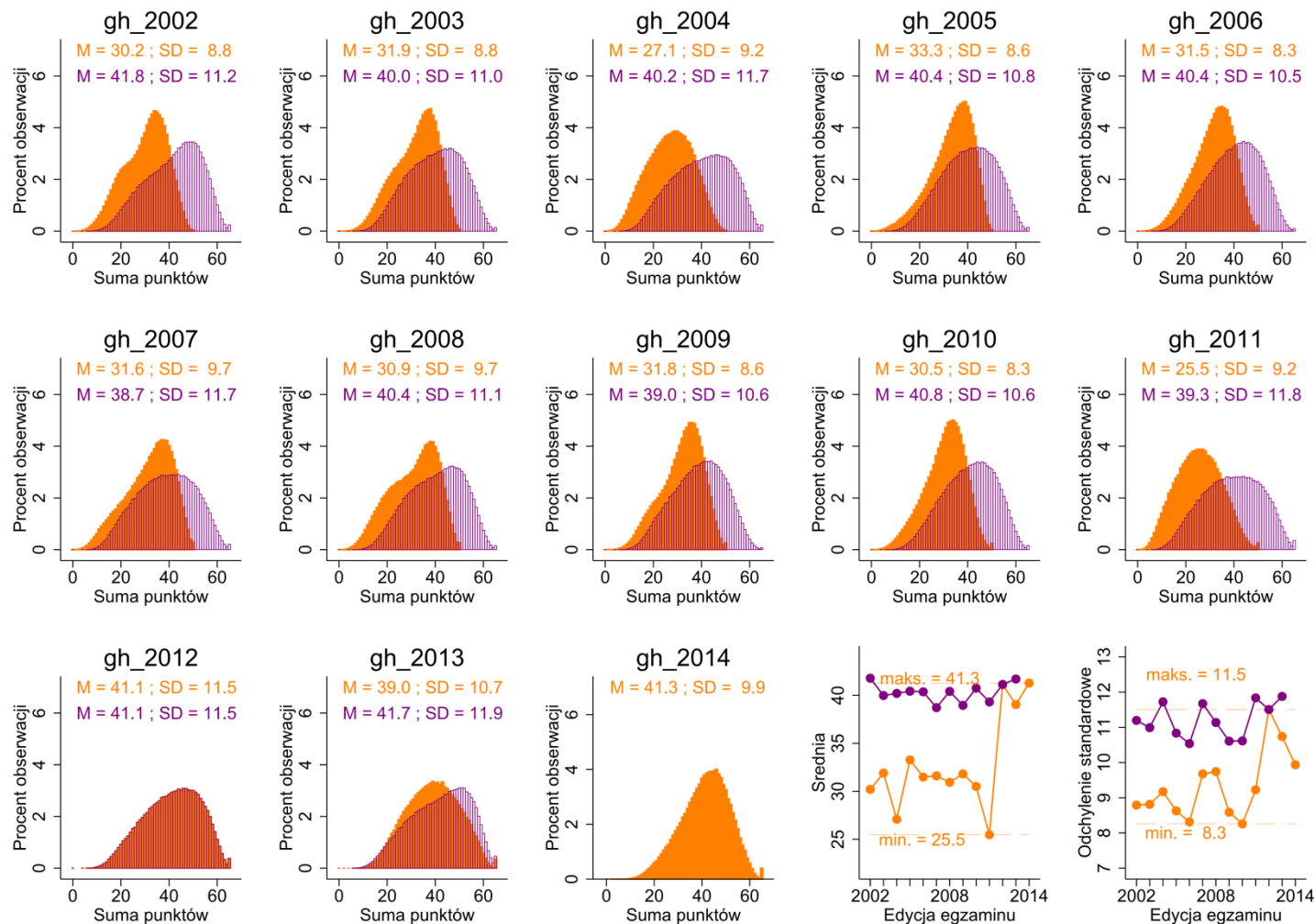


Rysunek 6.5. Średnie zrównane wyniki dla części matematyczno-przyrodniczej egzaminu gimnazjalnego na skali 100;15 wraz z 95% przedziałem ufności

6.3. Egzamin gimnazjalny – część humanistyczna

6.3.1. Zmiany trudności egzaminu w latach 2002–2013

Podobnie jak to miało miejsce w przypadku części matematyczno-przyrodniczej egzaminu gimnazjalnego, tak i w części humanistycznej w roku 2012 wprowadzono zmianę, która zwiększała liczbę punktów, jakie można uzyskać z tej części egzaminu w porównaniu z latami 2002–2011. Zmiana w przypadku części humanistycznej była jednak inna niż dla części matematyczno-przyrodniczej. Po pierwsze, rozszerzenie skali było wyraźniejsze – punktacja wzrosła o 30%, z 50 punktów przed 2012 rokiem do 65 punktów od 2012 roku. Po drugie maksymalna punktacja w części humanistycznej od 2012 roku jest stała – gdy w części matematyczno-przyrodniczej zmienia się oscylując wokół 56 punktów. Opisane różnice jednak nie mają znaczenia w kontekście tego jak wpływają one na interpretację różnic między rozkładami wyników obserwowanych otrzymanych na egzaminach w latach 2002–2011, a rozkładami wyników w tych latach wyrażonymi na skali zrównanej do roku referencyjnego 2012. Trzeba pamiętać, że wspomniane wyniki są wyrażone na skalach o różnej maksymalnej możliwej do uzyskania liczbie punktów i średnie nie przekładają się w sposób bezpośredni na porównywalną informację o łatwości testu – aby wnioskować o łatwości należałoby podzielić średnią przez odpowiadające jej maksimum skali.



Rysunek 6.6. Rozkłady wyników obserwowanych z części humanistycznej egzaminu gimnazjalnego w latach 2002-2014. Pełne słupki odnoszą się do niezrównanych wyników egzaminacyjnych, zrównane wyniki przedstawiono za pomocą słupków pustych

Zaznaczywszy wątek o różnicy rozpiętości skal między latami 2002–2011 a 2012–2013 (oraz 2014), możemy bez większych wątpliwości stwierdzić, porównując różnice między średnimi obserwowanych wyników zrównanych do roku 2012 oraz średnimi surowych wyników obserwowanych (zob. Rysunek 6.6), że obserwowane znaczne fluktuacje w surowych średnich wynikach z egzaminu między latami są przede wszystkim odzwierciedleniem różnic w ich łatwości. Jest to obserwacja analogiczna do wcześniej omawianych przypadków sprawdzianu oraz części matematyczno-przyrodniczej egzaminu gimnazjalnego.

Należy zauważyć, że dwa roczniki z analizowanego okresu 2002–2011, które najbardziej odstawały na tle pozostałych lat, czyli rok 2004 ze średnią 27,1 oraz 2011 ze średnią 25,5, przestają się w ogóle wyróżniać na tle pozostałych, gdy ich poziom umiejętności porównuje się z innymi latami na skali części humanistycznej z roku 2012. Na skali surowych wyników 0–50 punktów, średnia w roku 2004 była o 4,8 punktu niższa niż w roku 2003 oraz aż o 6,2 punktu niższa niż w roku 2005. Informacja ta kontrastuje z obrazem jaki wyłania się dla tych trzech roczników, gdy wyniki są wyrażone na szerszej, 0–65 punktów, skali wyników zrównanych do roku 2012, gdzie w 2004 mamy średnią wyższą o 0,2 punktu niż w 2003 oraz o tyle samo niższą niż w 2005. Podobnie w roku 2011, na krótszej, niezrównanej skali mamy wynik niższy o 5 punktów względem roku 2010 oraz niższy aż o 6,3 punktu względem roku 2009, natomiast na dłuższej, zrównanej skali mamy różnice rzędu 1,5 punktu oraz 0,3 punktu, przy czym ostatnią o przeciwnym zwrocie. Należy wyciągnąć stąd wniosek, że procedury kontroli trudności egzaminów w latach 2002–2011 nie działały skutecznie, dopuszczając do znacznych różnic poziomu trudności części humanistycznej egzaminu między latami.

Pozytywnym obrazem wyłaniającym się z analizy przedstawionych przez Rysunek 6.6 rozkładów i ich parametrów jest względna stabilność średniego surowego wyniku w latach 2012–2014²⁴, która sugeruje, że wraz z nową formułą egzaminu w przypadku części humanistycznej egzaminu gimnazjalnego wprowadzono skuteczne procedury kontroli trudności egzaminu. Na tle części matematyczno-przyrodniczej egzaminu gimnazjalnego za pozytywną cechę rozkładów wyników egzaminacyjnych w latach 2012–2014 należy uznać również ich bardziej zbliżony do normalnego kształt (por. Rysunek 6.3 oraz Rysunek 6.4 wraz z komentarzem).

Czytelnika zainteresowanego przeliczeniem wyników z dowolnej części humanistycznej egzaminu gimnazjalnego w latach 2002–2013 na wyniki w teście z roku referencyjnego odsyłamy do Aneksu 1.

6.3.2. Wyniki w latach 2002–2013 na skali zmiennej ukrytej

Tabela 6.4 oraz Rysunek 6.7 przedstawia informacje o parametrach charakteryzujących rozkład umiejętności mierzonych poprzez część humanistyczną egzaminu gimnazjalnego na skali zmiennej ukrytej. Analizując szerokości przedziałów ufności wokół średniej można zauważyć wiele różniących się, nawet z roku na rok, porównań. Wyróżnić należy okres od 2003 do 2006 roku, w którym poziom umiejętności jest stabilny, co kontrastuje z obrazem jaki wyłaniałby się gdyby analizować jedynie surowe wyniki egzaminacyjne (por. przedostatni wykres – Rysunek 6.6). Uczniowie zdający pierwszy egzamin w 2002 odznaczają się względem analizowanego okresu wyższym poziomem umiejętności

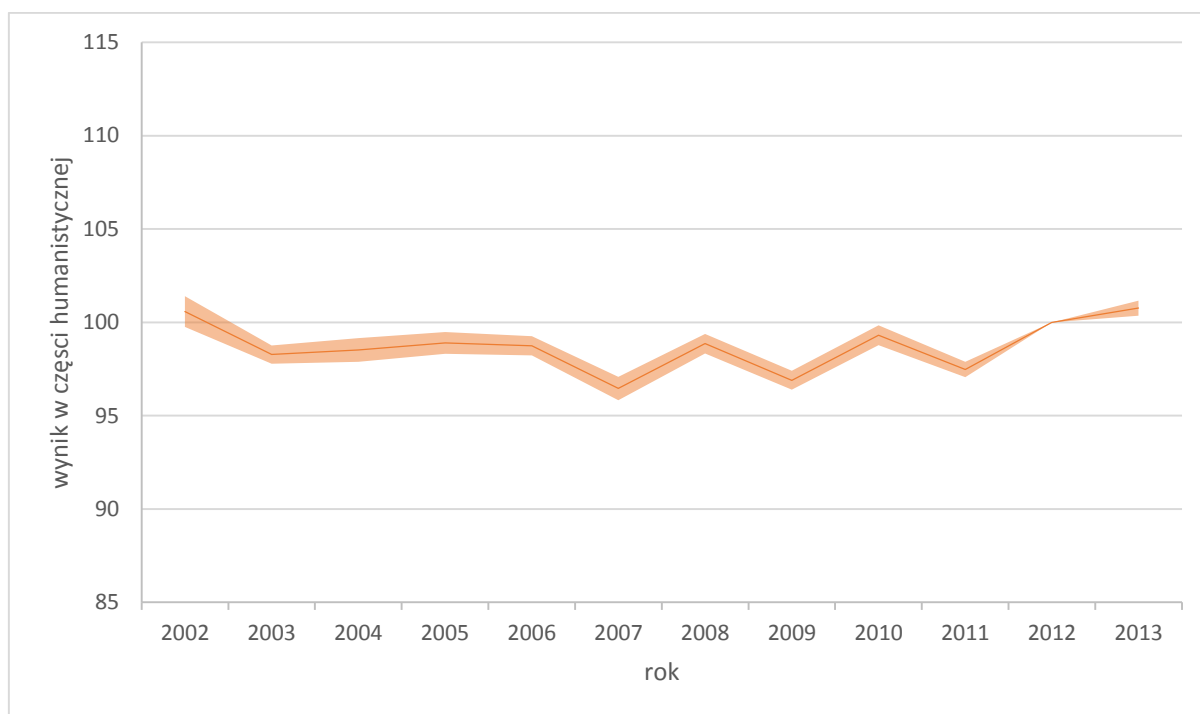
²⁴ Mimo iż rok 2014 nie jest objęty procedurą zrównywania, to analizując wielkości zmian średniej testu referencyjnego między latami 2002–2013 można przypuszczać, że łatwość testu 2014 jest zbliżona do łatwości testów w latach 2013 oraz 2012. Takie przypuszczenie byłoby nieprawdziwe jedynie gdyby zmiana poziomu umiejętności uczniów w roku 2014 była wyjątkowo silna na tle tych zaobserwowanych w latach 2002–2013, jednak przesłanek do zakładania takiej hipotetycznej, silnej zmiany poziomu umiejętności w roku 2014 brak.

humanistycznych, o około 2 punkty na skali o SD=15, jednak trzeba mieć na względzie iż był to pierwszy egzamin i nie do końca wiadomo na ile wszystkie procedury jego przeprowadzenia były zachowane w sposób rzetelny. Następnie można wyróżnić okres corocznych wahań poziomu umiejętności przypadających na lata 2007–2012, w którym to okresie obserwujemy naprzemienny spadek i wzrost poziomu umiejętności o około 2 punkty, który przekracza próg istotności statystycznej, jednak jest dość trudny do wytłumaczenia. Poziom umiejętności ostatnich trzech roczników (2011–2013) tworzy natomiast trend wzrostowy. Uczniowie z każdym rokiem uzyskują istotnie statystycznie wyższy wynik, poczynając od średniej 97,5 w roku 2011, na średniej 100,8 w roku 2013 kończąc. Patrząc na długookresowe zmiany na przedziale 2002–2013 należy jednak zasadniczo uznać poziom mierzonej umiejętności za względnie stały.

Zauważamy również, iż z kilkoma wyjątkami (lata 2004 oraz 2007), rozproszenie wyników zmiennej ukrytej (zob. Tabela 6.4) przed rokiem 2011 było statystycznie istotnie niższe niż w ostatnich trzech analizowanych rocznikach, gdzie notujemy wartości rzędu 15–15,7.

Tabela 6.4. Średnia oraz odchylenie standardowe zmiennej umiejętności dla części humanistycznej egzaminu gimnazjalnego w latach 2002–2013, wyniki zrównane na skali 100; 15 zakotwiczone w roku 2012

Rok egzaminu	Średnia	Błąd standard.	95% przedział ufn.		Odch. stand.	Błąd standard.	95% przedział ufn.	
			dolna gr.	górną gr.			dolna gr.	górną gr.
2002	100,6	0,42	99,8	101,4	14,2	0,48	13,3	15,2
2003	98,3	0,25	97,8	98,8	13,7	0,32	13,1	14,3
2004	98,5	0,32	97,9	99,2	15,2	0,44	14,4	16,1
2005	98,9	0,30	98,3	99,5	13,6	0,36	12,9	14,3
2006	98,7	0,26	98,2	99,2	13,0	0,32	12,4	13,6
2007	96,5	0,32	95,8	97,1	15,4	0,43	14,6	16,3
2008	98,9	0,26	98,3	99,4	14,0	0,34	13,3	14,6
2009	96,9	0,26	96,4	97,4	13,2	0,32	12,6	13,8
2010	99,3	0,27	98,8	99,8	13,1	0,33	12,5	13,8
2011	97,5	0,21	97,1	97,9	15,7	0,29	15,1	16,3
2012	100,0	–	–	–	15,0	–	–	–
2013	100,8	0,21	100,4	101,2	15,7	0,30	15,1	16,3

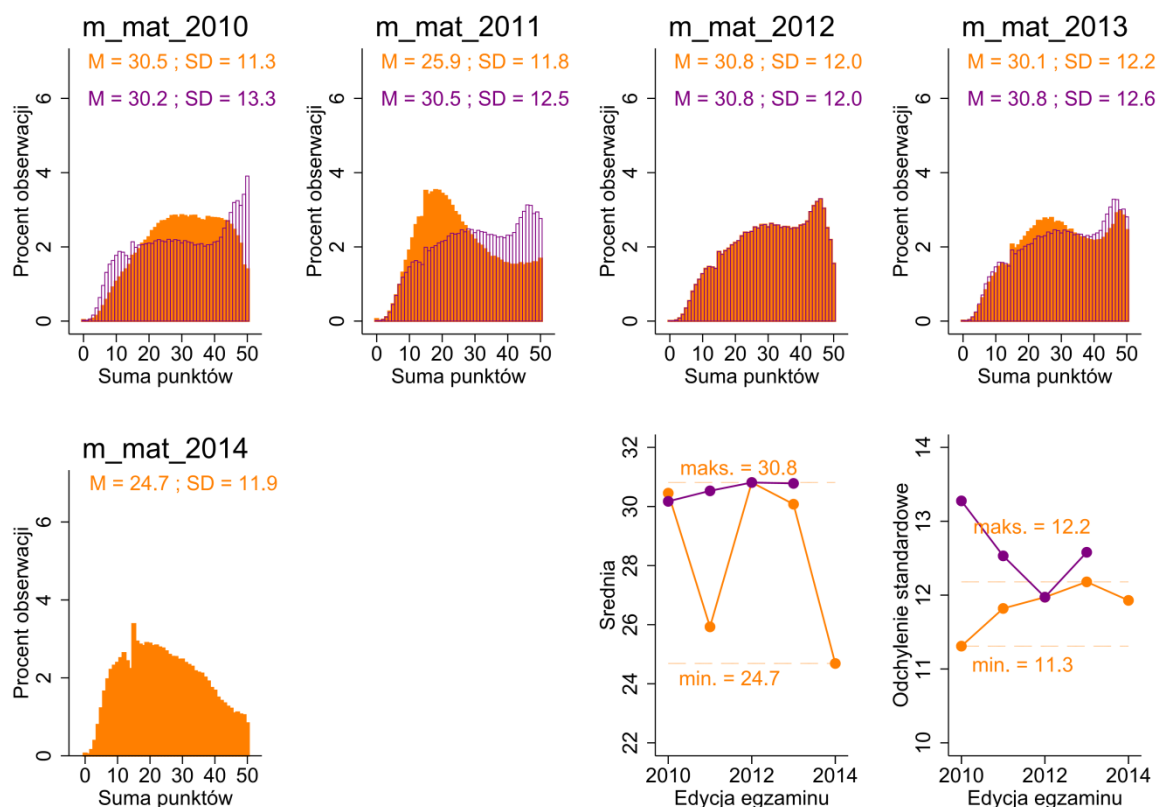


Rysunek 6.7. Średnie zrównane wyniki dla części humanistycznej egzaminu gimnazjalnego na skali 100;15 wraz z 95% przedziałem ufności

6.4. Egzamin maturalny z matematyki na poziomie podstawowym

6.4.1. Zmiany trudności egzaminu w latach 2010–2013

Rysunek 6.8 porównuje rozkłady wyników z pierwszego terminu egzaminu maturalnego z matematyki na poziomie podstawowym, jakie faktycznie uzyskali uczniowie podchodzący do egzaminu w latach 2010–2013 z rozkładami wyników, jakie uzyskaliby ci sami uczniowie gdyby rozwiązywali zadania arkusza z roku 2012. Średnie wyniki uzyskiwane na egzaminie w latach 2010–2013 wynoszą kolejno 30,5, 25,9, 30,8 oraz 30,1 punktu, natomiast na skali zrównanej odpowiednio 30,2, 30,5, 30,8 oraz ponownie 30,8 punktu. Zatem, podobnie jak to miało miejsce w przypadku egzaminów z niższego szczebla, zmiany średniego wyniku na wspólnej skali egzaminu z roku 2012 są o wiele mniejsze niż zmiany średniego wyniku faktycznie zaobserwowanego podczas różnych edycji egzaminu. Oznacza to, że obserwowane z roku na rok fluktuacje średniego wyniku z matury z matematyki są w głównej mierze odzwierciedleniem niezwiązanych z poziomem umiejętności uczniów zmian w jej trudności. Zasadniczo odstaje tu przede wszystkim egzamin w roku 2011, który na tle reszty jest zauważalnie trudniejszy. Gdyby uczniowie w roku 2010 zdawali egzamin z roku 2012 to ich średni wynik byłby o 0,3 punktu niższy niż faktycznie uzyskany na egzaminie, w przypadku uczniów z roku 2013 różnica wynosi 0,7 punktu, natomiast dla uczniów piszących egzamin w roku 2011 jest to już aż 4,6 punktu (na skali o rozpiętości 0–50 przekłada się to na 9,2%).



Rysunek 6.8. Rozkłady wyników obserwowanych z matury podstawowej z matematyki w latach 2010–2014. Pełne słupki odnoszą się do nierównanych wyników egzaminacyjnych, zrównane wyniki przedstawiono za pomocą słupków pustych

Porównanie kształtów rozkładów wyników matury z matematyki w latach 2010–2013, jakie przedstawia Rysunek 6.8 uświadamia jednak, że sama analiza zrównanej i faktycznej miary tendencji centralnej nie jest w stanie opisać całej złożoności różnic pomiędzy pomiarowymi właściwościami tych egzaminów. Widzimy, że nierównane wyniki obserwowane uzyskane w latach 2012 oraz 2013 mają dość zbliżony kształt rozkładu – lewostronnie skośny z dwoma modami, pierwszą przypadającą w okolicach środka skali, a drugą przypadającą w rejonie najwyższych wyników i odzwierciedlającą efekt sufitowy, który zawdzięczać należy uczniom podchodzącym również do egzaminu z matematyki na poziomie rozszerzonym. Natomiast rozkład nierównanych wyników w roku 2011 jest silnie prawostronnie skośny z modą w okolicach progu zdawalności i nieznacznie zaznaczonym efektem sufitowym. Dla odmiany, w przypadku rozkładu wyników z 2010 roku obserwujemy w miarę symetryczny platykurtyczny rozkład wyników. Jeżeli jednak przyjrzymy się kształtowi rozkładu wyników egzaminu z roku 2012, jaki uzyskiwaliby uczniowie z populacji 2010–2013, gdyby ten egzamin rozwiązywali, to widzimy, iż w każdym z rocznika zachowuje on swoje zasadnicze właściwości – lewoskośność oraz występowanie efektu sufitowego dla najzdolniejszych uczniów. Taki obraz sugeruje, że obserwowane różnice w skośności oraz kurtozie między wynikami uzyskiwanymi w kolejnych edycjach egzaminu maturalnego z matematyki są odzwierciedleniem psychometrycznych właściwości tychże, a nie charakterystyką zdających egzamin kolejnych kohort uczniów (w szczególności, uwaga ta dotyczy również referencyjnego egzaminu z roku 2012).

Interesujące jest, że przy minimalnych wahaniach średniego wyniku na wspólnej skali egzaminu z roku 2012 obserwujemy relatywnie znaczne zmiany w rozproszeniu uczniów – odchylenie standardowe

wyników zrównanych do roku 2012 dla uczniów w roku 2010 wynosi 13,3 punktu, w latach 2011 oraz 2013 12,5 oraz 12,6 punktu, a najmniej w roku referencyjnym, tj. 12 punktów. Odpowiedź na pytanie czy zróżnicowanie średnich oraz odchyłeń standardowych umiejętności mierzonych na egzaminie maturalnym z matematyki przekracza próg istotności statystycznej znajduje się przy omówieniu wyników zrównania na poziomie zmiennej ukrytej, jakie ma miejsce w dalszej części rozdziału.

Opisane powyżej różnice we właściwościach rozkładów egzaminów z matury z matematyki w latach 2010–2013 przy zestawieniu z wynikami na wspólnej skali egzaminu z roku 2012 wskazują, że surowe wyniki egzaminu nie stanowią porównywalnej między latami miary uczniowskich umiejętności. Dla przykładu, Tabela 6.5 przedstawia jak wyglądała faktyczna zdawalność egzaminu w analizowanych czterech rocznikach uczniów i jak wyglądałaby zdawalność dla tych samych roczników, gdyby wszyscy rozwiązywali ten sam test – egzamin referencyjny z roku 2012. Możemy zauważyć, że zdawalność wyliczona przy uwzględnieniu wyników zrównanych jest o wiele bardziej stabilna między latami, niż ta faktycznie zaobserwowana w danych rocznikach. Jest to obserwacja zgodna z prezentowaną przez Rysunek 6.8 bardzo dużą stabilnością średniej z wyników zrównanych na przestrzeni lat, w porównaniu z średnią surowych wyników obserwowanych. Jednocześnie, zanotowanie najniższej zdawalności na zrównanych wynikach w roku 2010 wynoszącej 84,4% (bez wykorzystania wyników zrównanych mamy w tym roku zdawalność najwyższą – 90,3%) idzie w parze z informacją o najwyższej w tym roku wariancji wyników (zob. ostatni wykres – Rysunek 6.8).

Tabela 6.5. Zdawalność egzaminu maturalnego z matematyki wyliczona na niezrównanych wynikach egzaminacyjnych oraz na wynikach wyrażonych na skali egzaminu z roku 2012

Rok egzaminu	Procent uczniów uzyskujących 15 punktów lub więcej	
	niezrównane egzaminacyjne	wyniki na skali egzaminu z 2012 roku
2010	90,3%	84,4%
2011	82,1%	87,2%
2012	88,8%	88,8%
2013	88,2%	87,2%

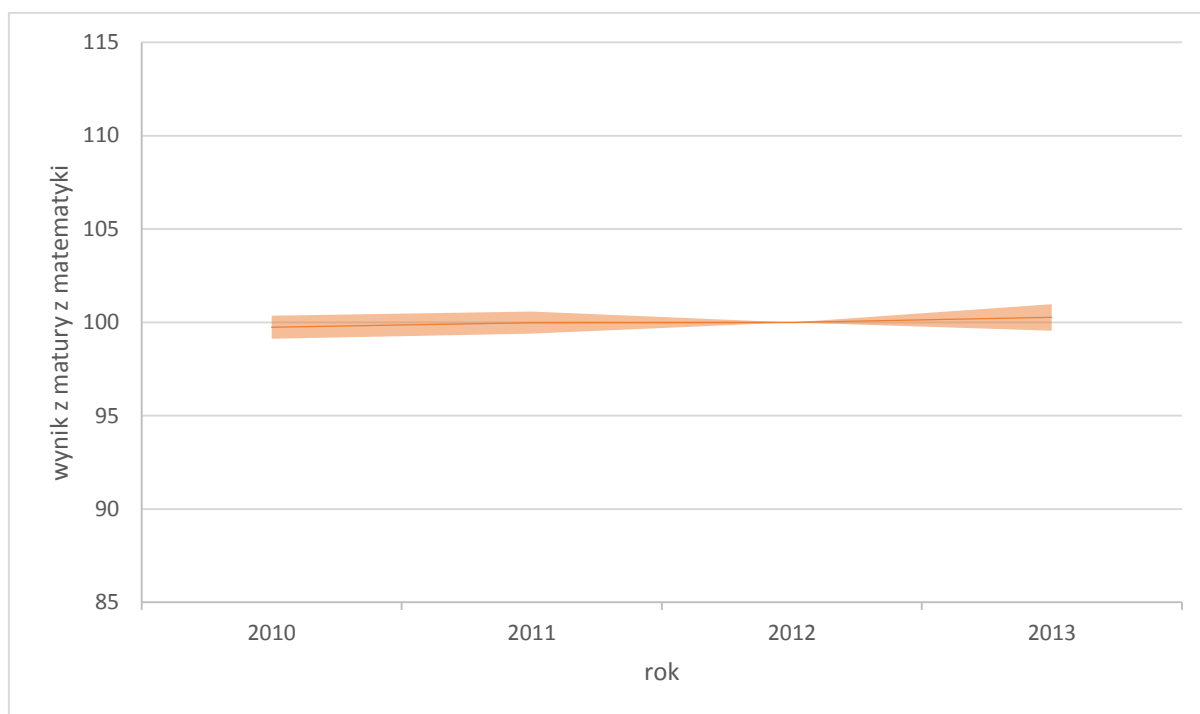
Informację o progu zdawalności uzupełnia zamieszczona w Aneksie 1 tabela pozwalająca przeliczyć wynik surowy uzyskany w latach 2010–2013 na wynik w teście z roku 2012. Można z niej odczytać, ile należałoby w każdej edycji egzaminu maturalnego uzyskać punktów, żeby poziom umiejętności odpowiadał progowi zdawalności egzaminu ustalonemu na maturze z roku 2012. I tak, w roku 2010, aby zdać maturę trzeba byłoby uzyskać 18 (a nie 15) punktów, aby zdać maturę według tych samych kryteriów co w roku 2012, natomiast w roku 2011 wystarczyłoby do tego zdobycie jedynie 13 (a nie 15) punktów. Tylko lata 2012 oraz 2013 wydają się mieć w pełni porównywalny próg zdawalności dla matury podstawowej z matematyki.

6.4.2. Wyniki w latach 2010–2013 na skali zmiennej ukrytej

Uzyskane w wyniku dopasowania do danych wielogrupowego modelu IRT oszacowania parametrów rozkładu umiejętności na skali o średniej 100 oraz odchyleniu standardowym 15 w roku referencyjnym 2012 przedstawia Tabela 6.6, ponadto Rysunek 6.9 ukazuje w sposób graficzny dynamikę zmian średniego poziomu umiejętności wraz z 95% przedziałami ufności. Okazuje się, że niewielkie różnice między średnimi zrównanych wyników obserwowanych, jakie omówiono wcześniej (zob. Rysunek 6.8 – przedostatni wykres) przekładają się na minimalne różnice na skali zmiennej ukrytej (od 99,7 w 2010 do 100,3 w 2013 roku), które nie przekraczają w żadnej konfiguracji progu istotności statystycznej. Natomiast obserwujemy istotne różnice w zakresie odchylenia standardowego. Odchylenie standardowe umiejętności matematycznych mierzonych testem maturalnym w roku 2010, oszacowane na 18,7, okazuje się istotnie statystycznie wyższe niż we wszystkich pozostałych latach, natomiast zbliżone odchylenia standardowe dla lat 2011 oraz 2013 (odpowiednio 16,7 oraz 16,8) są istotnie statystycznie wyższe od wartości 15, jaką ustalono dla roku referencyjnego 2012. Jest to wynik zgodny z tym, jaki uzyskano dla skali wyników obserwowanych (zob. Rysunek 6.8 – ostatni wykres), przy czym jest uzupełniony o informację o precyzji oszacowania.

Tabela 6.6. Średnia oraz odchylenie standardowe zmiennej umiejętności dla egzaminu maturalnego z matematyki w latach 2010–2013, wyniki zrównane na skali 100; 15 zakotwiczone w roku 2012

Rok egzaminu	Średnia	Błąd standard.	95% przedział ufn.		Odch. stand.	Błąd standard.	95% przedział ufn.	
			dolna gr.	górna gr.			dolna gr.	górna gr.
2010	99,7	0,31	99,1	100,3	18,7	0,42	17,9	19,6
2011	100,0	0,30	99,4	100,6	16,7	0,38	15,9	17,4
2012	100,0	–	–	–	15,0	–	–	–
2013	100,3	0,36	99,6	101,0	16,8	0,46	15,9	17,7



Rysunek 6.9. Średnie zrównane wyniki dla egzaminu maturalnego z matematyki na skali 100;15 wraz z 95% przedziałem ufności

6.5. Egzamin maturalny z języka polskiego na poziomie podstawowym

6.5.1. Specyficzne problemy związane ze zrównywaniem egzaminu maturalnego z języka polskiego

Zanim przejdziemy do omówienia wyników analizy zrównującej uzyskanych dla matury z języka polskiego na poziomie podstawowym poświęcimy trochę miejsca na omówienie specyficznych trudności związanych z przeprowadzeniem takiej analizy. Zasadniczą trudnością związaną z przeprowadzeniem zrównania tego egzaminu, jest fakt, że zdający mogą wybrać jeden z dwóch różnych tematów, na który formułują rozbudowaną wypowiedź pisemną podlegającą ocenie. Tematy różnią się znacznie zakresem lektur, jakie potencjalnie mogą się okazać użyteczne przy formułowaniu argumentacyjnego tekstu, ponadto jeden z tematów (częściej wybierany) wymaga stworzenia wypowiedzi o charakterze rozprawki problemowej, a drugi (rzadziej wybierany) wiąże się z koniecznością dokonania interpretacji tekstu literackiego. W terminach psychometrycznych oznacza to, że tematy zadań rozszerzonej odpowiedzi (wypracowań) z tej samej edycji egzaminu mogą potencjalnie różnić się trudnością, ale również mogą potencjalnie mierzyć nieco odmienne konstrukty. Należy się spodziewać, że całkowita wariancja wyników uzyskanych poprzez realizację konkretnego tematu będzie zawierała pewną porcję specyficzną dla tego zadania, która znajduje odzwierciedlenie w uniwersum treści jakie pragniemy testem maturalnym z języka polskiego mierzyć, ale nie będzie miała reprezentacji w innych zadaniach maturalnych. O ile problem różnic w trudności tematów jest dość prosty do kontroli statystycznej przy zastosowaniu podczas zrównywania modeli IRT, o tyle problem równoważności mierzonych konstruktyw jest zdecydowanie trudniejszy, przez co stanowi znaczne zagrożenie dla trafności przeprowadzonej analizy.

Konstrukcja egzaminu maturalnego z języka polskiego na poziomie podstawowym powoduje, że pojedyncza edycja egzaminu stanowi *de facto* mieszankę dwóch różnych testów, które zawierają część

zadań wspólnych (zadania nie będące wypracowaniem), oraz które nie są rozprowadzone między uczniami w sposób losowy. W pewnym sensie stanowi to klasyczny przykład schematu nierównoważnych grup z testem kotwiczącym (NEAT, por. Tabela 4.8, Rozdział 4), gdyż mamy dwie grupy uczniów, które potencjalnie różnią się poziomem umiejętności i każda z grup rozwiązuje test złożony z części zadań wspólnych (zadania kotwiczące) oraz części zadań, które są specyficznie wykorzystane tylko w jednej z grup. Mimo iż schemat zbierania danych dla pojedynczej edycji egzaminu przyjmuje formalnie postać schematu NEAT, jest istotna różnica między tym jak pomiar umiejętności poprawnie skonstruowany w schemacie NEAT powinien wyglądać, a tym co ma miejsce w przypadku matury z języka polskiego. Mianowicie, aby prawidłowo skontrolować różnice w trudności między dwoma egzaminami oraz różnice w poziomie umiejętności między dwoma grupami uczniów w schemacie NEAT, zadania kotwiczące powinny być reprezentatywną próbką zadań mierzących badane umiejętności. W przypadku matury z języka polskiego założenie to nie jest spełnione. Kotwiczące zadania wspólne dla dwóch grup uczniów nie będą w stanie uchwycić pewnych specyficznych aspektów umiejętności językowych jakie znajdują swoją reprezentację w realizacji tematów wypracowania. Zrównywanie wyników dwóch grup uczniów, wybierających różne tematy w ramach tej samej edycji egzaminu maturalnego będzie się wiązało z utratą pewnej porcji informacji o prawdziwym poziomie umiejętności uczniów. Porcji informacji, która jest specyficzna dla umiejętności realizowanych w ramach formułowania wypowiedzi pisemnej, a w szczególności która jest specyficzna dla umiejętności realizowanych w ramach formułowania wypowiedzi pisemnej jeden z wybranych tematów.

Dla dalszego zilustrowania problemów związanych z możliwością wyboru tematu wypowiedzi pisemnej, jakie mogą wpływać na zrównanie egzaminu maturalnego z języka polskiego, Tabela 6.7 przedstawia informacje o średniej oraz odchyleniu standardowym wyników uzyskiwanych za zadanie rozszerzonej odpowiedzi (wypracowanie), pozostałe zadania oraz za całość egzaminu. Tabela 6.7 zawiera także wyrażone na wspólnej skali parametry rozkładu uzyskane przez dopasowanie do danych wielogrupowego modelu IRT, które są zakotwiczone ze średnią 0 oraz odchyleniem standardowym 1 w grupie uczniów piszących temat drugi (II) w roku 2012 – wyniki te są podstawą analiz na zrównanych wynikach przedstawianych w dalszej części.

Tabela 6.7. Wyniki matury z języka polskiego w zależności od wybranego tematu*

Rok i temat			% uczniów	Suma punktów za wypracowanie	Suma punktów za pozostałe zadania		Suma za całą pracę		Parametry wielogrupowego modelu irt			
				Średnia	SD	Średnia	SD	Średnia	SD	Średnia	Błąd st.	SD
2010	I	8,4%	32,6	8,1	14,8	2,9	47,4	9,7	1,10	0,058	1,05	0,037
	II	15,2%	23,9	8,6	13,8	3,2	37,6	10,5	0,44	0,044	1,24	0,043
	?	76,4%	27,4	9,7	14,2	3,1	41,6	11,5	0,71	0,048	1,20	0,041
2011	I	18,2%	29,0	9,9	13,8	3,3	42,8	12,1	0,63	0,033	0,95	0,026
	II	62,5%	24,3	8,4	13,4	3,3	37,7	10,5	0,45	0,030	0,90	0,025
	?	19,4%	24,9	9,4	13,3	3,3	38,2	11,6	0,42	0,030	0,93	0,026
2012	I	8,6%	26,18	8,7	13,7	3,0	39,9	10,6	0,70	0,009	0,94	0,008
	II	91,1%	26,22	8,2	12,2	3,3	38,4	10,2	0,00	–	1,00	–
	?	0,4%	28,6	16,1	13,0	4,8	41,5	20,2	0,47	0,109	3,35	0,108
2013	I	21,6%	27,1	9,2	14,1	3,2	41,3	11,3	0,45	0,030	0,95	0,024
	II	78,0%	25,0	8,5	13,9	3,1	38,8	10,4	0,31	0,028	0,86	0,021
	?	0,3%	27,4	17,5	14,1	5,0	41,5	21,7	1,01	0,115	2,88	0,129

* „?” oznacza przypadki braku informacji o tym, jaki temat został wybrany; SD – odchylenie standardowe

Tabela 6.7 wskazuje na pierwszą informację ilustrującą dalsze problemy, które mogą wystąpić podczas zrównywania wyników: nie w każdym przypadku informacja o wyborze tematu przez ucznia została zebrana. Dla takich uczniów wyniki za realizację wypowiedzi pisemnej są modelowane jako kolejny, trzeci zestaw, unikalnych dla tej grupy zadań, mimo iż wiadomo, że jest to mieszanka wyników z realizacji tematów I oraz II. W pierwszych dwóch latach problem braku informacji o wyborze tematu był systemowy – tylko niektóre okręgowe komisje egzaminacyjne zbierały takie dane – w dwóch ostatnich latach przyczyna nie jest jasna, jednocześnie braki w tej zmiennej mają już tylko marginalny charakter.

Przyjmując, że brak informacji o wyborze tematu ma charakter losowy, możemy oszacować zróżnicowaną wybieralność tematu I oraz II między latami. Temat I (interpretacyjny) w roku 2010 był wybierany przez 35,7% uczniów, w roku 2011 przez 22,5% uczniów, w roku 2012 przez 8,6% uczniów, a w roku 2013 przez 21,7% uczniów. Widzimy zatem, że rok 2012 znacznie odstaje na tle reszty, przez to że interpretacyjny temat był wybierany przez uczniów bardzo rzadko. Specyficzne zachowanie się wyników w roku 2012 na tym się nie kończy. Dla wszystkich lat, oprócz 2012, możemy zaobserwować, że uczniowie wybierający temat I uzyskiwali wyższe wyniki na części egzaminu nie będącej wypracowaniem, co jest sygnałem, że ich poziom umiejętności jest wyższy niż uczniów wybierających temat II, i jednocześnie uzyskiwali wyższe wyniki za część egzaminu będącą rozbudowaną wypowiedzią pisemną. Natomiast w roku 2012, 8,6% uczniów wybierających temat I również uzyskiwało wyższe wyniki za część egzaminu nie będącą wypracowaniem, co również przemawia za tym, że ich ogólny poziom umiejętności językowych był wyższy niż uczniów wybierających temat II w tym roku. Jednak średni wynik z realizacji części pisemnej dla grupy wybierającej temat I w roku 2012 był nawet nieznacznie niższy (na drugim miejscu po przecinku) niż w grupie, która wybrała temat II. Obserwacja, że uczniowie o wyższym poziomie ogólnej umiejętności językowej (za czym przemawia wyższy wynik na części egzaminu nie będącej wypowiedzią pisemną) uzyskali za wypowiedź pisemną w 2012 niższy wynik nakazuje wyciągnąć wniosek, że realizacja tematu I była na pewno trudniejsza od realizacji tematu II w tym roku.

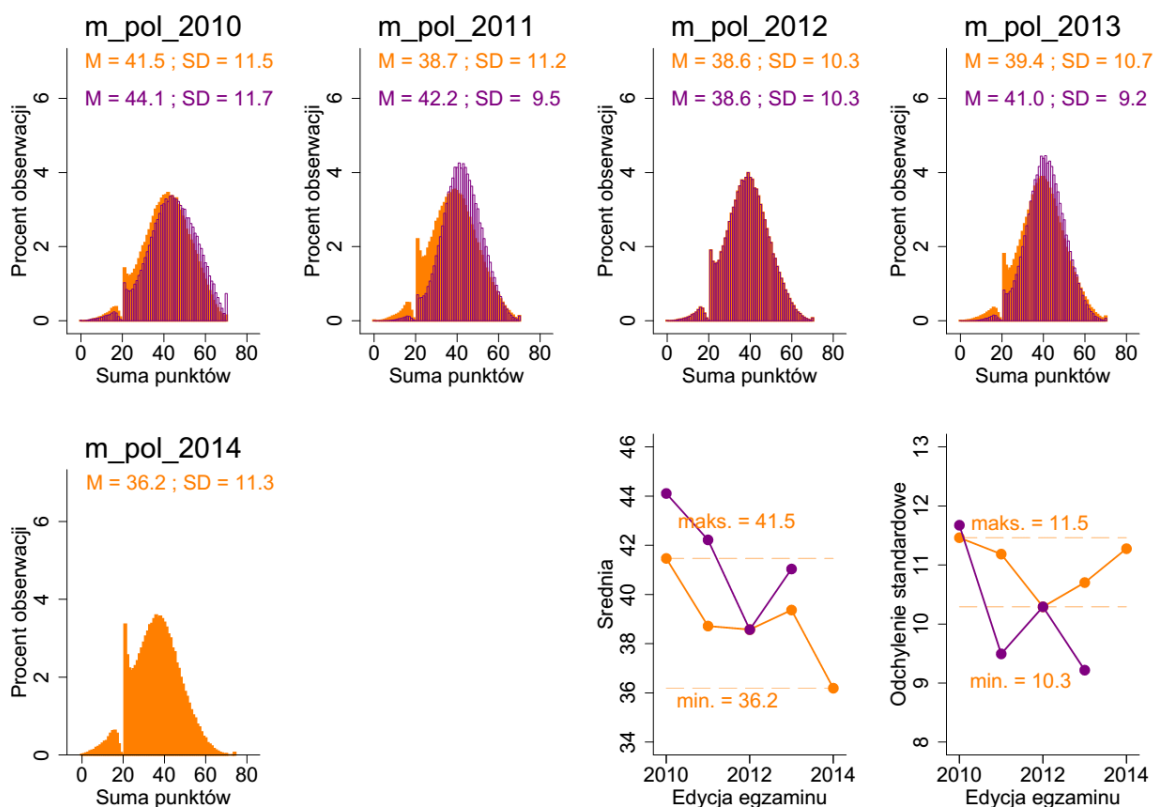
Choć różnice w trudności zadań mogą zostać poddane kontroli poprzez zastosowanie wielogrupowego modelu IRT, rodzi się pytanie czy obserwowany wzór różnic w wynikach w dwóch wyróżnionych częściach egzaminu maturalnego z języka polskiego nie jest związany z wspomnianymi wcześniej specyficznymi aspektami umiejętności językowych wymaganymi do poprawnego napisania wypowiedzi pisemnej na poszczególne tematy. Skutki takich hipotetycznych różnic w specyficzności tematów I oraz II są trudne do przewidzenia oraz nie mogły być kontrolowane w trakcie przeprowadzonej analizy. Problematyka ta jest związana nie tylko z oceną trafności prezentowanych wyników zrównanych z matury z języka polskiego, ale także, w całej ogólności, z możliwością porównywania wyników uzyskiwanych przez uczniów wybierających poszczególne tematy wypracowania podczas danej edycji matury z języka polskiego oraz z możliwością porównywania wyników tego egzaminu między latami. Jest to doniosłe zagadnienie, wymagające dalszej skrupulatnej analizy, która jednak wykracza poza zakres niniejszej monografii. Niemniej, trzeba zaznaczyć, że prezentowane w dalszej części wyniki zrównania matury z języka polskiego mogą być obciążone pewnym błędem związanym z brakiem kontroli specyficzności tematów wypowiedzi pisemnej. Uwaga ta jest istotna tym bardziej, że prezentowane dalej wyniki dla matury z języka polskiego w znacznym stopniu odstają swoim wzorem od tych jakie obserwujemy we wszystkich innych zrównywanych egzaminach.

Ze względu na specyficzną dla tego egzaminu maturalnego konstrukcję testu, polegającą na tym, że uczniowie wybierają jeden z dwóch tematów wypowiedzi pisemnej, pewnego komentarza wymaga sposób, w jaki obserwowane wyniki na skali testu z roku 2012 zostały oszacowane dla innych roczników. W celu oszacowania rozkładów na skali z roku 2012 najpierw dopasowano do zbioru danych zrównujących model IRT, w którym wyniki uzyskiwane za realizację różnych tematów były traktowane jako odrębne zadania, a dla grup uczniów wybierających różne tematy włączono do modelu możliwość

niezależnego oszacowania rozkładu umiejętności (parametry modelu IRT dla grup przedstawia Tabela 6.7). Wielogrupowe modelowanie w obrębie pojedynczej edycji egzaminu miało na celu kontrolę nielosowego (skorelowanego z poziomem umiejętności) mechanizmu leżącego za wyborem tematu. Następnie, wyliczono warunkowe prawdopodobieństwo uzyskania określonej liczby punktów w maturze z 2012 roku w zależności od poziomu zmiennej ukrytej uczniów rozwiązujących test w roku 2012, zupełnie ignorując informację o tym, czy suma punktów w teście została uzyskana przy wyborze tematu I lub tematu II. Ostatecznie, takie nieparametryczne krzywe charakterystyczne dla całego testu, zostały wykorzystane do oszacowania rozkładów wyników obserwowanych w innych latach niż 2012 przy uwzględnieniu rozkładu ukrytej zmiennej umiejętności w tych latach. Zastosowana procedura oznacza przyjęcie implícite założenia, że gdyby uczniowie w latach innych niż 2012 mieli do wyboru takie same tematy jak para pojawiająca się w egzaminie z 2012, to zależność pomiędzy poziomem umiejętności uczniów z tych lat, a dokonywanym przez nich wyborem tematu byłaby identyczna jak w 2012 roku. W szczególności, gdyby, hipotetycznie, między latami nie było różnic w zakresie umiejętności językowych, to w każdym z roczników temat I, zamiast tematu II, z 2012 roku wybierałoby zawsze około 8% uczniów. Jeżeli przyjęte założenie o stałej między latami zależności między poziomem umiejętności językowych a prawdopodobieństwem wyboru tematu I zamiast tematu II w roku 2012 jest błędne, to prezentowane wyniki będą tym błędem w odpowiednim stopniu obciążone. Należy zauważyć, że mechanizmy leżące u podstaw podejmowanego przez uczniów wyboru tematu spośród pojawiającej się w każdym roku pary są problemem mało zbadanym, a mającym bardzo istotne znaczenie zarówno przy interpretacji (zrównanych, jak i niezrównanych) wyników z matury z języka polskiego.

6.5.2. Zmiany trudności egzaminu w latach 2010–2013

Rysunek 6.10 przedstawia rozkłady wyników z pierwszego terminu egzaminu maturalnego z języka polskiego na skali egzaminu z roku 2012 oszacowane na podstawie analizy zrównującej (puste słupki) oraz faktycznie uzyskane przez uczniów (pełne słupki) w latach 2010–2013. Okazuje się, że w przypadku matury z języka polskiego mamy do czynienia z odmienną niż dla innych egzaminów zależnością między obserwowanymi wynikami zrównanymi, a faktycznym rozkładem wyników egzaminacyjnych. Mianowicie, matura z języka polskiego jest jedynym egzaminem, dla którego obserwujemy silniejsze zmiany w zakresie średniej, jak i odchylenia standardowego wyników zrównanych do roku referencyjnego, niż fluktuacje analogicznych parametrów między latami dla surowych wyników obserwowanych (dwa ostatnie wykresy – Rysunek 6.10). Taki wynik odpowiada sytuacji, w której przy znacznych zmianach umiejętności uczniów między latami, trudność egzaminów ma względem tych zmian charakter buforujący, tj. populacja uczniów bardziej zdolnych podchodzi do egzaminu trudniejszego (2010, w dalszej kolejności 2011 oraz 2013), a populacja uczniów mniej zdolnych – do egzaminu łatwiejszego (rok 2012).



Rysunek 6.10. Rozkłady wyników obserwowanych z matury podstawowej z języka polskiego w latach 2010–2014. Pełne słupki odnoszą się do niezrównanych wyników egzaminacyjnych, zrównane wyniki przedstawiono za pomocą słupków pustych

Tabela 6.8 pokazuje jak różnice we właściwościach rozkładów egzaminów z matury z języka polskiego w latach 2010–2013 przy zestawieniu z wynikami na wspólnej skali egzaminu z roku 2012 przekładają się na zdawalność tego egzaminu. W tabeli tej przedstawiono jak wyglądała faktyczna zdawalność egzaminu w analizowanych czterech rocznikach uczniów i jakby zdawalność dla tych samych roczników wyglądała, gdyby wszyscy rozwiązywali ten sam test – egzamin referencyjny z roku 2012. Zgodnie z tym, że zrównane do roku 2012 obserwowane wyniki egzaminacyjne podniosły w latach innych niż 2012 średni wynik z egzaminu, w porównaniu do wyniku faktycznie uzyskanego, możemy zaobserwować wzrost zdawalności egzaminu w latach 2010 (z 97,3 do 98,1), 2011 (z 96,8 do 99,2) oraz 2013 (z 97,4% do 99,1%). Egzamin z roku 2010, nawet po wyrażeniu na skali zrównanej do 2012 charakteryzuje się niższą w porównaniu z latami 2011 oraz 2013 zdawalnością, mimo iż średni poziom umiejętności na wspólnej skali w tym roku oszacowano na najwyższy. Dzieje się tak, ponieważ wariancja wyników zrównanych w tym roku jest zdecydowanie wyższa niż dla innych lat. Aneks 1 zawiera tabelę pozwalającą przeliczyć surowy próg zdawalności w roku 2012 na równoważne mu poziomem umiejętności językowych odpowiedniki w innych latach. I tak, próg zdawalności równoważny temu z 2012 roku oszacowano dla roku 2010 przypada w zakresie 17–20 punktów (zamiast 21), dla roku 2011 w zakresie 12–14 punktów (zamiast 21) a w roku 2013 w zakresie 13–15 punktów. Zamiast pojedynczej reprezentacji punktowej progu zdawalności pojawiają się tu zakresy, gdyż wartość na progu zdawalności w roku 2012 (i w każdym innym dla matury z języka polskiego) jest nadreprezentowana względem innych wartości, zwłaszcza tych poniżej progu.

Tabela 6.8. Zdawalność egzaminu maturalnego z języka polskiego wyliczona na niezrównanych wynikach egzaminacyjnych oraz na wynikach wyrażonych na skali egzaminu z roku 2012

Rok egzaminu	Procent uczniów uzyskujących 21 punktów lub więcej		
	niezrównane egzaminacyjne	wyniki	wyniki na skali egzaminu z 2012 roku
2010	97,3%		98,1%
2011	96,8%		99,2%
2012	97,6%		97,6%
2013	97,4%		99,1%

6.5.3. Wyniki w latach 2010–2013 na skali zmiennej ukrytej

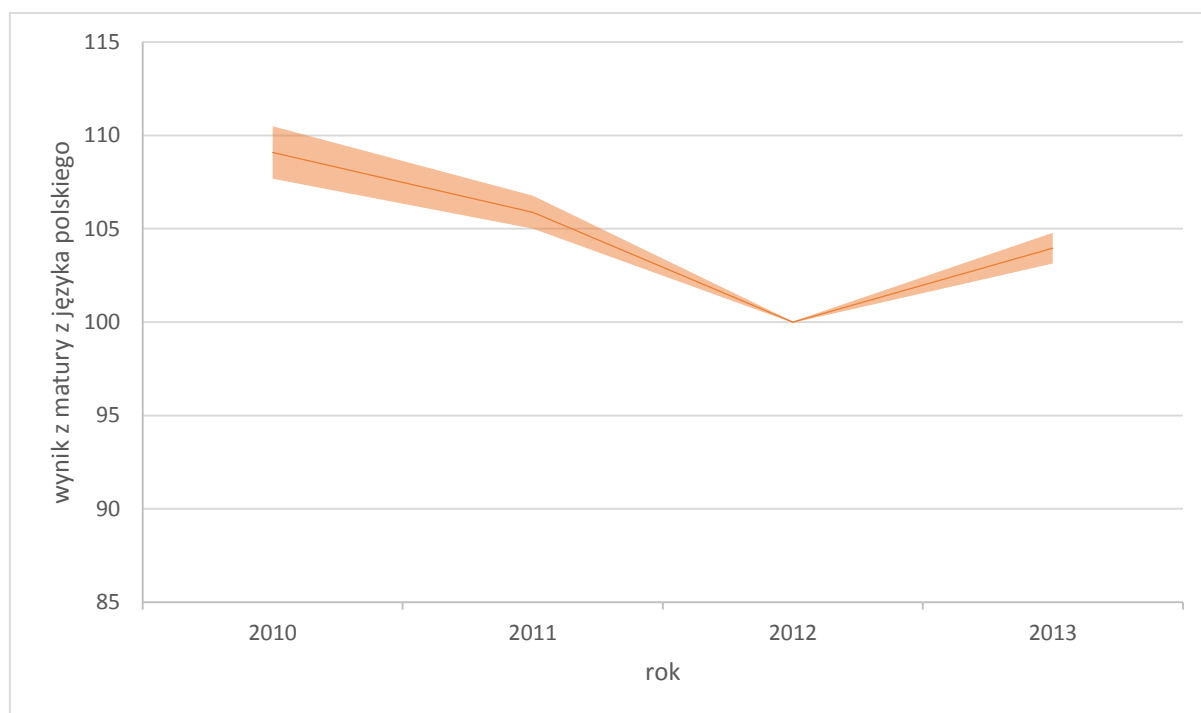
Tabela 6.7 przedstawia uzyskane w wyniku bezpośredniego dopasowania do danych wielogrupowego modelu IRT oszacowania parametrów rozkładu umiejętności, zakotwiczone na skali o średniej 0 oraz odchyleniu standardowym 1 w grupie uczniów piszących temat II w roku 2012. Aby wyrazić wyniki zrównania w sposób analogiczny do zestawień dla innych egzaminów, w których wyniki zmiennej ukrytej są zakotwiczone na skali o średniej 100 oraz odchyleniu standardowym 15 w całej populacji uczniów podchodzących do egzaminu, wartości (zob. Tabela 6.7) uśredniono uwzględniając wybieralność tematów w poszczególnych rocznikach, a następnie przekształcono, tak aby uzyskać docelowy rozkład w grupie referencyjnej. Ostateczne wyniki zrównania na skali zmiennej ukrytej dla matury z języka polskiego przedstawia Tabela 6.9, oraz graficznie Rysunek 6.11.

Tabela 6.9. Średnia oraz odchylenie standardowe zmiennej umiejętności dla egzaminu maturalnego z języka polskiego w latach 2010–2013, wyniki zrównane na skali 100; 15 zakotwiczone w roku 2012

Rok egzaminu	Średnia	Błąd standard.	95% przedział ufn.		Odch. stand.	Błąd standard.	95% przedział ufn.	
			dolna gr.	górna gr.			dolna gr.	górna gr.
2010	109,1	0,72	107,7	110,5	17,9	0,61	16,7	19,1
2011	105,9	0,45	105,0	106,8	13,5	0,37	12,8	14,2
2012	100,0	0,02	100,0	100,0	15,0	0,02	15,0	15,0
2013	104,0	0,42	103,1	104,8	13,0	0,33	12,3	13,6

Różnice w średnich wynikach umiejętności językowych mierzonych testem maturalnym okazują się na tyle znaczne, że średni poziom umiejętności między wszystkimi analizowanymi rocznikami różni się w

sposób istotny statystycznie, nawet mimo relatywnie dużych błędów standardowych. Największą różnicę obserwujemy między rokiem 2010, ze średnią 109,1 a rokiem referencyjnym 2012. Tak wysokich różnic (rzędu 60% odchylenia standardowego) między populacjami uczniów rozwiązującymi egzamin w różnych latach nie obserwowano dla żadnego innego egzaminu. Jeżeli dodatkowo uwzględnimy wszystkie wcześniej opisane wątpliwości związane z trafnością zrównywania matury z języka polskiego, to należy podejść do uzyskanego rezultatu z dużą ostrożnością. Oprócz różnic w średnim poziomie umiejętności zanotowano istotne statystycznie różnice dla parametru odchylenia standardowego (zob. Tabela 6.9). Rozproszenie wyników jest silniejsze w roku 2010 w porównaniu z rokiem referencyjnym, natomiast w latach 2011 oraz 2013 jest niższe niż w 2012.

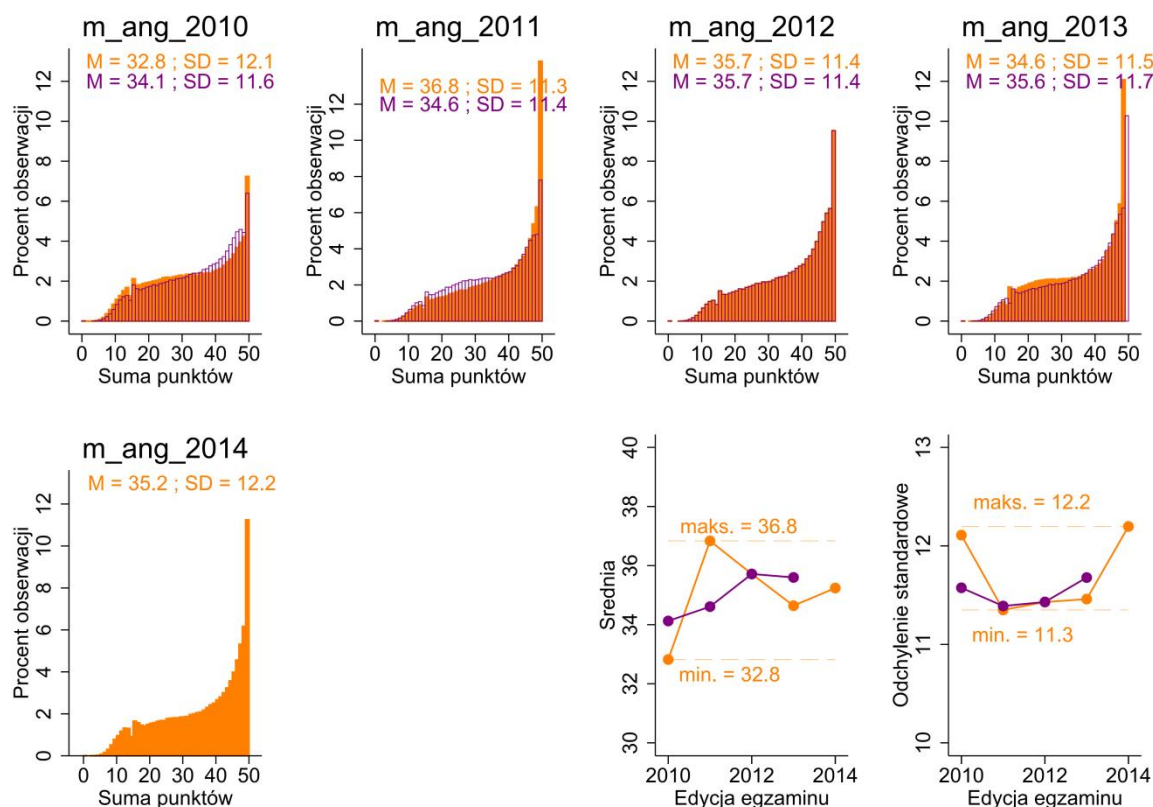


Rysunek 6.11. Średnie zrównane wyniki dla egzaminu maturalnego z języka polskiego na skali 100;15 wraz z 95% przedziałem ufności

6.6. Egzamin maturalny z języka angielskiego na poziomie podstawowym

6.6.1. Zmiany trudności egzaminu w latach 2010–2013

Wyniki obserwowane z matury z języka angielskiego na poziomie podstawowym mają bardzo charakterystyczny kształt, który utrzymuje się we wszystkich analizowanych edycjach egzaminu (zob. Rysunek 6.12). Rozkłady są bardzo silnie lewostronnie skośne, w związku z wyraźnym efektem sufitowym. Porównując rozkłady wyników egzaminacyjnych faktycznie uzyskane w poszczególnych rocznikach z rozkładami wyników na skali referencyjnego roku 2012 obserwujemy większą stabilność tych drugich na przestrzeni lat. Średnia zrównanych do 2012 roku wyników zmienia się w latach 2010–2013 w zakresie 34,1–35,7 punktu, natomiast oryginalne średnie mają w tym okresie dwuipółkrotnie większą rozpiętość: od 32,8 do 36,8 punktu. Większą stabilność parametrów zrównanych rozkładów obserwujemy również dla odchylenia standardowego. Taki obraz skłania do sformułowania wniosku, że obserwowane między latami zmiany na surowych wynikach egzaminu maturalnego z języka angielskiego są głównie konsekwencją różnic w trudności między egzaminami.



Rysunek 6.12. Rozkłady wyników obserwowanych z matury podstawowej z języka angielskiego w latach 2010–2014. Pełne słupki odnoszą się do niezrównanych wyników egzaminacyjnych, zrównane wyniki przedstawiono za pomocą słupków pustych

Przełożenie efektu różnic we właściwościach psychometrycznych egzaminów matury z języka angielskiego na próg zdawalności tego egzaminu przedstawia Tabela 6.10. Efekt wygładzenia różnic między rozkładami towarzyszący zastosowaniu wyników zrównanych do skali z 2012 roku, ilustrowany przez Rysunek 6.12, znajduje swoje odzwierciedlenie również przy analizie progów zdawalności. Zdawalność egzaminu w latach 2010–2013 przedstawiona na oryginalnej skali obejmuje zakres od 91,1% (rok 2010) do 95,5% (rok 2011), natomiast wyrażona na porównywalnej między latami skali egzaminu z roku referencyjnego zawęża się do zakresu od 93,2% (rok 2010) do 94,6% (rok 2011). Posiłkując się tablicą przeliczeniową wyników surowych na wyniki zrównane do skali z roku 2012, jaką zamieszczono w Aneksie 1, możemy stwierdzić jaka liczba punktów zdobytych w egzaminie z innych lat odpowiada progowi zdawalności w roku referencyjnym. Aby zdać egzamin w roku 2010 trzeba było na porównywalnej skali z roku 2012 zdobyć 13,5 punktu²⁵, w roku 2011 15,5–16 punktów, natomiast w roku 2013 15–15,5 punktu. Zatem, w porównaniu z rokiem 2012, próg zdawalności był zauważalnie zawyżony w edycji egzaminu z 2010 roku, a w latach 2011 oraz 2013 był nieznacznie zawyżony.

²⁵ W przypadku niektórych zadań na egzaminie maturalnym z języka angielskiego na poziomie podstawowym istnieje możliwość zdobycia 0,5 punktu, co przekłada się na możliwość zdobywania wyników połowkowych z całego egzaminu.

Tabela 6.10. Zdawalność egzaminu maturalnego z języka angielskiego wyliczona na niezrównanych wynikach egzaminacyjnych oraz na wynikach wyrażonych na skali egzaminu z roku 2012

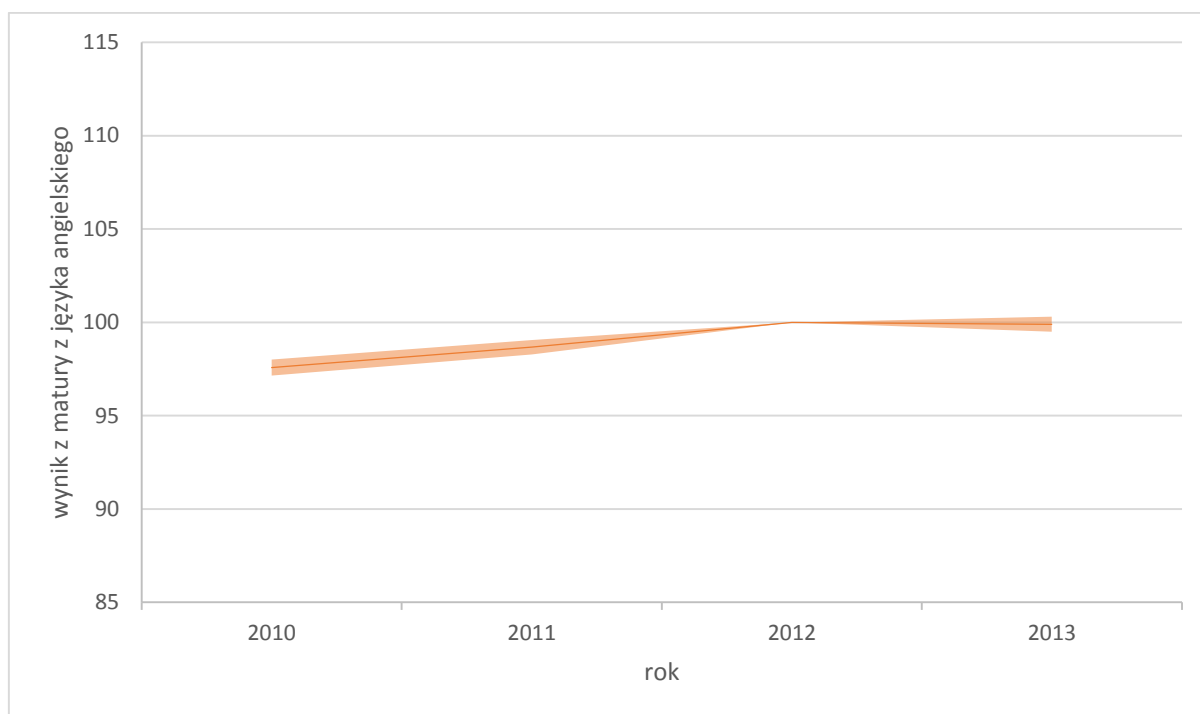
Rok egzaminu	Procent uczniów uzyskujących 15 punktów lub więcej		
	niezrównane egzaminacyjne	wyniki	wyniki na skali egzaminu z 2012 roku
2010	91,1%		93,2%
2011	95,5%		94,5%
2012	94,6%		94,6%
2013	94,4%		94,1%

6.6.2. Wyniki w latach 2010–2013 na skali zmiennej ukrytej

Analizując zmiany w poziomie umiejętności mierzonych testem maturalnym z języka angielskiego wyrażonymi na skali zmiennej ukrytej (zob. Tabela 6.11 oraz Rysunek 6.13) zauważamy istotny statystycznie wzrost umiejętności między latami 2010 oraz 2011 (o 1,1 punktu) oraz między latami 2011 oraz 2012 (o 1,3 punktu). Średni poziom umiejętności w latach 2012 oraz 2013 należy natomiast uznać za stały o czym świadczy nie tylko brak istotności statystycznej różnic, ale również minimalna wartość różnicy między średnimi w tych dwóch latach wynosząca 0,1 punktu na skali o odchyleniu standardowym równym 15 w roku 2012. Analizując zmiany w parametrze rozproszenia, obserwujemy stabilność odchylenia standardowego wyników w przedziale czasowym, w którym zanotowano wzrost średnich (2010–2012) oraz istotną statystycznie zmianę wartości odchylenia standardowego w roku 2013.

Tabela 6.11. Średnia oraz odchylenie standardowe zmiennej umiejętności dla egzaminu maturalnego z języka angielskiego w latach 2010–2013, wyniki zrównane na skali 100; 15 zakotwiczone w roku 2012

Rok egzaminu	Średnia	Błąd standard.	95% przedział ufn.		Odch. stand.	Błąd standard.	95% przedział ufn.	
			dolna gr.	górną gr.			dolna gr.	górną gr.
2010	97,6	0,22	97,1	98,0	15,1	0,33	14,4	15,7
2011	98,7	0,20	98,3	99,1	14,5	0,27	14,0	15,1
2012	100,0	–	–	–	15,0	–	–	–
2013	99,9	0,20	99,5	100,3	15,8	0,34	15,2	16,5



Rysunek 6.13. Średnie zrównane wyniki dla egzaminu maturalnego z języka angielskiego na skali 100;15 wraz z 95% przedziałem ufności

6.7. Zrównanie ekwicytylowe

Poniżej zaprezentowano wyniki zrównania sprawdzianu (2002–2012), egzaminu gimnazjalnego części humanistycznej i matematyczno-przyrodniczej (2002–2012) oraz egzaminu maturalnego z języka polskiego, angielskiego i matematyki (2010–2013) metodą ekwicytylową. Prezentowane tu wyniki służą jedynie porównaniu i walidacji wyników zrównania opartego o IRT, stąd nie będziemy tu skupiać się na omówieniu występujących w nich trendów związanych z umiejętnościami uczniów. Do zrównania użyto autorskiego oprogramowania (Wołodźko, 2015) działającego w oparciu o środowisko statystyczne R (R Core Team, 2014).

W przypadku sprawdzianu i egzaminu gimnazjalnego użyto łańcuchów zrównań w planie pojedynczej grupy (*single group*, SG; zob. podrozdział 4.5), gdzie testem bazowym był dla sprawdzianu test z roku 2004, a dla egzaminu gimnazjalnego z 2003 roku. Każdy z testów został zrównany do testu bazowego przy pomocy kilku łańcuchów funkcji zrównujących, które następnie zostały uśrednione w celu minimalizacji błędów zrównania. Struktura danych dotyczących tych egzaminów zebranych podczas badań pozwalała co prawda na zrównanie przy użyciu prostszego planu grup równoważnych (*equivalent groups*, EG; zob. podrozdział 4.5 4.5), należy jednak zaznaczyć, że zrównanie w planie EG jest obciążone nieco większym błędem zrównania, co pokazały przeprowadzone analizy. Procedura zrównania tych egzaminów oraz szczegółowy opis przeprowadzonych analiz zostały szczegółowo omówione w artykule Tymoteusza Wołodźko, Bartosza Kondratka i Henryka Szaleńca (2014).

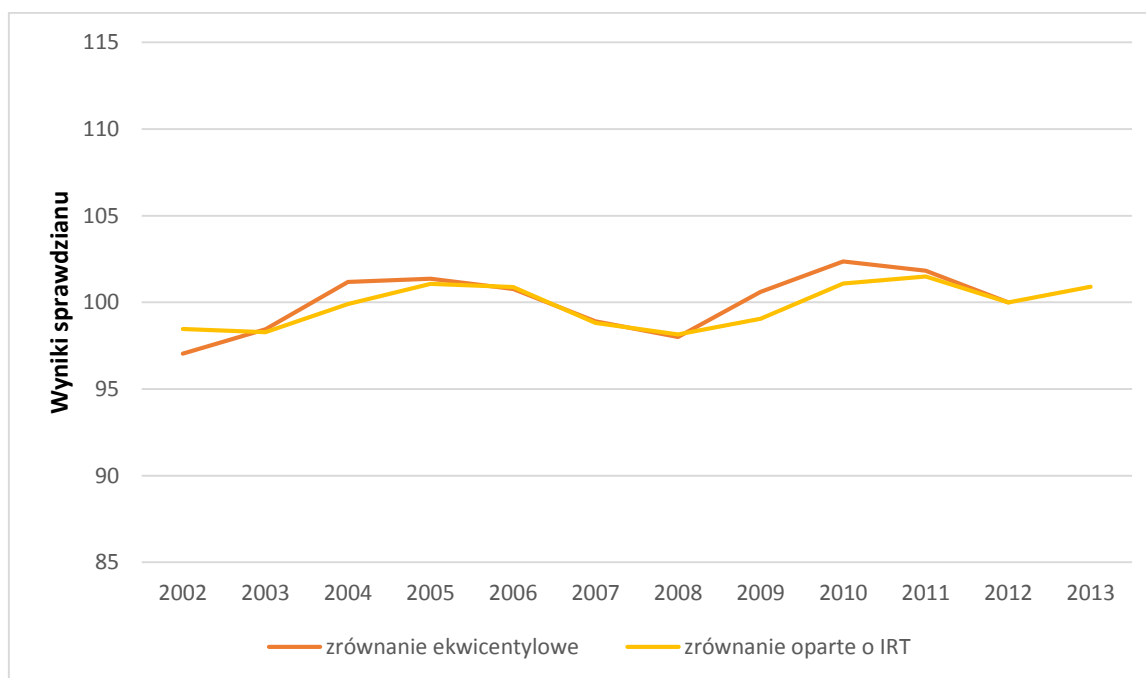
Dla zrównanych wyników egzaminów maturalnych przyjęto jako bazowy test z roku 2012. W przypadku zrównania wyników egzaminu maturalnego z języka angielskiego zastosowano plan SG, gdzie poszczególne testy zrównane zostały parami do testu z roku 2012.

Dane użyte przy zrównaniu wyników egzaminu maturalnego z matematyki pochodziły z dwóch badań przeprowadzonych w latach 2013 i 2014, gdzie pierwsze z nich pozwoliło na zrównania parami testów z lat 2010 i 2011 z testem z 2012 roku, a drugie, na zrównanie testu z 2013 roku z testem z 2012 roku. Zrównania odbyły się w planie SG.

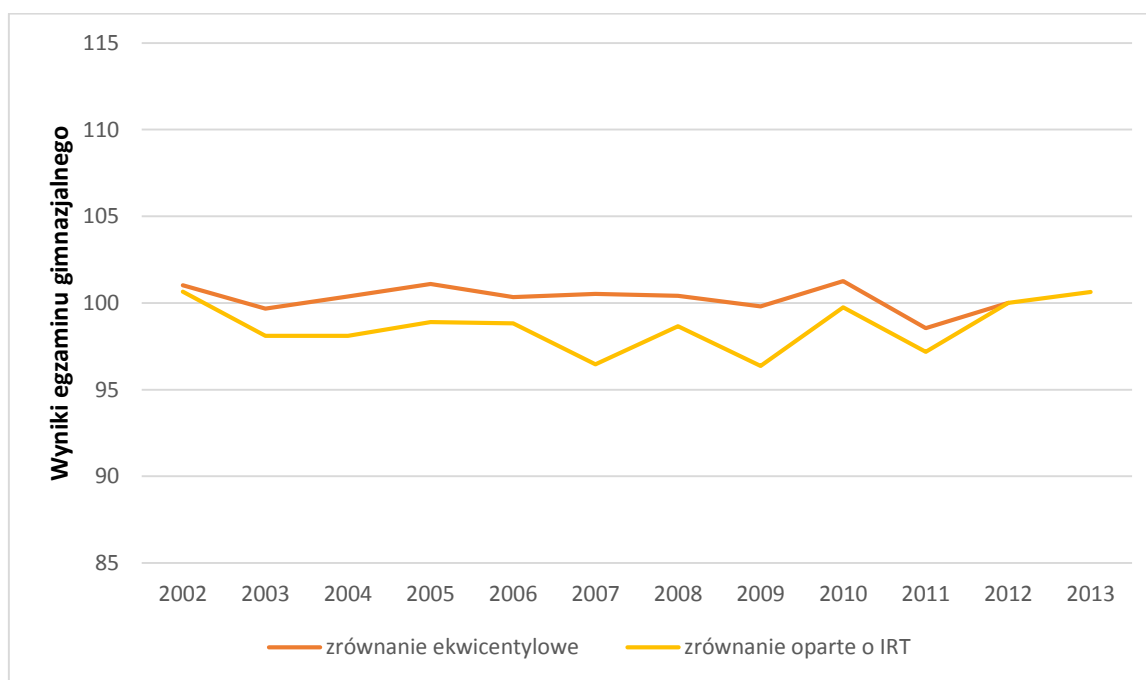
Najbardziej skomplikowany plan zrównania zastosowany został w przypadku egzaminu z języka polskiego. Zauważmy, że jest to egzamin, w którym uczniowie samodzielnie mogą wybrać jeden z dwóch tematów, na który muszą przygotować rozszerzoną wypowiedź pisemną. W przypadku zrównywania tych wyników oznacza to *de facto*, że mamy do czynienia z dwoma różnymi testami rozwiązywanymi przez dwie grupy uczniów, które potencjalnie mogą różnić się poziomem umiejętności. Co więcej, egzamin maturalny z języka polskiego składa się zarówno z zadań zamkniętych lub krótkiej wypowiedzi, jak i rozszerzonej wypowiedzi, prezentowane w tym podrozdziale wyniki dotyczą jednak wyłącznie zrównania ocen przyznawanych uczniom za rozszerzoną wypowiedź pisemną. By zrównać takie wyniki konieczne było w pierwszej kolejności zrównanie ze sobą obu wersji testu z jednego roku (obu tematów rozszerzonej wypowiedzi), tak by w efekcie końcowym wyniki uzyskane przez uczniów rozwiązujących testy w poszczególnych edycjach egzaminu maturalnego zawsze wyrażone były w skali jednego z tematów w tych testach (wybrano temat drugi). Zrównanie wyników dotyczących dwóch tematów dla jednego roku przeprowadzone zostało w planie nierównoważnych grup z testem kotwiczącym (*nonequivalent groups with anchor test*, NEAT; zob. podrozdział 4.5), gdzie wyniki dotyczące pierwszego tematu zakotwiczone zostały przy użyciu zadań zamkniętych lub krótkiej wypowiedzi z danego roku przy użyciu danych egzaminacyjnych, a następnie zrównane do skali tematu drugiego. Tak uzyskane wyniki w skali tematu drugiego w następnej kolejności zrównane zostały w planie EG do testu-tematu drugiego z roku 2012, który pełnił rolę testu bazowego.

Wszystkie wyniki zrównań, niezależnie od rodzaju egzaminu, zostały przekształcone liniowo do skali, w której w roku 2012 średnia wynosiła 100 i odchyleniu standardowym 15, co miało na celu ułatwienie Czytelnikowi porównania uzyskanych wyników ze zrównaniem metodą opartą na IRT.

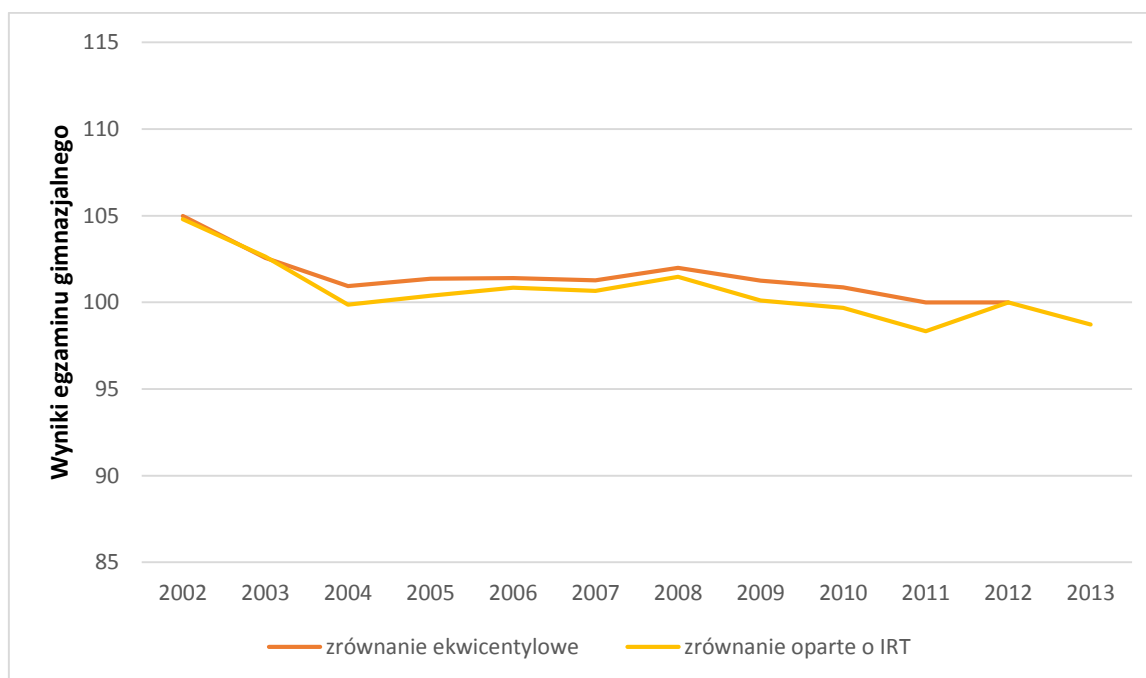
Zestawione tu wyniki zrównania ekwicytylowego i opartego o IRT nie różnią się między sobą w sposób istotny statystycznie (por. Wołodźko, Kondrątek i Szaleniec, 2014). Jak możemy się przekonać na podstawie prezentowanych poniżej rysunków, większe, choć nieistotne, różnice między tymi metodami zrównania można zaobserwować w przypadku sprawdzianu (por. Rysunek 6.14) i egzaminu gimnazjalnego (por. Rysunek 6.15 i Rysunek 6.16), niż w przypadku egzaminów maturalnych (por. Rysunek 6.17, Rysunek 6.18 i Rysunek 6.19). Wyniki zrównania ekwicytylowego charakteryzują się mniejszymi wahaniami wokół średniej, niż zrównanie oparte o IRT. Różnice te możemy potencjalnie wyjaśniać zastosowaniem dla sprawdzianu i egzaminu gimnazjalnego bardziej skomplikowanego schematu badawczego, prowadzącego do użycia dłuższych łańcuchów funkcji zrównujących, co nieuchronnie prowadzi do sumowania się błędów jakimi obarczone są poszczególne zrównania (Wołodźko, Kondrątek i Szaleniec, 2014). W przypadku egzaminów maturalnych zastosowany schemat badawczy pozwolił na zrównanie poszczególnych testów parami w planie SG lub EG dla egzaminu z języka polskiego, dzięki czemu procedura obarczona jest mniejszym błędem. Wyniki zrównań dla wszystkich egzaminów prezentują podobne trendy, a różnice między nimi są pomijalne, wyjątkiem są zrównania części humanistycznej egzaminu gimnazjalnego i egzaminu maturalnego z języka polskiego. W przypadku części humanistycznej egzaminu gimnazjalnego i egzaminu maturalnego z języka polskiego ujawniły się różnice między oszacowaniami uzyskanymi metodą ekwicytylową oraz opartą o IRT. W przypadku części humanistycznej egzaminu gimnazjalnego największa różnica między zrównaniami przy użyciu dwóch metod, wynosi 4,07 punktu w roku 2007, a dla pozostałych lat od 0,37 do 3,43 punktów, natomiast dla egzaminu maturalnego z języka polskiego różnica dla lat 2010 i 2011 wynosi około pięciu punktów. Równocześnie należy zaznaczyć, że w przypadku zrównań wszystkich testów kształty obserwowanych trendów są podobne w przypadku obu zrównań.



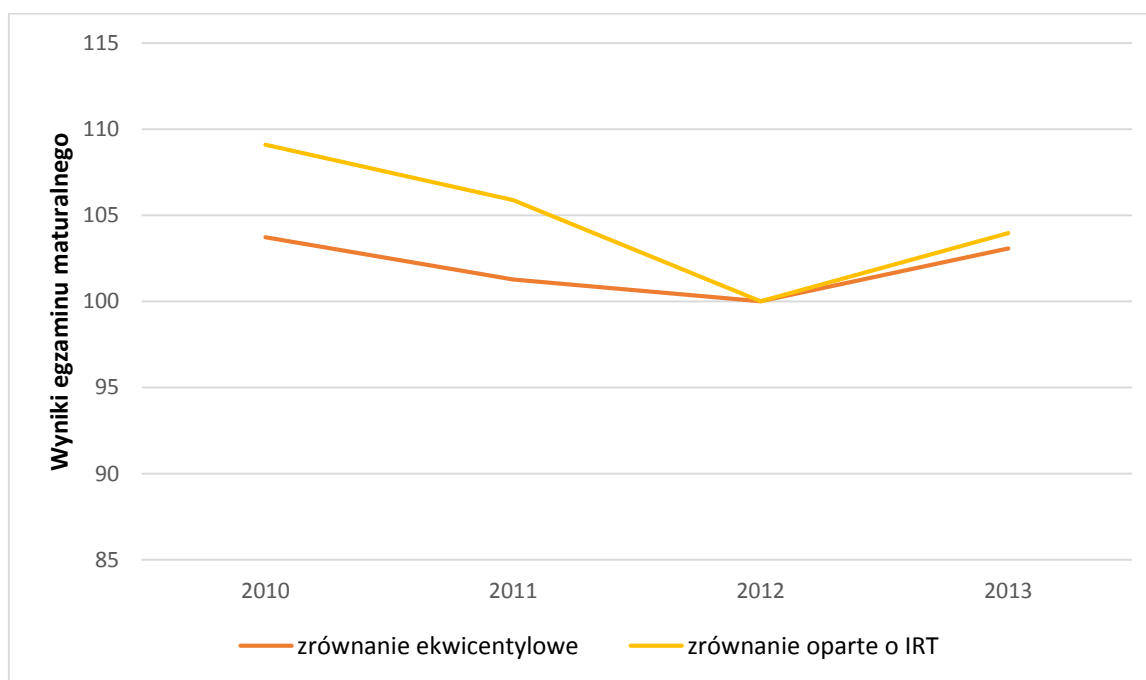
Rysunek 6.14. Wyniki zrównania ekwicyntylowego sprawdzianu w zestawieniu z wynikami zrównania opartego o IRT



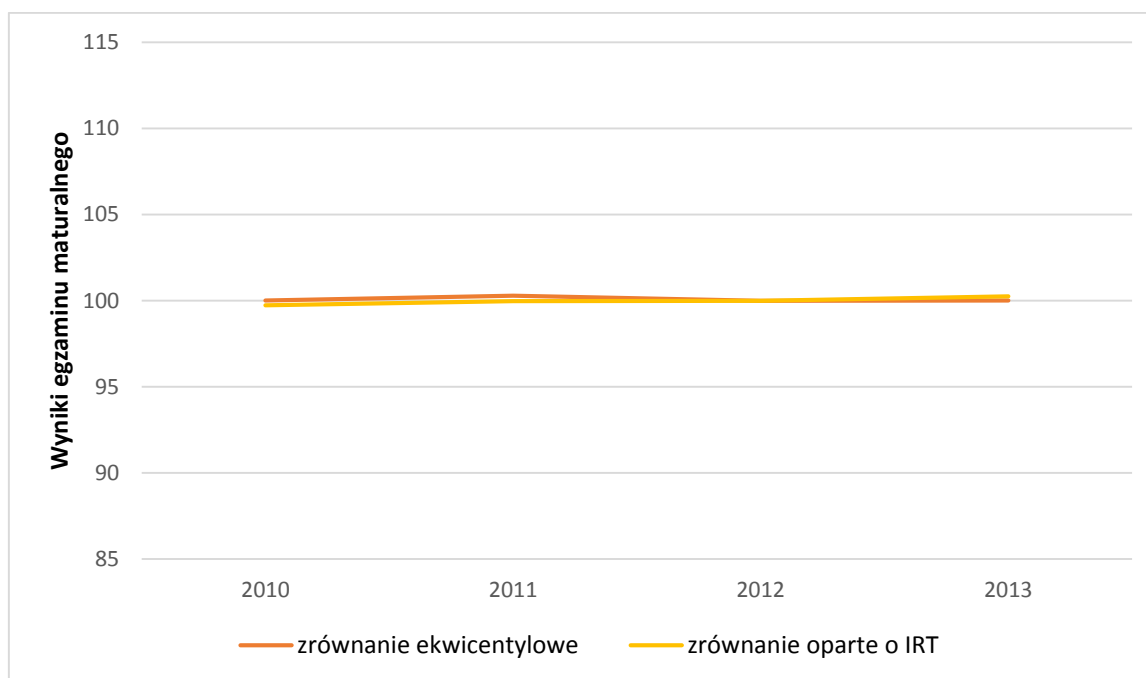
Rysunek 6.15. Wyniki zrównania ekwicyntylowego części humanistycznej egzaminu gimnazjalnego w zestawieniu z wynikami zrównania opartego o IRT



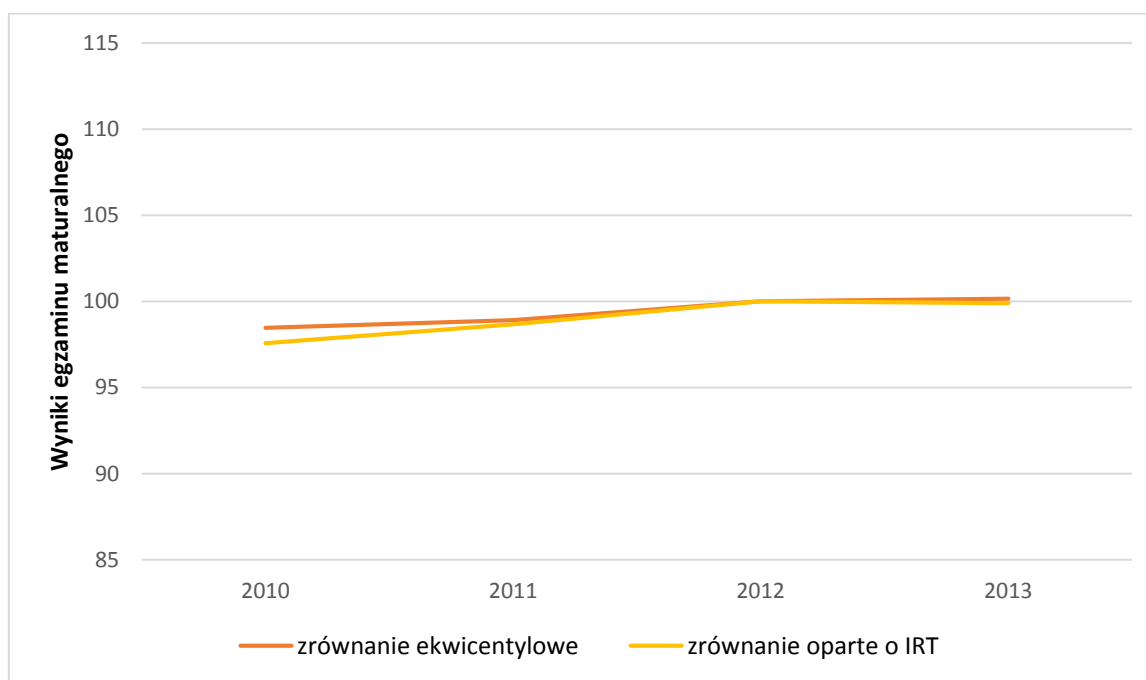
Rysunek 6.16. Wyniki zrównania części matematyczno-przyrodniczej egzaminu gimnazjalnego w zestawieniu z wynikami zrównania opartego o IRT



Rysunek 6.17. Wyniki zrównania egzaminu maturalnego z języka polskiego w zestawieniu z wynikami zrównania opartego o IRT



Rysunek 6.18. Wyniki zrównania egzaminu maturalnego z matematyki w zestawieniu z wynikami zrównania opartego o IRT



Rysunek 6.19. Wyniki zrównania egzaminu maturalnego z języka angielskiego w zestawieniu z wynikami zrównania opartego o IRT

Przedstawione tu oraz w artykule Wołodźko i współpracowników (2014) wyniki pokazały, że zrównania wyników sprawdzianu, egzaminu gimnazjalnego i egzaminu maturalnego przy użyciu różnych metod (ekwicytylowej i opartej o IRT), dały zbliżone rezultaty. Oznacza to, że walidacja oszacowań metodą opartą o IRT przebiegła pomyślnie. Z drugiej strony, badanie pokazało, że metoda ekwicytylowa może być skutecznie stosowana w praktyce, wymaga jednak zaplanowania procedury badań zrównujących

pod kątem takiego zrównania – uwzględniającego odpowiedni dobór zadań do testów zrównujących (możliwość kotwiczenia testów w planie SG lub NEAT) oraz prób biorących udział w badaniu (założenie o ekwiwalentności przy planie EG). W przypadku zrównania ekwicytylowego, zastosowanie dużej i dobrze dobranej próby sprawia, że zrównania z zastosowaniem wstępnego wygładzania i uciągania rozkładów przy użyciu estymatora jądrowego daje rezultaty porównywalne do zrównań przy użyciu prostszych metod. Istotnym rezultatem prowadzonych analiz jest zwrócenie uwagi na fakt jak dużą rolę w planowaniu badań zrównujących gra stworzenie takiego schematu badawczego, który pozwoli na zrównywanie poszczególnych testów parami, bądź przy użyciu możliwie krótkich łańcuchów funkcji zrównujących. Planowanie badań zrównujących powinno uwzględniać odpowiedni wybór planu zrównania, ponieważ prowadzi to do ograniczenia błędów zrównania i uproszczenia procedury analizy danych. Zrównanie w planie jednej grupy (SG) jest najprostsze, ale wymaga by ci sami uczniowie rozwiązywali zadania pochodzące z obu testów. Taka sytuacja jednak nie zawsze jest możliwa, czego przykładem mogą być rozszerzone wypowiedzi z egzaminu maturalnego z języka polskiego. Nie ma możliwości by ci sami uczniowie tworzyli teksty na dwa różne tematy, ponieważ tak przeprowadzone badanie w sposób znaczący odbiegałoby od prawdziwej sytuacji egzaminacyjnej. Alternatywnym rozwiązaniem jest zrównanie w planie NEAT, gdzie dwie grupy uczniów rozwiązują pewną liczbę wspólnych zadań lub zrównanie w planie grup ekwiwalentnych (EG). Oba te rozwiązania prowadzą jednak do zwiększenia błędów zrównania. Skomplikowany plan zrównania, jak w przypadku sprawdzianu i egzaminu gimnazjalnego, może wymagać tworzenia łańcuchów funkcji zrównujących, lecz w praktyce wydaje się to jednak nie być rozwiązaniem optymalnym. W tym miejscu należy zaznaczyć, że skomplikowany schemat badawczy użyty w tym badaniu miał na celu zrównanie wsteczne kilku testów z różnych lat. Tak skomplikowany schemat badawczy nie byłby konieczny gdyby nie potrzeba zrównania wielu testów równocześnie. Przykładem prostszego schematu badawczego były zrównania egzaminu maturalnego, co równocześnie skutkowało mniejszymi błędami zrównania.

6.8. Skalowanie wertykalne

W niniejszym podrozdziale przedstawiono specyfikę danych zebranych podczas studium zrównującego przez Zespół Analiz Osiągnięć Uczniów, wynikające z niej uzasadnienie wyboru do skalowania modelu podwójnego czynnika (*bi-factor model*), opis niniejszego modelu oraz wyniki skalowania. O specyfice i tej metody w niniejszym badaniu można przeczytać w podrozdziale 4.10.

Budowę skali pionowej zaczniemy od przedstawienia schematu kotwiczenia zadań mierzących umiejętności z zakresu języka polskiego oraz matematyki. Tabela 6.12 (język polski) i Tabela 6.13 (matematyka) przedstawiają liczbę zadań rozwiązywanych w poszczególnych falach badania zrównującego na etapie sprawdzianu, egzaminu gimnazjalnego oraz maturalnego oraz liczbę zadań kotwiczących pomiędzy sąsiadującymi między sobą etapami.

Tabela 6.12. Schemat kotwiczenia zadań dla języka polskiego w latach 2012–2014

Test \ Rok badania	Liczba zadań z danego etapu edukacyjnego		
	Zrównywanie 2012	Zrównywanie 2013	Zrównywanie 2014
Sprawdzian (SP)	66	15	12
Kotwica dla etapów SP i EG (K1)	6	4	3
Egzamin gimnazjalny (EG)	32	25	22
Kotwica dla etapów EG i M (K2)	nd	nd	5
Matura (M)	nd	nd	42

nd – w danym zrównaniu nie badano danego etapu edukacyjnego

Schemat kotwiczenia zadań z języka polskiego (zob. Tabela 6.12) pozwala na zbudowanie skali wertykalnej dla dwóch etapów edukacyjnych w przypadku badań zrównujących z roku 2012 i 2013 oraz trzech etapów edukacyjnych dla studium przeprowadzonego w roku 2014. Można zauważyć, że skalowanie wertykalne umiejętności z zakresu języka polskiego opiera się na niewielkiej liczbie zadań kotwiczących (oznaczanych w tabelach jako K1 i K2). Pomiedzy etapami sprawdzianu i egzaminu gimnazjalnego (K1) w roku 2012 można było wyróżnić sześć zadań wspólnych, natomiast w roku 2013 cztery takie zadania. W roku 2014 w schemacie skalowania zastosowano trzy zadania wspólne dla etapu sprawdzianu i egzaminu gimnazjalnego (K1) oraz cztery takie zadania dla etapu egzaminu gimnazjalnego i matury (K2).

Tabela 6.13. Schemat kotwiczenia zadań dla matematyki w latach 2012–2014

Test \ Rok badania	Liczba zadań z danego etapu edukacyjnego		
	Zrównywanie 2012	Zrównywanie 2013	Zrównywanie 2014
Sprawdzian (SP)	62	11	15
Kotwica dla etapów SP i EG (K1)	5	3	4
Egzamin gimnazjalny (EG)	17	18	25
Kotwica dla etapów EG i M (K2)	nd	15	5
Matura (M)	nd	61	33

nd – w danym zrównaniu nie badano danego etapu edukacyjnego

Zastosowana w badaniach zrównujących liczba zadań kotwiczących jest niewielka zarówno podczas pomiaru umiejętności z języka polskiego jak i matematyki. W literaturze (Kolen i Brennan, 2004) wskazuje się, że powinna ona wynosić co najmniej 20% całego testu, ze względu na stabilność oszacowań. Oznacza to, że procedura skalowania wertykalnego w przypadku danych zebranych przez ZAOU może być problematyczna (trzeba pamiętać jednak, że przeprowadzone badania były podporządkowane przede wszystkim zrównywaniu horyzontalnemu).

Biorąc pod uwagę specyficzny schemat kotwiczenia zastosowany w badaniach zrównujących oraz potencjalne obciążenie związane z tradycyjnymi metodami skalowania wertykalnego rozwiązaniem

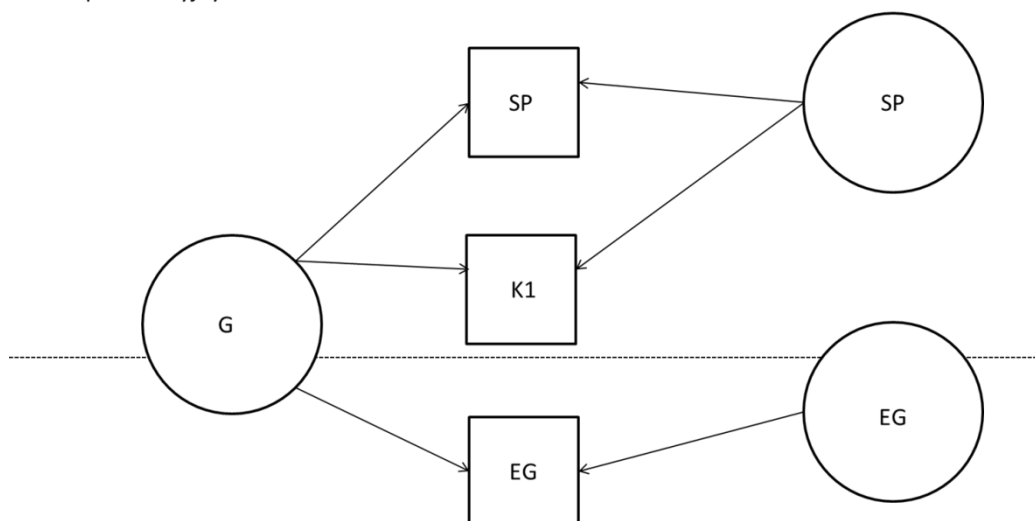
pozwalającym na bardziej rzetelne oszacowanie skal pionowych może być wykorzystanie modelu podwójnego czynnika (*bi-factor model*) (Li, 2011; Li i Lissitz, 2012).

Model podwójnego czynnika to model, w którym zakładamy, że struktura kowariancji pomiędzy odpowiedziami udzielonymi przez badanego może być wyjaśniona przez jeden czynnik ogólny, który odzwierciedla wariację wspólną dla wszystkich pytań/zadań oraz czynniki specyficzne (grupowe), które odzwierciedlają dodatkową wspólną wariację pomiędzy zbiorami/skupieniami zadań, zazwyczaj tych o podobnej treści (Gibbons i Hedeker, 1992; Reise, 2012; Świst, Skórska i Pokropek, w druku). Czynnik główny ma za zadanie adekwatnie odzwierciedlać związki pomiędzy pozycjami/zadaniami oraz mierzonym konstruktem (np. jakością życia, a w kontekście edukacyjnym np. umiejętnościami matematycznymi), poprzez kontrolę wielowymiarowości związanej z czynnikami specyficznymi (Reise, 2012). Model ten musi spełnić dwa założenia. Po pierwsze każde zadanie musi posiadać niezerowy ładunek na czynniku wspólnym i tylko jeden niezerowy ładunek na czynniku specyficznym. Po drugie, czynniki specyficzne muszą być ortogonalne wobec siebie oraz wobec czynnika głównego (Li, 2011).

Stosując model podwójnego czynnika w kontekście skalowania wertykalnego otrzymujemy wspólną skalę (czynnik główny) dla poszczególnych etapów edukacyjnych, biorąc także pod uwagę czynniki specyficzne odzwierciedlające konkretne, unikalne umiejętności nabywane na poszczególnych etapach edukacyjnych (czynniki specyficzne). Wszystkie zadania z kolejnych etapów edukacyjnych²⁶ ładują czynnik wspólny (odpowiadający ogólnie rozumianym umiejętnościom). Zadania specyficzne dla poszczególnych klas ładują czynniki specyficzne, odwołujące się do specyficznych umiejętności nauczanych na poszczególnych etapach edukacyjnych. Schematy modeli podwójnego czynnika dla dwóch i trzech etapów edukacyjnych przedstawia Rysunek 6.20 oraz Rysunek 6.21.

²⁶ Zadania zostały rozwiązywane przez uczniów tak jak wynika to ze schematu doboru próby oraz planów zrównania.

I etap edukacyjny

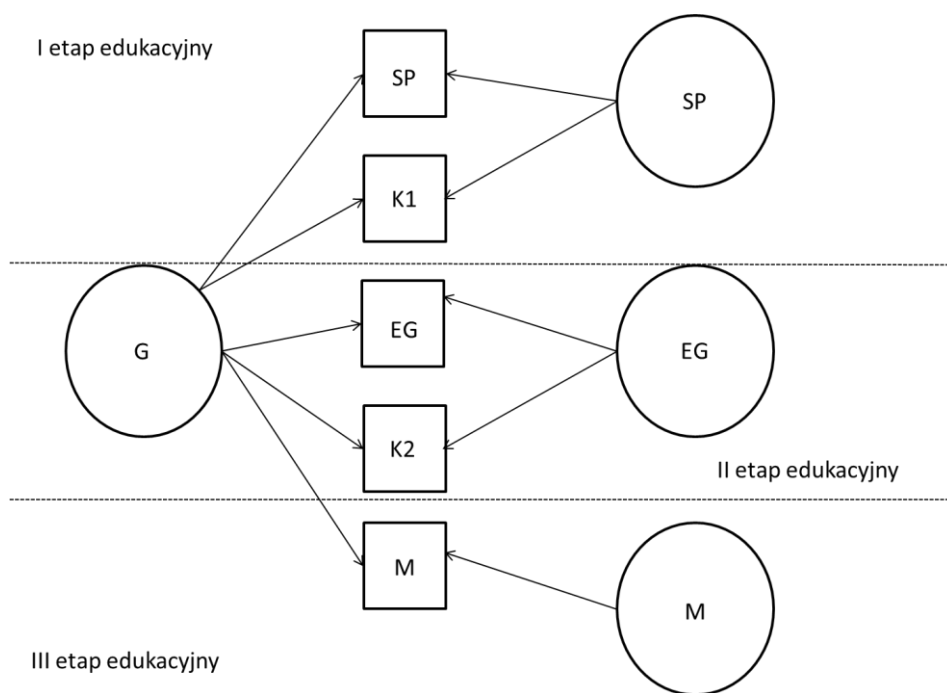


II etap edukacyjny

Rysunek 6.20. Model podwójnego czynnika dla dwóch etapów edukacyjnych²⁷

Model podwójnego czynnika dla dwóch etapów edukacyjnych (zob. Rysunek 6.20) zakłada, że wszystkie zadania mierzące umiejętności z zakresu sprawdzianu (SP), egzaminu gimnazjalnego (EG) oraz zadania kotwiczące (K1) pomiędzy tymi dwoma etapami ładują czynnik główny (oznaczony jako G). Zadania odpowiadające poziomowi szkoły podstawowej (SP) oraz zadania kotwiczące (K1) ładują także czynnik specyficzny SP, natomiast zadania unikalne dla etapu egzaminu gimnazjalnego ładują czynnik specyficzny EG.

²⁷ SP – szkoła podstawowa, K1 – zadania kotwiczące pomiędzy szkołą podstawową, a egzaminem gimnazjalnym, EG – egzamin gimnazjalny, G – czynnik wspólny (poziom umiejętności).



Rysunek 6.21. Model podwójnego czynnika dla trzech etapów edukacyjnych²⁸

Jeśli skonstruujemy skalę pionową dla trzech etapów edukacyjnych w modelu podwójnego czynnika (zob. Rysunek 6.21), oznaczać to będzie, że wszystkie typy zadań: mierzące umiejętności z zakresu sprawdzianu (SP), egzaminu gimnazjalnego (EG), matury (M), kotwiczące pomiędzy szkołą podstawową a gimnazjum (K1) oraz gimnazjum a maturą (K2) będą ładować czynnik główny (G). Zadania z poziomu sprawdzianu oraz kotwiczące K1 ładują czynnik specyficzny SP (umiejętności unikalne dla szkoły podstawowej), zadania z poziomu egzaminu gimnazjalnego oraz kotwiczące K2 ładują czynnik specyficzny EG (analogicznie – umiejętności unikalne dla gimnazjum) oraz zadania z poziomu matury – czynnik specyficzny M (czyli odpowiadające specyfice nauczania na ostatnim etapie edukacyjnym).

Wykazano, że modele podwójnego czynnika cechują się lepszą dobrocią dopasowania do danych z różnorodnych testów osiągnięć niż modele jednowymiarowe bądź też bardziej złożone modele wielowymiarowe (Gibbons, Bock, Hedeker, Weiss, Segawa i Bhaumik, 2007; Gibbons i Hedeker, 1992; Reise, Morizot i Hays, 2007). Ying Li (2011) wskazuje, że struktura modelu podwójnego czynnika w sposób naturalny oddaje ideę skalowania wertykalnego. Czynnik wspólny może modelować skalę pionową, natomiast czynniki specyficzne mogą modelować przesunięcie się konstruktów.

Do oceny uzyskanych skal pionowych wykorzystane zostały jej dwie własności:

- średni przyrost umiejętności z klasy do klasy/jednego etapu edukacyjnego do drugiego (*average grade-to-grade growth*), czyli różnica pomiędzy średnimi (lub medianami/wybranymi punktami percentylowymi) dla konkretnych etapów nauczania,

²⁸ SP – szkoła podstawowa, K1 – zadania kotwiczące pomiędzy szkołą podstawową, a egzaminem gimnazjalnym, EG – egzamin gimnazjalny, K2 – zadania kotwiczące pomiędzy egzaminem gimnazjalnym a maturalnym, M – egzamin maturalny, G – czynnik wspólny (poziom umiejętności).

- zróżnicowanie przyrostu pomiędzy klasami/etapami edukacyjnymi (*grade-to-grade variability*), czyli różnica pomiędzy odchyleniami standardowymi (lub innymi miarami zmienności) pomiędzy sąsiadującymi etapami nauczania.

Analizy wykonano w programie Mplus 7.1 (Muthén i Muthén, 1992-2010), z wykorzystaniem estymacji MLR. Poniżej przedstawiono wyniki w podziale na język polski oraz na matematykę. W przypadku rozszerzonej wypowiedzi pisemnej z języka polskiego (matura), część kryteriów jest oceniana na skali posiadającej więcej niż 10 punktów. Mplus nie ma możliwości potraktowania takich zmiennych jako kategoryalnych, więc zadania te zostały uznane za mierzone na skali ciągłej.

Podczas interpretacji wyników należy przede wszystkim pamiętać, że nie istnieje żadna interpretowalna jednostka przyrostu umiejętności (Beard, 2008) – dlatego porównanie przyrostu umiejętności matematycznych lub humanistycznych odbywa się wyłącznie pomiędzy kolejnymi falami badania.

6.8.1. Język polski

Przyrost umiejętności humanistycznych mierzono w trzech falach badania. W roku 2012 i 2013 mierzono przyrost pomiędzy poziomem szkoły podstawowej oraz gimnazjum, w roku 2014 dodano pomiar przyrostu umiejętności pomiędzy gimnazjum a szkołami ponadgimnazjalnymi. W roku 2012 i 2013 przyjęto poziom szkoły podstawowej za poziom odniesienia (średnia dla czynnika głównego została ustalona jako 0, wariancja jako 1). W roku 2014, celem łatwiejszych porównań, przyjęto poziom gimnazjum za poziom odniesienia i dla niego ustalono średnią czynnika głównego jako 0, a wariancję jako 1.

Tabela 6.14 oraz Tabela 6.15 przedstawiają wyniki skalowania wertykalnego dla umiejętności humanistycznych²⁹. Tendencja przyrostu oraz zróżnicowania umiejętności wygląda podobnie w roku 2012 i 2013 – umiejętności rosną wraz z przejściem ze szkoły podstawowej do gimnazjum, co jest zgodne z oczekiwaniami. Zróżnicowanie przyrostu umiejętności jest podobne dla roku 2012 i 2013 – większe na trzecim niż na drugim etapie edukacyjnym. Zaobserwowano większy przyrost umiejętności w roku 2012 niż 2013 – należy również zauważyć, że w roku 2013 wariancja czynnika specyficznego dla gimnazjum jest wyższa niż w roku 2012. Oznacza to, że istnieje część wariancji odpowiadająca za specyficzne aspekty umiejętności z języka polskiego, mierzone przez egzamin gimnazjalny. W roku 2012 uzyskano spójną pionową skalę umiejętności – wariancja czynników specyficznych jest niewielka. Rozstrzygnięcie o przyczynie pojawienia się wariancji czynnika specyficznego dla etapu EG w roku 2013 wymagałoby jakościowej analizy treści zeszytów testowych z danego roku oraz porównania jej wyników z analizą treści zeszytów wykorzystanych w roku 2012 (gdzie uzyskano spójną skalę pionową).

W roku 2014, zaobserwowano podobne natężenie przyrostu umiejętności pomiędzy gimnazjum a szkołą podstawową, oraz szkołą ponadgimnazjalną a gimnazjum. Należy jednak zauważyć, że zróżnicowanie przyrostu umiejętności pomiędzy gimnazjum a szkołą ponadgimnazjalną drastycznie spada. Świadczy to o tym, że uczniowie stają się na etapie szkół ponadgimnazjalnych coraz bardziej podobni w nabywaniu umiejętności. Dla roku 2014 wariancja czynników specyficznych jest niewielka, tak więc pionowa skala umiejętności jest spójna.

²⁹ Uzyskano następujące miary dopasowania: a) rok 2012: AIC = 250272,409 i BIC = 253455,907 b) rok 2013: AIC = 85456,916 i BIC = 86529,095 c) rok 2014: AIC = 320417,959 i BIC = 328379,872.

Tabela 6.14. Oszacowania średnich oraz wariancji dla przyrostu umiejętności uczniów z języka polskiego dla modelu podwójnego czynnika w latach 2012–2014

Rok	Wariancje czynników specyficznych			Średnie dla czynnika głównego			Wariancja dla czynnika głównego		
	SP	EG	M	I etap ed.	II etap ed.	III etap ed.	I etap ed.	II etap ed.	III etap ed.
2012	0,10	0,05	nd	0,00	3,98	nd	1,00	1,70	nd
2013	0,02	0,34	nd	0,00	1,09	nd	1,00	1,53	nd
2014	0,02	0,02	0,07	-0,15	0,00	0,13	0,07	1,00	0,07

nd – w danej fali badań zrównujących nie badano tego etapu

Tabela 6.15. Oszacowania średniego przyrostu umiejętności, jego zróżnicowania dla języka polskiego w latach 2012–2014

Rok	Średni przyrost umiejętności (wartość bezwzględna)		Zróżnicowanie umiejętności (wartość bezwzględna)	
	SP-GIM	GIM-MAT	SP-GIM	GIM-MAT
2012	3,98	nd	0,70	nd
2013	1,09	nd	0,53	nd
2014	0,15	0,13	0,93	0,93

nd – w danej fali badań zrównujących nie badano tego etapu

6.8.2. Matematyka

Przyrost umiejętności matematycznych mierzono również w trzech falach badania. W roku 2012 mierzono przyrost pomiędzy poziomem szkoły podstawowej oraz gimnazjum, w roku 2013 i 2014 również pomiędzy gimnazjum a szkołami ponadgimnazjalnymi. Podobnie, jak dla języka polskiego, w roku 2012 przyjęto poziom szkoły podstawowej za poziom odniesienia, a w roku 2013 i 2014 poziom gimnazjum.

Tabela 6.16 i Tabela 6.17 obrazują wyniki skalowania wertykalnego dla umiejętności matematycznych³⁰. Dla roku 2012 istnieje przyrost umiejętności matematycznych pomiędzy pierwszym i drugim etapem edukacyjnym oraz wzrasta zróżnicowanie umiejętności dla drugiego etapu edukacyjnego. Wariancja czynnika specyficznego dla egzaminu gimnazjalnego jest dosyć znaczna, co oznacza nabywanie specyficznych umiejętności na danym etapie.

W roku 2013 wariancje czynników specyficznych są znaczne dla każdego z analizowanych etapów edukacyjnych (szczególnie w przypadku szkoły podstawowej). Zaobserwowano większy wzrost umiejętności matematycznych pomiędzy szkołą podstawową a gimnazjum, natomiast stosunkowo mniejszy przyrost pomiędzy gimnazjum a liceum.

³⁰ Uzyskano następujące miary dopasowania: a) rok 2012: AIC = 234408,808 i BIC = 237256,844 b) rok 2013: AIC = 240554,79 i BIC = 245346,203 c) rok 2014: AIC = 167466,611 i BIC = 170332,675.

W roku 2014 uzyskano spójną pionową skalę umiejętności – wariancja czynników specyficznych jest niewielka. Można zaobserwować przyrost umiejętności pomiędzy szkołą podstawową a gimnazjum oraz mniejszy przyrost umiejętności pomiędzy gimnazjum a szkołą ponadgimnazjalną. Zróżnicowanie poziomu umiejętności rośnie pomiędzy drugim i trzecim etapem edukacyjnym, a następnie maleje (jak we wszystkich analizowanych zrównaniach) między trzecim a czwartym etapem edukacyjnym.

Tabela 6.16. Oszacowania średnich oraz wariancji dla przyrostu umiejętności uczniów z matematyki w przypadku modelu podwójnego czynnika w latach 2012–2014

Rok	Wariancje czynników specyficznych			Średnie dla czynnika głównego			Wariancja dla czynnika głównego		
	SP	EG	M	I etap ed.	II etap ed.	III etap ed.	I etap ed.	II etap ed.	III etap ed.
2012	0,13	0,53	nd	0,00	0,16	nd	1,00	1,73	nd
2013	1,41	0,27	0,65	-3,51	0,00	1,36	0,53	1,00	0,65
2014	0,00	0,00	0,08	-0,91	0,00	0,34	0,86	1,00	0,75

nd – w danej fali badań zrównujących nie badano tego etapu

Tabela 6.17. Oszacowania średniego przyrostu umiejętności oraz jego zróżnicowania dla matematyki w latach 2012–2014

Rok	średni przyrost umiejętności (wartość bezwzględna)		zróżnicowanie umiejętności (wartość bezwzględna)	
	SP-GIM	GIM-MAT	SP-GIM	GIM-MAT
2012	0,16	nd	0,73	nd
2013	3,51	1,36	0,47	0,33
2014	0,91	0,34	0,14	0,26

nd – w danej fali badań zrównujących nie badano tego etapu

Można zauważyć, że zróżnicowanie przyrostu umiejętności uczniów między gimnazjum a maturą zmniejsza się we wszystkich analizowanych falach badania. Jest to zgodne z tendencjami obserwowanymi w literaturze – zjawiskiem znanym jako kurczenie się skali (*scale shrinkage*) (Yen, 1985; Camilli, Yamamoto i Wang, 1993). Prawdopodobnie do szkół ponadgimnazjalnych dostaje się skutek selekcji zbliżona do siebie poziomem grupa uczniów, natomiast w przypadku gimnazjów i szkół podstawowych, w Polsce obowiązkowych, zróżnicowanie poziomu umiejętności uczniów jest zdecydowanie większe. Podobne wnioski płyną z badań nad zróżnicowaniem międzyszkolnym uczniów, czyli tym w jakim stopniu można przewidywać wynik egzaminacyjny na podstawie tego, jaką szkołę skończył dany uczeń (Dolata, 2012). Wskazuje się, że selekcyjność szkół wzrasta wraz z kolejnymi etapami edukacyjnymi – do gimnazjów uczęszczają uczniowie bardziej podobni do siebie pod względem poziomu umiejętności niż do szkół podstawowych.

Przy interpretacji modeli podwójnego czynnika oraz przedstawionych wyników skalowania wertykalnego należy pamiętać o niewielkiej liczbie zadań kotwiczących, która determinuje jakość uzyskanej skali wertykalnej. Nie należy też zapominać, że brakuje kryteriów pozwalających na ocenę czy przyrost umiejętności uczniów mierzony w danej fali badania jest „duży” czy „mały” – można je porównywać

wyłącznie względem siebie. Można zaobserwować także, że w zależności od fali badania, udało się zbudować mniej lub bardziej spójne skale wertykalne. Należy zauważyć, że wariancja czynników specyficznych potrafi być znaczna. Oznacza to, że na poszczególnych etapach edukacyjnych nabywane są specyficzne aspekty umiejętności matematycznych i w zakresie języka polskiego. Z powyższych powodów model podwójnego czynnika wydaje się być koncepcyjnie dobrym wyborem do analizy tego typu danych. Rozwiązanie to wymaga jednak przeprowadzenia badań symulacyjnych, pozwalających na rozstrzygnięcie o stabilności parametrów uzyskiwanych przy pomocy tej metody (skoro właściwości skal uzyskanych przy pomocy modelu podwójnego czynnika różnią się zarówno pomiędzy przedmiotami jak i falami zrównania). Analizy symulacyjne powinny być szczególnie ważne w przypadku skali pionowej dla języka polskiego (rok 2014), dla której to wyniki wydają się być szczególnie niestabilne. W związku z tym, że w zeszytach zrównujących wykorzystano szereg jakościowo różnych tematów wypracowań z lat poprzednich, mogą one odpowiadać szeregowi jakościowo różnych umiejętności.

6.9. Badania ankietowe

6.9.1. Motywacja testowa uczniów

W niniejszym rozdziale przedstawiono wyniki analizy wpływu motywacji testowej na wyniki uzyskiwane w testach umiejętności wykorzystanych w badaniu zrównującym z wykorzystaniem Skali Motywacji Testowej (lata 2011-2013) oraz Skali Opinii Uczniów (rok 2014). Motywacja do rozwiązywania zadań testowych (*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 może być w sposób oczywisty uwarunkowana konsekwencjami, jakie dla ucznia niesie rozwiązanie danego testu, tj. jego doniosłością. Testy wykorzystywane w badaniach zrównujących można określić mianem testów o niskiej doniosłości (*low-stakes tests*) – ich 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)³¹. Szacowanie poziomu umiejętności uczniów w takich testach, bez wzięcia pod uwagę motywacji testowej uczniów, może prowadzić do uzyskania wyników o obniżonej trafności, gdyż motywacja do rozwiązywania zadań testowych może być źródłem wariancji w wynikach niezwiązanej z mierzonym konstruktem (*construct-irrelevant variance*, CIV) (Messick, 1989; Messick 1995). W takiej sytuacji poziom umiejętności uczniów mógłby zostać niedoszacowany – niski wynik w teście niekoniecznie musiałby odzwierciedlać niski poziom umiejętności, a jedynie niską motywację ucznia (Swerdzewski, Finney i Harmes, 2007). Dodatkowo motywacja testowa może wpływać na oszacowanie własności statystycznych zadań składających się na test, na przykład na oszacowania ich parametru trudności (Barry i Finney, 2009; DeMars, 2000).

Ze względu na poważne konsekwencje nieuwzględnienia motywacji testowej podczas oszacowania umiejętności uczniów na podstawie wyników testów niskiej doniosłości, w kolejnych edycjach badania zrównującego kontrolowano poziom motywacji testowej ucznia. W pierwszych trzech falach badania zrównującego korzystano z autorskiej Skali Motywacji Testowej (zob. Aneks 4.1). Własności tej skali zostały zbadane m.in. przez Dorotę Węziak-Białowolską (2011). Dziesięć pytań tej skali mierzy

³¹ Do przykładów testów niskiej doniosłości można zaliczyć znane międzynarodowe badania, takie jak PISA (Programme for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study), czy PIRLS (Progress in International Reading Literacy Study).

motywację testową, pozostałe cztery – znajomość pytań z arkuszy zrównujących³². W analizach związku motywacji testowej z wynikami uczniów z badania zrównującego wykorzystano tylko pytania dotyczące motywacji, która została potraktowana w tej analizie jako zjawisko jednowymiarowe. W ostatniej edycji badań zrównujących (rok 2014), skala ta została zastąpiona polskim tłumaczeniem Skali Opinii Uczniów (*Student Opinion Scale*, SOS), stworzonej przez Donnę Sundre (2007) (zob. Aneks 4.2). Tłumaczenie kwestionariusza z języka angielskiego na język polski odbyło się za zgodą autorki oraz zostało przeprowadzone przez pracowników Zespołu Analiz Osiągnięć Uczniów, we współpracy z anglistą. Analiza czynnikowa wskazuje na dwa wymiary tego narzędzia – ważność i wysiłek (Sundre i Finney, 2002). Pozycje testowe w Skali Opinii Uczniów są punktowane na skali przyjmującej wartości 1–5. Najwyższy możliwy do uzyskania wynik na danej podskali wynosi 25. Wynik 25 w skali wysiłku oznacza, że uczniowie włożyli cały swój wysiłek w rozwiązywanie testu, wynik 25 w skali ważności – że konsekwencje danego testu były niezmiernie ważne dla ucznia.

Wpływ motywacji testowej na standaryzowane wyniki uzyskane w teście umiejętności oszacowano przy pomocy następującego modelu regresji liniowej:

$$Y_i = \beta_0 + \beta_1 X_i + e_i, \quad (6.1)$$

gdzie:

Y_i – standaryzowany wynik w teście uzyskany przez ucznia i ,

X_i – suma punktów uzyskana w kwestionariuszu motywacji (przy czym w latach 2011–2013 była to suma punktów dla wszystkich 10 pytań mierzących poziom motywacji testowej, a w 2014 roku osobne sumy punktów zliczane dla dwóch wymiarów motywacji – wysiłku i ważności),

β_0 – stała regresji,

β_1 – współczynnik regresji odzwierciedlający siłę związku motywacji ze standaryzowanym wynikiem testu,

e_i – błąd resztowy, odzwierciedlający różnicę między zaobserwowanym wynikiem testu a przewidywanym na podstawie modelu regresji.

Wpływ motywacji testowej na standaryzowane wyniki testów umiejętności wykorzystywanych w badaniach zrównujących 2011–2013 w większości testów okazał się istotny statystycznie, co może być jednak wynikiem dużych prób w badaniach. Wielkość efektu nie ma jednak znaczenia praktycznego dla studium zrównującego. W najbardziej skrajnym przypadku zmiana wskaźnika motywacji testowej o jeden punkt wiąże się ze zmianą wyniku testu umiejętności o 0,05 odchylenia standardowego (dla egzaminów gimnazjalnych z 2011 roku). Potwierdza to współczynnik R^2 , wskazujący, że motywacja testowa odpowiada za najwyżej kilka procent wariancji wyników testów w poszczególnych latach. Dokładne wartości prezentuje Tabela 6.18.

³² Zeszyty testowe w badaniu zrównującym składały się m.in. z tzw. zadań „starych”, występujących w arkuszach egzaminacyjnych z poprzednich edycji danego egzaminu.

Tabela 6.18. Współczynniki regresji i procent wyjaśnionej wariancji dla motywacji testowej w badaniach zrównujących 2011-2013

typ egzaminu		β (błąd standardowy ³³)	R ²
sprawdzian	2012	0,03* (0,00)	0,04
	2013	-0,00 (0,02)	0,00
egzamin gimnazjalny	język polski 2012	0,03* (0,00)	0,02
	język polski 2013	0,03* (0,00)	0,02
	historia 2012	0,01* (0,01)	0,00
	historia 2013	0,02* (0,00)	0,02
	humanistyczny 2011	0,05* (0,00)	0,06
	matematyczno–przyrodniczy 2011	0,05* (0,00)	0,09
	matematyka 2012	-0,01* (0,01)	0,00
	matematyka 2013	0,02* (0,01)	0,01
	przyroda 2012	0,00 (0,00)	0,00
	przyroda 2013	0,02* (0,01)	0,01
matura	matematyka 2013	0,04* (0,00)	0,05

* współczynniki istotne statystycznie

Podobne wnioski płyną z analizy wpływu motywacji testowej na wyniki testów w badaniu zrównującym 2014, gdzie dwa wymiary motywacji do rozwiązywania zadań testowych są istotnymi statystycznie predyktorami standaryzowanych wyników testów osiągnięć. Efekt jest niewielki, najsilniejszy w przypadku sprawdzianu, gdzie zmiana ważności testu o jedną jednostkę wiąże się ze zmianą wyniku testu o 0,05 odchylenia standardowego, a wzrost wskaźnika wysiłku ucznia o jeden punkt, wiąże się ze zmianą wyniku testu o 0,09 odchylenia standardowego. Warto zauważyć, że rozróżniając dwa wymiary motywacji testowej, wysiłek ucznia silniej wiąże się z wynikiem testu umiejętności niż ważność testu. Relacje te prezentuje Tabela 6.19.

³³ W wykorzystanym modelu regresji, szacowano odporne błędy standardowe (robust standard errors).

Tabela 6.19. Współczynniki regresji i procent wyjaśnionej wariancji dla motywacji testowej w badaniach zrównujących 2014

typ egzaminu	wymiar motywacji	β (błąd standardowy)	R^2
sprawdzian	ważność	0,05* (0,01)	0,03
	wysiłek	0,09* (0,01)	0,10
egzamin gimnazjalny	historia	ważność	0,01* (0,00)
		wysiłek	0,02* (0,00)
	język polski	ważność	0,02* (0,00)
		wysiłek	0,03* (0,00)
	matematyka	ważność	0,02* (0,00)
		wysiłek	0,03* (0,00)
	przyroda	ważność	0,01* (0,00)
		wysiłek	0,01* (0,00)
matura	język polski	ważność	0,03* (0,01)
		wysiłek	0,04* (0,00)
	matematyka	ważność	0,03* (0,01)
		wysiłek	0,07* (0,01)
	język angielski	ważność	0,03* (0,00)
		wysiłek	0,04* (0,00)

* współczynniki istotne statystycznie

Zakłócenie wyników zrównywania może być spowodowane także tym, że motywacja testowa uczniów jest zróżnicowana względem poszczególnych zeszytów testowych. Należy pamiętać, że we wszystkich zrównaniach każdy zeszyt testowy posiadał wersję A i B różniące się kolejnością odpowiedzi w pytaniach zamkniętych. Tabela 6.20 wskazuje, czy poziom motywacji dla zeszytów testowych w danym zrównaniu z lat 2011–2013 jest istotny statystycznie. Analizę przeprowadzono przy pomocy techniki ANOVA z poprawką Bonferroniego, ze względu na znaczną liczbę porównań parami. Należy pamiętać, że w latach 2011–2013 motywację mierzono przy pomocy kwestionariusza zakładającego jednowymiarowość tego konstrukt, natomiast w roku 2014, mierzono dwa konstrukty – ważność i wysiłek.

Tabela 6.20. Różnice w motywacji między zeszytami testowymi

typ egzaminu		Liczba wykorzystanych zeszytów (wliczając wersje A i B)	Czy średni poziom motywacji różni się pomiędzy zeszytami?	Pomiędzy którymi zeszytami różni się średni poziom motywacji?	Wartość bezwzględna różnicy
sprawdzian	2012	24	TAK	A5 i B10	1,97
				A10 i B2	1,62
				B10 i A8	1,56
				B10 i B2	2,00
				B10 i B3	1,80
				B10 i B4	1,71
				B10 i B5	1,83
				B10 i B7	1,62
egzamin gimnazjalny	2013	4	NIE	–	–
	język polski 2012	2	NIE	–	–
	język polski 2013	2	NIE	–	–
	historia 2012	2	NIE	–	–
	historia 2013	2	NIE	–	–
	humanistyczny 2011	22	NIE	–	–
	matematyczno–przyrodniczy 2011	22	TAK	11B i 4B	1,48
	matematyka 2012	2	NIE	–	–
	matematyka 2013	2	NIE	–	–
	przyroda 2012	2	NIE	–	–
	przyroda 2013	2	NIE	–	–
matura	matematyka 2013	14	NIE	–	–

W większości przypadków poziom motywacji uczniów nie różni się pomiędzy zeszytami testowymi. Jedyne różnice w poziomie motywacji można wyróżnić pomiędzy zeszytami 11B i 4B dla zrównania testu matematyczno-przyrodniczego z 2011 roku oraz ośmiu par zeszytów z roku 2012. Ciekawym wydaje się, że uczniowie rozwiązujący zeszyt B10 wydali się szczególnie zmotywowani. Jednakże brakuje podstaw by stwierdzić, że różnice te przekładają się na jakość zrównania. Po pierwsze, różnica wynosi maksymalnie dwa punkty (w kwestionariuszu można było uzyskać maksymalnie 50 punktów). Po drugie, jak wskazano wcześniej, wpływ motywacji na wynik uzyskany w teście jest minimalny. Po trzecie, trafność wykorzystanego narzędzia badawczego jest nieznana. Po czwarte, im więcej przeprowadzonych parami porównań, tym większe prawdopodobieństwo, że różnice pomiędzy średnimi okażą się artefaktem badawczym. Choć niniejsze analizy przeprowadzono z korektą Bonferroniego, przyjęty próg dla porównań parami może nie być wystarczający.

Tabela 6.21. Różnice w motywacji między zeszytami testowymi (wymiar ważności i wysiłku)

typ egzaminu		Liczba wykorzystanych zeszytów (wersje A i B)	Czy poziom motywacji różni się pomiędzy zeszytami?		Pomiędzy którymi zeszytami różni się poziom motywacji?	
			Wymiar ważności	Wymiar wysiłku	Wymiar ważności	Wymiar wysiłku
sprawdzian		2	NIE	NIE	–	–
egzamin gimnazjalny	historia	2	NIE	NIE	–	–
	język polski	2	NIE	TAK	–	1A i 1B (0,42)
	matematyka	2	NIE	NIE	–	–
	przyroda	2	NIE	NIE	–	–
matura	język polski	10	TAK	TAK	Z1 i Z8 (0,77)	Z1 i Z4 (0,91) Z1 i Z8 (0,94)
					Z1 i Z4 (0,72)	Z1 i Z9 (1,31)
					Z1 i Z9 (1,01)	Z4 i Z10 (0,87) Z5 i Z9 (0,81)
					A1 i B3 (0,92)	A1 i B3 (1,24)
	matematyka	8	TAK	TAK	B1 i B3 (1,28)	B2 i B3 (1,22)
					B2 i B3 (1,14)	B3 i B4 (1,07)
						1B i 5A (1,58) 1B i 5B (1,31) 11A i 1A (1,40)
	język angielski	20	NIE	TAK	–	11A i 1B (1,78) 11A i 2B (1,46) 11B i 1B (1,41) 11A i 7B (1,34)

Dla etapu gimnazjum i szkoły podstawowej można zauważyć, że w większości przypadków poziom motywacji uczniów, na żadnym z mierzonych przez Skalę Opinii Uczniów wymiarów, nie różni się pomiędzy zeszytami testowymi (poza językiem polskim, lecz bezwzględna różnica średnich wynosi zaledwie 0,42 punktu) (zob. Tabela 6.21). Zróżnicowanie można zauważyć na etapie maturalnym. Dla języka angielskiego różnice pojawiają się wyłącznie na wymiarze raportowanego wysiłku. Co ciekawe, poziom wysiłku jest wyższy zwłaszcza dla obu wersji zeszytu 11. Różnice wynoszą jednak niecałe dwa punkty. Dla matury z matematyki różnice w raportowanym poziomie ważności i wysiłku pojawiają się dla

trzech par zeszytów. Dla zeszytu B3 poziom wysiłku i ważności jest niższy niż dla pozostałych zeszytów, różnice nie przekraczają jednak 1,5 punktu. Dla matury z języka polskiego więcej różnic pojawia się na wymiarze raportowanego wysiłku niż ważności. Pomiędzy zeszytami 1 i 8 oraz 1 i 9 można zauważyć niższe wyniki dla zeszytu 1 w obydwu analizowanych wymiarach (jednak wynoszą one około punktu). Ze względu na brak praktycznego znaczenia tych różnic oraz wykazany wcześniej niewielki wpływ poziomu raportowanego wysiłku i ważności na wynik uzyskiwany w teście można przypuszczać, że zaobserwowane różnice w niewielki sposób wpływają na poprawność zrównania. Należy mieć również na uwadze, że nawet jeśli pojawiają się istotne różnice, mogą być one artefaktem związanym ze znaczną liczbą porównań parami między zeszytami.

6.9.2. Badania ankietowe nauczycieli

Celem badań ankietowych przeprowadzonych wśród nauczycieli było m.in. poznanie ich opinii na temat poziomu motywacji uczniów do udziału w badaniu zrównującym, stopnia ich przygotowania do danego egzaminu zewnętrznego oraz wpływu udziału w badaniu zrównującym na wyniki uzyskiwane we właściwym egzaminie zewnętrznym. Nauczycieli pytano także, czy wykorzystują poprzednie arkusze egzaminacyjne w celu lepszego przygotowania uczniów do egzaminu oraz czy uważają, że trudność danego egzaminu jest porównywalna między latami. Oryginalne brzmienie pytań zawartych w kwestionariuszu nauczycieli zawiera Aneks 4 (4.3 – szkoła podstawowa, 4.4 – gimnazjum oraz 4.5 – szkoła ponadgimnazjalna). Wyniki przedstawiono w podziale na etapy edukacyjne oraz etapy badania zrównującego.

6.9.2.1. Sprawdzian

W badaniu ankietowym zadano nauczycielom m.in. pytanie o ocenę przygotowania uczniów do sprawdzianu. Jak można zauważyć (Tabela 6.22), większość nauczycieli (około 86%) ocenia stopień przygotowania uczniów jako „raczej dobrze”, co może też świadczyć o wybieraniu bezpiecznej, wymijającej odpowiedzi. W roku 2013 nauczyciele oceniali odrobinę wyżej stopień przygotowania uczniów do sprawdzianu niż w pozostałych latach (odsetek nauczycieli oceniających przygotowanie uczniów jako złe wynosił zaledwie 3,87). Sam rozkład procentowy opinii nauczycieli jest podobny we wszystkich analizowanych latach.

Tabela 6.22. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, jak nauczyciel ocenia przygotowanie uczniów do sprawdzianu

Rok:	2012 (n=1764)	2013 (n=310)	2014 (n=318)
Bardzo dobrze	7,94	9,35	6,60
Raczej dobrze	85,88	86,77	85,85
Raczej źle	5,90	3,87	7,23
Bardzo źle	0,28	0,00	0,31

W związku z tym, że sprawdzian jest egzaminem wysokiej doniosłości, natomiast badanie zrównujące jest testem niskiej doniosłości, nauczycieli poproszono również o ocenę stopnia motywacji uczniów do rozwiązywania zadań z zeszytów wykorzystanych w badaniu zrównującym. Opinie nauczycieli w analizowanych latach były podobne (zob. Tabela 6.23): około 1/3 nauczycieli twierdziła, że uczniom bardzo zależy na jak najlepszym rozwiązaniu testu, ponad połowa, że uczniom zależy jedynie trochę, natomiast ponad 10% twierdziło, że uczniom zależy w niewielkim stopniu lub nie zależy wcale na wynikach uzyskiwanych w teście. Analiza wyników ankiety przeprowadzonej wśród uczniów wskazuje

jednak na niewielki związek pomiędzy sumą punktów uzyskiwaną w kwestionariuszu motywacji a wynikiem uzyskiwanym w teście.

Tabela 6.23. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel uważa, że uczniom zależy na jak najlepszym rozwiązaniu testu w badaniu zrównującym (sprawdzian)

Rok:	2012 (n=1764)	2013 (n=309)	2014 (n=318)
Bardzo zależy	35,88	35,28	37,11
Trochę zależy	51,76	54,37	51,89
Niewiele zależy	11,56	10,36	10,38
Wcale nie zależy	0,79	0,00	0,63

Nauczycieli zapytano także, czy ich zdaniem udział w badaniu zrównującym sprawi, że uczniowie uzyskają wyższy wynik na rzeczywistym egzaminie (sprawdzianie) niż pozostali uczniowie. W analizowanych latach, nauczyciele odpowiadali najczęściej „raczej tak” lub „ani tak, ani nie”, choć tendencja do odpowiadania „raczej tak” jest najniższa w roku 2014 (zob. Tabela 6.24.). W roku 2014 zwiększyła się także tendencja do odpowiadania „raczej nie”. Ponownie można jednak zauważyć dosyć podobny rozkład odpowiedzi na pytanie w poszczególnych latach. Ciekawie może stosunkowo spora wybieralność kategorii środkowej we wszystkich analizowanych latach – ponad 1/3 nauczycieli nie jest w stanie stwierdzić, czy udział w badaniu zrównującym w jakikolwiek sposób wpłynie na wyniki uzyskiwane przez uczniów.

Tabela 6.24. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel uważa, że udziału w badaniu zrównującym może mieć wpływ na wyniki sprawdzianu

Rok:	2012 (n=1763)	2013 (n=309)	2014 (n=315)
Zdecydowanie tak	4,25	3,56	4,76
Raczej tak	40,39	43,37	37,46
Ani tak, ani nie	37,27	34,95	38,73
Raczej nie	16,51	16,83	18,41
Zdecydowanie nie	1,59	1,29	0,63

Nauczycielom zadano także pytanie o korzystanie z arkuszy z poprzednich edycji sprawdzianu podczas zajęć z uczniami wybranej klasy. Można zauważyć, że większość (około 2/3) nauczycieli korzysta ze starych arkuszy, choć w 2014 roku zaobserwowano spadek w wybieralności kategorii „tak” o około 4 punkty procentowe (zob. Tabela 6.25). Może (lecz nie musi) to świadczyć o pewnej tendencji do uczenia pod egzamin.

Tabela 6.25. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel wykorzystywał arkusze z poprzednich edycji sprawdzianu podczas zajęć z uczniami (sprawdzian)

Rok:	2012 (n=1680)	2013 (n=296)	2014 (n=310)
Tak	67,08	63,51	62,90
Nie	28,45	31,08	31,61
Nie pamiętam	4,46	5,41	5,48

Kolejne pytanie dotyczyło opinii nauczycieli na temat porównywalności trudności arkuszy egzaminacyjnych wykorzystywanych podczas sprawdzianu. Ponad połowa nauczycieli uważa, że trudność sprawdzianu pomiędzy latami jest porównywalna, natomiast nieco ponad jedna trzecia, że nie (por. Tabela 6.26). Pomiędzy kolejnymi edycjami badania rośnie procent nauczycieli, którzy na to pytanie odpowiadają „nie wiem”. Gdyby badania kwestionariuszowe nauczycieli w przyszłości były pogłębione należałoby się zastanowić na jakich przesłankach opiera się przekonanie ponad połowy nauczycieli o tym, że trudności arkuszy jest porównywalna między latami.

Tabela 6.26. Procentowy rozkład odpowiedzi na pytanie o to, czy nauczyciele uważają, że trudność arkuszy sprawdzianu jest porównywalna między latami

Rok:	2012 (n=1755)	2013 (n=308)	2014 (n=315)
Tak	57,04	52,92	56,51
Nie	35,56	37,99	33,65
Nie wiem	7,41	9,09	9,84

6.9.2.2. Egzamin gimnazjalny

Większość nauczycieli raczej dobrze ocenia poziom przygotowania uczniów do egzaminu gimnazjalnego (zob. Tabela 6.27). Był on oceniany jako najwyższy w roku 2013 – można więc zauważyć spadek dla roku 2014, jednak jego wystąpienie wymaga dalszej eksploracji. Spadek odsetka osób oceniających w 2012 przygotowanie uczniów jako bardzo dobre lub raczej dobre w stosunku do roku 2013 mógł być związany ze zmianą formuły egzaminu. Hipoteza ta wymagałaby jednak sprawdzenia. Rozkłady opinii dla poszczególnych edycji zrównania są do siebie podobne.

Tabela 6.27. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, jak nauczyciel ocenia przygotowanie uczniów do egzaminu gimnazjalnego

Rok:	2011 (n=3252)	2012 (n=618)	2013(n=608)	2014 (n=620)
Bardzo dobrze	4,55	2,27	4,61	4,52
Raczej dobrze	80,72	81,39	85,03	79,68
Raczej źle	14,27	16,34	9,87	15,65
Bardzo źle	0,46	0,00	0,49	0,16

Rozkład odpowiedzi na pytanie o to, czy nauczyciel uważa, że uczniom zależy na jak najlepszym rozwiązaniu testu w badaniu zrównującym jest podobny dla wszystkich analizowanych lat (zob. Tabela

6.28). Można zauważyć, że w stosunku do etapu sprawdzianu (por. Tabela 6.23) mniejszy procent nauczycieli uważa, że uczniom zależy w znacznym stopniu na wynikach badania zrównującego. Procent nauczycieli uważających, że uczniom zależy trochę na wynikach badania zrównującego jest podobny, jak wśród nauczycieli szkół podstawowych. Większa niż dla szkół podstawowych jest natomiast wybieralność kategorii „niewiele zależy” oraz „wcale nie zależy”.

Tabela 6.28. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel uważa, że uczniom zależy na jak najlepszym rozwiązaniu testu w badaniu zrównującym

Rok:	2011 (n=3268)	2012 (n=617)	2013 (n=604)	2014 (n=620)
Bardzo zależy	26,50	20,26	20,36	21,15
Trochę zależy	51,47	49,11	56,95	55,45
Niewiele zależy	20,07	28,85	21,19	20,03
Wcale nie zależy	1,96	1,78	1,49	3,37

Opinia nauczycieli na temat tego, czy udział w badaniu zrównującym może mieć wpływ na wyniki egzaminu gimnazjalnego jest podobna w analizowanych latach (zob. Tabela 6.29). Najczęściej wybraną kategorią jest odpowiedź „ani tak, ani nie”. Kategorie skrajne są stosunkowo najrzadziej wybierane.

Tabela 6.29. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel uważa, że udziału w badaniu zrównującym mógł mieć wpływ na wyniki egzaminu gimnazjalnego

Rok:	2011 (n=3278)	2012 (n=607)	2013 (n=595)	2014 (n=623)
Zdecydowanie tak	2,44	2,80	2,69	2,61
Raczej tak	32,09	30,64	35,97	36,54
Ani tak, ani nie	39,35	38,22	39,33	41,60
Raczej nie	23,83	26,85	20,67	17,46
Zdecydowanie nie	2,29	1,48	1,34	1,79

Jak wskazuje Tabela 6.30, w latach 2013 i 2014 zwiększył się procent nauczycieli wykorzystujących arkusze z poprzednich edycji egzaminu gimnazjalnego podczas zajęć z uczniami (różnica rzędu 3–4 punktów procentowych). Może to się wiązać ze zmianą formuły egzaminu gimnazjalnego od 2012 roku. Można też zauważyć, że proporcja nauczycieli, która korzysta ze starych arkuszy gimnazjalnych podczas zajęć jest dla wszystkich analizowanych lat większa niż w przypadku sprawdzianu (zob. Tabela 6.25). To z kolei może potwierdzać, że wraz z kolejnymi etapami egzaminacyjnymi rośnie doniosłość wyników egzaminów i nasilenie praktyk związanych z nauczaniem pod egzaminy.

Tabela 6.30. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel wykorzystywał arkusze z poprzednich edycji egzaminu gimnazjalnego podczas zajęć z uczniami

Rok:	2011 (n=3276)	2012 (n=629)	2013 (n=603)	2014 (n=628)
Tak	76,86	79,65	83,14	82,32
Nie	19,51	18,76	13,72	14,81
Nie pamiętam	3,63	1,59	3,14	2,87

Tabela 6.31 obrazuje opinie na temat porównywalności trudności arkuszy egzaminów gimnazjalnych. Możemy zauważyć, że w roku 2014 zwiększyła się proporcja nauczycieli, twierdząca, że trudność arkuszy egzaminu gimnazjalnego jest porównywalna pomiędzy latami. Wyniki są dosyć niespójne – te z roku 2011 przypominają wyniki z roku 2013. W roku 2012 procent nauczycieli uważających, że trudność arkuszy jest nieporównywalna między latami był największy, natomiast w roku 2014 tendencja ta uległa odwróceniu. Być może niższy odsetek nauczycieli uważających trudność arkuszy za porównywalną w 2012 roku wynikał ze zmiany formuły egzaminu gimnazjalnego. Do momentu przeprowadzenia egzaminu nauczyciele mogli spodziewać się, że w nowej formule egzamin będzie trudniejszy niż w poprzedniej formule. W roku 2013 i 2014 odsetek nauczycieli uważających trudność arkuszy za porównywalną ponownie wzrasta.

Tabela 6.31. Procentowy rozkład odpowiedzi na pytanie o to, czy nauczyciele uważają, że trudność arkuszy egzaminu gimnazjalnego jest porównywalna między latami

Rok:	2011 (n=3278)	2012 (n=617)	2013 (n=612)	2014 (n=627)
Tak	63,33	52,19	62,91	71,61
Nie	31,45	42,14	28,43	22,33
Nie wiem	5,22	5,67	8,66	6,06

6.9.2.3. Matura

Zdecydowana większość nauczycieli ocenia, że uczniowie są raczej dobrze przygotowani do matury, przy czym największy odsetek dla tej kategorii odpowiedzi charakteryzuje ocenę przygotowania do egzaminu z języka angielskiego i polskiego (zob. Tabela 6.32). W przypadku egzaminu maturalnego z matematyki około 30% nauczycieli w 2014 roku i 26% nauczycieli w 2013 roku ocenia, że uczniowie są raczej źle przygotowani do egzaminu. W odniesieniu do matury z języka angielskiego, największy odsetek (11,4) nauczycieli jest przekonany o bardzo dobrym przygotowaniu uczniów do egzaminu, podczas gdy w odniesieniu do egzaminu z matematyki, w 2013 roku tylko 1% nauczycieli wyraził takie przekonanie. Stosunkowo najwyższe odsetki nauczycieli matematyki uważających, że uczniowie są źle przygotowani do matury może wynikać ze świadomości, że matematyka od lat jest przedmiotem sprawiającym na maturze najwięcej problemów (niska zdawalność w porównaniu do innych przedmiotów).

Tabela 6.32. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, jak nauczyciel ocenia przygotowanie uczniów do matury

Przedmiot i rok:	matematyka 2013 (n=193)	matematyka 2014 (n=123)	j. angielski 2014 (n=535)	j. polski 2014 (n=295)
Bardzo dobrze	1,04	4,07	11,40	3,73
Raczej dobrze	72,54	65,04	78,69	82,37
Raczej źle	25,91	30,08	9,72	13,56
Bardzo źle	0,52	0,81	0,19	0,34

Zdecydowana większość nauczycieli uważa, że uczniom zależało na jak najlepszym rozwiązaniu testu w badaniu zrównującym. W przypadku każdego egzaminu maturalnego, co najmniej ¼ badanych nauczycieli jest przekonana o bardzo wysokim poziomie zaangażowania uczniów (zob. Tabela 6.33).

Bardzo mała liczba nauczycieli podziela opinię, że uczniom wcale nie zależało na rozwiązaniu testu w badaniu zrównującym – było to w każdym analizowanym egzaminie maksymalnie 1% przebadanych nauczycieli.

Tabela 6.33. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel uważa, że uczniom zależy na jak najlepszym rozwiązaniu testu w badaniu zrównującym (matura)

Przedmiot i rok:	matematyka 2013 (n=199)	matematyka 2014 (n=124)	j. angielski 2014 (n=540)	j. polski 2014 (n=302)
Bardzo zależy	27,14	25,81	32,84	24,83
Trochę zależy	53,27	55,65	50,75	59,60
Niewiele zależy	18,59	18,55	15,30	14,24
Wcale nie zależy	1,01	0,00	1,12	1,32

Opinie nauczycieli na temat możliwego wpływu udziału uczniów w badaniu zrównującym na późniejsze wyniki egzaminu maturalnego są podzielone. Najliczniejszą grupę stanowili dla wszystkich analizowanych egzaminów nauczyciele, którzy nie wiedzą, czy taki wpływ mógłby mieć miejsce – uważa tak około 40% badanych (zob. Tabela 6.34). Ciekawe wydaje się to, że największy odsetek nauczycieli zdecydowanie nie zgadzający się ze stwierdzeniem, że udział w badaniach może mieć wpływ na wyniki matury występuje wśród anglistów. W zestawieniu z najwyższym odsetkiem wśród anglistów przekonanych o bardzo dobrym przygotowaniu uczniów do matury, można zastanawiać się, czy nie uważają oni, że uczniowie są tak dobrze przygotowani do egzaminu, że rozwiązanie kolejnego testu nic nie zmienia. Hipoteza ta wymagałaby jednak potwierdzenia w kolejnych badaniach.

Tabela 6.34. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel uważa, że udział w badaniu zrównującym mógł mieć wpływ na wyniki matury

Przedmiot i rok:	matematyka 2013 (n=199)	matematyka 2014 (n=126)	j. angielski 2014 (n=541)	j. polski 2014 (n=307)
Zdecydowanie tak	6,53	2,38	4,44	4,89
Raczej tak	36,18	41,27	32,72	35,83
Ani tak, ani nie	41,21	41,27	41,40	37,79
Raczej nie	15,08	14,29	18,85	19,87
Zdecydowanie nie	1,01	0,79	2,59	1,63

Zdecydowana większość nauczycieli deklaruowała wykorzystanie arkuszy z poprzednich edycji matury podczas zajęć z uczniami, choć wśród nauczycieli języka angielskiego ich odsetek jest najmniejszy (zob. Tabela 6.35). Proporcja nauczycieli wykorzystujących stare arkusze egzaminacyjne jest zdecydowanie wyższa dla wszystkich przedmiotów maturalnych, niż w przypadku egzaminu gimnazjalnego czy sprawdzianu. Rosnące między etapami edukacyjnymi wykorzystanie arkuszy egzaminacyjnych z poprzednich edycji egzaminu w nauczaniu potwierdza coraz większy stopień doniosłości egzaminów zewnętrznych na zakończenie kolejnych typów szkoły.

Tabela 6.35. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel wykorzystywał arkusze z poprzednich edycji matury podczas zajęć z uczniami

Przedmiot i rok:	matematyka 2013 (n=198)	matematyka 2014 (n=126)	j. angielski 2014 (n=539)	j. polski 2014 (n=309)
Tak	94,97	92,86	86,81	96,12
Nie	4,52	6,35	10,07	2,27
Nie pamiętam	0,50	0,79	3,11	1,62

Nauczyciele uważają w większości, że trudność arkuszy maturalnych jest porównywalna pomiędzy latami – szczególnie w przypadku matematyki (tendencja ta wzrosła w roku 2014 w porównaniu do roku 2013) oraz języka angielskiego (Tabela 6.36). W przypadku języka polskiego, procent nauczycieli zgadzających się z tym stwierdzeniem jest mniejszy. Proporcja nauczycieli zgadzających się z tym twierdzeniem jest wyższa niż w odnośnie egzaminu gimnazjalnego i sprawdzianu.

Tabela 6.36. Procentowy rozkład odpowiedzi na pytanie o to, czy nauczyciele uważają, że trudność arkuszy maturalnych jest porównywalna między latami

Przedmiot i rok:	matematyka 2013 (n=198)	matematyka 2014 (n=126)	j. angielski 2014 (n=539)	j. polski 2014 (n=309)
Tak	75,25	80,95	86,08	76,80
Nie	23,74	16,67	10,81	21,57
Nie pamiętam	1,01	2,38	3,11	1,63

6.9.3. Podsumowanie

Najważniejsze wnioski z badań ankietowych można przedstawić w następujących punktach:

- Poziom motywacji testowej ma niewielki wpływ na wyniki osiągnięte w badaniu zrównującym, niezależnie od wykorzystanej miary motywacji.
- Różnice w poziomie motywacji pomiędzy zeszytami wykorzystanymi w zrównaniu są w większości nieistotne statystycznie. Występowanie różnic istotnych statystycznie pomiędzy zeszytami testowymi przejawia się w różnicach rzędu jednego do dwóch punktów, tak więc ich praktyczne znaczenie jest pomijalne.
- Nauczyciele oceniają poziom przygotowania uczniów do podejmowania egzaminów zewnętrznych jako „raczej dobry”, przy czym nieco wyżej oceniają przygotowanie uczniów do sprawdzianu i egzaminu gimnazjalnego niż matury.
- Nauczyciele w większości wypadków uważają, że uczniowie są zmotywowani do udziału w badaniach zrównujących i zależy im na jak najlepszych wynikach. Przy czym, należy zauważyć, że jako najwyższy poziom motywacji uczniów oceniają nauczyciele uczący w szkołach podstawowych.

- Nauczyciele w większości wykorzystują arkusze z poprzednich edycji egzaminów podczas zajęć, przy czym tendencja ta rośnie wraz z kolejnymi etapami edukacyjnym – na etapie matury ponad 90% nauczycieli wykorzystuje stare arkusze podczas zajęć.
- Nauczyciele w większości uznają, że trudność egzaminów jest porównywalna między latami. Przy czym proporcja nauczycieli zgadzająca się ze stwierdzeniem, że trudność egzaminów jest porównywalna jest najniższa w szkole podstawowej, wyższa w gimnazjum, a najwyższa w szkołach ponadgimnazjalnych.

Do wyników ankietowych należy jednak podchodzić z pewnego rodzaju ostrożnością. Przede wszystkim, należy pamiętać, że miary samoopisowe są podatne na zniekształcenia. Na wyniki mogą wpływać zmienne, które są niezwiązane z mierzonym konstruktem (Baumgartner i Steenkamp, 2001) – przede wszystkim związane z nastawieniem odpowiedzi (*response set*) i stylem odpowiedzi (*response style*) (zob. Rorer, 1965). Nastawienie odpowiedzi polega na intencjonalnym manipulowaniu udzielanymi odpowiedziami, powodowanym np. przez zmienną aprobaty społecznej (*social desirability*). Styl odpowiedzi oznacza różnice indywidualne w używaniu i wykorzystywaniu kategorii odpowiedzi niezależnie od treści danego pytania oraz mierzonej przez nie cechy ukrytej (np. Paulhus, 1991). Na prezentowane wyniki badań samoopisowych uczniów i nauczycieli mogła wpływać w znacznym stopniu tendencja do zaprezentowania się w lepszym świetle (np. przy pytaniach o poziom motywacji zadawanych uczniom oraz przy pytaniach, jak nauczyciele oceniają ich poziom przygotowania do właściwego egzaminu).

7. Analizy z użyciem porównywalnych wyników egzaminacyjnych

Główną zaletą porównywalnych wyników egzaminacyjnych (PWE) jest możliwość odnoszenia do siebie wyników egzaminów z poszczególnych lat. Do tej pory, w związku z brakiem kontroli nad trudnością arkuszy egzaminacyjnych w poszczególnych edycjach egzaminu, przeprowadzanie takich porównań na surowych wynikach nie było metodologicznie uprawnione. Dzięki zrównaniu wyników poszczególnych egzaminów i umieszczeniu ich na wspólnej skali możemy bez przeszkód wypowiadać się o zmianach wyników kształcenia na poszczególnych etapach edukacyjnych (mierzonych na odpowiednich egzaminach) w czasie.

W badaniach edukacyjnych często porusza się kwestię różnic w osiągniętych wynikach pomiędzy grupami uczniów wydzielonymi ze względu na:

- uwarunkowania regionalne (zob. Herbst, 2004; Herbst i Herczyński, 2005; Herczyński i Herbst, 2002),
- płeć uczniów (zob. EACEA, 2010),
- posiadanie (lub nie) przez uczniów diagnozy dysleksji rozwojowej (zob. Wejner, 2009),
- lokalizację szkoły (zob. Dolata, 2008; Dolata, Jasińska i Modzelewski, 2012; Żółtak, 2011),
- typ szkoły (niepubliczna vs publiczna) (zob. Putkiewicz i Wiłkomirska, 2004).

W tym rozdziale będziemy analizować zrównane wyniki egzaminacyjny pod kątem wymienionych powyżej zmiennych. Istotnym warunkiem trafności egzaminów jest to, aby wyniki pomiaru nie były zależne od powyżej wymienionych czynników. Same porównywalne wyniki egzaminacyjne nie pozwolą jednak na wyjaśnienie istniejącego zróżnicowania – potrzebna jest do tego wiedza z odpowiedniego zakresu, a często również pogłębione badania. Zanim jednak nastąpi wyjaśnianie, konieczne jest zidentyfikowanie i opisanie danego zjawiska, do czego porównywalne wyniki egzaminacyjne z powodzeniem mogą posłużyć.

7.1. Ogólne wyniki egzaminów w poszczególnych latach

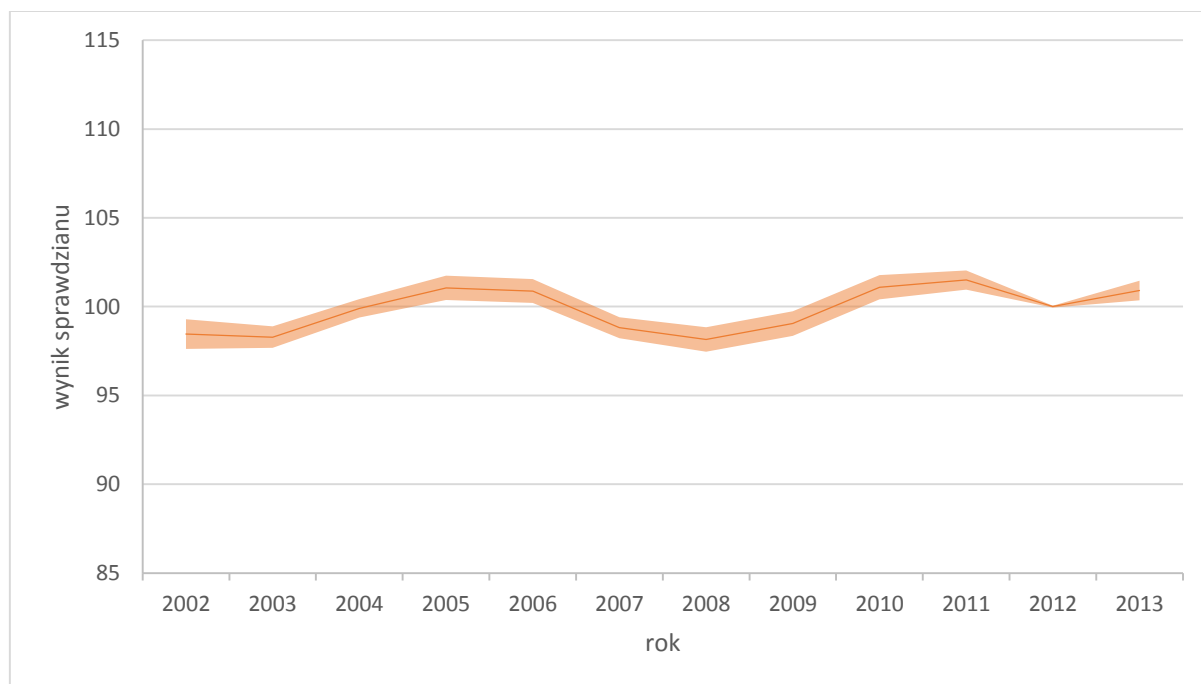
W tej części rozdziału omówione zostaną średnie porównywalne wyniki egzaminacyjne z poszczególnych egzaminów na przestrzeni czasu dla całej populacji uczniów. Porównanie tych wyników w kolejnych edycjach egzaminu informuje nas o tym, czy z biegiem lat zmieniały się umiejętności przystępujących doń uczniów. Należy jednak pamiętać o tym, że osiągnięte przez uczniów wyniki są zależne od formy przeprowadzanego egzaminu (np. zmian struktury testu czy badanych umiejętności). Gwałtowne zmiany umiejętności w czasie mogą być zatem tylko odzwierciedleniem istotnych zmian w opisywanej powyżej specyfice przeprowadzanego egzaminu, a nie wynikać z faktycznych różnic w umiejętnościach odpowiednich kohort uczniów.

Wykresy prezentowane w dalszej części przedstawiają średnie porównywalne wyniki egzaminacyjne (oś pionowa), w kolejnych edycjach egzaminu (oś pozioma). Wartość 100 odpowiada średniemu wynikowi egzaminu w 2012 roku (ten rok wybrano jako rok bazowy), a 15 punktów odpowiada jednemu odchyleniu standardowemu wyników egzaminu w 2012 roku. Średni porównywalny wynik egzaminu oznaczono ciągłą linią, a jaśniejszy pasek wokół niej oznacza 95% przedział ufności, który służy do

zobrazowania niepewności, jaką obarczone jest szacowanie średniego porównywalnego wyniku egzaminu dla danej grupy uczniów.

7.1.1. Sprawdzian

Rysunek 7.1 przedstawia wykres średnich porównywalnych wyników egzaminacyjnych dla sprawdzianu dla całego kraju. Można zaobserwować, że wyniki sprawdzianu ulegają niewielkim fluktuacjom na przestrzeni czasu. Różnica pomiędzy najwyższymi i najniższymi wynikami wynosi poniżej pięciu punktów, czyli mniej niż jedną trzecią odchylenia standardowego. Jeśli założymy, że w kolejnych edycjach sprawdzianu mierzone są podobne umiejętności, nie należy oczekiwać znacznych zmian ich poziomu dla całej populacji uczniów w krótkim okresie. Pomimo istotnych statystycznie różnic pomiędzy skrajnymi wynikami w okresach ich wzrostów i spadków, oscylują one wokół wartości 100, w granicach $\pm 2,5$ punktu. Nie są to duże odchylenia, można zatem uznać, że w przypadku sprawdzianu mamy do czynienia z dość stabilnym poziomem umiejętności uczniów w kolejnych latach. Porównywalne wyniki egzaminacyjne sprawdzianu obejmują wyniki uzyskane z całego testu, bez podziału na poszczególne umiejętności. Nie możemy zatem stwierdzić, czy zaobserwowane wahania wyników są spowodowane przez zmianę poziomu wszystkich mierzonych umiejętności, czy zmianę poziomu konkretnej umiejętności (np. czytania lub pisania).



Rysunek 7.1. Porównywalne wyniki sprawdzianu w latach 2002–2013

Uzyskane wyniki można zweryfikować poprzez porównanie z wynikami uzyskanymi w badaniach uczniów w podobnym wieku, zwłaszcza o dużej trafności i rzetelności. Porównywalne wyniki egzaminacyjne dla sprawdzianu można zestawiać z wynikami badania PIRLS (*Progress in International Reading Literacy Study*). Badano w nim dzieci nieco młodsze, około dziesięcioletnie (będące w czwartej klasie szkoły podstawowej), podczas gdy dzieci podchodzące do sprawdzianu w szóstej klasie szkoły podstawowej mają około dwunastu lat. Pomiedzy dwoma edycjami badania PIRLS, w 2006 i 2011 roku, nastąpił nieznaczny wzrost umiejętności językowych uczniów (Konarzewski i in., 2012). Wynosił on jednak zaledwie 7 punktów w skali o średniej 500 i odchyleniu standardowym 100 i był nieistotny statystycznie. W porównywalnych wynikach egzaminacyjnych również mamy do czynienia z

nieznacznym i nieistotnym statystycznie wzrostem umiejętności uczniów pomiędzy tymi latami (o 0,6 punktu).

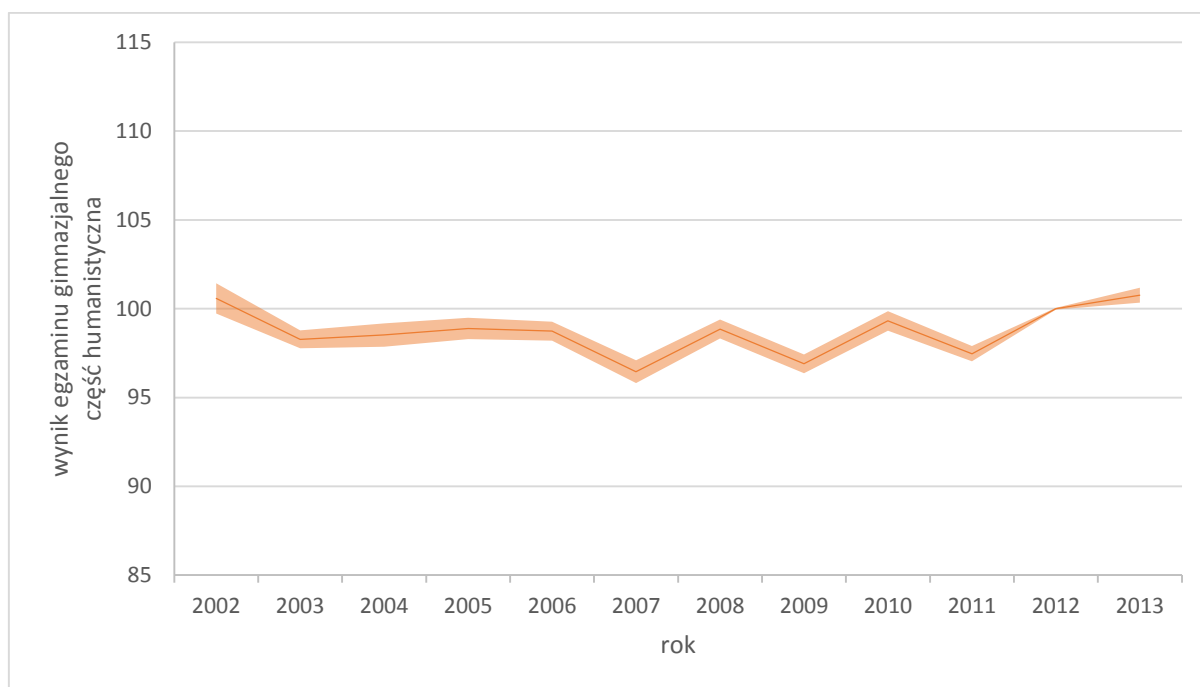
Niestety w badaniu PIRLS dysponujemy pomiarami jedynie z wskazanych dwóch lat i nie jesteśmy w stanie stwierdzić, jaki poziom umiejętności prezentowałyby uczniowie w latach pośrednich. Nie wiadomo więc, czy spadek porównywalnych wyników egzaminacyjnych w latach 2006–2008, a następnie wzrost w latach 2008–2011 zostałyby potwierdzone w badaniu PIRLS. Ponadto badanie PIRLS dotyczy jedynie umiejętności językowych, podczas gdy sprawdzian jest egzaminem interdyscyplinarnym. Ewentualne różnice w trendach mogłyby zatem wynikać z faktu mierzenia innych umiejętności w obydwu testach. Należy więc zachować ostrożność przy weryfikacji wyników zrównywania przy pomocy badań międzynarodowych, choć wydaje się, że jest to najlepszy sposób dla oceny trafności zrównywania.

7.1.2. Egzamin gimnazjalny

Od momentu wprowadzenia, egzamin gimnazjalny podzielony był na dwie, osobne części: humanistyczną i matematyczno-przyrodniczą. Dzięki temu istnieje możliwość obserwowania zmian w czasie w zakresie poszczególnych rodzajów umiejętności, czego nie możemy uczynić dla sprawdzianu w związku z jego ponadprzedmiotowym charakterem. W roku 2012 dodatkowo w ramach obu części wydzielono egzaminy z języka polskiego, historii i wiedzy o społeczeństwie (w miejsce wcześniejszej części humanistycznej), oraz matematyki i przedmiotów przyrodniczych (w miejsce wcześniejszej części matematyczno-przyrodniczej). Podobnie jak w przypadku sprawdzianu, zrównanie wyników obydwu części egzaminu gimnazjalnego przeprowadzono do roku 2012. W celu zachowania porównywalności wyników między wszystkimi edycjami egzaminu, od roku 2012 wyniki z języka polskiego oraz historii i wiedzy o społeczeństwie łączone są w część humanistyczną, a z matematyki i przedmiotów przyrodniczych w część matematyczno-przyrodniczą. Egzamin gimnazjalny z języka obcego nie podlegał zrównaniu.

7.1.2.1. Część humanistyczna

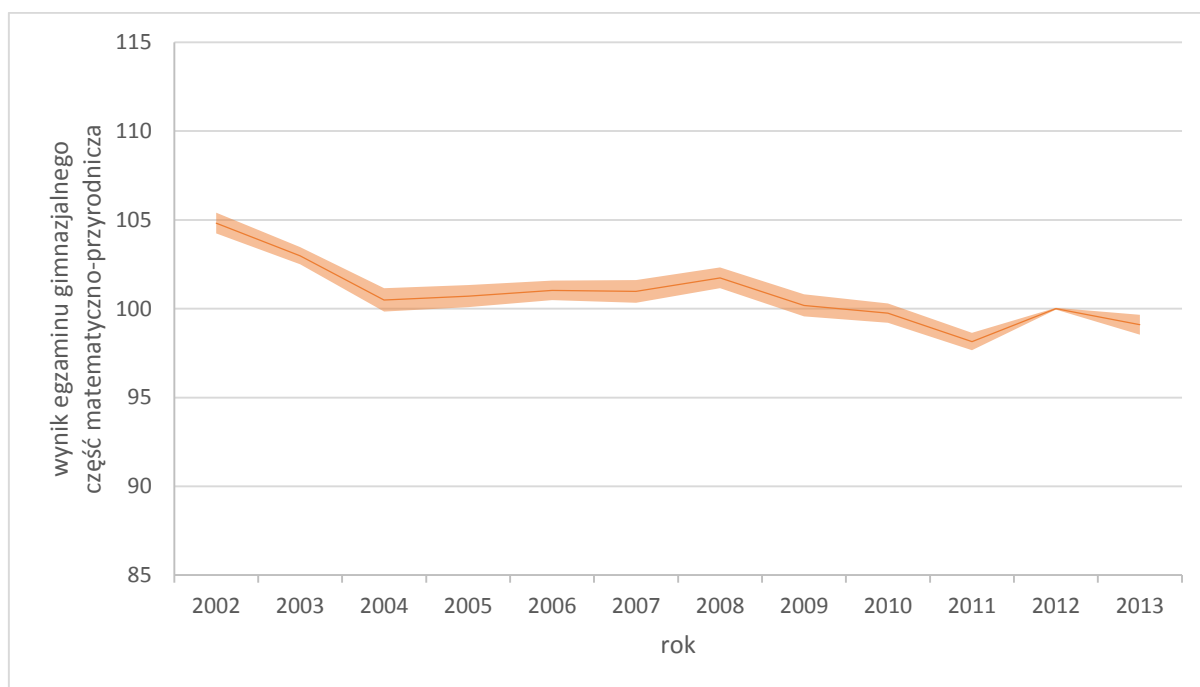
Wyniki części humanistycznej egzaminu gimnazjalnego na poziomie kraju, które ilustruje Rysunek 7.2, również ulegają niewielkim fluktuacjom na przestrzeni czasu, podobnie jak wyniki sprawdzianu. Spadek wyników między 2002 i 2003 rokiem mógł być spowodowany zawyżonymi wynikami egzaminu w 2002 roku. Była to pierwsza edycja egzaminu, więc należy podchodzić ostrożnie do jego wyników, gdyż nie wszystkie procedury mogły być właściwie stosowane. Zaobserwować można także następujące po sobie spadki i wzrosty wyników w latach 2006–2012. Różnice pomiędzy poszczególnymi latami w tym okresie wynosiły od dwóch do dwóch i pół punktu. Są one istotne statystycznie, lecz na tyle małe, że nie powinny budzić niepokoju. Zmuszają one jednak do zastanowienia nad ich przyczyną, gdyż, jak wspomniano wcześniej w kontekście sprawdzianu, nie należy spodziewać się znaczących różnic pomiędzy sąsiednimi latami w wynikach populacji uczniów, w sytuacji, gdy zakres treściowy egzaminu nie ulega znaczącym zmianom. Wnioskowanie o przyczynach tego zjawiska musiałoby obejmować dokładną analizę struktury i formatu egzaminu w opisywanym powyżej przedziale czasowym.



Rysunek 7.2. Porównywalne wyniki części humanistycznej egzaminu gimnazjalnego w latach 2002–2013

7.1.2.2. Część matematyczno-przyrodnicza

Porównywalne wyniki egzaminacyjne dla części matematyczno-przyrodniczej egzaminu gimnazjalnego na poziomie kraju są bardziej stabilne w czasie, niż w przypadku części humanistycznej. W latach 2002–2004 można zauważyć znaczący spadek wyników (o ponad cztery punkty, czyli prawie jedną trzecią odchylenia standardowego). Z podobnych spadkiem, choć mniejszym, mieliśmy do czynienia w części humanistycznej egzaminu, gdzie sygnalizowano, że może to mieć związek ze stopniowym wdrażaniem procedur nowego egzaminu i nie powinno budzić niepokoju. Następnie wyniki stopniowo rosły do 2008 roku, a później malały do 2011 roku (zob. Rysunek 7.3). W 2012 roku wprowadzono nową formułę egzaminu, co mogło zaowocować niewielkim wzrostem wyników (uczniowie mogli lepiej przygotować się do egzaminu w związku z niepewnością co do jego formy), w 2013 roku wyniki spadły jednak do poziomu z roku 2011 (przedziały ufności dla tych lat nie są rozłączne).



Rysunek 7.3. Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013

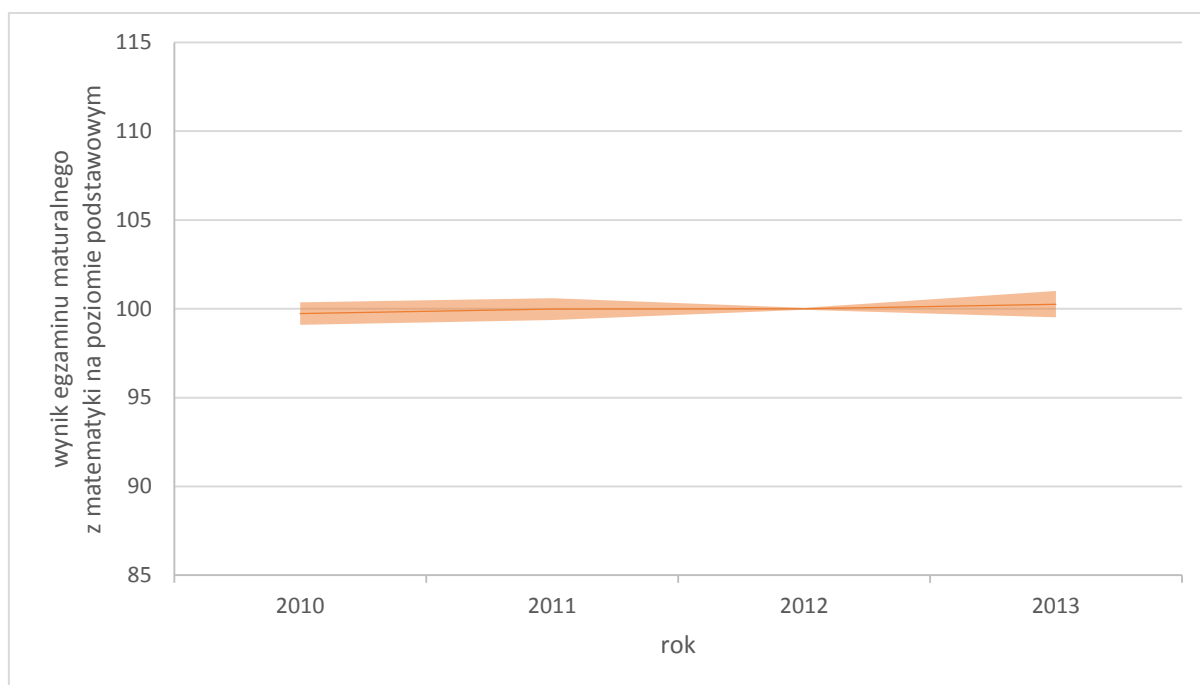
7.1.3. Egzamin maturalny

Różnorodność typów szkół ponadgimnazjalnych wymaga precyzyjnego określenia grupy uczniów, dla których przeprowadzane są analizy na tym etapie kształcenia. Najczęściej kiedy mowa o szkołach ponadgimnazjalnych, chodzi o licea ogólnokształcące i technika dla młodzieży, gdyż są to dominujące typy szkół ponadgimnazjalnych. Nie można jednak zapominać, że istnieją także licea profilowane, w których uczniowie kształcą się także zawodowo (oprócz kształcenia ogólnego) oraz szkoły dla dorosłych, w których specyfika nauczania różni się znacząco od szkół dla młodzieży. Szkoły te zostały utworzone w różnych celach, zatem trudno oczekiwać, że osiągnięte w nich wyniki kształcenia będą podobne. Uczniowie liceów ogólnokształcących, profilowanych i techników stanowią zdecydowaną większość na tym etapie kształcenia. W innych typach szkół uczy się stosunkowo niewielu uczniów, co stanowi problem przy próbach analizy w podziale na podgrupy z powodu ich niewielkiej liczebności. Musimy także uwzględnić kompletność danych, którymi dysponujemy. W przypadku szkół uzupełniających i dla dorosłych, częściej niż w liceach i technikach dla młodzieży, spotykamy się z brakami danych. Badania zrównujące wyniki egzaminu maturalnego przeprowadzono tylko dla uczniów z liceów ogólnokształcących, profilowanych i techników dla młodzieży. Powodem były po części wymienione wyżej niedogodności, a także ograniczenia logistyczne i finansowe. Więcej informacji na ten temat zawiera podrozdział 4.2, w którym opisano poszczególne populacje i dobór prób do badań zrównujących.

Poniżej przedstawione zostały średnie porównywalne wyniki dla poziomu podstawowego egzaminu maturalnego z matematyki i języka angielskiego. Dysponujemy takimi wynikami tylko dla lat 2010–2013, co ze względu na dość krótki okres utrudnia porównanie stabilności wyników z matematyki w stosunku do wcześniejszych etapów edukacyjnych. Język angielski na etapie gimnazjum nie został zrównany, więc takie porównanie nie jest w ogóle możliwe.

7.1.3.1. Matematyka

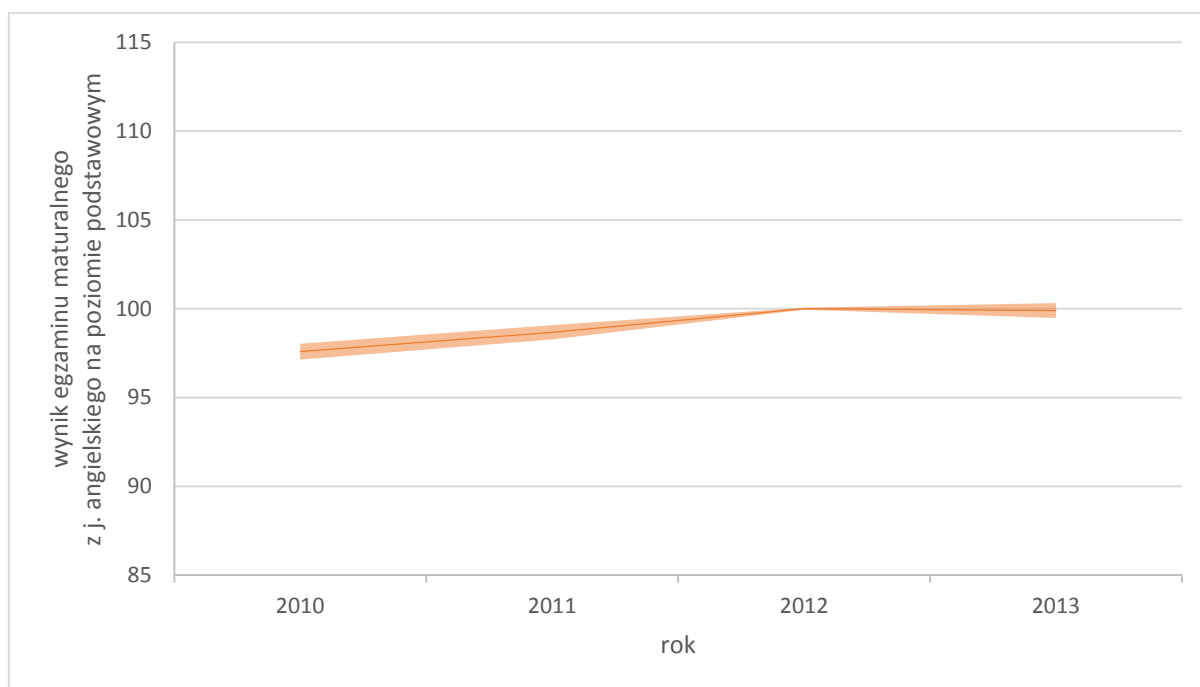
W latach 2010–2013 porównywalne wyniki poziomu podstawowego egzaminu maturalnego z matematyki utrzymują się na stałym poziomie (zob. Rysunek 7.4). Jak już wcześniej wspomniano, jest to zgodne z oczekiwaniami w stosunku do pomiaru poziomu umiejętności populacji uczniów w krótkim okresie.



Rysunek 7.4. Porównywalne wyniki egzaminu maturalnego z matematyki w latach 2010–2013

7.1.3.2. Język angielski

Najczęściej wybieranym przez uczniów na egzaminie maturalnym językiem nowożytnym jest język angielski. We wszystkich badanych latach (2010–2013) język ten wybrało ponad 80% zdających (Centralna Komisja Egzaminacyjna, 2010; 2011; 2012; 2013). W latach 2010–2012 następował niewielki wzrost porównywalnych wyników z poziomu podstawowego egzaminu maturalnego z języka angielskiego, natomiast w latach 2012 i 2013 wyniki te nie różnią się od siebie (zob. Rysunek 7.5).



Rysunek 7.5. Porównywalne wyniki egzaminu maturalnego z języka angielskiego w latach 2010–2013

7.2. Regiony kraju

Prowadząc analizy związane z funkcjonowaniem systemu edukacji możemy być zainteresowani zróżnicowaniem umiejętności uczniów w poszczególnych regionach kraju. Istnienie zależności pomiędzy wynikami w testach umiejętności oraz obserwowanym poziomem rozwoju gospodarczego na poziomie makroekonomicznym (Hanushek i Kim, 1995) prowadzi do wniosku, że im wyższy poziom edukacji w danym regionie, tym większy potencjał jego rozwoju gospodarczego oraz poziom życia mieszkańców (Herbst, 2004; Kurek, 2010). Porównywalne wyniki egzaminacyjne dają szansę prześledzenia zmian w umiejętnościach uczniów w czasie i sprawdzenia czy zmiany te przebiegają we wszystkich regionach w podobny sposób. Poniżej zostanie przedstawiona analiza PWE dla poszczególnych etapów kształcenia w podziale na regiony wg klasyfikacji zgodnej z Nomenklaturą Jednostek Terytorialnych do Celów Statystycznych (NTS)³⁴. Tabela 7.1 zawiera nazwy regionów wraz z wyszczególnieniem województw wchodzących w ich skład.

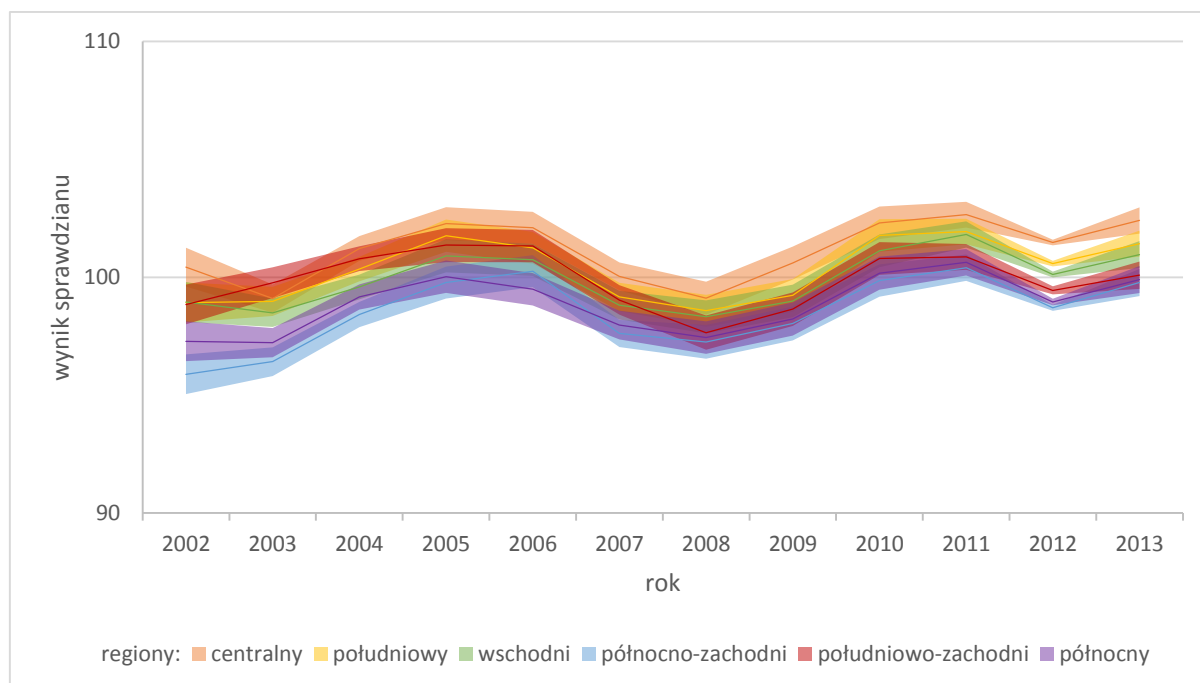
³⁴ Rozporządzenie Rady Ministrów z dnia 13 lipca 2000 r. w sprawie wprowadzenia Nomenklatury Jednostek Terytorialnych do Celów Statystycznych (NTS) (Dz. U. z 2000 r. Nr 58, poz. 685).

Tabela 7.1. Regiony i województwa wchodzące w ich skład (wg NTS)

Region	Województwa
centralny	łódzkie i mazowieckie
południowy	małopolskie i śląskie
wschodni	lubelskie, podkarpackie, świętokrzyskie i podlaskie
północno-zachodni	wielkopolskie, zachodniopomorskie i lubuskie
południowo-zachodni	dolnośląskie i opolskie
północny	kujawsko-pomorskie, warmińsko-mazurskie i pomorskie

7.2.1. Sprawdzian

Porównywalne wyniki sprawdzianu dla poszczególnych regionów kraju nie różnią się znacząco, choć w centralnej, południowej i wschodniej części Polski często były one wyższe niż w północnej i zachodniej części kraju (zob. Rysunek 7.6). Herczyński i Herbst (2002) oraz Herbst (2004) zaobserwowali podobne związki dla wyników sprawdzianu i egzaminu gimnazjalnego. Wskazują oni, że to zróżnicowanie można tłumaczyć wieloma uwarunkowaniami społeczno-ekonomicznymi, jak np. stopa bezrobocia czy poziom urbanizacji oraz historycznymi, pośród których wymieniają przynależność danego obszaru do różnych zaborów i nierównomierne obciążenie rolnictwem opartym na PGR-ach. Od roku 2011 można zaobserwować nieco wyraźniejsze różnice pomiędzy tymi grupami regionów niż we wcześniejszych latach. Warto zwrócić uwagę na nieco odmienny trend analizowanych wyników w regionie południowo-zachodnim. W latach 2002–2007 jego średnie wyniki były bliższe wynikom regionów: centralnego, południowego i wschodniego, lecz w latach 2008–2013 spadły one do poziomu wyników regionu północnego i północno-zachodniego.

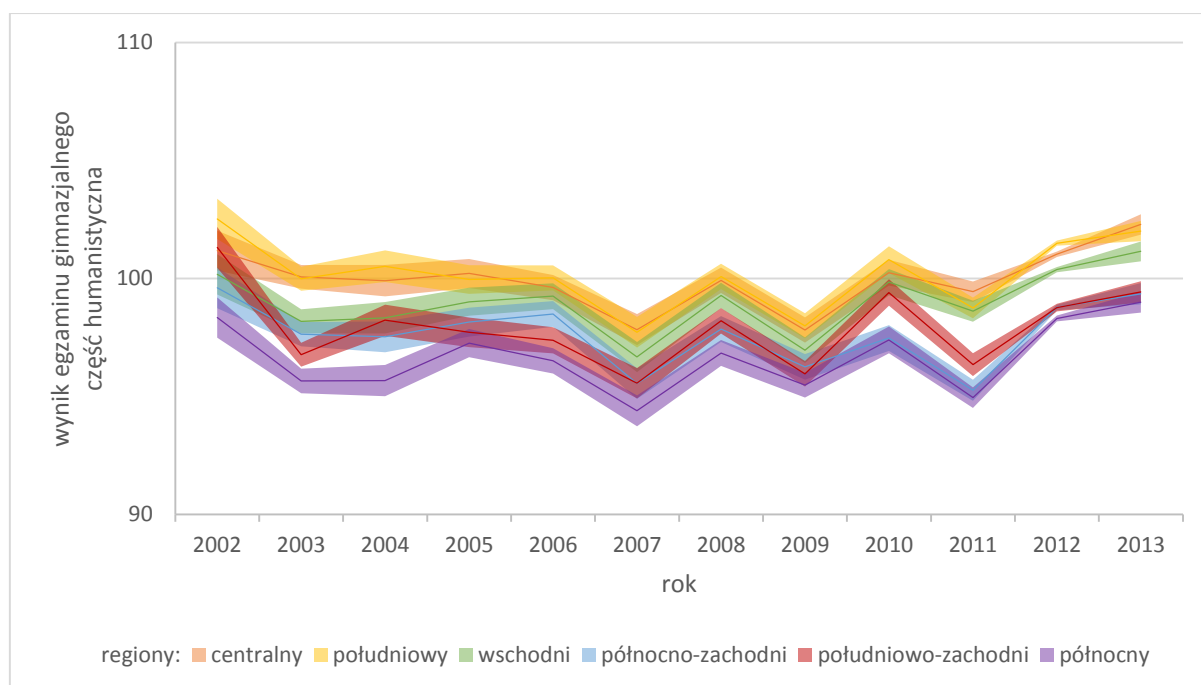


Rysunek 7.6. Porównywalne wyniki sprawdzianu w latach 2002–2013 w podziale na regiony wg NTS

7.2.2. Egzamin gimnazjalny

7.2.2.1. Część humanistyczna

Zaobserwowane dla sprawdzianu zróżnicowanie porównywalnych wyników pomiędzy centralną, południową i wschodnią częścią kraju a północną i zachodnią można również zauważyć dla części humanistycznej egzaminu gimnazjalnego. Od roku 2011 występuje bardzo wyraźna różnica średnich wyników pomiędzy tymi obszarami Polski (zob. Rysunek 7.7). W regionach centralnej, południowej i wschodniej części kraju można zaobserwować wyższe wyniki egzaminacyjne, natomiast w północnej i zachodniej części Polski są one niższe. Choć jest to tendencja zauważalna już na etapie sprawdzianu, to na etapie gimnazjum jest ona o wiele bardziej widoczna. Podobnie jak w przypadku sprawdzianu, w początkowych edycjach egzaminu uczniowie w regionie południowo-zachodnim osiągalni relatywnie wyższe wyniki, niż miało to miejsce w późniejszych latach. Jedynie w tym regionie trend odbiega nieco od trendu krajowego. Obserwowane różnice są istotne statystycznie, lecz nie są one duże – w latach 2011–2013 wynoszą nieco ponad dwa punkty. W okresie tych trzech lat nie ulegają one zmianom, lecz trudno ocenić czy w przyszłości pozostaną stałe.

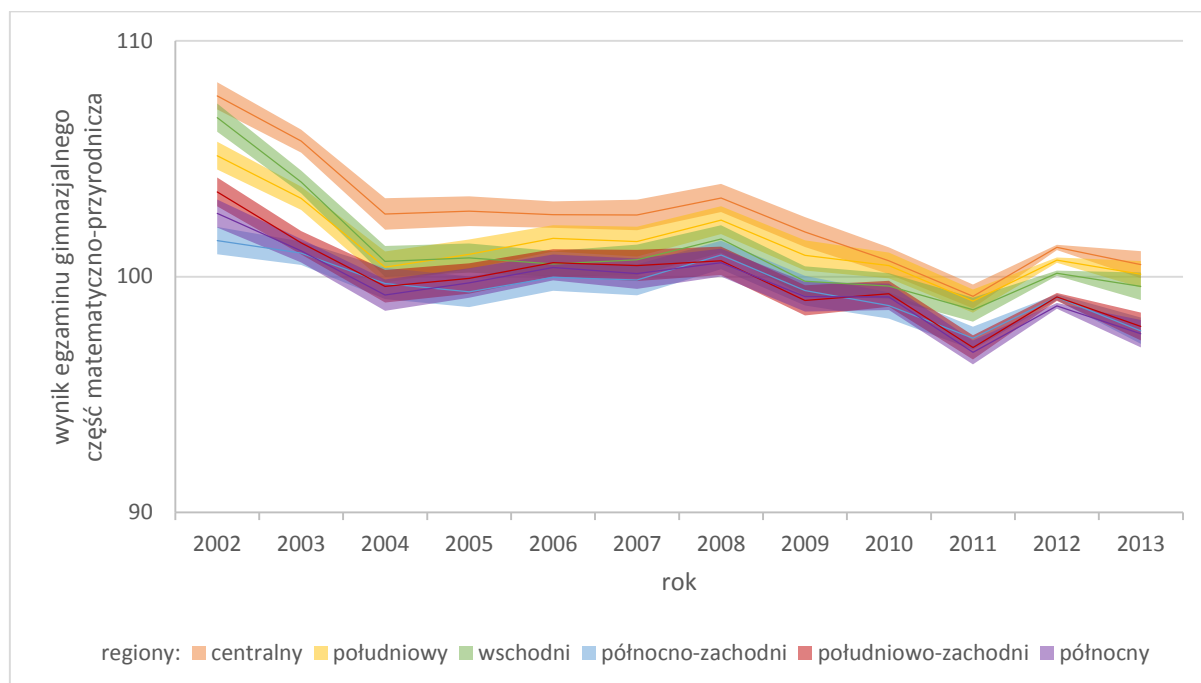


Rysunek 7.7. Porównywalne wyniki części humanistycznej egzaminu gimnazjalnego w latach 2002–2013 w podziale na regiony wg NTS

7.2.2.2. Część matematyczno-przyrodnicza

Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego również uwidaczniają geograficzny podział kraju, pod względem umiejętności uczniów (zob. Rysunek 7.8). Inaczej jednak niż w części humanistycznej egzaminu, w części matematyczno-przyrodniczej różnice między tymi obszarami kraju widoczne były już w początkowych edycjach egzaminu w latach 2002–2003, a nie tylko ostatnio, w latach 2011–2013. Jak dla sprawdzianu i części humanistycznej egzaminu gimnazjalnego, wyższe wyniki egzaminacyjne zaobserwowano w regionach: centralnym, południowym i wschodnim, natomiast niższe w regionach: północnym, północno-zachodnim i południowo-zachodnim. W części matematyczno-przyrodniczej bardziej wyraźna jest przewaga w osiągniętych wynikach uczniów z regionu centralnego nad pozostałymi uczniami – ich średni wynik jest we wszystkich latach najwyższy.

W latach 2004–2010 można dostrzec zmniejszenie się różnic pomiędzy poszczególnymi regionami kraju, jednak obserwowany wcześniej podział wzmacnia się po tym okresie. Wyniki regionu południowo-zachodniego kształtują się odmiennie niż dla sprawdzianu i części humanistycznej egzaminu gimnazjalnego – od pierwszej edycji egzaminu uczniowie z tego obszaru osiągają wyniki zbliżone do wyników uczniów z regionu północno-zachodniego i północnego.

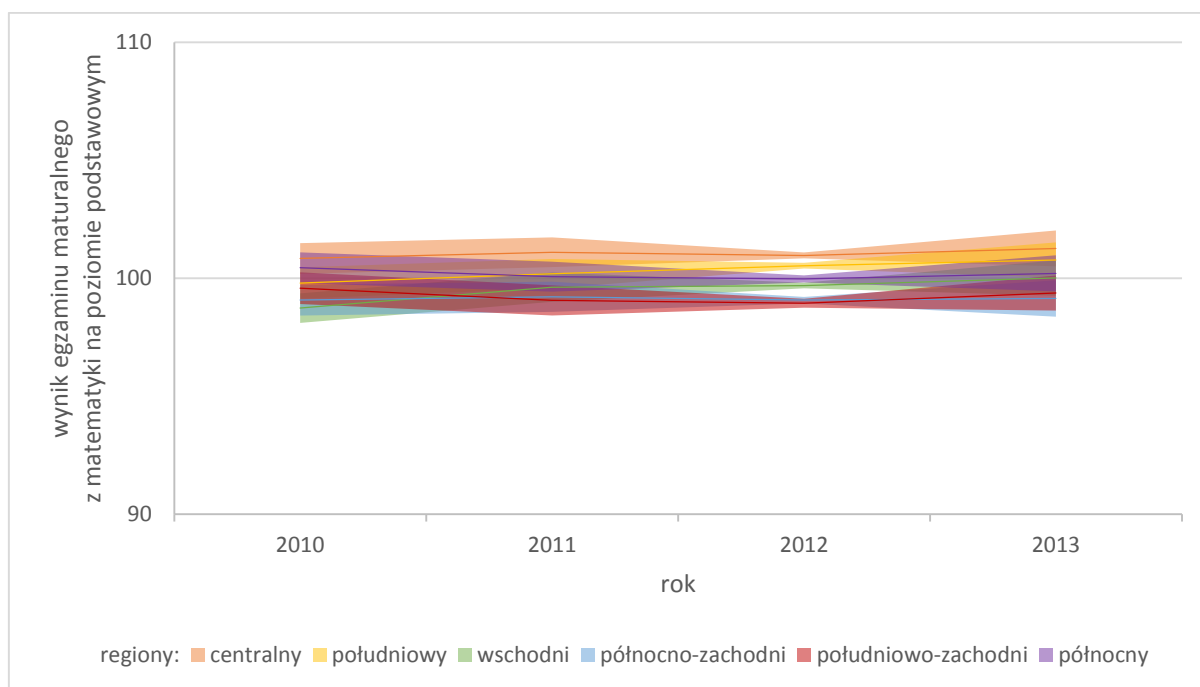


Rysunek 7.8. Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013 w podziale na regiony wg NTS

7.2.3. Egzamin maturalny

7.2.3.1. Matematyka

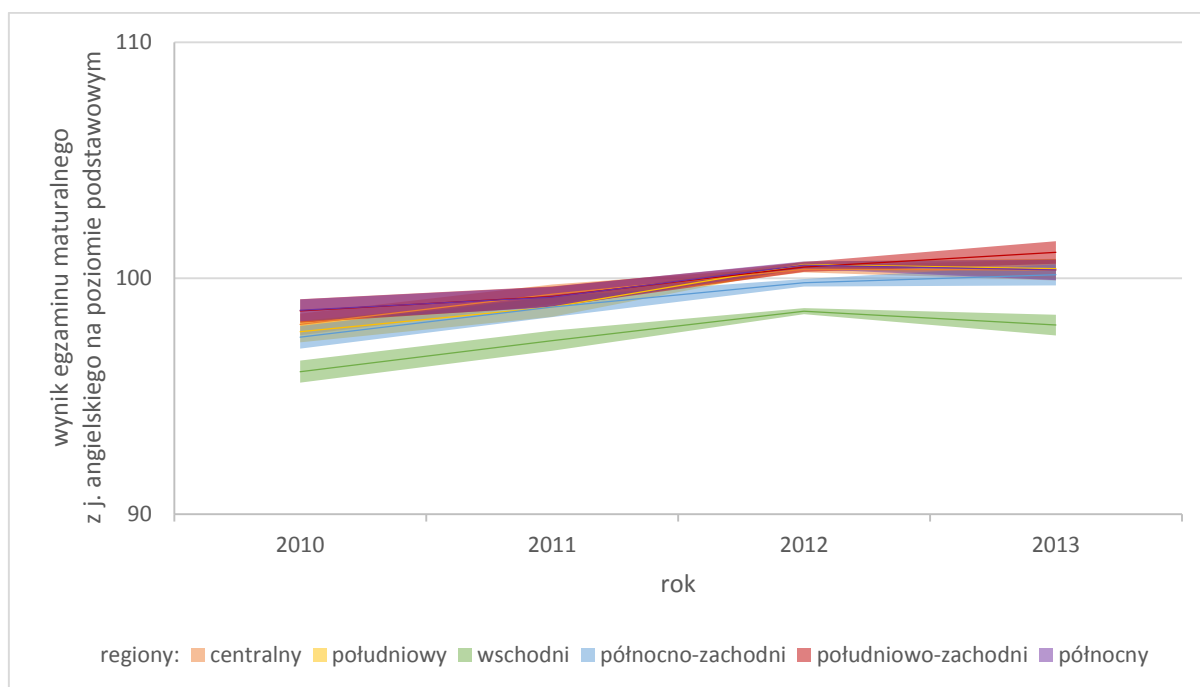
Porównywalne wyniki egzaminu maturalnego z matematyki nie wskazują na istnienie znaczących różnic pomiędzy poszczególnymi regionami kraju (zob. Rysunek 7.9). Dysponujemy danymi jedynie dla lat 2010–2013, lecz podział kraju pod względem wyników sprawdzianu i egzaminu gimnazjalnego można zaobserwować od początku ich istnienia (czyli 2002 roku). W przypadku matury z matematyki brak jest tak wyraźnych różnic, jak na wcześniejszych etapach edukacji. Obserwowane wcześniej różnice w toku kształcenia w szkołach ponadgimnazjalnych są zatem wyrównywane. Poziom umiejętności uczniów mierzonych na egzaminie maturalnym z matematyki nie jest zależny od regionu kraju, w którym uczniowie się kształcą. Jest to dobrą cechą systemu edukacyjnego, gdyż pozwala na zachowanie równych szans w procesie rekrutacji na uczelnie wyższe dla uczniów z różnych części kraju i nie ogranicza mobilności w dążeniu do uzyskania wyższego wykształcenia.



Rysunek 7.9. Porównywalne wyniki egzaminu maturalnego z matematyki w latach 2010–2013 w podziale na regiony wg NTS

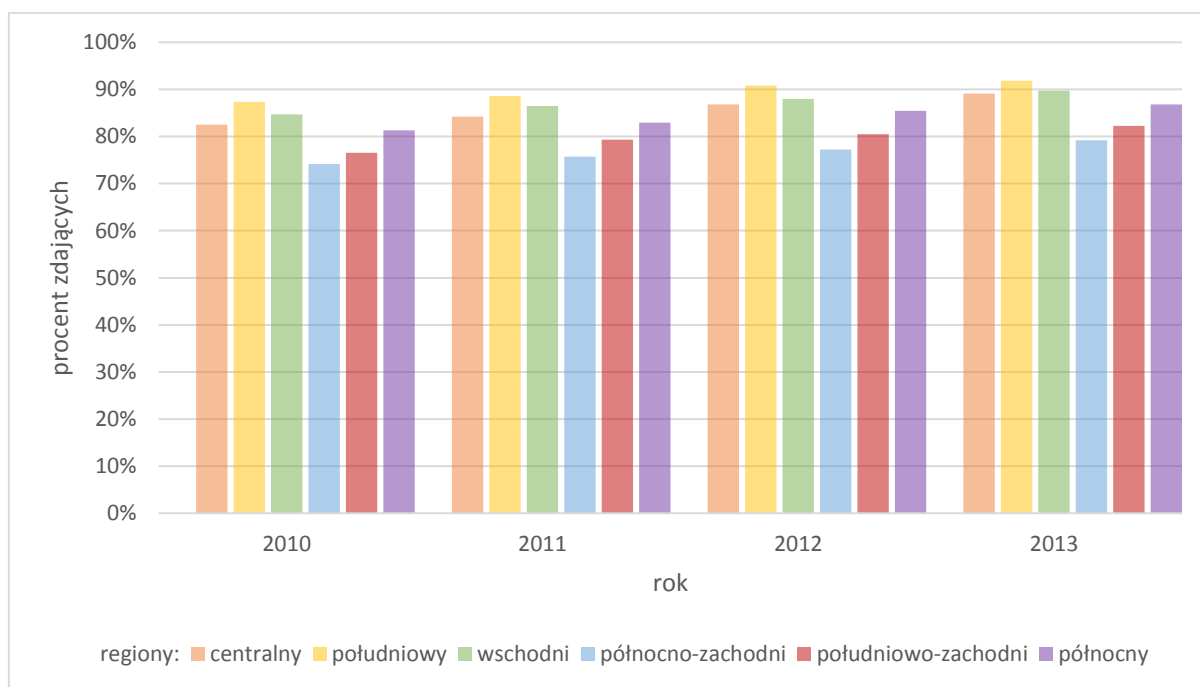
7.2.3.2. Język angielski

Analizując porównywalne wyniki egzaminu maturalnego z języka angielskiego w podziale na regiony kraju można zauważyć, że w regionie wschodnim uczniowie osiągalni wyniki niższe niż w pozostałych częściach kraju (zob. Rysunek 7.10). Różnica ta nie jest duża i wynosi około dwóch punktów dla poszczególnych lat, lecz jest ona istotna statystycznie w każdym analizowanym roku. Dodatkowo pomiędzy rokiem 2012 i 2013 wyniki w regionie wschodnim minimalnie spadają, a w pozostałych regionach nieznacznie rosną lub pozostają na tym samym poziomie.



Rysunek 7.10. Porównywalne wyniki egzaminu maturalnego z języka angielskiego w latach 2010–2013 w podziale na regiony wg NTS

Niższy poziom umiejętności z języka angielskiego w regionie wschodnim można by łączyć z bliskością krajów posługujących się językami wschodniosłowiańskimi, co może owocować stosunkowo wyższymi umiejętnościami z języka rosyjskiego. Gdy jednak przyjrzymy się wynikom w podziale na województwa, to okazuje się, że najniższe wyniki z języka angielskiego w skali kraju osiągają uczniowie z województwa świętokrzyskiego. Natomiast wyniki województw bezpośrednio sąsiadujących z Rosją, Białorusią, czy Ukrainą nie odbiegają tak bardzo od średniej ogólnopolskiej, co zdaje się przeczyć takiej próbie wyjaśnienia. Ponadto wybieralność języka angielskiego w regionie wschodnim nie różni się znacząco od wybieralności tego języka w regionie centralnym czy południowym (zob. Rysunek 7.11). Wyjaśnienie przyczyn zaobserwowanego zjawiska wymaga bliższego przyjrzenia się nauczaniu języków obcych w tej części Polski, a głównie w województwie świętokrzyskim. Przydatna byłaby też analiza wyników egzaminu gimnazjalnego z języka obcego, aby stwierdzić czy takie zróżnicowanie istnieje już na wcześniejszym etapie edukacyjnym.



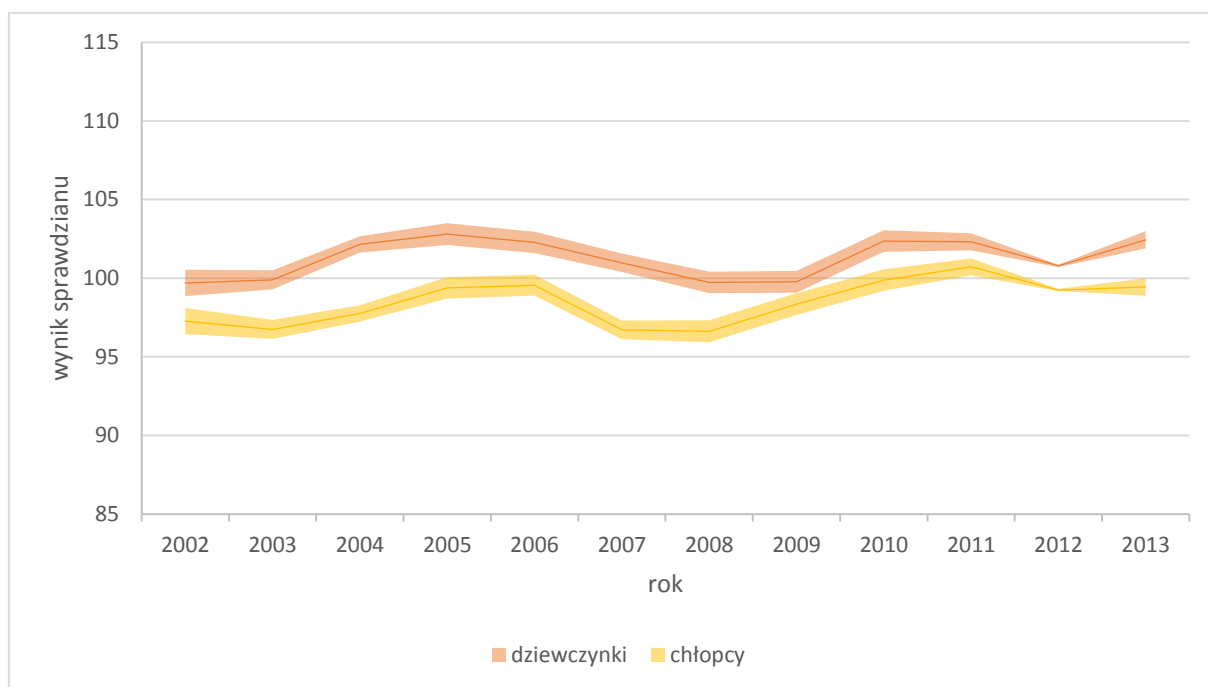
Rysunek 7.11. Wybieralność języka angielskiego na maturze (procent uczniów) w latach 2010–2013 w podziale na regiony

7.3. Płeć uczniów

Równy dostęp do edukacji dla obu płci jest kluczowym zagadnieniem dla polityki równościowej (zob. EACEA, 2010; World Economic Forum, 2013). Choć wskazuje się, że iloraz inteligencji kobiet nie różni się w sposób istotny statystycznie od ilorazu inteligencji mężczyzn, kobiety otrzymują zazwyczaj wyższe niż mężczyźni oceny szkolne (zob. Konarzewski, 1996). Tendencja ta ulega odwróceniu w przypadku standaryzowanych egzaminów zewnętrznych – mężczyźni uzyskują zazwyczaj wyższe wyniki (zob. Perkins, Kleiner, Roey i Brown, 2004; Nofle i Robins, 2007). Efekt ten określa się mianem niedoszacowania predykcji wyników egzaminacyjnych dla kobiet (*female underprediction effect*, FUE) i przeszacowania predykcji wyników egzaminacyjnych dla mężczyzn (zob. Hyde i Kling, 2001). Na dalszych etapach edukacyjnych pojawia się także rozbieżność w wynikach ze względu na przedmiot nauczania – chłopcy osiągają wyższe wyniki z matematyki, a dziewczęta w czytaniu i pisaniu (Willingham i Cole, 1997). Należy więc sprawdzić, czy tendencja ta jest zauważalna w polskim systemie egzaminów zewnętrznych.

7.3.1. Sprawdzian

Rysunek 7.12 obrazuje porównywalne wyniki egzaminacyjne w poszczególnych latach w grupach wyróżnionych ze względu na płeć. Różnica w wynikach egzaminacyjnych chłopców i dziewczynek utrzymuje się na stałym poziomie na przestrzeni czasu i wynosi średnio nieco poniżej trzech punktów. Dziewczynki uzyskują wyższe wyniki niż chłopcy. W latach 2009, 2011 i 2012 różnica jest nieco mniejsza – wynosi mniej niż dwa punkty. Można zatem dojść do wniosku, że w sprawdzianie nie mamy do czynienia z przeszacowaniem poziomu umiejętności dla chłopców – ich wyniki są w analizowanym okresie systematycznie niższe niż wyniki dziewczynek. Analizując wyniki sprawdzianu, należy jednak pamiętać o jego ponadprzedmiotowym charakterze. W związku z tym, że egzamin ten mierzył zarówno umiejętności czytania i pisania, jak i umiejętności matematyczne, ocena zróżnicowania wyników ze względu na poszczególne umiejętności w grupach chłopców i dziewcząt nie jest możliwa.

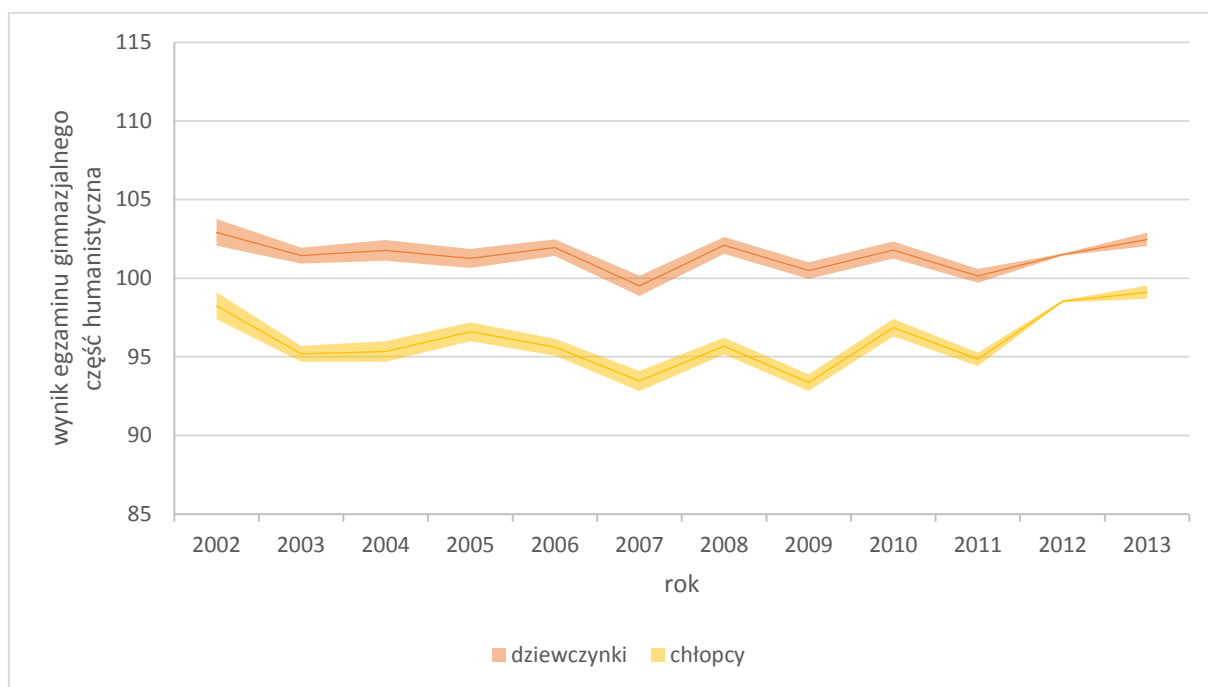


Rysunek 7.12. Porównywalne wyniki sprawdzianu w latach 2002–2013 w podziale na płeć uczniów

7.3.2. Egzamin gimnazjalny

7.3.2.1. Część humanistyczna

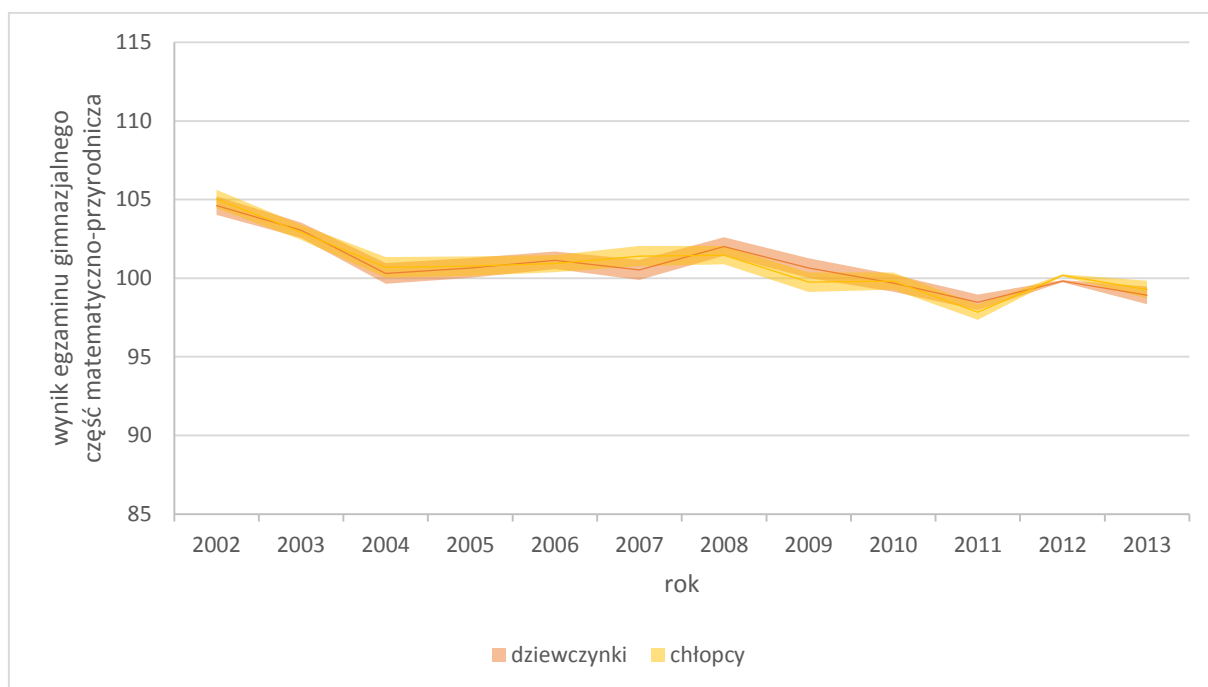
Różnica w wynikach części humanistycznej egzaminu gimnazjalnego pomiędzy chłopcami i dziewczynkami utrzymuje się na zbliżonym poziomie na przestrzeni czasu (zob. Rysunek 7.13). Obserwuje się podobną tendencję jak w przypadku sprawdzianu – we wszystkich latach dziewczynki uzyskują więcej punktów niż chłopcy. Zależność zaobserwowana w danych znajduje swoje odzwierciedlenie w literaturze – dziewczynki uzyskują wyższe wyniki dla przedmiotów humanistycznych (zob. von Schrader i Ansley, 2006; Skórska i Świst, 2014; Willingham i Cole, 1997). Różnica w wynikach pomiędzy dziewczynkami i chłopcami zmniejszyła się w ostatnich latach (2011–2013). W latach 2002–2011 różnice wynosiły między pięć a siedem punktów, a w latach 2012 i 2013 spadły do około trzech punktów, co może być pocieszające. Wskazuje się (OECD, 2009), że zróżnicowanie wyników dziewcząt i chłopców może być skorelowane z zainteresowaniem czytaniem. Indeks zainteresowania czytaniem w badaniach PISA jest we wszystkich badanych krajach zdecydowanie niższy u chłopców niż u dziewcząt (około pół odchylenia standardowego różnicy). Podobny wniosek zaobserwowano również w badaniach przeprowadzonych przez IBE – piętnastoletnie dziewczęta czytają więcej niż chłopcy (Zasacka, 2014).



Rysunek 7.13. Porównywalne wyniki części humanistycznej egzaminu gimnazjalnego w latach 2002–2013 w podziale na płeć uczniów

7.3.2.2. Część matematyczno-przyrodnicza

Różnice w wynikach dziewczynek i chłopców z części matematyczno-przyrodniczej egzaminu gimnazjalnego w poszczególnych latach są nieistotne statystycznie, co ilustruje Rysunek 7.14. Jest to sytuacja diametralnie inna niż w części humanistycznej oraz sprawdzianu, gdzie wyniki dziewcząt są zdecydowanie wyższe niż chłopców we wszystkich analizowanych latach. Wyniki egzaminu gimnazjalnego w części matematyczno-przyrodniczej nie potwierdzają jednak istnienia zjawiska niedoszacowania predykcji wyników egzaminacyjnych dla kobiet (zob. Hyde i Kling, 2001). Również w badaniach PISA różnice między chłopcami, a dziewczynkami w dziedzinie umiejętności matematycznych okazały się nieistotne statystycznie (OECD, 2014).

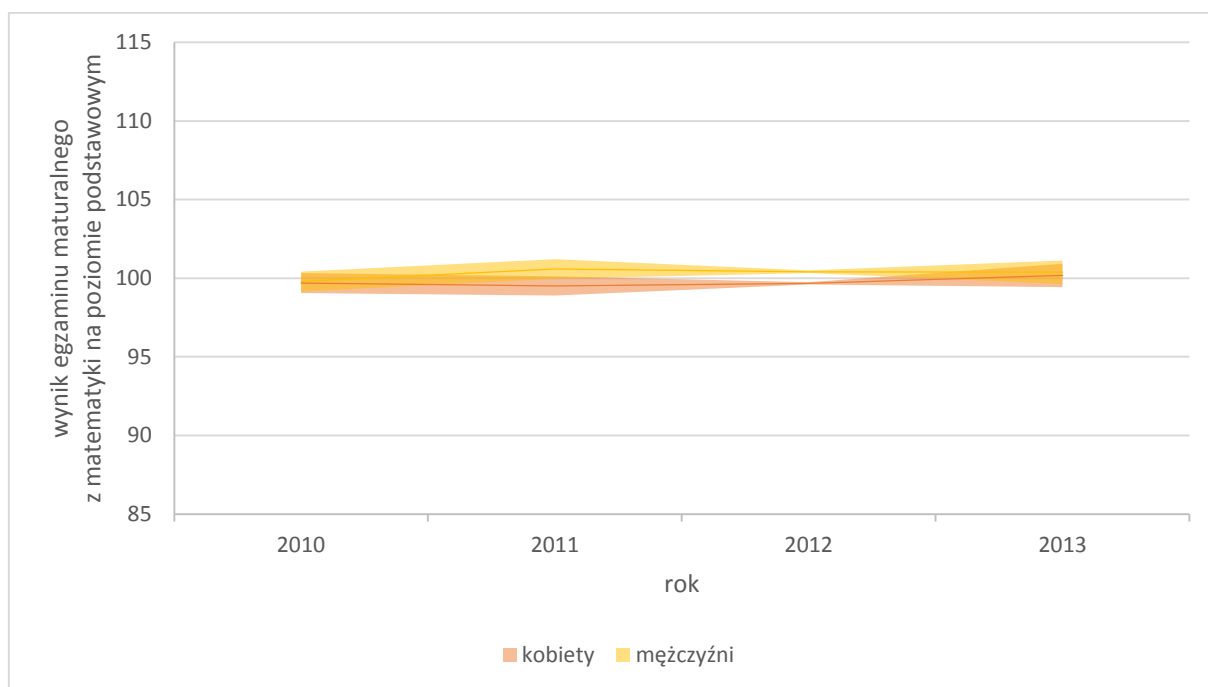


Rysunek 7.14. Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013 w podziale na płeć uczniów

7.3.3. Egzamin maturalny

7.3.3.1. Matematyka

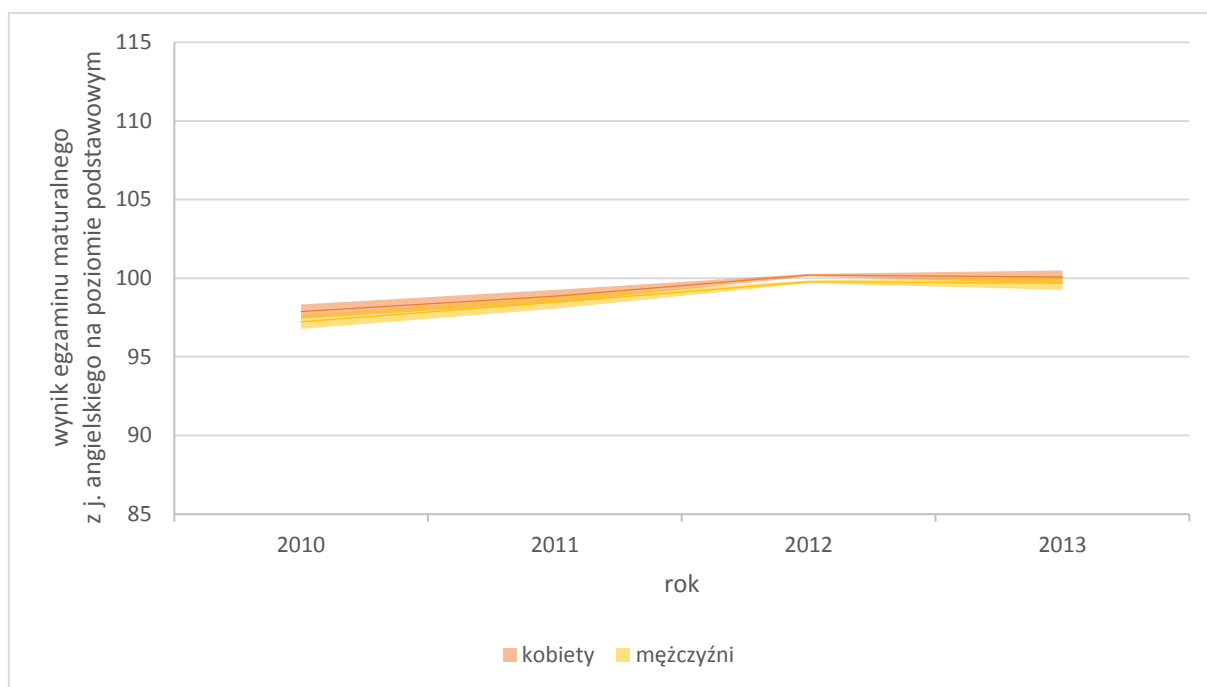
Płeć uczniów nie różnicuje znacząco osiąganych na egzaminie maturalnym z matematyki wyników (zob. Rysunek 7.15). Jest to zgodne z tendencją obserwowalną w części matematyczno-przyrodniczej egzaminu gimnazjalnego. W latach 2011–2012 mężczyźni osiągnęli co prawda nieco wyższe wyniki niż kobiety, lecz ich różnica wynosi około jednego punktu, a ponadto w roku 2011 jest ona nieistotna statystycznie.



Rysunek 7.15. Porównywalne wyniki egzaminu maturalnego z matematyki w latach 2010–2013 w podziale na płeć uczniów

7.3.3.2. Język angielski

Analiza porównywalnych wyników egzaminu maturalnego z języka angielskiego w podziale na płeć uczniów nie wskazuje na różnice w poziomie umiejętności kobiet i mężczyzn (zob. Rysunek 7.16). W okresie od 2010 do 2013 roku obserwuje się minimalny wzrost wyników dla obydwu grup odzwierciedlający trend krajowy. Przyjmując zatem założenie, że obydwie płcie nie różnią się między sobą pod kątem sprawdzanych umiejętności, można wnioskować, że egzamin maturalny z języka angielskiego nie dawał przewagi żadnej z nich.



Rysunek 7.16. Porównywalne wyniki egzaminu maturalnego z języka angielskiego w latach 2010–2013 w podziale na płeć uczniów

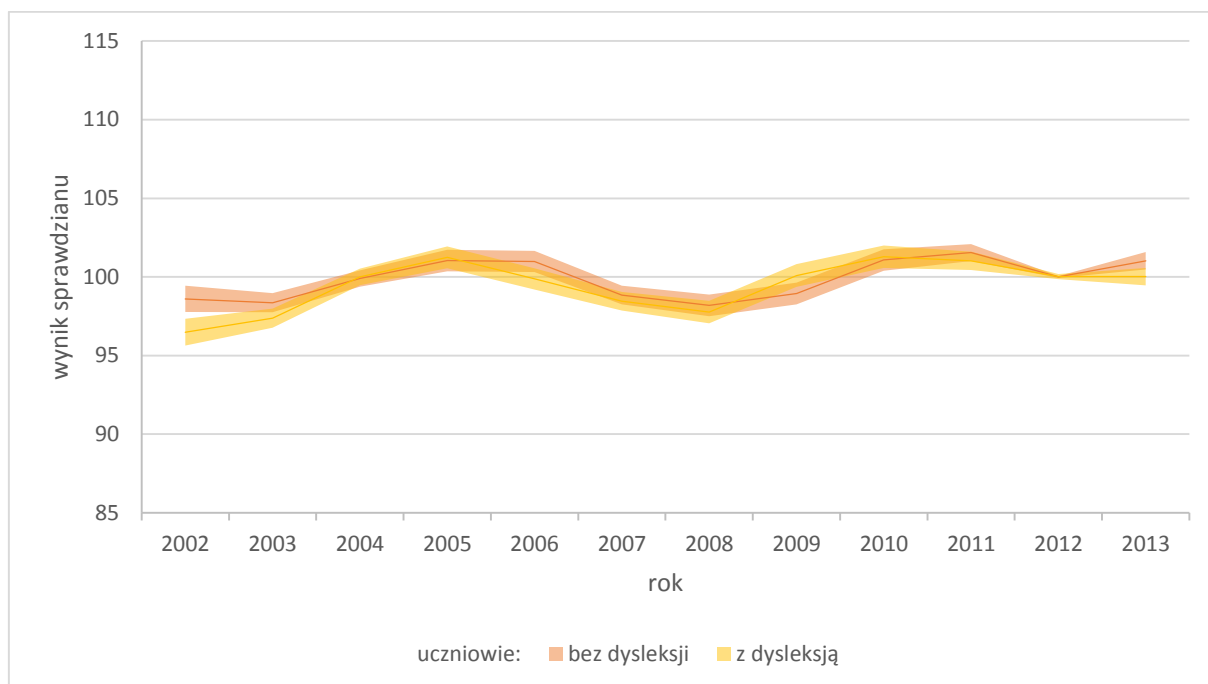
7.4. Diagnoza dysleksji

W związku z tym, że uczniowie ze zdiagnozowaną dysleksją rozwojową mogą korzystać ze specjalnych dostosowań podczas egzaminów, należy zbadać efekty oraz skuteczność takiego postępowania. Dysleksja jest definiowana przez Światową Federację Neurologów jako „zaburzenia manifestujące się trudnościami w nauce czytania, mimo stosowania obowiązujących metod nauczania, normalnej inteligencji i sprzyjających warunków społeczno-kulturowych. Jest spowodowana zaburzeniami podstawowych funkcji poznawczych” (Bogdanowicz, 1994, s. 32). Wskazuje się (CKE, 2009), że podczas sytuacji stresującej (którą z pewnością jest egzamin zewnętrzny), uczniowie z diagnozą specyficznych trudności w uczeniu się (np. z dysleksją rozwojową) mogą reagować zaburzeniami uwagi i pamięci. Tacy uczniowie charakteryzują się również wolniejszym tempem pracy, trudnościami z czytaniem oraz mało czytelnym pismem zawierającym błędy. Wprowadzone dla nich dostosowania zostały omówione w częściach poświęconych konkretnym egzaminom.

7.4.1. Sprawdzian

Dostosowanie wprowadzone dla uczniów piszących sprawdzian obejmuje m. in. wydłużenie czasu pisania egzaminu do 50% (zgodnie z zaleceniami zawartymi w orzeczeniu wydanym przez poradnię psychologiczno-pedagogiczną), odczytywanie zadań przez członków komisji, czy specjalnie dostosowane kryteria oceniania. Uczniowie ci nie muszą też przenosić odpowiedzi na karty skanerowe – robią to za nich egzaminatorzy (CKE, 2009). Porównywalne wyniki egzaminacyjne w poszczególnych latach dla grup uczniów z diagnozą dysleksji i bez tej diagnozy ilustruje Rysunek 7.17. Wyniki egzaminacyjne uczniów z diagnozą dysleksji i bez diagnozy dysleksji przeważnie nie różnią się od siebie w sposób istotny statystycznie. Taki stan rzeczy może być spowodowany wprowadzeniem wspomnianych dostosowań i stworzeniem uczniom z dysleksją rozwojową dogodnych warunków do zaprezentowania zdobytych umiejętności oraz wiedzy. Jedynie w roku 2002 można zaobserwować niewielkie różnice między wynikami tych dwóch grup. Różnicę tę należy jednak interpretować bardzo

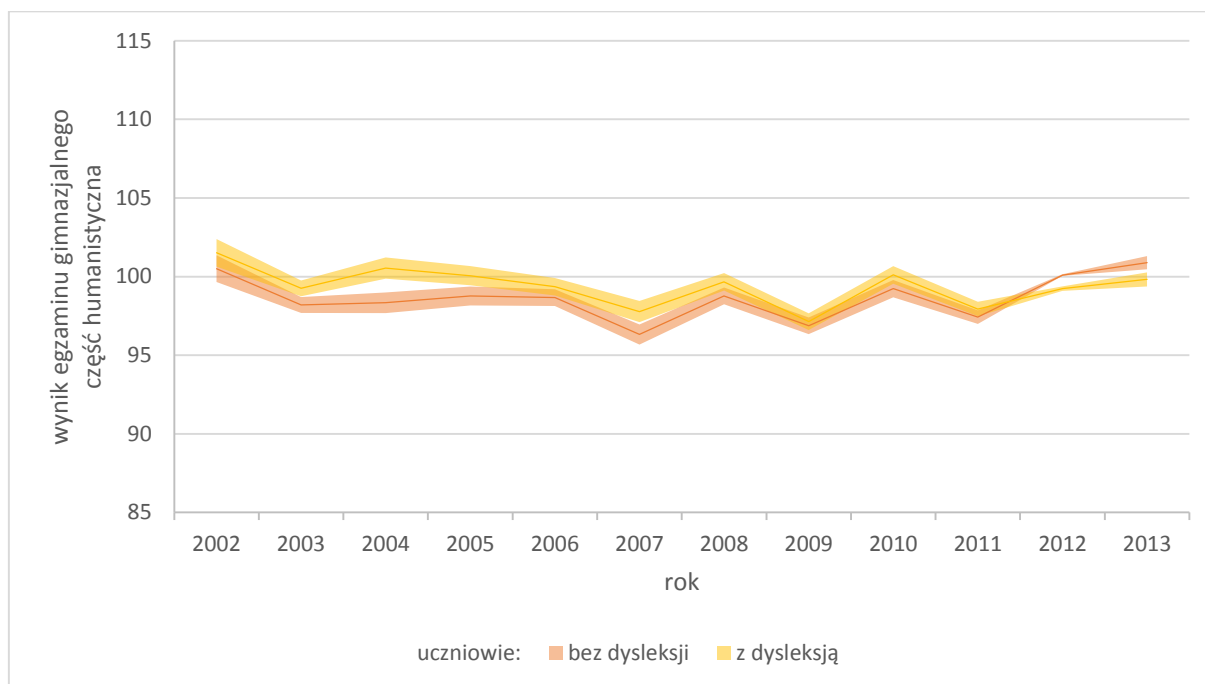
ostrożnie, z uwagi na to, że była to pierwsza edycja sprawdzianu. Można więc stwierdzić, że wyniki sprawdzianu są stabilne, ze względu na brak różnic w wynikach uczniów z diagnozą dysleksji oraz bez tej diagnozy.



Rysunek 7.17. Porównywalne wyniki sprawdzianu w latach 2002–2013 w podziale na grupy bez dysleksji rozwojowej i z dysleksją rozwojową

7.4.2. Egzamin gimnazjalny

7.4.2.1. Część humanistyczna



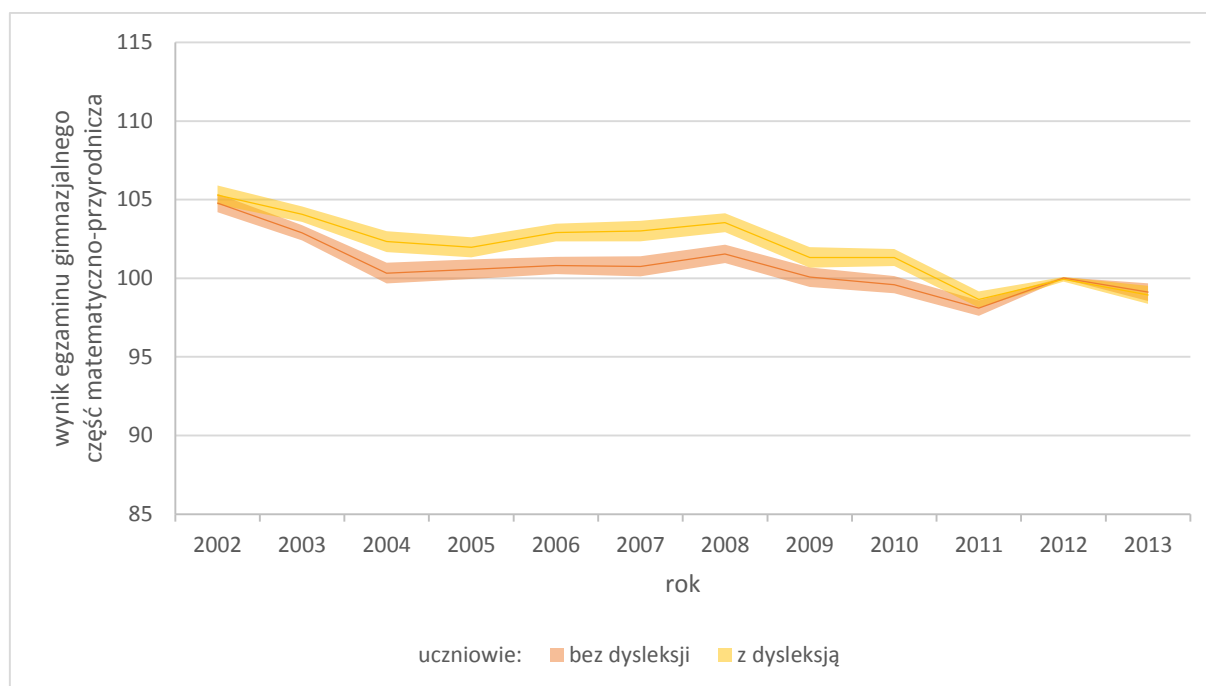
Rysunek 7.18. Porównywalne wyniki części humanistycznej egzaminu gimnazjalnego w latach 2002–2013 w podziale na grupy uczniów bez dysleksji rozwojowej i z dysleksją rozwojową

Podobnie jak w przypadku sprawdzianu, uczniowie z diagnozą dysleksji mogą korzystać z szeregu dostosowań podczas pisania egzaminu gimnazjalnego, np. zaznaczania odpowiedzi w zadaniach zamkniętych bez konieczności przenoszenia ich na kartę, wydłużenia czasu pisania, korzystania z pomocy nauczyciela wspomagającego, czy zastosowania szczególnych kryteriów podczas oceny zadań otwartych (CKE, 2013). W początkowych latach funkcjonowania egzaminu gimnazjalnego, częściej niż w latach późniejszych, wyższe wyniki uzyskiwali uczniowie ze zdiagnozowaną dysleksją (zob. Rysunek 7.18). Różnice te nie mają jednak systematycznego charakteru. Począwszy od 2012 roku to uczniowie bez zdiagnozowanej dysleksji rozwojowej osiągają średnio minimalnie wyższe wyniki niż uczniowie z dysleksją. Utrzymanie tego wyniku w następnych latach powinno świadczyć o stabilności konstrukcji egzaminu i kryteriów oceniania względem zjawiska dysleksji rozwojowej. Jednakże ze względu na kompensacyjną naturę dostosowań dla uczniów z diagnozą dysleksji, wskazane byłoby, aby obydwie te grupy osiągały takie same wyniki, jak miało to miejsce we wcześniejszych latach.

7.4.2.2. Część matematyczno-przyrodnicza

Uczniowie z diagnozą dysleksji rozwojowej osiągalni do roku 2010 wyższe wyniki z części matematyczno-przyrodniczej tego egzaminu niż uczniowie bez takiej diagnozy (z wyjątkiem pierwszej edycji egzaminu gimnazjalnego w roku 2002). Od roku 2011 nie obserwuje się statystycznie istotnych różnic w wynikach obydwu tych grup. Można wnioskować, że dostosowania dla uczniów z diagnozą dysleksji mogły stwarzać tej grupie niewielką przewagę w części matematyczno-przyrodniczej (prawdopodobnie głównie ze względu na wydłużony czas na udzielanie odpowiedzi). Przewaga ta była większa niż w części humanistycznej egzaminu gimnazjalnego. Może być to spowodowane tym, że zadania otwarte z matematyki i przyrody mogą sprawiać uczniom z diagnozą dysleksji mniej trudności niż zadania z części humanistycznej. W zdecydowanej większości nie wymagają one bowiem budowania długich

wypowiedzi pisemnych. Jeśli uznać, że celem wprowadzenia dostosowań dla uczniów z dysleksją rozwojową jest wyrównanie ich szans w stosunku do uczniów bez takiej diagnozy, to ich efektem powinien być raczej brak różnic w średnich wynikach tych grup, który możemy obserwować od roku 2011. Omawiane wyniki ilustruje Rysunek 7.19.



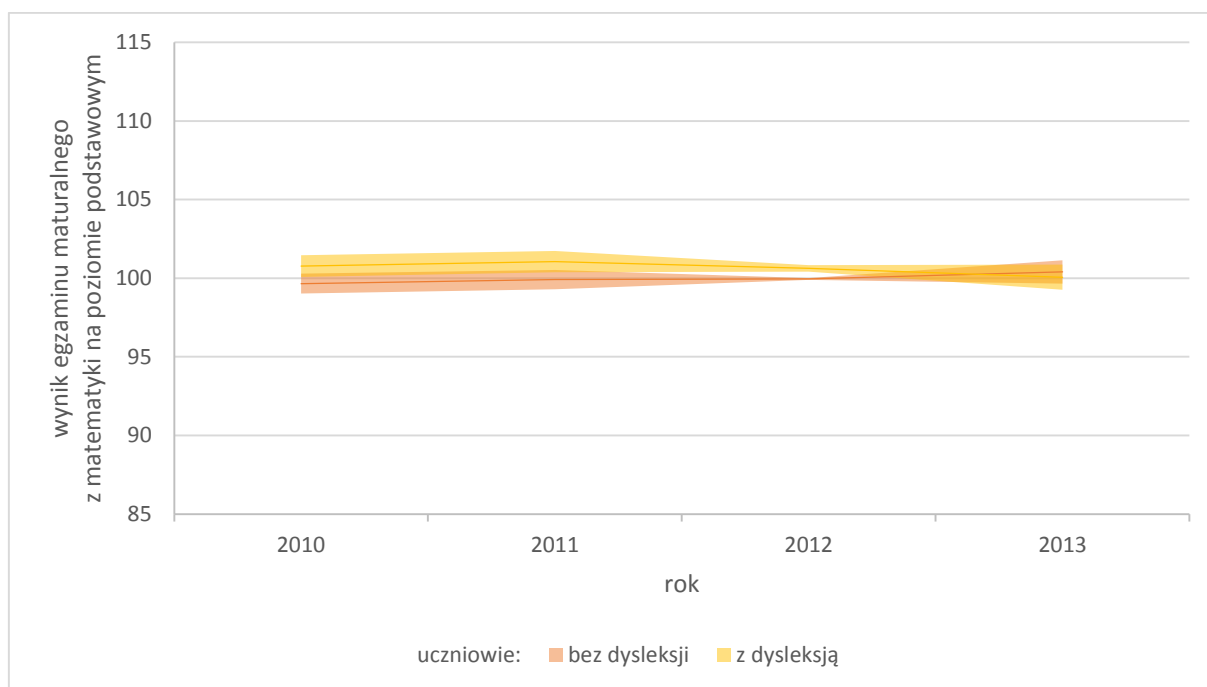
Rysunek 7.19. Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013 w podziale na grupy uczniów bez dysleksji rozwojowej i z dysleksją rozwojową

7.4.3. Egzamin maturalny

7.4.3.1. Matematyka

Dostosowania dla uczniów z dysleksją rozwojową na egzaminie maturalnym obejmowały w analizowanym okresie wyłącznie zastosowanie szczegółowych kryteriów w ocenianiu³⁵. W przeciwieństwie do wcześniejszych etapów edukacyjnych uczniowie ci nie dysponowali np. dodatkowym czasem na rozwiązanie zadań. Jak wskazuje Rysunek 7.20, nie ma znaczących różnic pomiędzy wynikami egzaminu maturalnego z matematyki w grupach uczniów z diagnozą dysleksji oraz nie posiadających takiej diagnozy. Trend ten odzwierciedla wyniki w części matematyczno-przyrodniczej egzaminu gimnazjalnego (od 2011 roku). Jedynie dla roku 2012 różnica pomiędzy tymi grupami uczniów jest istotna statystycznie, lecz wynosi mniej niż jeden punkt.

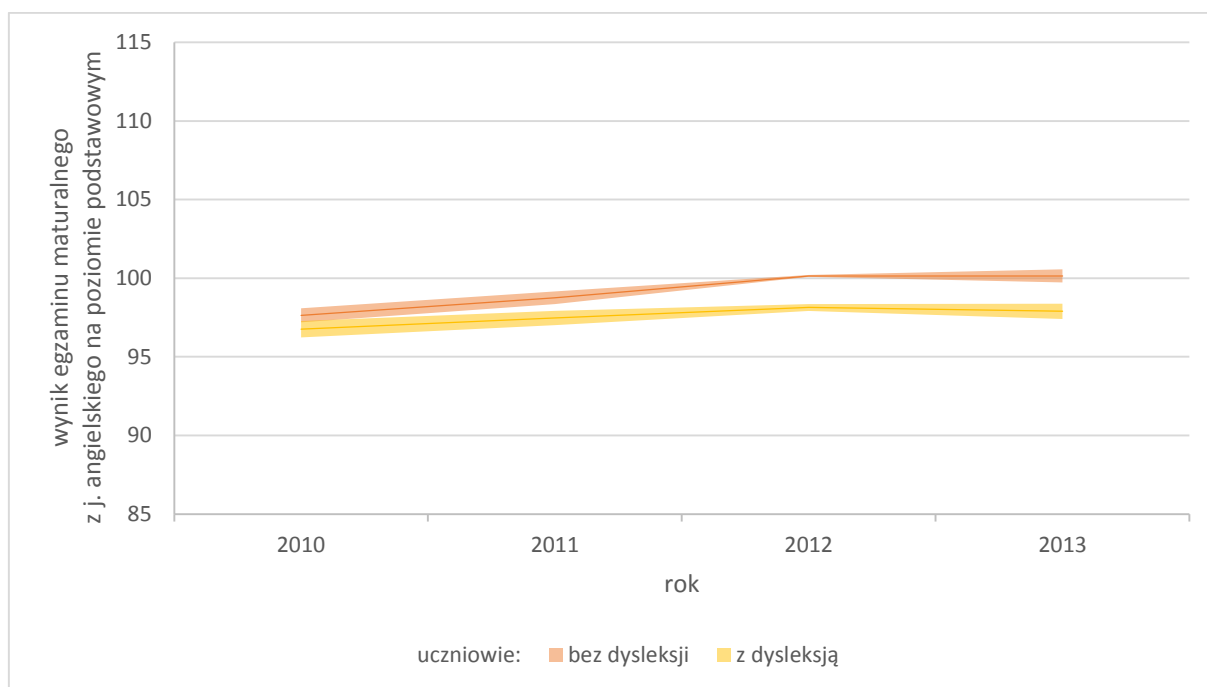
³⁵ Pomijając możliwość pisania pracy na komputerze w szczególnych przypadkach głębokiego zaburzenia grafii.



Rysunek 7.20. Porównywalne wyniki egzaminu maturalnego z matematyki w latach 2010–2013 w podziale na grupy uczniów bez dysleksji rozwojowej i z dysleksją rozwojową

7.4.3.2. Język angielski

W przeciwieństwie do egzaminu maturalnego z matematyki, dla języka angielskiego możemy zaobserwować różnice porównywalnych wyników egzaminacyjnych pomiędzy uczniami z dysleksją rozwojową oraz bez niej. Choć w 2010 roku różnice te były minimalne i nieistotne statystycznie, to w następnych latach systematycznie rosły (zob. Rysunek 7.21). W roku 2013 uczniowie bez dysleksji rozwojowej osiągnęli średnio dwa punkty więcej niż ci, którzy posiadali taką diagnozę. Może to świadczyć o tym, że dostosowania dla uczniów z dysleksją stosowane podczas egzaminu z języka angielskiego nie są wystarczające do wyrównania szans tej grupy w stosunku do uczniów bez zaburzeń. Ze względu na następujące zjawiska – niemożność identyfikacji i rozróżniania fonemów, niezbędnych podczas rozumienia ze słuchu oraz spowolnienie funkcjonowania pamięci długotrwałej, utrudniające przyswojenie materiału (Zawadzka, 2010), dysleksja może przysparzać szczególnych problemów podczas nauki języków obcych. Ich przyczyną może być też trudność w opanowaniu języka ojczystego, również spowodowana niską świadomością fonologiczną (Nijakowska, 2009).



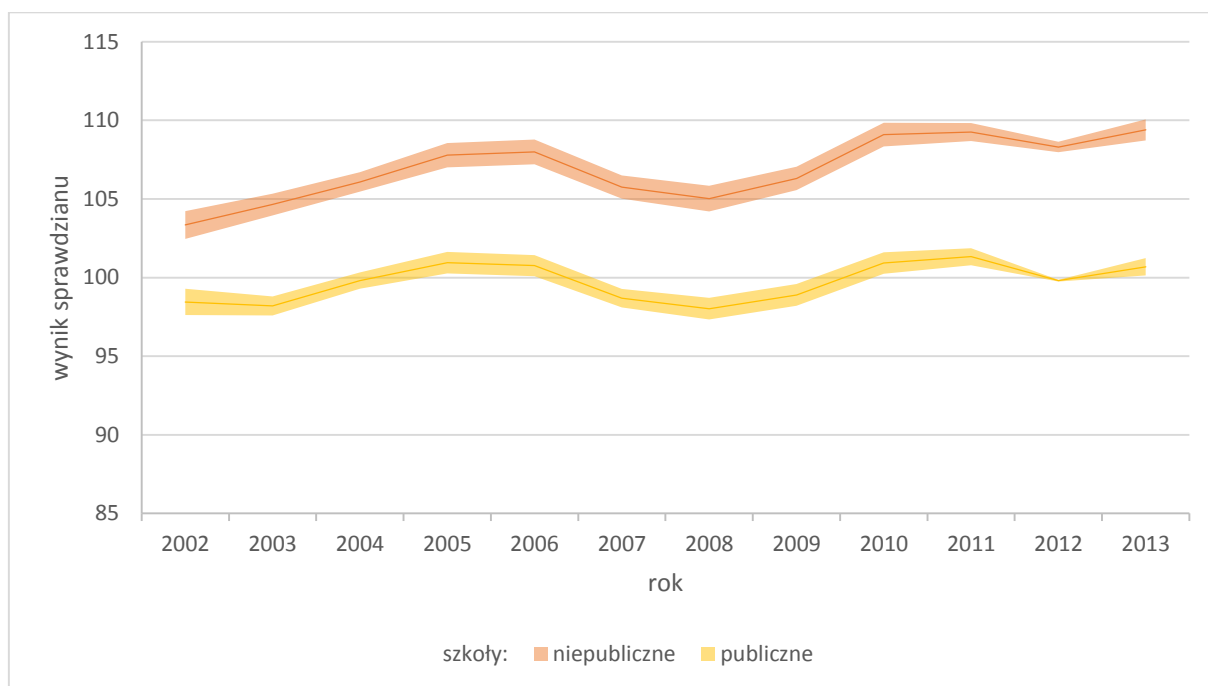
Rysunek 7.21. Porównywalne wyniki egzaminu maturalnego z języka angielskiego w latach 2010–2013 w podziale na grupy uczniów bez dysleksji rozwojowej i z dysleksją rozwojową

7.5. Typ szkoły

Bez wątpienia na wyniki osiągane przez uczniów ma wpływ szkoła do której uczęszczają. W szczególności uczęszczanie do szkoły publicznej lub niepublicznej może różnicować wyniki egzaminacyjne uczniów. Różnice w wynikach uzyskiwanych przez uczniów uczęszczających do szkół publicznych i niepublicznych mogą świadczyć o coraz większym rozwarstwieniu społecznym. Jak wskazują Putkiewicz i Wiłkomirska (2004), szkoły niepubliczne prowadzą specyficzny rodzaj kształcenia: w mniejszych oddziałach, z naciskiem na rozwój indywidualnych zainteresowań uczniów czy też możliwością zaimplementowania autorskich programów nauczania. Według Piekarczyk (2014) szkoły niepubliczne mają zapewnić lepszy rozwój zdolności i zainteresowań, gwarantować lepsze traktowanie uczniów i większe bezpieczeństwo. Dodatkowo, uczniowie uczęszczający do szkół publicznych i niepublicznych mogą różnić się ze względu na status społeczno-ekonomiczny (SES) rodziny. Starsze badania (Jung-Miklaszewska i Rusakowska, 1995; Zahorska-Bugaj, 1994), wskazują, że rodzice dzieci uczęszczających do szkół niepublicznych to zazwyczaj specjaliści, przedstawiciele wolnych zawodów, czy przedsiębiorcy, posiadający wyższe wykształcenie. Dysproporcja pomiędzy szkołami publicznymi i niepublicznymi, związana z wykształceniem oraz statusem społeczno-ekonomicznym rodziców, widoczna jest zwłaszcza w przypadku szkół podstawowych³⁶, na etapie ponadpodstawowym była ona mniejsza. Niestety brakuje aktualnych danych pozwalających na analizę struktury społecznej szkół publicznych i niepublicznych dla analizowanego okresu.

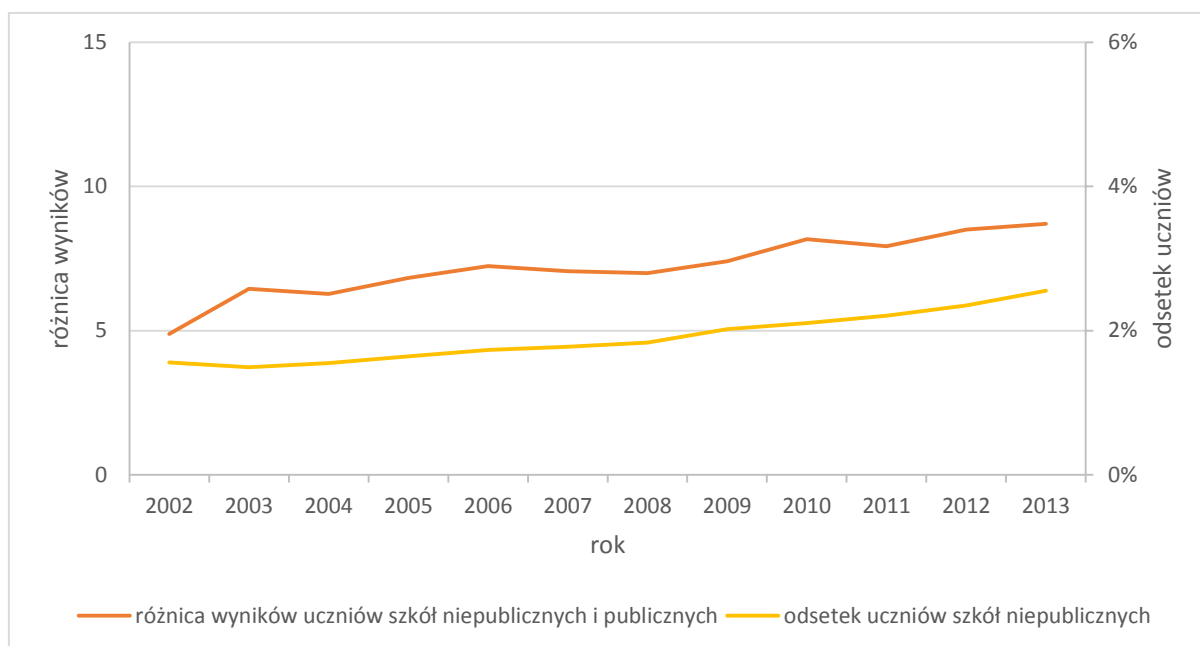
³⁶ Sformułowanie szkoły podstawowe i ponadpodstawowe dotyczy tutaj dwustopniowego systemu szkolnictwa przed reformą systemu oświaty realizowaną w Polsce od 1 września 1999, która doprowadziła do przekształcenia obowiązującego od 1968 r. dwustopniowego systemu szkolnictwa w strukturę trzystopniową.

7.5.1. Sprawdzian



Rysunek 7.22. Porównywalne wyniki sprawdzianu w latach 2002–2013 w podziale na typ szkoły (niepubliczna vs publiczna)

Dla wszystkich analizowanych lat porównywalne wyniki sprawdzianu dla szkół niepublicznych są wyższe niż wyniki dla szkół publicznych (zob. Rysunek 7.22). Różnica ta z roku na rok stawała się większa – w roku 2003 wynosiła 6,4 punktu, a w roku 2013 już 8,7 punktu. W roku 2002 nie udało się określić typu szkoły dla prawie 7% uczniów. Wyniki z tego roku należy traktować ostrożnie, nie wiadomo bowiem, czy braki te mają losowy charakter. W miarę upływu lat zmieniały się proporcje uczniów uczęszczających do szkół niepublicznych i publicznych – odsetek uczniów szkół niepublicznych wynosił w 2003 roku 1,5, a w roku 2013 2,6 (zob. Rysunek 7.23).

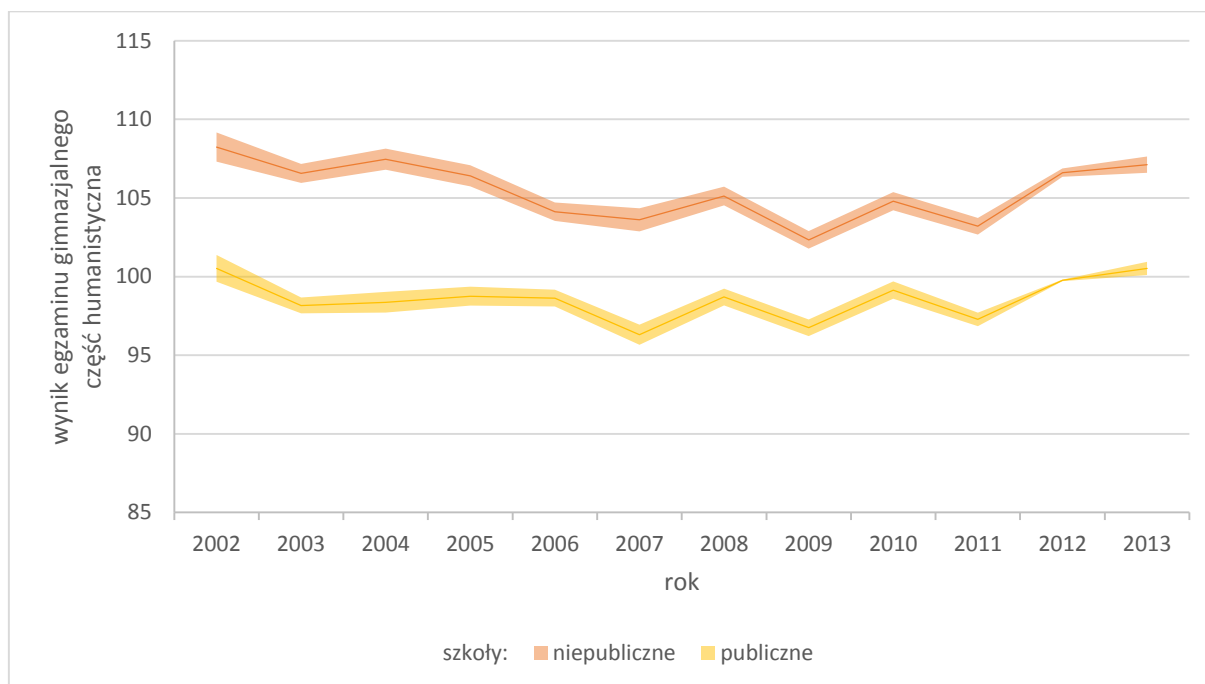


Rysunek 7.23. Różnica porównywalnych wyników sprawdzianu oraz odsetek uczniów szkół niepublicznych w latach 2002–2013

Herczyński i Sobotka (2014) w raporcie dotyczącym zmian w sieci szkół podstawowych i gimnazjów w latach 2007–2012 opisują zjawisko przejmowania przez stowarzyszenia małych szkół gminnych, co często wiąże się ze zmianą statusu szkoły z publicznej na niepubliczną. Wzrost liczby uczniów ze szkół niepublicznych w stosunku do publicznych jest także powodowany przez powstawanie szkół, gdzie organem prowadzącym są podmioty komercyjne. Jest to jednak zjawisko dużo rzadsze niż przekazywanie szkół przez gminy. Pomimo coraz większej różnicy w średnich wynikach sprawdzianu pomiędzy uczniami ze szkół niepublicznych i publicznych, obydwo typach szkół obserwujemy podobny trend. Prawdopodobnie sprawdzian nie zmienia się zatem na korzyść lub niekorzyść uczniów z któregoś typu szkoły, a w szkołach niepublicznych osiąga się wyższą efektywność kształcenia. Biorąc pod uwagę przejmowanie szkół przez stowarzyszenia, można uznać, że tego typu organy prowadzące osiągają wyższe wyniki kształcenia niż gminy. Weryfikacja tej hipotezy powinna jednak nastąpić przy użyciu danych dotyczących konkretnych szkół, w których doszło do zmiany organu prowadzącego.

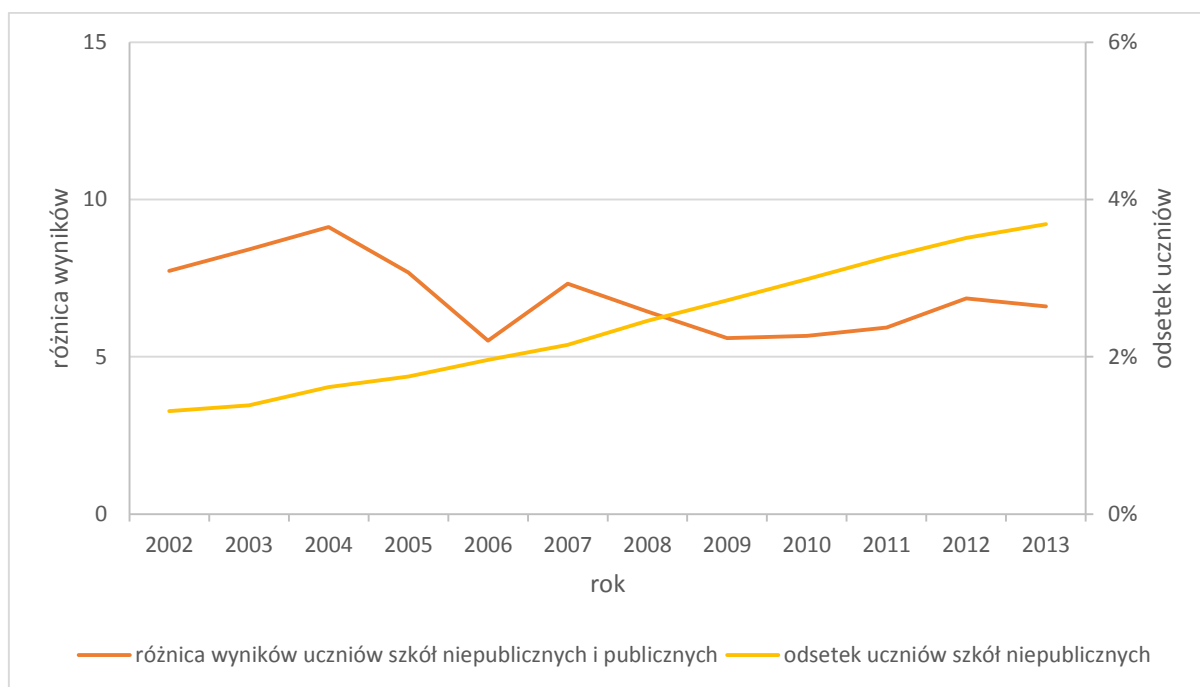
7.5.2. Egzamin gimnazjalny

7.5.2.1. Część humanistyczna



Rysunek 7.24. Porównywalne wyniki części humanistycznej egzaminu gimnazjalnego w latach 2002–2013 w podziale na typ szkoły (niepubliczna vs publiczna)

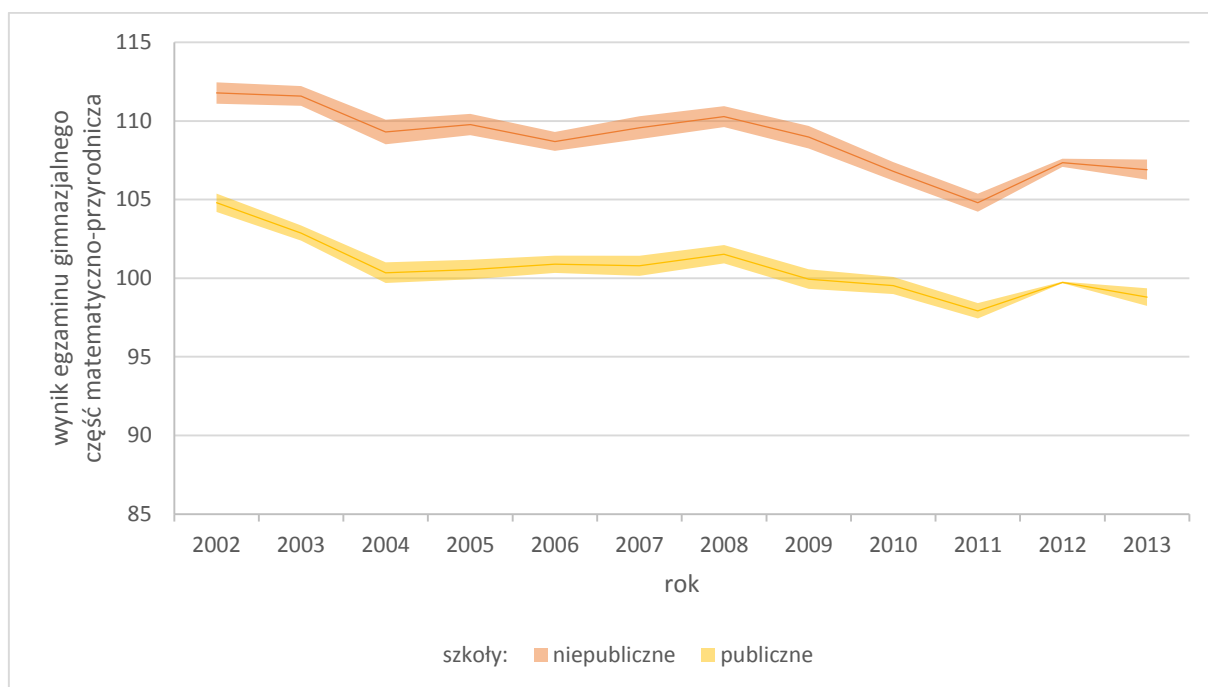
Podobnie jak w przypadku sprawdzianu, uczniowie gimnazjów niepublicznych uzyskali w latach 2010–2013 wyższe wyniki z części humanistycznej egzaminu gimnazjalnego niż uczniowie z gimnazjów publicznych. Różnica ta jest wyraźna w przypadku wszystkich analizowanych lat, lecz odwrotnie niż w przypadku sprawdzianu, od 2004 do 2009 roku można zaobserwować zmniejszenie się tych różnic, z pojedynczym wzrostem w roku 2007 (zob. Rysunek 7.24). W całym okresie od 2004 do 2013 roku różnica średnich wyników pomiędzy omawianymi typami szkół zmalała z 8,4 do 6,6 punktu, jednakże od 2009 roku, kiedy była ona najmniejsza, nastąpił jej niewielki wzrost (o jeden punkt). Trudno dopatrywać się w otrzymanych wynikach analiz systematyczności. Być może przyjrzenie się efektywności nauczania (np. EWD) w podziale na typy szkół mogłoby rzucić nieco światła na zaobserwowane zależności. Podobnie jak w przypadku szkół podstawowych, systematycznie wzrasta odsetek uczniów z gimnazjów niepublicznych (zob. Rysunek 7.25). Herczyński i Sobotka (2014) wskazują, że inaczej niż w przypadku szkół podstawowych, przekazywanie przez gminy w latach 2007–2012 gimnazjów innym organom prowadzącym było zjawiskiem rzadkim. W tym okresie otwarto jednak sporo nowych gimnazjów, których organem prowadzącym był inny podmiot niż gmina. Jeśli ich absolwenci osiągnęli wyższe niż przeciętne wyniki na egzaminie, to mogłoby to tłumaczyć rosnące wyniki uczniów szkół niepublicznych w ostatnich latach analizowanego okresu.



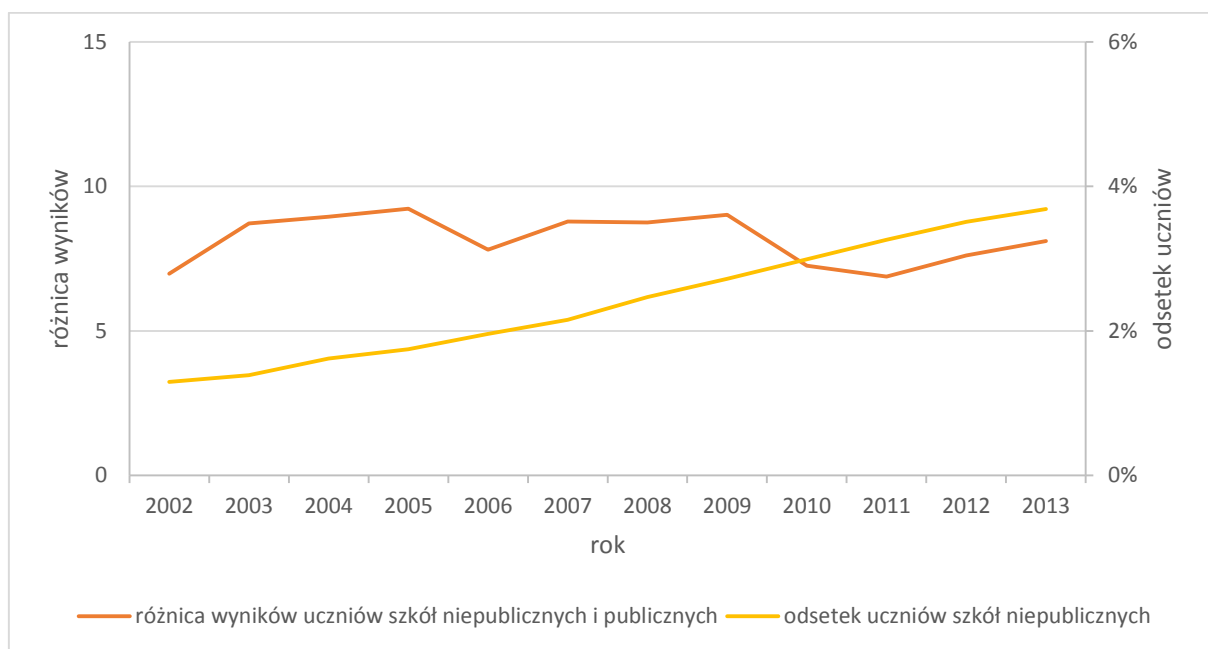
Rysunek 7.25. Różnica porównywalnych wyników części humanistycznej egzaminu gimnazjalnego oraz odsetek uczniów szkół niepublicznych w latach 2002–2013

7.5.2.2. Część matematyczno-przyrodnicza

W przypadku porównywalnych wyników części matematyczno-przyrodniczej egzaminu gimnazjalnego w każdym z analizowanych lat uczniowie gimnazjów niepublicznych również osiągnęli średnio wyższe wyniki niż uczniowie gimnazjów publicznych (zob. Rysunek 7.26). Różnica w wynikach uczniów z tych typów szkół waha się w granicach od siedmiu do dziewięciu punktów (zob. Rysunek 7.27). Podobnie jak w części humanistycznej egzaminu gimnazjalnego, zmianom tej różnicy brak systematycznego charakteru.



Rysunek 7.26. Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013 w podziale na typ szkoły (niepubliczna vs publiczna)



Rysunek 7.27. Różnica porównywalnych wyników części matematyczno-przyrodniczej egzaminu gimnazjalnego oraz odsetek uczniów szkół niepublicznych w latach 2002–2013

7.5.3. Egzamin maturalny

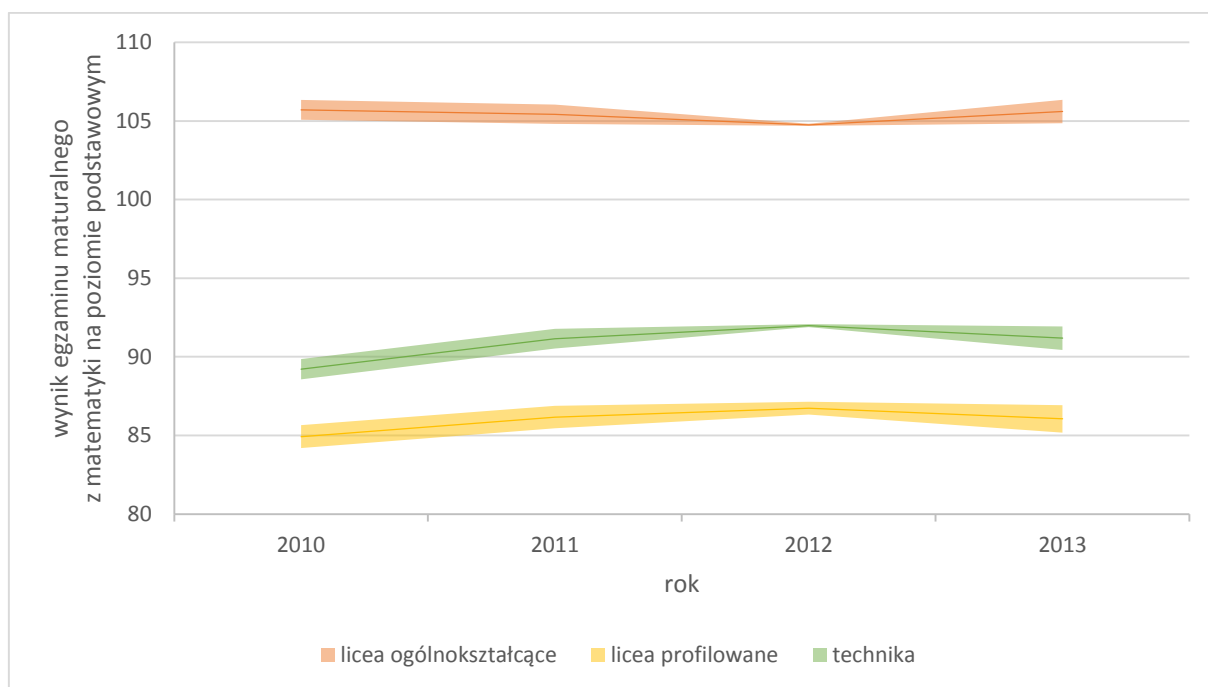
Wyniki osiągane przez uczniów szkół ponadgimnazjalnych trudno analizować w podziale na szkoły publiczne i niepubliczne. O wiele bardziej istotną cechą omawianych szkół jest to, czy jest to liceum ogólnokształcące, profilowane czy też technikum. Spośród uczniów szkół tego rodzaju zdecydowana większość (96–97% w zależności od roku) uczy się w szkołach publicznych. Jeśli uwzględnimy podział

na licea i technika, to liczebność niektórych grup będzie bardzo mała, co prowadzi do niezbyt rzetelnych porównań. Z podobnymi problemami mamy do czynienia w przypadku rodzaju gminy, w której położona jest szkoła. Większość szkół ponadgimnazjalnych zlokalizowana jest w gminach miejskich lub miejsko-wiejskich. Do szkół zlokalizowanych na terenie gmin wiejskich uczęszczało zaledwie cztery procent uczniów z opisywanych liceów i techników. Ponadto procentowy udział techników na wsi jest zdecydowanie wyższy w porównaniu z liceami. Według danych GUS dla roku 2012/2013 na wsiach znajdowało się 14% techników (GUS, 2013, s. 229), podczas gdy dla tego samego roku na wsi zlokalizowanych było tylko 8% ogółu liceów ogólnokształcących (w tym uzupełniających) (GUS, 2013, s. 214).

W związku z powyższym analiza porównywalnych wyników dla szkół ponadgimnazjalnych zostanie przedstawiona w podziale na licea ogólnokształcące, profilowane i technika. Pomimo możliwości przystąpienia do egzaminu maturalnego w każdym z tych rodzajów szkoły, nie wolno zapominać, że utworzono je w różnych celach i odmienne wyniki osiągane na egzaminach nie powinny dziwić. Różnice pomiędzy szkołami publicznymi i niepublicznymi, czy w zależności od rodzaju gminy, w której mieści się szkoła powinno się rozpatrywać wewnątrz tak wydzielonych grup – trudno porównywać np. publiczne licea z niepublicznymi technikami. Tak szczegółowa analiza nie zostanie jednak zaprezentowana w niniejszym opracowaniu.

7.5.3.1. Matematyka

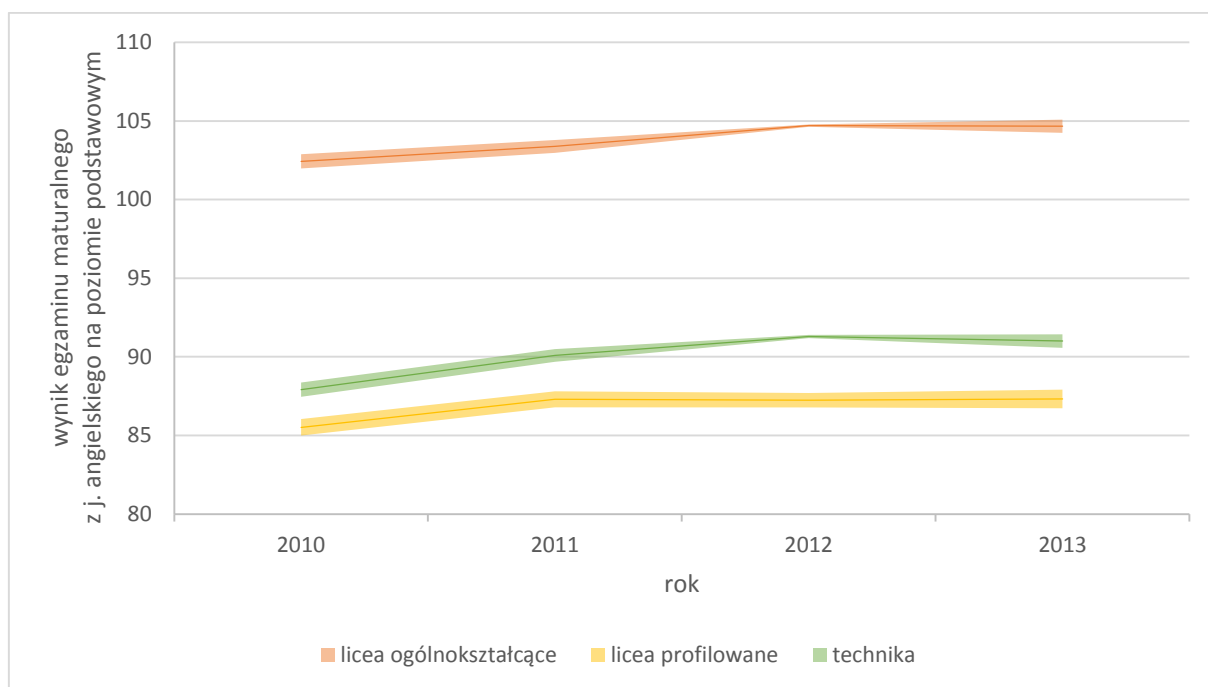
Rysunek 7.28 prezentuje porównywalne wyniki egzaminu maturalnego z matematyki dla lat 2010–2013 w podziale na rodzaje szkół. We wszystkich analizowanych latach najwyższe wyniki osiągają uczniowie liceów ogólnokształcących (uczy się w nich prawie dwie trzecie uczniów). W zależności od roku, uczniowie techników osiągają wyniki niższe od uczniów liceów ogólnokształcących o 13–16 punktów, czyli jedno odchylenie standardowe (są oni drugą pod względem liczebności grupą). Tendencja ta może być niepokojąca – większość przedmiotów zawodowych w technikum bazuje na dobrej znajomości matematyki. Uczniowie liceów profilowanych osiągają najniższe wyniki – o ponad jedno odchylenie standardowe niższe niż uczniowie liceów ogólnokształcących (około 20 punktów). Jest to najmniej liczna grupa uczniów – w roku 2010 stanowili oni cztery procent wszystkich uczniów, w 2011 trzy procent, a w latach 2012 i 2013 tylko dwa procent. W poszczególnych rodzajach szkół trendy wyników nieco różnią się od siebie. Średnie wyniki uczniów liceów ogólnokształcących prawie nie zmieniały się pomiędzy latami, natomiast w pozostałych rodzajach szkół w latach 2010–2012 nastąpił delikatny wzrost, a następnie spadek w roku 2013. Niezależnie od rodzaju szkoły, uczniowie szkół publicznych uzyskiwali wyższe wyniki niż uczniowie szkół niepublicznych.



Rysunek 7.28. Porównywalne wyniki egzaminu maturalnego z matematyki w latach 2010–2013 w podziale na rodzaje szkół

7.5.3.2. Język angielski

Porównywalne wyniki egzaminu maturalnego z języka angielskiego w podziale na licea ogólnokształcące, profilowane i technika kształtują się podobnie, jak dla matury z matematyki (zob. Rysunek 7.29). Uczniowie liceów ogólnokształcących uzyskują najwyższe wyniki, uczniowie technik uzyskują wyniki o ponad dwie trzecie odchylenia standardowego niższe od nich, a uczniowie liceów profilowanych uzyskują wyniki najniższe (o ponad jedno odchylenie standardowe niższe niż uczniowie liceów ogólnokształcących). Również podobnie jak w przypadku egzaminu z matematyki, uczniowie szkół publicznych uzyskiwali wyższe wyniki niż uczniowie szkół niepublicznych, niezależnie od rodzaju szkoły.



Rysunek 7.29. Porównywalne wyniki egzaminu maturalnego z języka angielskiego w latach 2010–2013 w podziale na rodzaje szkół

7.6. Lokalizacja szkoły

Lokalizacja szkoły, rozumiana jako rodzaj jednostki samorządu terytorialnego, na terenie której leży, może różnicować osiągane przez uczniów wyniki. W Polsce wyróżnia się trzy typy gmin³⁷: gminy miejskie (o statusie miasta), miejsko-wiejskie (jedna z miejscowości na ich terenie ma status miasta) i wiejskie (na ich terenie znajdują się wyłącznie wsie). Podział ten niezbyt dobrze odzwierciedla zróżnicowanie składu społecznego mieszkańców gmin, co niewątpliwie jest jego wadą. Gminy wiejskie to zarówno gminy ościenne wielkich aglomeracji miejskich, jak i gminy peryferyjne bez ośrodka miejskiego (Dolata i in., 2014). Rodzaj gminy jest powiązany z takimi czynnikami jak: poziom bezrobocia (powiązany ze średnim poziomem zarobków rodziców oraz relacjami panującymi w rodzinach), struktura gospodarki (liczba podmiotów gospodarczych zatrudniających wykwalifikowanych pracowników) czy powszechność edukacji przedszkolnej (zob. Dolata i in., 2014). Z kolei czynniki te mogą mieć związek z wynikami osiąganymi przez uczniów na egzaminach.

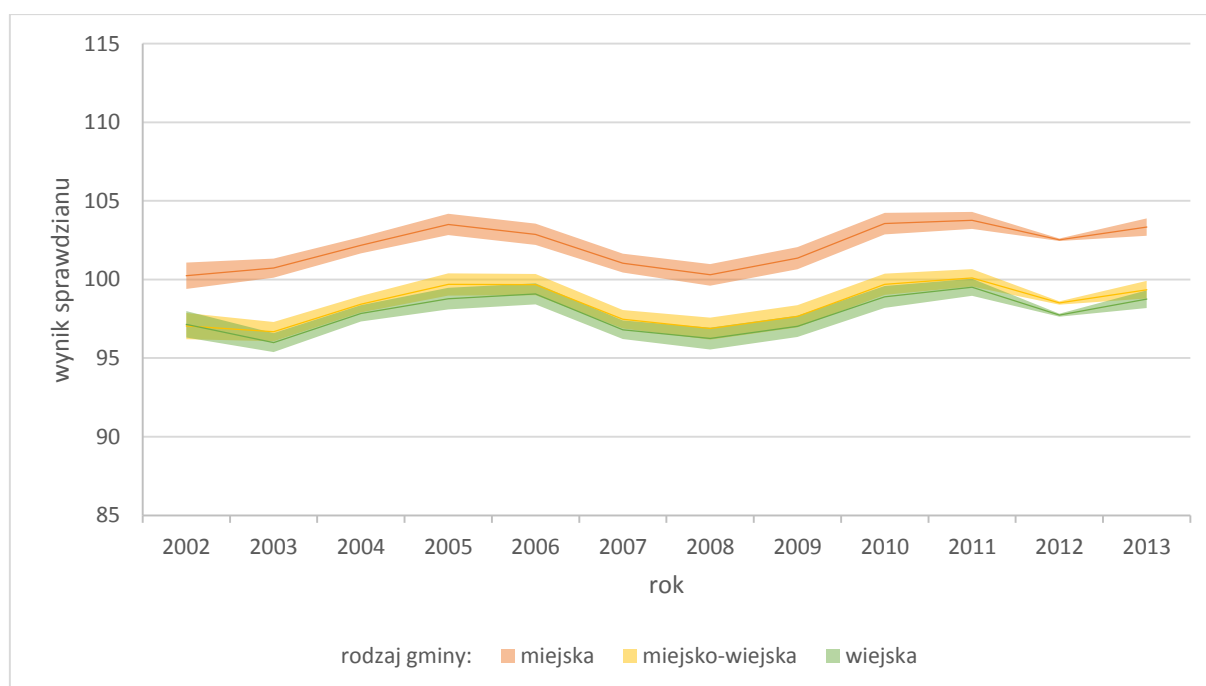
Analizując wyniki sprawdzianu z 2002 roku, Herbst (2004) zwraca uwagę, że średnie wyniki osiągane w gminach miejskich są wyższe niż w gminach wiejskich (różnica pomiędzy średnimi wynosi 3,5%), jednakże wskaźnik wartości dodanej szkoły jest wyższy w gminach wiejskich niż miejskich. Dolata (2008) wskazuje, że analizując zbiorczo wyniki szkół miejskich i wiejskich należy pamiętać, że choć różnica między średnimi wynikami uzyskiwanymi przez szkoły wiejskie i miejskie jest niewielka, to wewnątrzgrupowe zróżnicowanie szkół wiejskich oraz miejskich może być diametralnie inne. Warto

³⁷ Rozporządzenie Rady Ministrów z dnia 15 grudnia 1998 r. w sprawie szczegółowych zasad prowadzenia, stosowania i udostępniania krajowego rejestru urzędowego podziału terytorialnego kraju oraz związanych z tym obowiązków organów administracji rządowej i jednostek samorządu terytorialnego (Dz.U. 1998 nr 157 poz. 1031).

mieć to na uwadze rozpatrując różnice w porównywalnych wynikach egzaminacyjnych pomiędzy poszczególnymi rodzajami gmin.

7.6.1. Sprawdźian

Porównywalne wyniki sprawdzianu są wyższe w gminach miejskich niż w gminach miejsko-wiejskich i wiejskich we wszystkich analizowanych latach (zob. Rysunek 7.30). Dla części uczniów z 2002 roku, podobnie jak w przypadku typu szkoły, nie udało się ustalić rodzaju gminy, w której znajdowała się szkoła, do której uczęszczali. Wyniki z 2002 roku należy zatem traktować ostrożnie, gdyż jeśli braki danych nie były losowe, to mogą one być nieco inne niż w rzeczywistości. Wcześniej wskazywano również uwagę na to, że podczas pierwszej edycji sprawdzianu, w związku z jego nowością, mogło nie udać się zastosować w pełni wszystkich przewidzianych procedur. Poza rokiem 2002 trend wyników w każdym rodzaju gminy kształtuje się podobnie. Różnica wyników między szkołami miejsko-wiejskimi i wiejskimi jest nieistotna statystycznie i wynosi mniej niż jeden punkt. Nie należy jednak wyższych wyników uczniów ze szkół miejskich jednoznacznie przypisywać dużej efektywności kształcenia w tych szkołach. Opisane różnice, jak zauważa Konarzewski i in. (2012, s.7) „nie wynikają z bardziej skutecznych metod kształcenia, lecz wyłącznie stąd, że w rejonie tych szkół mieszka więcej rodzin zamożnych, z wykształconymi i ustabilizowanymi zawodowo rodzicami. Rzekome upośledzenie szkół wiejskich to w istocie upośledzenie polskiej wsi – relatywnie biednej i źle wykształconej”.



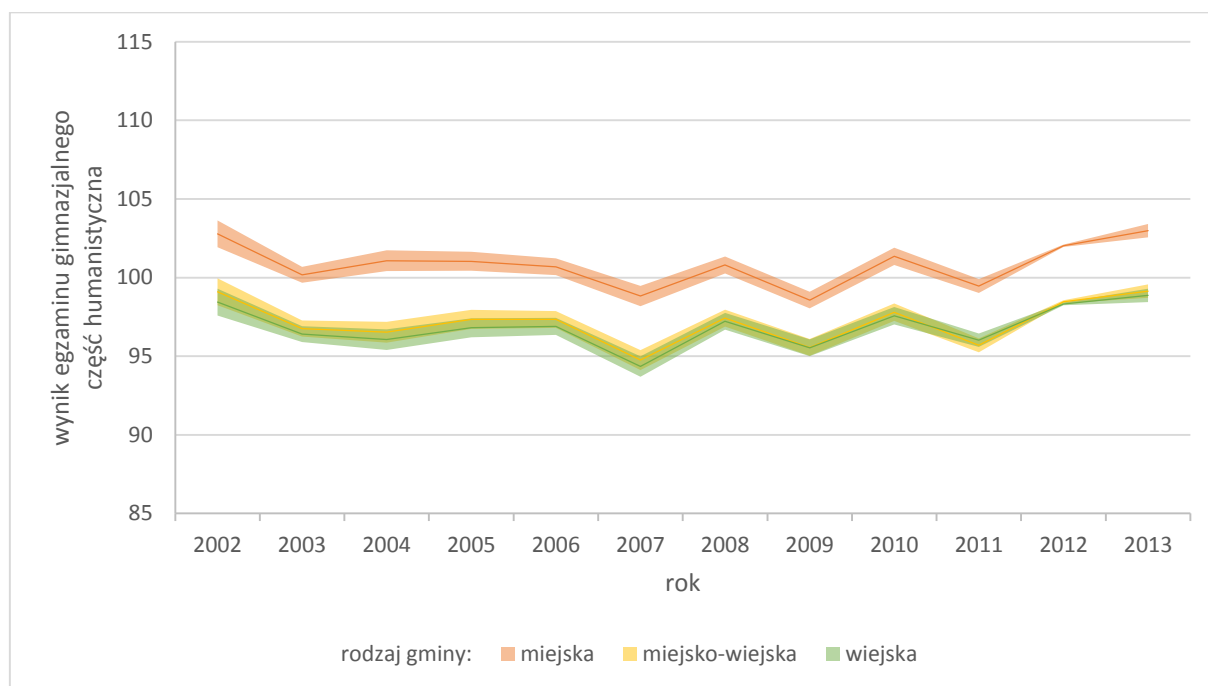
Rysunek 7.30. Porównywalne wyniki sprawdzianu w latach 2002–2013 w podziale na rodzaj gminy, w której znajduje się szkoła

7.6.2. Egzamin gimnazjalny

7.6.2.1. Część humanistyczna

Porównywalne wyniki części humanistycznej egzaminu gimnazjalnego, które ilustruje Rysunek 7.31, kształtują się podobnie jak w przypadku sprawdzianu. Uczniowie z gmin miejskich uzyskują wyniki wyższe (średnio o trzy do pięciu punktów) niż uczniowie z pozostałych typów gmin, a różnice pomiędzy

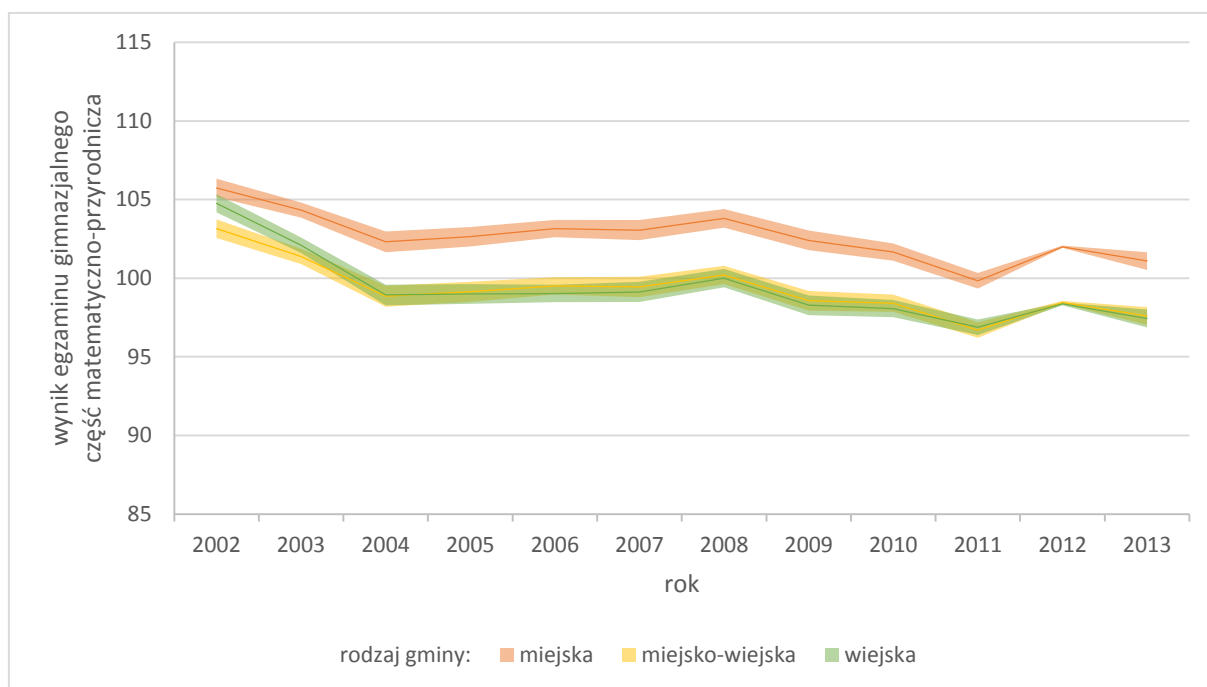
uczniami z gmin miejsko-wiejskich i wiejskich są pomijalne. Również niezależnie od typu gminy, w której mieściła się szkoła, zmiany wyników w poszczególnych latach przebiegają podobnie.



Rysunek 7.31. Porównywalne wyniki części humanistycznej egzaminu gimnazjalnego w latach 2002–2013 w podziale na rodzaj gminy, w której znajduje się szkoła

7.6.2.2. Część matematyczno-przyrodnicza

Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego w poszczególnych rodzajach gmin kształtują się podobnie, jak w części humanistycznej (zob. Rysunek 7.32). Uczniowie ze szkół w gminach miejskich uzyskiwali średnio około dwa do czterech punktów więcej niż uczniowie ze szkół z gmin wiejskich i miejsko-wiejskich, pomiędzy którymi nie ma istotnych statystycznie różnic. Stosunkowo wyższe niż w późniejszych latach wyniki gmin wiejskich w 2002 roku (bliźsze średnim wynikom gmin miejskich) trudno wyjaśnić tylko na podstawie analizowanych danych. Być może jest to spowodowane nowością egzaminu w roku 2002, która mogła spowodować niepełne zastosowanie wszystkich procedur. Późniejsze wyniki szkół wiejskich sugerują taką przyczynę, lecz potwierdzenie tej hipotezy wymaga pogłębionych badań na ten temat. Brak początkowego spadku wyników uczniów ze szkół wiejskich w części humanistycznej egzaminu gimnazjalnego sugeruje jednak, że jest to zjawisko specyficzne tylko dla części matematyczno-przyrodniczej.



Rysunek 7.32. Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013 w podziale na rodzaj gminy, w której znajduje się szkoła

7.7. Podsumowanie

Główną zaletą PWE jest umożliwienie analizy zmian poziomu umiejętności uczniów na przestrzeni czasu. Trendy zmian analizowanych wyników egzaminacyjnych przedstawionych na wspólnej skali są względnie stałe. W przypadku sprawdzianu i obu części egzaminu gimnazjalnego można zaobserwować w poszczególnych latach niewielkie fluktuacje wyników, ale w dłuższej perspektywie czasowej wyniki tych egzaminów utrzymują się na podobnym poziomie. Dla matury z matematyki i języka angielskiego obserwujemy niewielki wzrost wyników w latach 2010–2013. Ze względu na to, że dysponujemy porównywalnymi wynikami egzaminacyjnymi wyłącznie dla kilku lat, nie możemy wnioskować o tym, czy trend ten będzie się utrzymywać w kolejnych latach. Poniżej podsumowano najważniejsze wnioski w podziale na analizowane zmienne: region kraju, płeć, diagnozę dysleksji, typ szkoły oraz lokalizację szkoły.

Prezentowane dane wskazują na istnienie zróżnicowania terytorialnego umiejętności uczniów głównie na trzecim etapie edukacyjnym – uczniowie, którzy podchodzili do egzaminu gimnazjalnego w centralnej, południowej i wschodniej części kraju, uzyskiwali wyższe wyniki niż uczniowie z pozostałych regionów. Podobne zależności można dostrzec również w ostatnich latach w odniesieniu do wyników sprawdzianu, lecz nie są one tak wyraźne. W literaturze wymienia się historyczne, ekonomiczne i społeczne korelaty tego zjawiska, ale prezentowane tu dane nie pozwalają na wnioskowanie o jego przyczynach. Dla wyników z matury z matematyki zanika wyraźna różnica pomiędzy regionem północnym i zachodnim a regionami centralnym, wschodnim i południowym, obserwowana na wcześniejszych etapach edukacyjnych. W przypadku matury z języka angielskiego tylko region wschodni charakteryzuje się niższymi wynikami (co jest powodowane niskimi wynikami w województwie świętokrzyskim).

Porównania wyników między płciami wskazały na przewagę dziewcząt w sprawdzianie i części humanistycznej egzaminu gimnazjalnego, natomiast pozostałe egzaminy nie różnicowały obu płci. Ze

względu na ponadprzedmiotowy charakter sprawdzianu nie można wnioskować o tym, czy chłopcy i dziewczynki różnią się ze względu na mierzone umiejętności (np. humanistyczne lub matematyczne). Zróznicowanie wyników ze względu na płeć w części humanistycznej egzaminu gimnazjalnego można próbować powiązać z poziomem czytelnictwa. Dziewczynki uzyskują wyższe wyniki od chłopców, co może być potencjalnie związane z ich wyższym poziomem czytelnictwa, raportowanym zarówno w polskich (Zasacka, 2014), jak i zagranicznych (OECD, 2009) badaniach. Może to być sugestią do wprowadzenia szczególnych inicjatyw zorientowanych na poprawę umiejętności w zakresie czytania wśród chłopców, podobnych do inicjatyw wprowadzonych w Wielkiej Brytanii (zob. EACEA, 2010). Różnica w wynikach dziewcząt i chłopców w części humanistycznej egzaminu gimnazjalnego zmniejsza się jednak wraz z upływem czasu. Dla części matematyczno-przyrodniczej różnica wyników pomiędzy płciami jest praktycznie pomijalna. Jest to zjawisko pozytywne, biorąc szczególnie pod uwagę zakrojone na szeroką skalę działania promujące karierę w nauce i technologii wśród dziewcząt.

Diagnoza dysleksji (lub jej brak) jest zmienną, która nie różnicuje w większości przypadków wyników egzaminacyjnych. Można więc wnioskować, że istniejące w systemie dostosowania pozwalające na wyrównywanie szans dla uczniów z dysleksją, spełniają swoją rolę. Jedynym egzaminem, którego wyniki różniły się na niekorzyść uczniów z diagnozą dysleksji, była matura z języka angielskiego. W związku z tym, że nauka języków obcych może być szczególnie trudna dla uczniów z dysleksją, dostosowania egzaminacyjne powinny dodatkowo uwzględniać tę specyfikę.

W wynikach sprawdzianu można zaobserwować zwiększającą się różnicę pomiędzy uczniami uczęszczającymi do szkół publicznych i niepublicznych, jednak trend wyników jest podobny. Tak jak dla sprawdzianu, również dla egzaminu gimnazjalnego, wyniki uczniów szkół niepublicznych są wyższe niż uczniów szkół publicznych. W rozstrzygnięciu o przyczynach zróznicowania wyników w szkołach publicznych i niepublicznych mogłaby pomóc analiza efektywności ich pracy, na przykład stosując wskaźniki edukacyjnej wartości dodanej.

Analizując wyniki ze względu na typ gminy, dla sprawdzianu i egzaminu gimnazjalnego można zaobserwować różnice na korzyść gmin miejskich, natomiast nie występują różnice pomiędzy gminami wiejskimi i miejsko-wiejskimi. Nie można natomiast wnioskować o lepszej efektywności pracy szkół w miastach, gdyż wyniki te mogą być związane np. z lepszym statusem społeczno-ekonomicznym rodzin w miastach lub procesem różnicowania się szkół (Dolata, Jasińska i Modzelewski, 2012).

W przypadku szkół ponadgimnazjalnych trudno analizować osiągnięte wyniki w podziale na szkoły publiczne i niepubliczne oraz na typ gminy (zdecydowana większość uczniów uczęszcza do szkół publicznych oraz do szkół znajdujących się w gminach miejskich lub miejsko-wiejskich). Właściwym kryterium wydaje się natomiast rodzaj szkoły: liceum ogólnokształcące, profilowane czy też technikum. Najwyższe wyniki matury osiągają uczniowie liceów ogólnokształcących (należy zauważyć, że uczy się w nich prawie dwie trzecie uczniów). Niższe wyniki uzyskują uczniowie techników (którzy są drugą pod względem liczebności grupą), a najniższe uczniowie liceów profilowanych.

8. Prezentacja porównywalnych wyników egzaminacyjnych

Począwszy od jesieni 2012 roku porównywalne wyniki egzaminacyjne (PWE) udostępniane są w serwisie internetowym <http://pwe.ibe.edu.pl>. Prezentacja wyników w postaci prostej w obsłudze strony internetowej zawierającej gotowe narzędzia do przeprowadzania podstawowych analiz z wykorzystaniem PWE ma na celu upowszechnienie wyników badania w jak najszerszym kręgu odbiorców. W szczególności rolą serwisu jest uczynienie PWE dostępnym także w środowiskach innych niż naukowe, np. wśród przedstawicieli administracji oświatowej, nauczycieli czy rodziców. W połowie 2015 roku, na podstawie zebranych w ciągu dwóch i pół roku doświadczeń, serwis został odświeżony, a jego budowa uproszczona. Pozwoliło to uczynić dostęp do PWE jeszcze łatwiejszym i jednocześnie przyspieszyć szybkość działania serwisu.

Opisana wyżej strona internetowa jest najważniejszym, lecz nie jedynym sposobem dostępu do PWE. Dla tych osób, które chciałyby zbiorowo pobrać wartości PWE dla dużej liczby szkół i/lub jednostek samorządu terytorialnego lub zgoła sięgnąć do danych na poziomie uczniów (np. w celu obliczenia PWE w pogrupowaniu innym niż te dostępne w serwisie internetowym PWE), istnieje możliwość bezpośredniego sięgnięcia do baz danych powstałych w toku badania, co zostało dokładniej omówione w Rozdziale 9.

8.1. Budowa serwisu

W serwisie wyróżnić można trzy główne moduły:

1. część opisową,
2. wyszukiwarkę szkół oraz jednostek samorządu terytorialnego,
3. moduł analiz.

The screenshot shows the website <http://pwe.ibe.edu.pl>. The header is orange with the logo 'PORÓWNYWALNE WYNIKI EGZAMINACYJNE' and 'IBE INSTYTUT BADAŃ EDUKACYJNYCH'. Below the header is a navigation bar with links: 'INSTITUT BADAŃ EDUKACYJNYCH', 'ENTUZJAŚCI EDUKACJI', 'BAZA NARZĘDZI DYDAKTYCZNYCH', and 'kontakt: pwe@ibe.edu.pl'. The main content area has a title 'Czym są Porównywalne Wyniki Egzaminacyjne (PWE)?' and a paragraph explaining the service. To the right of the main content is a sidebar with a search bar labeled 'Wyszukiwarka' and a list of links: 'Parametry wykresu', 'Opisy', 'Czym są PWE?', 'Procedura zrównywania', and 'Jak odczytywać wykresy?'.

Rysunek 8.1. Widok serwisu <http://pwe.ibe.edu.pl>

Przechodzenie między modułami możliwe jest za pomocą menu po prawej stronie ekranu (zob. Rysunek 8.1): kliknięcie na belce „Opisy”, a następnie wybranie interesującego nas zagadnienia otwiera wskazaną stronę opisową, kliknięcie na belce „Wyszukiwarka” powoduje otwarcie modułu wyszukiwarki, natomiast kliknięcie na belce „Parametry wykresu” otwiera moduł analiz.

Część opisowa w syntetyczny sposób przybliża użytkownikowi informacje o celu i metodologii zrównywania wyników egzaminacyjnych, opisuje również w przystępnej, ilustrowanej przykładami formie sposób interpretacji prezentowanych wyników. W tym celu podzielona została ona na trzy podstrony: *Czym są PWE?*, *Procedura zrównywania oraz Jak odczytywać wykresy?*

Wyszukiwarka umożliwia wybranie szkół i/lub jednostek samorządu terytorialnego (dalej zarówno szkoły, jak i jednostki samorządu terytorialnego będą dla uproszczenia nazywane *jednostkami*), których wyniki mają być analizowane. *Jednostki* wyszukiwać można wg kilku kryteriów, które można ze sobą łączyć:

- fragmentu nazwy i/lub adresu *jednostki*,
- typu *jednostki*, a więc określenia, czy ma nią być szkoła wskazanego typu, gmina, powiat czy województwo,
- jednostki samorządu terytorialnego, w której znajduje się *jednostka* – województwa, a później powiaty i gminy wybiera się z list rozwijalnych, które umożliwiają wybór dowolnej jednostki terytorialnej wg obowiązującego obecnie podziału Polski do poziomu gminy włącznie,
- dostępności wyników zrównanych *jednostki* we wskazanych latach.

Wyniki wyszukiwania prezentowane są w rozbiciu na typ *jednostki*, a więc szkoły, gminy, powiaty i województwa wyświetlane są oddzielnie. Każdą z wyszukanych *jednostek* można odpowiednim przyciskiem dodać do analizy. Istnieje również możliwość szybkiego dodania do analizy wszystkich wyszukanych *jednostek* danego typu.

Moduł analiz pozwala na wizualizację PWE jednej lub porównanie ze sobą PWE wielu *jednostek*. Z przyczyn technicznych możliwe jest jednoczesne porównywanie maksymalnie ok. 200 *jednostek*, w praktyce jednak wykresy stają się nieczytelne już przy ok. dziesięciu, ograniczenie to nie stanowi więc przy normalnym korzystaniu z serwisu problemu.

W belce „Parametry wykresu” po prawej stronie ekranu istnieje możliwość wybrania części egzaminu (wskaźnika PWE), który ma zostać przeanalizowany, okresu, tego czy mają zostać wyświetlone także średnie wyniki ogólnopolskie oraz typu wykresu. Dostępne są dwa rodzaje wykresów – podstawowy i zaawansowany. Pierwszy z nich to wykres liniowy prezentujący zmianę średniej PWE *jednostek* w czasie. Drugi to wykres skrzynkowy, z którego odczytać można najważniejsze parametry rozkładów wyników PWE *jednostek* w poszczególnych latach. Wykresy te omówione zostały dokładniej w dalszej części rozdziału.

Główną część ekranu modułu analiz zajmuje wykres, wyświetlany zgodnie z określonymi przez użytkownika (a opisanymi wyżej) parametrami. Nad nim znajduje się krótki komentarz opisujący, jakie dokładnie statystyki zobaczyć można na wykresie. Bezpośrednio pod wykresem dostępne są przyciski pozwalające:

- Wyświetlić dane wizualizowane na wykresie w postaci ujętych w tabelę wartości liczbowych.

- Pobrać wykres w postaci obrazu w formacie PNG, np. w celu utrwalenia wyników przeprowadzanej analizy, umieszczenia go w raporcie, itp.
- Pobrać wyniki PWE analizowanych *jednostek* w postaci pliku CSV np. w celu wykonania samodzielnych analiz lub wizualizacji w zewnętrznym oprogramowaniu. Pobrane zostaną zarówno informacje wyświetlane na wykresie podstawowym, jak i te prezentowane na wykresie zaawansowanym, a także liczbę uczniów w *jednostce*, błąd standardowy oszacowania średniej oraz kod TERYT *jednostki*.
- Usunąć wszystkie *jednostki* z analizy.

Poniżej przycisków wyświetlana jest lista wszystkich aktualnie analizowanych *jednostek*. Daje ona możliwość usunięcia wybranej *jednostki* z analizy (przycisk po prawej stronie), jak również czasowego pominięcia jej na wykresie (pole wyboru po lewej stronie).

8.2. Prezentacja wyników

8.2.1. Skala prezentacji wyników

Wyniki zrównane prezentowane są na skali, dla której wartości 100 oraz 15 odpowiadają, odpowiednio, średniemu wynikowi oraz odchyleniu standardowemu w populacji uczniów piszących daną część egzaminu w 2012 roku. Skala ta będzie w dalszej części nazywana skrótowo skalą (100; 15). Populację określono jako uczniów piszących arkusz standardowy egzaminu w szkołach dla młodzieży, które nie są placówkami specjalnymi ani związanymi, gdyż właśnie w takich szkołach prowadzone były badania zrównujące. W odniesieniu do egzaminu maturalnego populację zawężono dodatkowo do uczniów, którzy do egzaminu pochodzą po raz pierwszy³⁸. Rok 2012 jako punkt odniesienia dla skali wybrany został arbitralnie – z metodologicznego punktu widzenia mógłby to być dowolny rok spośród lat, dla których przeprowadzono zrównywanie.

Wyniki na skali (100; 15) należy interpretować jako *wynik, jaki osiągnąłby dany uczeń/szkoła/itd., gdyby napisał dany egzamin z roku 2012, w odniesieniu do wyników uczniów, którzy faktycznie pisali dany egzamin w roku 2012*. Interpretacja ta dobrze oddaje istotę zrównywania wyników egzaminacyjnych – wyrażamy rezultaty z różnych egzaminów za pomocą skali jednego, wybranego z nich.

To, co w sposób istotny różni skalę (100; 15) od skal, na jakich prezentowane są niezrównane wyniki egzaminów (liczba uzyskanych punktów, skala procentowa od 0 do 100%, centyl w rozkładzie populacyjnym) to brak minimum i maksimum. O ile np. na skali procentowej nie da się uzyskać wyniku niższego od 0% (minimum) ani wyższego niż 100% (maksimum), o tyle skala (100; 15) rozciąga się od minus nieskończoności do plus nieskończoności. Stąd dla skali (100; 15) możemy jedynie mówić o tym, jak prawdopodobne jest uzyskanie wskazanego wyniku. Np. wynik mniejszy od 40 (niższy o ponad 4 odchylenia standardowe od średniej) spodziewamy się zaobserwować tylko dla ok. 0,003% uczniów, możemy więc go uznać za bardzo mało prawdopodobny. Nie oznacza to jednak ani że nie może on wystąpić, ani że musi wystąpić. To, jaki odsetek uczniów spodziewamy się zaobserwować w jakim przedziale wyników odczytywane jest z rozkładu normalnego o średniej 100 i odchyleniu standardowym 15.

³⁸ Ma to szczególne znaczenie dla przedmiotów takich jak biologia czy chemia, gdzie osoby powtórnie podchodzące do egzaminu stanowią nawet kilkanaście procent wszystkich zdających.

8.2.2. Wykres liniowy

Wykres liniowy jest domyślnym sposobem prezentacji porównywalnych wyników egzaminacyjnych. Przedstawia on średni zrównany wynik egzaminacyjny z wybranej części egzaminu wraz z 95% przedziałem ufności. Prezentacja na wykresie przedziałów ufności pozwala w łatwy sposób stwierdzić, które różnice są istotne statystycznie – istotnie statystycznie różne są od siebie te wartości średnich, których przedziały ufności się nie przecinają. Korzystanie z tej reguły podczas porównywania ze sobą szkół, gmin, itp. jest kluczowe do tego, aby formułowane wnioski były poprawne metodologicznie.

Szerokość przedziału ufności zależy od trzech czynników:

- 1) liczby uczniów, którzy przystąpili do egzaminu w danej szkole, gminie, itp. – im większa liczba osób w danej grupie, tym węższy jest przedział ufności,
- 2) tego, jak spójne były wyniki egzaminacyjne osiągnięte przez uczniów w danej szkole, gminie, itp., – im wyniki w danej grupie mniej oddalone od średniej danej grupy, tym węższy jest przedział ufności,
- 3) niepewności związanej z faktem, że wyniki uzyskano w procesie modelowania statystycznego (procesie zrównywania).

Szczególnie wyraźnie widoczny jest wpływ pierwszego z wymienionych czynników – to on powoduje, że przedziały ufności dla gmin są na ogół³⁹ węższe od przedziałów ufności szkół, przedziały ufności powiatów są węższe od przedziałów ufności gmin, itd., a najwęższy jest przedział ufności dla średniego wyniku ogólnopolskiego. Z kolei wśród szkół, gmin, itp., w których do egzaminu gimnazjalnego w danym roku przystąpiła zbliżona liczba uczniów, o tym, która z nich będzie posiadać węższy, a która szerszy przedział ufności, decydować będzie zróżnicowanie wyników egzaminacyjnych w ramach danej szkoły, gminy, itp. Aby móc rozróżnić od siebie wpływ obydwu rozważanych czynników. w tabeli danych pod wykresem prezentowana jest liczba uczniów danej szkoły, gminy, itp. Trzeci z wymienionych czynników w sposób istotny wpływa na szerokość przedziałów ufności jedynie na poziomie ogólnopolskim.

8.2.3. Wykres skrzynkowy

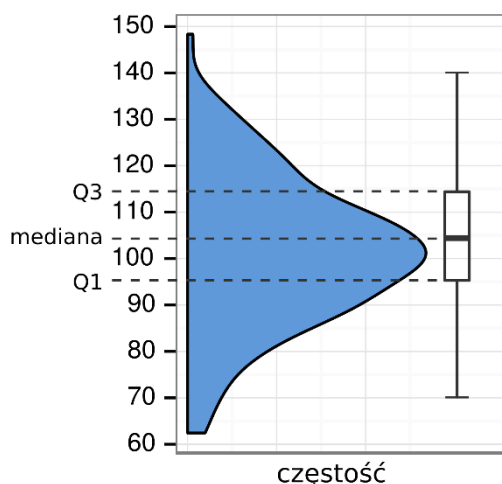
Wykres skrzynkowy, nazywany w serwisie wykresem zaawansowanym, jest alternatywną formą prezentacji porównywalnych wyników egzaminacyjnych. Prezentuje on bardziej szczegółowe informacje niż wykres liniowy, przez co daje możliwość przeprowadzenia bardziej wnikliwych analiz, lecz jest również trudniejszy w interpretacji. Na wykresie pokazywane są:

- Pierwszy (Q_1), drugi (mediana) i trzeci (Q_3) kwartył, oznaczane odpowiednio przez dolną krawędź skrzynki, linię w środku skrzynki oraz górną krawędź skrzynki na wykresie.
- Minimalny i maksymalny wynik nieodstający, oznaczone odpowiednio przez końcówki dolnego i górnego z „wąsów” odchodzących od skrzynki. Za obserwacje odstające uznawane są te wyniki egzaminacyjne, które są niższe od wartości pierwszego kwartyla pomniejszonej

³⁹ Wyjątek stanowią gminy, w których znajduje się tylko jedna szkoła danego typu – w takim wypadku wykres dla gminy pokrywa się z wykresem dla szkoły.

o półtorakrotność rozstępu międzykwartkowego⁴⁰ lub wyższe od wartości trzeciego kwartyła powiększonego o półtorakrotność rozstępu międzykwartkowego.

O ile wykres podstawowy skupia się na prezentacji średniego zrównanego wyniku egzaminacyjnego dla danej szkoły, gminy, itp., to wykres zaawansowany koncentruje się na tym, w jaki sposób rozkładają się wyniki poszczególnych grup uczniów w ramach danej szkoły, gminy, itp. Na każdą grupę kwartylową, z definicji, przypada 25% uczniów. Najślabi reprezentowani są przez koniec dolnego „wąsa” (najniższy wynik nieodstający), słabi przez pozycję dolnego końca skrzynki (1. kwartył), przeciętni przez medianę (2. kwartył), dobrzy przez pozycję górnego końca skrzynki (3. kwartył), najlepsi przez koniec górnego „wąsa” (najwyższy wynik nieodstający). Za najniższy wynik nieodstający przyjęto najniższy wynik nie mniejszy jednak od wartości 1. kwartyla pomniejszonej o 1,5-krotność różnicy pomiędzy 3. i 1. kwartylem. Analogicznie największy nieodstający wynik to najwyższy wynik nie większy jednak od wartości 3. kwartyla zwiększonego o 1,5-krotność różnicy pomiędzy 3. i 1. kwartylem. Im większa wysokość skrzynki i długość „wąsów”, tym bardziej zróżnicowane wyniki uczniów. Wykres skrzynkowy można interpretować jako uproszczoną prezentację rozkładu wyników (zob. Rysunek 8.2).



Rysunek 8.2. Relacja między wykresem skrzynkowym a rozkładem wyników na podstawie przykładowego rozkładu wyników w powiecie

8.2.4. Tabela danych

Prezentacji wykresu liniowego i skrzynkowego towarzyszy zawsze tabela danych. Zawiera ona te same informacje, które prezentowane są na danym wykresie oraz dodatkowo liczbę uczniów uwzględnionych przy obliczaniu wyników zrównanych dla danej szkoły, gminy, itp. za dany rok. Tabela danych pozwala odczytać dokładne wartości liczbowe prezentowanych na wykresie statystyk. Tabelę danych można również pobrać jako plik CSV w celu wykonania samodzielnych analiz i/lub wykresów, np. w arkuszu kalkulacyjnym lub programie statystycznym.

8.3. Przykłady

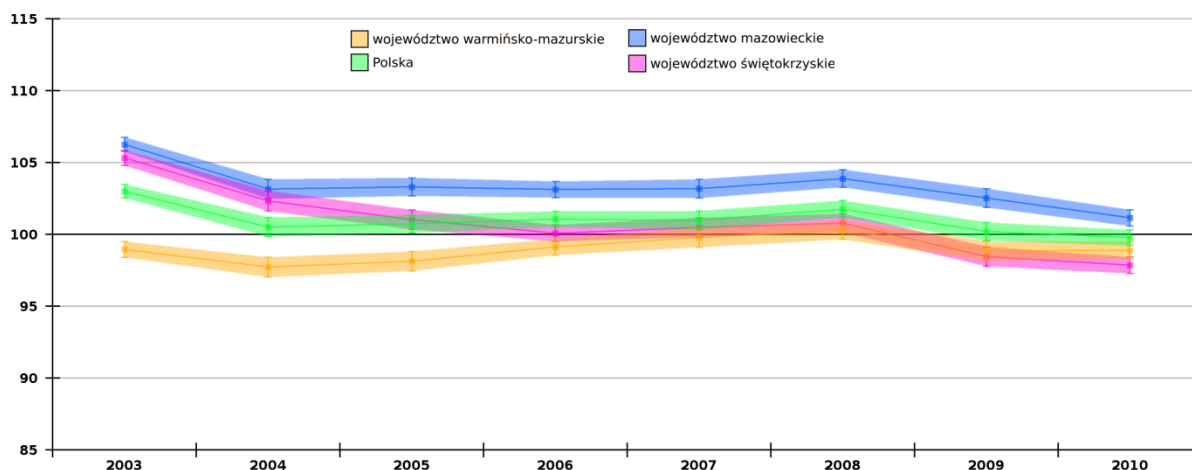
Poniżej przedstawione zostały przykłady analiz z wykorzystaniem serwisu <http://pwe.ibe.edu.pl> dla różnych poziomów agregacji (przykład 1 – województw, przykłady 2 i 3 – szkół), egzaminów (przykład 1

⁴⁰ Rozstęp międzykwartkowy jest równy różnicy wartości trzeciego i pierwszego kwartyla.

– egzamin gimnazjalny w części matematyczno-przyrodniczej, przykład 2 – sprawdzian, przykład 3 – matura na poziomie podstawowym z matematyki) oraz z wykorzystaniem zarówno wykresu podstawowego (przykłady 1 i 2), jak i zaawansowanego (przykład 3).

8.3.1. Przykład 1.

Jako pierwszy przykład posłużyć może analiza na poziomie makroskopowym. Porównane zostały średnie wyników zrównanych egzaminu gimnazjalnego w części matematyczno-przyrodniczej z lat 2003-2010 dla trzech wybranych województw (mazowieckiego, świętokrzyskiego, warmińsko-mazurskiego) oraz Polski (zob. Rysunek 8.3).



Rysunek 8.3. Średnie PWE (z 95% przedziałami ufności) z lat 2003-2010 dla trzech wybranych województw oraz Polski dla części matematyczno-przyrodniczej egzaminu gimnazjalnego

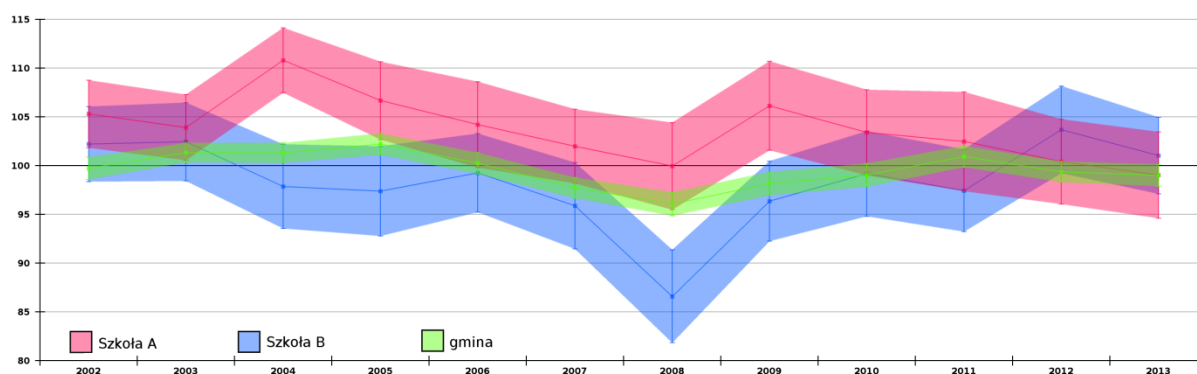
Przedstawiona analiza wskazuje na istnienie znacznego zróżnicowania średnich PWE pomiędzy rozważanymi województwami, maleje on jednak z upływem lat. W roku 2003 różnica pomiędzy średnim wynikiem w województwie warmińsko-mazurskim i mazowieckim wynosiła blisko 7,3, a więc prawie połowę odchylenia standardowego⁴¹, natomiast w latach 2007-2010 nie przekraczała już 3,7 punktów na skali, czyli zmalała blisko dwukrotnie. Widać też, że o ile województwa warmińsko-mazurskie oraz mazowieckie utrzymują w następujących po sobie latach podobną pozycję względem wyników ogólnopolskich, o tyle dla województwa świętokrzyskiego obserwowana jest wyraźna zmiana. W latach 2003-2004 średnie wyniki w województwie świętokrzyskim były wyraźnie wyższe niż średnie wyniki ogólnopolskie i tylko nieznacznie (nieistotnie statystycznie) niższe niż w województwie mazowieckim, jednak obserwowany w pozostałych województwach oraz Polsce trend spadkowy z lat 2003-2004 utrzymał się w województwie świętokrzyskim aż do roku 2006. Spowodowało to, że w latach 2007-2008 średni wynik w województwie świętokrzyskim był porównywalny już nie z województwem mazowieckim, lecz ze średnią krajową i województwem warmińsko-mazurskim (które w tym czasie zdążyło wyraźnie zmniejszyć dystans dzielący je od wyników ogólnopolskich). Co więcej, obserwowany we wszystkich rozważanych województwach oraz w całej Polsce trend spadkowy z lat 2008-2010 przebiegał z większą intensywnością dla województwa świętokrzyskiego, i w 2010 roku osiągnęło ono najniższy (istotnie statystycznie niższy wyników ogólnopolskich) średni wynik spośród rozpatrywanych jednostek.

⁴¹ Odchylenia standardowego wyników uczniów na egzaminie w roku bazowym (2012).

8.3.2. Przykład 2.

Analizy z wykorzystaniem PWE mogą również dotyczyć poziomu lokalnego. Jako przykład porównane zostały średnie wyniki zrównane na sprawdzianie szóstoklasisty w latach 2002-2013 w dwóch wybranych szkołach podstawowych oraz gminie, w której szkoły te się znajdują (zob. Rysunek 8.4).

Widoczna jest znaczna zmienność średnich wyników uzyskiwanych w obydwu szkołach w czasie. W latach 2002-2003, 2006-2007 oraz 2010-2013 średnie wyniki obydwu szkół różniły się od siebie w sposób nieistotny statystycznie. W rozdzielających je okresach 2004-2005 oraz 2008-2009 różnice okazały się być istotne statystycznie, przy czym w pierwszym wypadku decydujące znaczenie odgrywała wyraźna poprawa średnich wyników w roku 2004 w szkole A, zaś w drugim wyraźne obniżenie średnich wyników w roku 2008 w szkole B. Obniżenie to było przy tym wyjątkowo duże – blisko granicy istotności statystycznej pomiędzy rokiem 2007 i 2008 oraz z istotną statystycznie różnicą pomiędzy rokiem 2008 i 2009⁴². Niestety same wyniki PWE nie pozwalają dociec przyczyny zaobserwowanego zjawiska.



Rysunek 8.4. Średnie PWE (z 95% przedziałami ufności) dla dwóch przykładowych szkół podstawowych oraz gminy, w której się znajdują

Interesujące jest również zestawienie trendów wyników obserwowanych w szkołach ze średnimi wynikami w gminie. O ile w latach 2005-2008 obserwowany w gminie trend spadkowy odnotować można zarówno w szkole A, jak i w szkole B (w tym wypadku z wyłączeniem roku 2005), o tyle trend rosnący na poziomie gminy z lat 2008-2011 znajduje odzwierciedlenie jedynie w szkole B, zmiany w szkole A mają zaś odwrotny, spadkowy kierunek. Co więcej, rosnący trend w szkole B utrzymał się także w roku 2012, co pozwoliło jej po raz pierwszy od 10 lat osiągnąć średni wynik wyższy niż w gminie (choć nieistotnie statystycznie). Z kolei dość powolny, ale utrzymujący się przez 5 lat z rzędu trend spadkowy w szkole A w latach 2009-2013 sprawił, że w roku 2013 po raz pierwszy od momentu wprowadzenia egzaminów zewnętrznych jej średni wynik był niższy niż średni wynik w gminie (choć także nieistotnie statystycznie).

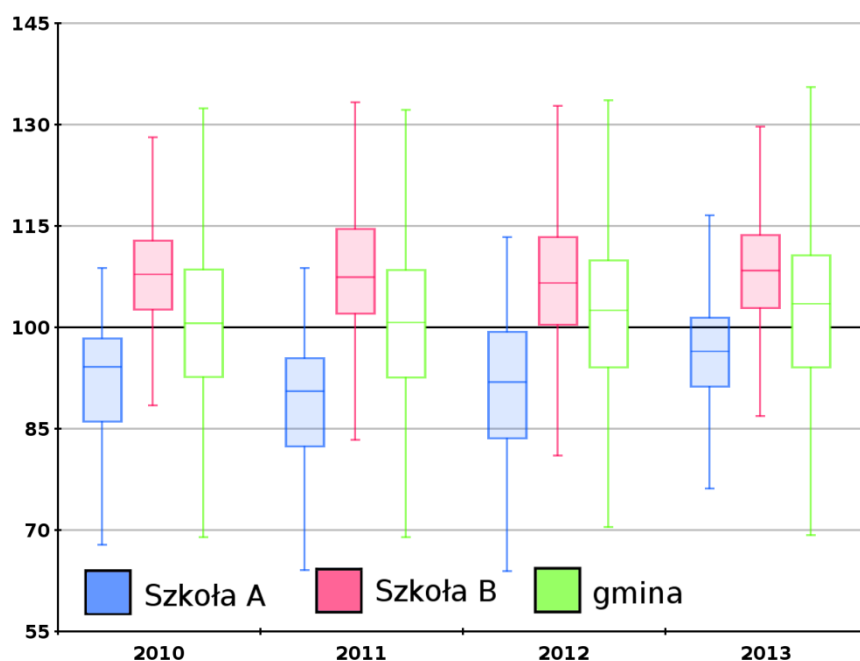
Największym ograniczeniem wykonywanych w serwisie pwe.ibe.edu.pl analiz na szczeblu lokalnym jest brak informacji o możliwych przyczynach obserwowanych zmian wyników. Aby dokonując analiz na tym szczeblu wyciągać wnioski o charakterze przyczynowo-skutkowym niezbędna jest dokładna znajomość

⁴² W praktyce bardzo rzadko spotyka się sytuacje, w której średnie wyniki PWE szkoły w kolejnych latach różniły się od siebie w sposób istotny statystycznie.

środowiska pracy porównywanych szkół, co wymaga prowadzenia dodatkowych badań, także jakościowych (z uwzględnieniem np. ewaluacji wewnętrznej pracy szkoły).

8.3.3. Przykład 3.

Dokładniejszych informacji dostarczyć może wykres zaawansowany. W tym wypadku jako przykład wykorzystane zostały wyniki zrównane z lat 2010-2013 z matury z języka angielskiego w dwóch przykładowych liceach ogólnokształcących oraz gminie, w której znajdują się te szkoły (zob. Rysunek 8.5).



Rysunek 8.5. Kwartyle oraz minimalne i maksymalne wyniki nieodstające PWE matury z języka angielskiego dla dwóch wybranych liceów ogólnokształcących oraz gminy, w której się znajdują

Analiza wykazała bardzo duże zróżnicowanie umiejętności pomiędzy uczniami w porównywanych szkołach. Uczniowie w szkole A osiągaliby wyraźnie słabsze wyniki zarówno od uczniów w szkole B, jak również od ogółu uczniów w gminie – w latach 2010-2011 najlepsi absolwenci liceum A osiągaliby wynik odpowiadający 3 kwartyłowi gminy, a więc w skali gminy lepsze wyniki od nich osiągnęła aż ¼ uczniów. Z kolei mediana wyników w szkole A odpowiadała we wszystkich analizowanych latach zlokalizowana jest w okolicach 1 kwartyła wyników w gminie, a więc przeciętny uczeń tej szkoły osiągał wyniki zbliżone do 25% najslabszych uczniów w całej gminie. Jeszcze bardziej jaskrawe jest zestawienie wyników liceum A z wynikami liceum B. W latach 2010-2011 najlepsi uczniowie szkoły A osiągaliby wynik tylko nieznacznie wyższy od mediany wyników w szkole B, a więc blisko połowa uczniów w szkole B osiągała wyniki lepsze niż najlepsi uczniowie w liceum A. Podobnie najslabsi uczniowie w szkole B cały czas osiągaliby wyniki lepsze od ponad ¼ uczniów w A liceum.

W latach 2012-2013 szkoła A poprawiła nieco uzyskiwane wyniki. Szczególnie rzuca się w oczy polepszenie się wyników uczniów najslabszych i ogólne zmniejszenie zróżnicowania wyników (wysokość skrzynki). Nie zmienia to jednak faktu, że cały czas jej absolwenci osiągaliby wyniki znacznie niższe niż w liceum B i w gminie.

Ciekawą cechą wyników uzyskiwanych w szkole A w latach 2010-2011 jest ich lewoskośność. Uczniowie o wynikach powyżej mediany byli więc znacznie mniej zróżnicowani od uczniów o wynikach poniżej mediany. Tą cechą rozkładu rozpoznać można po niesymetrycznym położeniu mediany w skrzynce szkoły oraz niesymetrycznej długości „wąsów” odchodzących od skrzynki.

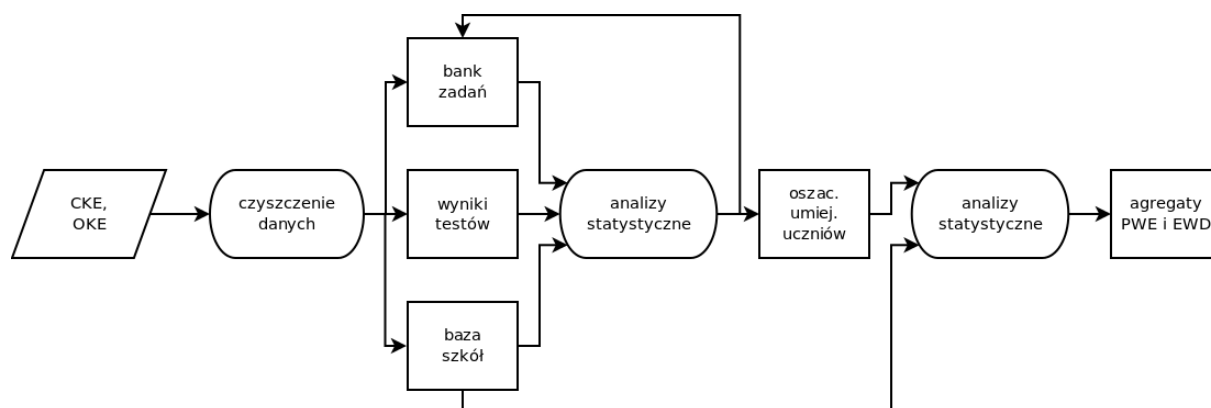
Na zakończenie warto podkreślić, że obserwowane różnice między szkołami B i A nie pozwalają jednoznacznie stwierdzić, która z nich lepiej naucza. Szkoły te mogą w sposób wyraźny różnić się profilem, środowiskiem pracy, poziomem uczniów przychodzących do tych szkół po gimnazjum, itp. Nie posiadając wszystkich tych informacji należy zachować powściągliwość przy formułowaniu wniosków.

9. Baza danych IBE

W celu przeprowadzenia zrównywania, jak również w jego wyniku, zgromadzone zostały bardzo duża liczba danych, m.in. wyniki uczniów, oszacowania parametrów zadań i umiejętności uczniów, wskaźniki PWE dla szkół oraz jednostek samorządu terytorialnego itp. Aby z jednej strony zapewnić spójność wszystkich tych informacji, z drugiej zaś umożliwić łatwy i elastyczny dostęp do nich, stworzona została relacyjna baza danych oraz zestaw narzędzi ułatwiających korzystanie z niej. Ponieważ zakres gromadzonych i struktura wytwarzanych w ramach badania informacji w dużej mierze pokrywał się z projektem Edukacyjnej Wartości Dodanej (EWD), baza danych zaprojektowana została i wdrożona w taki sposób, aby objąć również dane zbierane i generowane w ramach projektu EWD. Co istotne, informacje znajdujące się w bazie danych są publicznie dostępne za pomocą kilku interfejsów dostosowanych do różnych grup odbiorców.

9.1. Dane zawarte w bazie danych

Dane zawarte w bazie podzielić można tematycznie na kilka grup: wyniki testów pisanych przez uczniów wraz z podstawowymi informacjami o uczniach, baza szkół, bank zadań, oszacowania umiejętności uczniów uzyskane w wyniku modelowania statystycznego oraz zagregowane wartości wskaźników PWE i EWD (Rysunek 9.1). Dokładny zakres dostępnych informacji omówiony został poniżej.



Rysunek 9.1. Schemat przepływu danych w bazie (grupy danych w bazie oznaczone prostokątami)

Podczas konstruowania bazy danych dużo wysiłku włożono w to, aby umożliwić łączenie danych pochodzących z różnych lat, a więc np. łączenie ze sobą wyników ucznia z egzaminów kończących kolejne etapy edukacji czy śledzenie zmian wyników w danej szkole, pomimo zmiany identyfikatora szkoły używanego w systemie egzaminów zewnętrznych⁴³. W tym celu baza korzysta z własnych unikalnych identyfikatorów uczniów, szkół czy zadań, przy czym identyfikatory używane w innych systemach (np. kod OKE szkoły, identyfikator egzaminacyjny ucznia, itp.) także są w niej przechowywane i udostępniane.

⁴³ Np. wskutek zmian w strukturze administracyjnej kraju, zmiany organu prowadzącego, itp.

9.1.1. Wyniki uczniów i informacje o uczniach

W bazie przechowywane są wyniki wszystkich uczniów, którzy rozwiązywali arkusze standardowe następujących egzaminów:

- sprawdzianu szóstoklasisty z lat 2002-2014⁴⁴,
- egzaminu gimnazjalnego z lat 2002-2014⁴⁴,
- matury na poziomie podstawowym i rozszerzonym z biologii, chemii, fizyki i astronomii, geografii, fizyki, historii, informatyki, j. angielskiego, j. polskiego, matematyki oraz wiedzy o społeczeństwie z lat 2010-2014⁴⁴,
- matury poprawkowej z j. angielskiego, j. polskiego oraz matematyki z lat 2010-2014⁴⁴,

jak również wyniki testów przeprowadzonych w trakcie prowadzonych badań zrównujących.

Wyniki rejestrowane są na poziomie punktacji uzyskanej przez danego ucznia za każde z zadań, a w wypadku zadań ocenianych w podziale na kilka kryteriów (np. rozprawki z języka polskiego) punktacji uzyskanej za poszczególne kryteria. W odniesieniu do zadań zamkniętych wielokrotnego wyboru często dostępne są także informacje o tym, jaką dokładnie odpowiedź zaznaczył uczeń⁴⁵.

Poza samymi wynikami, zgromadzone zostały także dostępne w systemie egzaminów zewnętrznych dane kontekstowe o uczniach: płeć, rok urodzenia, posiadanie w trakcie danego egzaminu zaświadczenia o dysleksji oraz ew. informacja o byciu laureatem konkursu przedmiotowego⁴⁶.

Łączenie wyników uczniów pomiędzy latami i egzaminami możliwe jest:

- dla sprawdzianu od roku 2003,
- dla egzaminu gimnazjalnego od roku 2006,
- dla matury i matury poprawkowej od roku 2010.

9.1.2. Baza szkół

Zgromadzony w bazie wykaz szkół obejmuje:

- szkoły podstawowe z lat 2004-2014⁴⁴,
- gimnazja z lat 2004-2014⁴⁴,
- licea ogólnokształcące, technika oraz licea profilowane z lat 2010-2014⁴⁴.

⁴⁴ Do końca 2015 roku zostanie uzupełniona o wyniki z roku 2015.

⁴⁵ Dokładne informacje znaleźć można w tabeli na stronie <http://zpd.ibe.edu.pl/doku.php?id=obazie>.

⁴⁶ Co skutkuje przyznaniem uczniowi maksimum punktów za wszystkie zadania.

Dla każdej szkoły przechowywane są informacje o jej charakterystyce (szkoła dla młodzieży czy dla dorosłych, czy jest to placówka specjalna, czy jest to placówka związana⁴⁷), historii zmian identyfikatora OKE szkoły, danych adresowych oraz przypisania położenia szkoły do gminy, powiatu i województwa w czasie, jak również liczbie mieszkańców miejscowości, w której znajduje się szkoła.

9.1.3. Bank zadań

Bank zadań przechowuje informacje o kryteriach oceny, schemacie punktowania⁴⁸, schemacie odpowiedzi (dla zadań wielokrotnego wyboru), występowaniu na testach (egzaminacyjnych i zrównujących) wraz z oznaczeniem poprawnej odpowiedzi w danym teście (dla zadań wielokrotnego wyboru), parametrach statystycznych (KTT oraz IRT), stosowanych w modelowaniu statystycznym przekształceniach skali (np. usuwaniu luk w punktacji, łączeniu kilku zadań w jedno, itp.) oraz, w większości wypadków, treści zadań (w formatach DOCX i HTML). Wyżej wymienione informacje dostępne są dla wszystkich zadań, które wystąpiły w arkuszach standardowych na jakimkolwiek z egzaminów, które wyniki znajdują się w bazie, jak również dla wszystkich zadań nie pochodzących z egzaminów, które wykorzystane zostały w badaniach zrównujących (tzw. kotwica zewnętrzna).

9.1.4. Oszacowania umiejętności uczniów

Jednym z etapów zrównywania jest oszacowanie poziomu umiejętności uczniów na wspólnej skali za pomocą modelu IRT. Oszacowania takie są również wyliczane w procesie przygotowywania wskaźników EWD. Z racji znacznie lepszych od „surowych” wyników punktowych właściwości statystycznych tych oszacowań, polecane jest wykorzystywanie ich w miejsce sum punktów uzyskanych przez uczniów na egzaminach. Dostępne są oszacowania:

- estymatorami EAP oraz PV na zrównanej skali dla uczniów podchodzących do: sprawdzianu w latach 2002-2013, egzaminu gimnazjalnego w latach 2002-2013 oraz matury na poziomie podstawowym z j. angielskiego, j. polskiego i matematyki w latach 2010-2013.
- estymatorami EAP na oddzielnych skalach w każdym roku egzaminu dla: sprawdzianu z lat 2001–2011 oraz 2014⁴⁹, egzaminu gimnazjalnego z lat 2006-2014⁴⁴, jak również matury z j. polskiego, matematyki oraz łącznie wszystkich przedmiotów humanistycznych i wszystkich przedmiotów matematyczno-przyrodniczych z lat 2010-2014⁴⁴.

9.1.5. Zagregowane wartości wskaźników PWE i EWD

Przechowywane są wszystkie udostępniane w serwisach WWW (patrz podrozdział 9.2.1), jak również eksperymentalne i niepublikowane wartości zagregowanych wskaźników PWE i EWD. Wskaźniki PWE dostępne są dla szkół, gmin, powiatów, województw oraz Polski z lat 2002-2013 (dla sprawdzianu oraz egzaminu gimnazjalnego) lub 2010-2013 (dla matury na poziomie podstawowym z j. angielskiego,

⁴⁷ Typowym przykładem placówek związanych mogą być szkoły przyszpitalne.

⁴⁸ Także informacje o występowaniu luk w schemacie punktacji (np. język w ocenie wypracowania z j. polskiego na maturze punktowany na skali 0-1-3-6-9-12 punktów).

⁴⁹ Do końca 2015 roku zostanie uzupełniona o wyniki z roku 2012 oraz 2015.

j. polskiego i matematyki). Z kolei wskaźniki EWD obejmują okresy 2006-2014 dla gimnazjów⁵⁰ oraz 2010-2014 dla liceów ogólnokształcących i techników.

9.2. Dostęp do danych znajdujących się w bazie

Baza danych udostępniana jest na kilka sposobów (za pomocą kilku interfejsów). Różnią się one między sobą na wielu płaszczyznach: zakresem udostępnianych danych, łatwością użycia, możliwością (bądź jej brakiem) integracji z zewnętrznym oprogramowaniem. Zróżnicowanie to ma na celu dostosowanie do potrzeb bardzo zróżnicowanych grup odbiorców od osób zainteresowanych systemem edukacji (np. rodziców, nauczycieli i dyrektorów szkół, pracowników samorządów i administracji rządowej), poprzez naukowców, kończąc na programistach. Poniżej pokrótce omówiono każdy z nich.

9.2.1. Serwisy WWW

Serwisy <http://pwe.ibe.edu.pl> (omówiony w Rozdziale 8), <http://ewd.edu.pl/gimnazjum>, <http://ewd.edu.pl/matura>, umożliwiają samodzielne prowadzenie analiz na zagregowanych wskaźnikach Porównywalnych Wyników Egzaminacyjnych oraz Edukacyjnej Wartości Dodanej. Ich największa zaleta to prosty i intuicyjny sposób obsługi, który nie wymaga od odbiorcy umiejętności wykraczających ponad codzienne korzystanie ze stron internetowych. Dzięki temu są one dostępne dla wszystkich, także np. nauczycieli, rodziców czy przedstawicieli administracji, którzy nie posiadają zaawansowanych umiejętności statystycznych. Omawiane serwisy koncentrują się na analizie wartości wskaźników PWE i EWD za pomocą wykresów, natomiast dostęp do wizualizowanych danych, jakkolwiek możliwy, pełni rolę drugorzędną. Ich największa wada to ściśle ograniczony zakres udostępnianych danych – nie da się za ich pomocą pobrać z bazy danych nic więcej, niż zagregowane wartości wskaźników PWE i EWD.

Do tej kategorii zaliczyć należy również stronę WWW umożliwiającą przeglądanie banku zadań (w tym również pobieranie treści zadań czy podgląd ich parametrów statystycznych) znajdującą się pod adresem <http://zpd.ibe.edu.pl/doku.php?id=bazatestypytnia>. Także w tym wypadku łatwość korzystania z serwisu okupiona jest ograniczeniem zakresu udostępnianych danych do ściśle określonego fragmentu bazy.

9.2.2. API HTTP

Kolejną metodą dostępu do danych zgromadzonych w bazie to API⁵¹ HTTP. Jest to zestaw poleceń, dostępnych za pośrednictwem protokołu HTTP (tego samego, za pomocą którego przekazywana jest treść stron internetowych), które umożliwiają: przeszukiwanie banku szkół oraz informacji o podziale terytorialnym kraju, wyszukiwanie wskaźników PWE oraz EWD dla interesującego nas zakresu lat, egzaminu czy typu szkoły, pobieranie zbiorów danych z wartościami wskaźników PWE lub EWD dla zadanych szkół i/lub jednostek samorządu terytorialnego oraz wizualizację tych zbiorów danych w postaci wykresów (analogicznych do wykresów dostępnych w omawianych powyżej serwisach WWW). Interfejs ten powstał z myślą o programistach, którzy chcieliby w prosty i wygodny sposób zintegrować prezentację wskaźników PWE i/lub EWD z prowadzonymi przez siebie serwisami internetowymi. W oparciu o niego działają np. serwisy <http://pwe.ibe.edu.pl>, <http://ewd.edu.pl/gimnazjum>

⁵⁰ Do końca 2015 roku zostanie uzupełniona o okres 2013-2015.

⁵¹ Application Programming Interface, pol. interfejs programistyczny.

oraz <http://ewd.edu.pl/matura>. Dokładny opis interfejsu znaleźć można na stronie <http://zpd.ibe.edu.pl/doku.php?id=api> [http](http://zpd.ibe.edu.pl/doku.php?id=api).

9.2.3. Pakiet ZPD dla R

Poważnym ograniczeniem wymienionych powyżej sposobów dostępu do danych był stosunkowo wąski zakres możliwych do pobrania informacji. Pakiet ZPD dla R opracowany został z myślą o tym, aby udostępnić możliwie jak najszerszą część danych, jednocześnie nie wymagając od użytkownika nadmiernej wiedzy technicznej ani dokładnej znajomości fizycznej struktury bazy danych. Adresowany jest przede wszystkim do naukowców i analityków, ale także innych osób, dla których zakres danych i analiz udostępnianych w opisanych wyżej serwisach WWW jest niewystarczający.

Licząca blisko 100 tablic fizyczna struktura bazy danych została w pakiecie ZPD dla R uproszczona do 11 ułożonych tematycznie *grup danych*: wyniki testów (*grupa wyniki*), baza uczniów (*uczniowie*), informacje o uczniach w kontekście konkretnego egzaminu (*uczniowieTesty*), zastosowane modele statystyczne (*skale*), oszacowania umiejętności uczniów (*oszacowania*), baza szkół (*szkoły*), baza testów (*testy*), bank zadań (*kryteriaOceny*), parametry statystyczne zadań (*parametry*) oraz zagregowane wartości wskaźników PWE i EWD (*wartościWskaźników*). Dane z poszczególnych grup dają się ze sobą łatwo łączyć dzięki występowaniu w nich wspólnych identyfikatorów (uczniów, zadań, szkół, itd.).

Istotną zaletą pakietu ZPD dla R jest dostępność funkcji automatyzujących najczęściej wykonywane czynności, np. obliczających sumę punktów z testu, normalizujących wyniki egzaminu (np. do skali z czy skali staninowej), odnajdujących dla każdego ucznia jego pierwsze lub ostatnie podejście do wskazanego egzaminu czy poprawnie agregujących wskaźniki PWE. Możliwość skorzystania z tych funkcji pozwala nie tylko przyspieszyć pracę z danymi, ale także ustrzec się błędów, jakie mogłyby się wkręcić przy samodzielnym wykonywaniu tych, niekiedy dość złożonych, przekształceń.

Pakiet ZPD dla R został bardzo dobrze udokumentowany. Na stronie http://zpd.ibe.edu.pl/doku.php?id=r_zpd znaleźć można dokładne opisy poszczególnych *grup danych* i relacji między nimi wraz z przykładami użycia. Dostępny jest tam także wykaz wszystkich dostępnych zmiennych wraz ze wskazaniem, w jakich *grupach danych* występują oraz jakie dokładnie informacje przechowują. W końcu dostępne są tzw. samouczki, czyli rozbudowane przykłady na pobranie i przetworzenie danych znajdujących się w bazie do samodzielnego wykonania i przeanalizowania. Dokumentację uzupełnia omówienie bardziej zaawansowanych aspektów użycia pakietu ZPD dla R, takich jak samodzielne obliczanie zagregowanych wskaźników PWE czy dyskusję nad sposobami minimalizacji czasu pobierania danych z bazy.

Na koniec wypada wyjaśnić, dlaczego zdecydowano się na implementację tego interfejsu akurat w programie statystycznym R, a nie np. w bardziej popularnych w Polsce, jak SPSS, SAS albo Stata. Wymienić można dwie główne przyczyny. Pierwszą była chęć skorzystania z darmowego oprogramowania. Mamy nadzieję, że w ten sposób poszerza się grono potencjalnych użytkowników opisywanego interfejsu dostępu do bazy, nie każdego musi być bowiem stać na zakup licencji na komercyjne oprogramowanie statystyczne. Drugim czynnikiem była minimalizacja czasu niezbędnego do przygotowania interfejsu. Również to kryterium, dzięki doskonałej integracji z relacyjnymi bazami danych⁵², R spełniał najlepiej.

⁵² W szczególności dostępności pakietu dplyr.

9.2.4. Bezpośrednie wykonywanie zapytań SQL na bazie

Ostatnią możliwością dostępu do danych w bazie jest bezpośrednie połączenie z oprogramowaniem serwera bazy danych (używany system baz danych to PostgreSQL) i samodzielne formułowanie zapytań do bazy w języku SQL. Jakkolwiek daje to największe możliwości zarówno pod względem zakresu dostępnych danych (de facto cała baza), jak i potencjalnego zredukowania czasu ich pobrania, wymaga od użytkownika znacznej wiedzy z zakresu relacyjnych baz danych i biegłości w operowaniu językiem SQL. Nieumiejętne korzystanie z tego interfejsu rodzi niebezpieczeństwo pobrania innych danych, niż się zamierzało, jak również pobierania ich w taki sposób, który nie będzie się w stanie zakończyć w rozsądnym czasie. Dodatkowym utrudnieniem jest brak dokładnej dokumentacji. Dostępny jest co prawda diagram struktury fizycznej⁵³, na którym oznaczono wszystkie tablice wraz z kolumnami, relacje między tablicami, klucze podstawowe oraz indeksy, brak jednak opisów poszczególnych tablic i kolumn⁵⁴. W zdecydowanej większości wypadków te same dane da się łatwiej i w krótszym czasie pobrać używając pakietu ZPD dla R i zalecane jest korzystanie właśnie z niego. Pomimo wszystkich wymienionych obostrzeń istnienie tego interfejsu jest niezbędne, jest on bowiem wykorzystywany wewnątrz pakietu ZPD dla R oraz API HTTP. Warto nadmienić, że za pośrednictwem ODBC można w ten sposób uzyskać dostęp do bazy z większości programów używanych do przetwarzania danych (zarówno programów statystycznych jak SPSS, Stata czy R, jak i np. z arkuszy kalkulacyjnych, choćby MS Excel).

⁵³ http://zpd.ibe.edu.pl/lib/exe/fetch.php?media=struktura_bazy.svg

⁵⁴ Można posiłkować się wykazem zmiennych przygotowanych dla pakietu ZPD dla R (http://zpd.ibe.edu.pl/doku.php?id=r_zmienne), jednak analogiczny wykaz dla wszystkich kolumn w fizycznej strukturze bazy danych nie istnieje.

10. Rekomendacje – możliwe scenariusze rozwoju zrównywania wyników polskich egzaminów

Wprowadzenie do systemu egzaminacyjnego mechanizmów pozwalających na bieżące działania umożliwiające poprawne porównywanie wyników egzaminacyjnych jest kwestią niezwyklej wagi. Zauważyła to Najwyższa Izba Kontroli, która wśród najważniejszych rekomendacji raportu *System egzaminów zewnętrznych w oświacie* wnioskuje do Dyrektora CKE o *wprowadzenie na etapie tworzenia testów mechanizmu umożliwiającego coroczne zrównywanie wyników egzaminów zewnętrznych z kolejnych lat, tak aby zapewnić możliwość ich porównywania* (NIK, 2015: s. 9–10). Korzyści płynące z takiego rozwiązania były wielokrotnie podkreślane w niniejszej monografii, lecz powtórzmy je raz jeszcze. Zapewnienie corocznej porównywalności wyników umożliwi sprawiedliwe działanie systemów selekcyjnych. Uczniowie zdający egzaminy w różnych sesjach będą mieli równe szanse dostania się do wybranych placówek edukacyjnych, adekwatne do ich poziomu umiejętności (a nie własności rozwiązywanego arkusza egzaminacyjnego). Porównywalne wyniki egzaminacyjne są ponadto wskaźnikami, które mogą być wykorzystane do monitorowania i ewaluacji jakości pracy szkoły, gminy, powiatu lub województwa w zakresie realizacji zadań edukacyjnych. Inaczej niż inne dostępne wskaźniki (choćby EWD), porównywalne między latami wyniki egzaminacyjne mogą służyć do ewaluacji całego systemu oświatowego, poprzez monitorowanie zmian poziomu umiejętności pełnych populacji. Wprowadzenie procedur zapewniających porównywalność wprowadzi też mechanizmy umożliwiające kontrolowanie różnej trudności arkuszy egzaminacyjnych (która zawsze występuje do pewnego stopnia, niezależnie od działań podejmowanych w celu zapewniania porównywalności własności egzaminów), a tym samym może przyczynić się do podniesienia ich jakości.

W niniejszym rozdziale opisane zostaną rozwiązania stosowane na świecie w zakresie zrównywania wyników egzaminacyjnych oraz warianty, które mogą zostać wykorzystane w Polsce do konstrukcji porównywalnych między latami (kolejnymi sesjami) wyników dla egzaminów powszechnych, czyli dla: sprawdzianu, egzaminu gimnazjalnego oraz egzaminu maturalnego z przedmiotów obowiązkowych. W przypadku egzaminów obejmujących mniejsze grupy uczniów (jak np. egzaminy zawodowe, bądź egzaminy maturalne z przedmiotów dodatkowych) należałoby poszukać nieco odmiennych strategii zapewniania porównywalności wyników. Należałoby ocenić relację pomiędzy nakładami niezbędnymi do zapewnienia porównywalności wyników do ewentualnych korzyści wdrożenia takich rozwiązań, biorąc pod uwagę fakt, że egzaminy te dotyczą niewielkiej części populacji uczniów. W skrajnych przypadkach, tj. egzaminów rozwiązywanych przez bardzo nieliczne grupy uczniów, zapewnienie porównywalności ich wyników może być uzyskanie jedynie poprzez odwołanie się do opinii ekspertów na temat relatywnej trudności.

Rozpatrywane do wdrożenia w Polsce warianty obejmują trzy rozwiązania, które można traktować jako docelowe do wprowadzenia w Polsce systemowo. Warianty te jednak w mniejszym lub większym stopniu wiążą się z koniecznością zmiany funkcjonującego w Polsce systemu egzaminów zewnętrznych i związanego z tym systemu legislacyjnego. Te trzy warianty zostaną przedstawione w kolejności ze względu na najbardziej satysfakcjonujące własności metodologiczne:

- 1) Wariant A – włączenie do egzaminu niejawnych zadań kotwiczących. Z metodologicznego i pomiarowego punktu widzenia jest to rozwiązanie optymalne, dające najbardziej precyzyjne oszacowanie funkcji zrównującej i jest najmniej narażone na potencjalne czynniki zakłócające. Rozwiązanie to wymaga jednak istotnych zmian w sposobie przeprowadzania egzaminów.

- 2) Wariant B – jednoczesna standaryzacja zadań z wielu edycji egzaminu. Rozwiązanie odznacza się mniejszą efektywnością wyznaczania funkcji zrównującej niż w wariantcie A. Jest też obciążone potencjalnie zakłócającym czynnikiem związanym z innym poziomem motywacyjnym uczniów na sesji standaryzacyjnej niż na sesji egzaminacyjnej. Jeżeli czynnik zróżnicowanej motywacji można by uznać za zaniedbywalny lub podatny kontroli, to wariant B umożliwiałby poprawne zrównywanie egzaminów między latami.
- 3) Wariant C – dodatkowe badania na reprezentatywnej próbie. Jest to rozwiązanie nie wprowadzające żadnych zmian w systemie egzaminacyjnym, przez co byłoby najprostsze w implementacji. Jednocześnie wariant C wyróżnia się największą liczbą potencjalnych problemów i możliwości zakłóceń procesu, w którym ustalana będzie porównywalność między egzaminami. Ponadto odznacza się najniższą efektywnością. Ze względu na znaczne błędy standardowe przydatność tego rozwiązania do zrównywania wyników pojedynczych uczniów jest ograniczona. Zaletą tego rozwiązania jest możliwość śledzenia zmian w poziomie umiejętności populacji między latami oraz zbieranie informacji o stabilności trudności testów egzaminacyjnych między latami.

Analizując proponowane rozwiązania, uwzględnione zostaną takie czynniki warunkujące ich wdrożenie jak: koszty finansowe, możliwa do osiągnięcia precyzja zrównywania wyników, ograniczenia czasowe i organizacyjne czy konieczność przekonania opinii publicznej. W kontekście precyzji zrównywania, w każdym z wariantów podane zostaną wielkości błędu standardowego przy porównywaniu średnich między dwoma latami, przy założeniu że ich prawdziwa różnica wynosi 0,1 na skali o odchyleniu standardowym 1. Taka różnica średnich między latami odpowiada rzędowi zmian średniego poziomu umiejętności zaobserwowanemu w przeprowadzonych badaniach zrównujących. Uzyskane w symulacjach wielkości błędów standardowych pozwolą na porównanie relatywnej efektywności każdego z wariantów. Szczegóły badania symulacyjnego w każdym z wariantów zostaną opisane osobno.

Należy również pamiętać, że w roku 2015 zmieniono formułę standaryzowanych egzaminów zewnętrznych (sprawdzian w szóstej klasie szkoły podstawowej i matura), tak więc rekomendowane rozwiązania dotyczą wdrożenia procedury zrównywania od 2016.

10.1. Rozwiązania stosowane na świecie

Wiele narodowych systemów egzaminacyjnych zawiera mechanizmy pozwalające na zrównywanie wyników z różnych sesji testowych. Jednymi z najczęściej opisywanych w literaturze testów są np. SAT (*Scholastic Assessment Test*), ACT (*American College Testing*), PET (*Psychometric Entrance Test*) oraz SweSAT (*Swedish Scholastic Assessment Test*)⁵⁵.

SAT i ACT są testami wysokiej stawki, których wyniki brane są pod uwagę podczas rekrutacji na uczelnie wyższe w Stanach Zjednoczonych. Rozwiązania wykorzystywane w SAT i ACT opierają się na założeniu, że zarówno zadania kotwiczące, jak i zadania będące podstawą punktowania uczniów są niejawnie, co w znacznym stopniu ułatwia zastosowanie procedury zrównywania.

Obecnie SAT składa się z dziewięciu sekcji testowych oraz jednej sekcji zrównującej. Trzy sekcje mierzą umiejętność czytania ze zrozumieniem, kolejne trzy – umiejętności matematyczne, a następne – umiejętności wypowiedzi pisemnej. Sekcja zrównująca obejmuje zagadnienia z jednej dziedziny wiedzy (czytanie ze zrozumieniem, matematyka lub pisanie). Uczniowie nie są świadomi, która sekcja

⁵⁵ Dokładny opis tych testów znajduje się w artykule Artura Pokropka i Bartosza Kondratka (2012).

jest sekcją zrównującą. Końcowy wynik ucznia określa się wyłącznie na podstawie 170 pytań z sekcji testowej - odpowiedzi na zadania sekcji zrównującej nie są brane pod uwagę podczas raportowania tego wyniku. Obecnie rokiem bazowym, do którego przyrównuje się zrównane wyniki SAT jest rok 1995. Mimo zmian w procedurze zrównywania, cały czas stosuje się schemat dla planu nierównoważnych grup z testem kotwiczącym (NEAT) (zob. podrozdział 4.5). Wyniki są skalowane i zrównywane metodą ekwicyntylową (zob. podrozdział 4.7), a następnie przedstawiane na jednej, zagregowanej skali z przedziału 600–2400 punktów oraz na trzech osobnych podskalach: dla czytania ze zrozumieniem, matematyki i pisanie z przedziału 200–800.

ACT mierzy cztery rodzaje umiejętności: posługiwania się językiem angielskim, matematyczne, czytania ze zrozumieniem oraz rozumowania w naukach przyrodniczych. Uczeń może podejść także do dodatkowej sekcji: pisanie. Zrównywanie odbywa się metodą ekwicyntylową. Wyskalowane wyniki testu przedstawiane są na skali od 1 do 36 punktów, gdzie punkty są liczbami całkowitymi. Wyniki są publikowane w podskalach odpowiadających poszczególnym dziedzinom wiedzy. Zrównywanie w teście ACT odbywa się na podstawie schematu z równoważnymi grupami. Aby przeprowadzić zrównanie dwóch testów z różnych lat, spośród wszystkich uczniów losowana jest reprezentatywna próba. Uczniowie należący do wylosowanej próby, oprócz aktualnej edycji testu, rozwiązują też kilka nowych, wcześniej niepublikowanych arkuszy egzaminacyjnych. Jako że w populacji zrównującej uczniowie rozwiązywali zadania z testu właśnie przeprowadzonego oraz zadania z testów, które dopiero mają się odbyć w kolejnych sesjach, możliwe jest zrównanie wyników z testu już przeprowadzonego z kolejnymi edycjami. W teście ACT do zrównywania używana jest metoda ekwicyntylowa wykorzystująca analityczne metody wygładzania rozkładów (Kolen, 1984; ACT, 2007).

W Izraelu funkcjonuje *Psychometric Entrance Test* (PET), którego wynik jest brany pod uwagę przy rekrutacji na uczelnie wyższe. Od roku 1990 PET składa się z trzech sekcji: rozumowania werbalnego, rozumowania ilościowego oraz znajomości języka angielskiego (Beller, 1994; Allalouf i Ben Shakhar, 1998). Schemat zrównywania jest analogiczny do schematu amerykańskiego SAT. W danej edycji izraelscy uczniowie rozwiązują sześć takich samych sekcji testowych, ale za to różne sekcje zrównujące, co oznacza, że zrównywanie prowadzi się według planu nierównoważnych grup z testem kotwiczącym. Jedna sekcja zrównująca testu PET rozwiązywana jest zawsze przez około 1000 egzaminowanych. W sekcji zrównującej mogą zawierać się pytania ze starych testów. Dla każdego testu i dla każdej umiejętności wykorzystuje się proste zrównywanie liniowe (Beller, 1994; Rapp 1999).

SweSAT (*Swedish Scholastic Assessment Test*) jest egzaminem decydującym o przyjęcie na szwedzkie uczelnie wyższe. Test składa się z sześciu części: znajomości słownictwa; czytania ze zrozumieniem; czytania ze zrozumieniem tekstów angielskich; umiejętności matematycznych, umiejętności interpretowania danych (głównie wykresów, tabel i map) oraz wiedzy ogólnej. Wszystkie zadania w teście są zadaniami zamkniętymi, punktowanymi na skali 0–1 (Stage i Ögren, 2004). Surowy wynik skalowany jest za pomocą metody ekwicyntylowej i przekształcany na skalę z punktacją z przedziału od 0,0 do 2,0 punktów. Test zrównywany jest przy założeniu, iż populacje z roku na rok się nie zmieniają. Zrównywanie polega na przekształceniu wyników surowych metodą ekwicyntylową przy uwzględnieniu płci, wieku oraz pochodzenia społecznego uczniów. Funkcja zrównująca wybierana jest w taki sposób, by z roku na rok wyniki egzaminacyjne w poszczególnych podgrupach utworzonych ze względu na wymienione zmienne nie różniły się (Stage, 2004).

10.2. Warianty docelowo rekomendowane dla polskiego systemu egzaminów zewnętrznych

10.2.1. Wariant A – włączenie do egzaminu niejawnych zadań kotwiczących

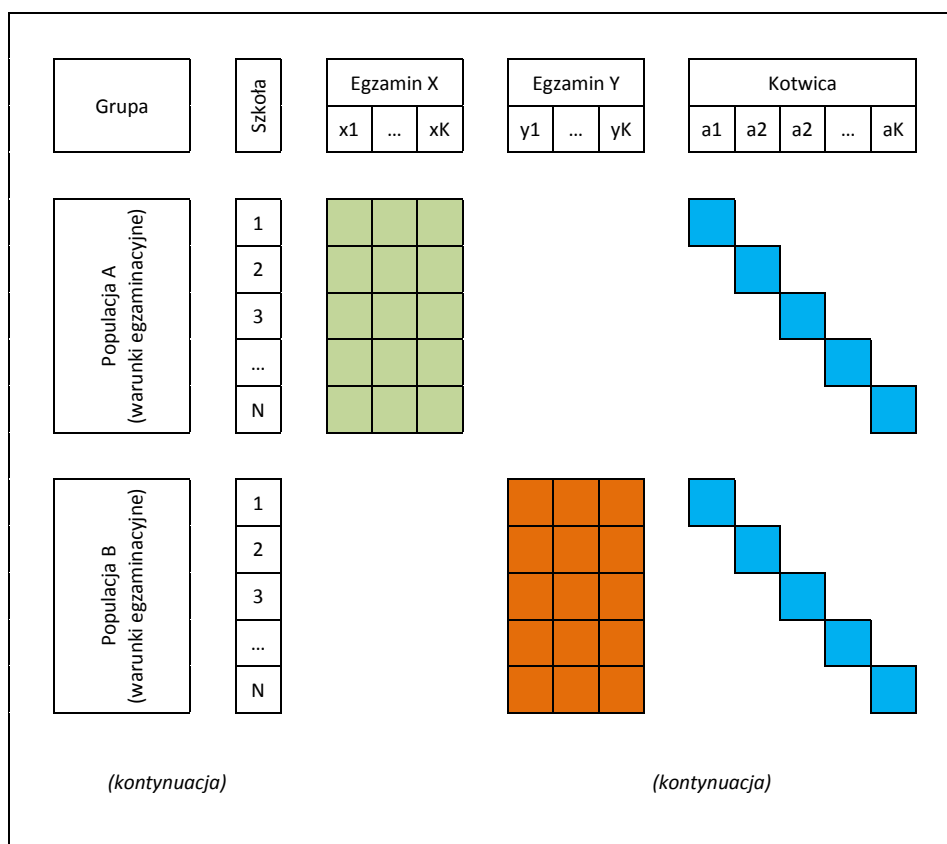
Optymalne rozwiązanie dla zapewnienia systemowej porównywalności wyników zakłada wykorzystanie schematu łączenia wyników opartego na wspólnych zadaniach kotwiczących (zob. Rysunek 10.1). Oznacza to konieczność włączenia do arkuszy egzaminacyjnych pewnej liczby zadań, które byłyby niejawne oraz powtórnie wykorzystywane w kolejnych edycjach egzaminów. W tym scenariuszu wyniki zadań niejawnych nie byłyby wliczane do rezultatu ucznia z danego egzaminu. Zdający nie mogliby jednak wiedzieć, które zadania pochodzą z właściwej części egzaminu, wpływającej na uzyskany wynik, a które są elementem części kotwiczącej.

Zadań niejawnych powinno być możliwie dużo w całej populacji zdających, lecz jeden uczeń rozwiązywałby stosunkowo niewiele z nich. Można to osiągnąć poprzez stworzenie wielu wariantów arkuszy egzaminacyjnych, które różniłyby się tylko zadaniami kotwiczącymi. Takie rozwiązanie jest podyktowane kilkoma czynnikami. Po pierwsze, czym mniej zadań niejawnych będzie rozwiązywanych przez pojedynczego ucznia (przez uczniów z pojedynczej szkoły), tym mniejsze będzie ryzyko ich ujawnienia na przestrzeni lat. Dążenie do ograniczenia liczby zadań kotwiczących przypadających na pojedynczego ucznia (nawet do jednego zadania) jest również wskazane w celu zminimalizowania kosztów (np. czasu), jakie ponosi uczeń rozwiązując zadania, których wynik nie wpływa na rezultat egzaminu. Zapewnienie jak największej liczby zadań niejawnych na poziomie całej populacji zmniejsza również ryzyko ujawnienia całej treści części kotwiczącej. Jednocześnie takie postępowanie zwiększa trafność pomiaru umiejętności, ponieważ możemy zbadać szersze spektrum z uniwersum treści, których opanowanie sprawdzane jest na egzaminie.

Jest to klasyczny schemat zrównywania oparty na wewnętrznej kotwicy stosowany między innymi w amerykańskim egzaminie SAT. Dwie populacje rozwiązują dwa różne egzaminy, jednak obydwa zawierają część wspólną – "kotwicę", czyli nieujawniane zadania, różne dla uczniów z różnych szkół. Dzięki takiemu podejściu możliwe jest uzyskanie dużej liczby zadań zrównujących przy minimalizacji liczby zadań kotwiczących dla każdego z ucznia.

Dla oszacowania precyzji z jaką w tym wariancie można szacować na wspólnej skali relatywny poziom umiejętności uczniów z różnych edycji egzaminu przeprowadzono eksperyment Monte Carlo odpowiadający schematowi ilustrującemu wariant A (zob. Rysunek 10.1). Rozkład umiejętności w populacji A miał średnią 0 i odchylenie standardowe 1, natomiast w populacji B, takie samo odchylenie standardowe i średnią 0,1. Dla symulowania odpowiedzi na zadania z egzaminu X oraz Y wykorzystano parametry IRT dla zadań maturalnych z matematyki z poziomu podstawowego z lat 2012 oraz 2013. Przyjęto, że w każdej szkole uczniowie oprócz zadań egzaminacyjnych w arkuszu będą mieli tylko jedno dodatkowe niejawne zadanie kotwiczące, które będzie oceniane 0–1. Zadanie kotwiczące wykorzystane w danej szkole było losowo wybierane z puli 200 zadań kotwiczących, parametry 200 dychotomicznych zadań kotwiczących wzięto z oszacowań IRT dla zadań zamkniętych z podstawowej matury z matematyki w latach 2010-2012. Przyjęto, iż w każdej populacji znajduje się 7000 szkół, w każdej szkole jest 50 uczniów, a współczynnik korelacji wewnątrzklasowej (ICC) wynosi 0,5. Należy zauważyć, że przyjęte warunki (tylko jedno zadanie kotwiczące na ucznia, takie samo w szkole i oceniane jedynie na skali 0-1), są skrajnie minimalistyczną wersją wariantu A, która odpowiadałaby najniższej możliwej do uzyskania w tym wariancie mocy wnioskowania statystycznego. Zwiększenie liczby zadań przypadających na ucznia, zwiększenie liczby punktów przypadających na zadanie lub zwiększenie liczby zadań wykorzystanych w pojedynczej szkole istotnie zmniejszałoby błąd zrównywania. Przy tak

wyznaczonych warunkach przeprowadzono 200 replikacji eksperymentu uzyskując błąd oszacowania średniej w populacji Y równy: $SE_A=0,014$. Wartość zostanie zestawiona z wynikami dla analogicznych symulacji dla wariantów A oraz B.



Rysunek 10.1. Wariant A – włączenie do egzaminu niejawnych zadań kotwiczących

Przedstawione w rozwiązane takie posiada następujące zalety:

- 1) Motywacja i poziom umiejętności jest taki sam dla zadań egzaminacyjnych i zadań kotwiczących, zatem motywacja uczniów nie wpływa na wyniki zrównywania.
- 2) Uzyskanie bardzo precyzyjnych oszacowań funkcji zrównujących w związku z dużą próbą.
- 3) Możliwość zrównania nie tylko pierwszego terminu egzaminu, ale też innych terminów oraz nie tylko poziomu podstawowego egzaminu maturalnego, ale także rozszerzonego.
- 4) Standaryzacja zadań do banku zadań – jako materiał do budowania arkuszy egzaminacyjnych w przyszłych kolejnych latach (zadania kotwiczące po pewnym czasie stają się „normalnymi” zadaniami).
- 5) Niższe koszty w porównaniu z innymi wariantami, gdyż nie jest wymagane przeprowadzanie dodatkowych, reprezentatywnych badań.

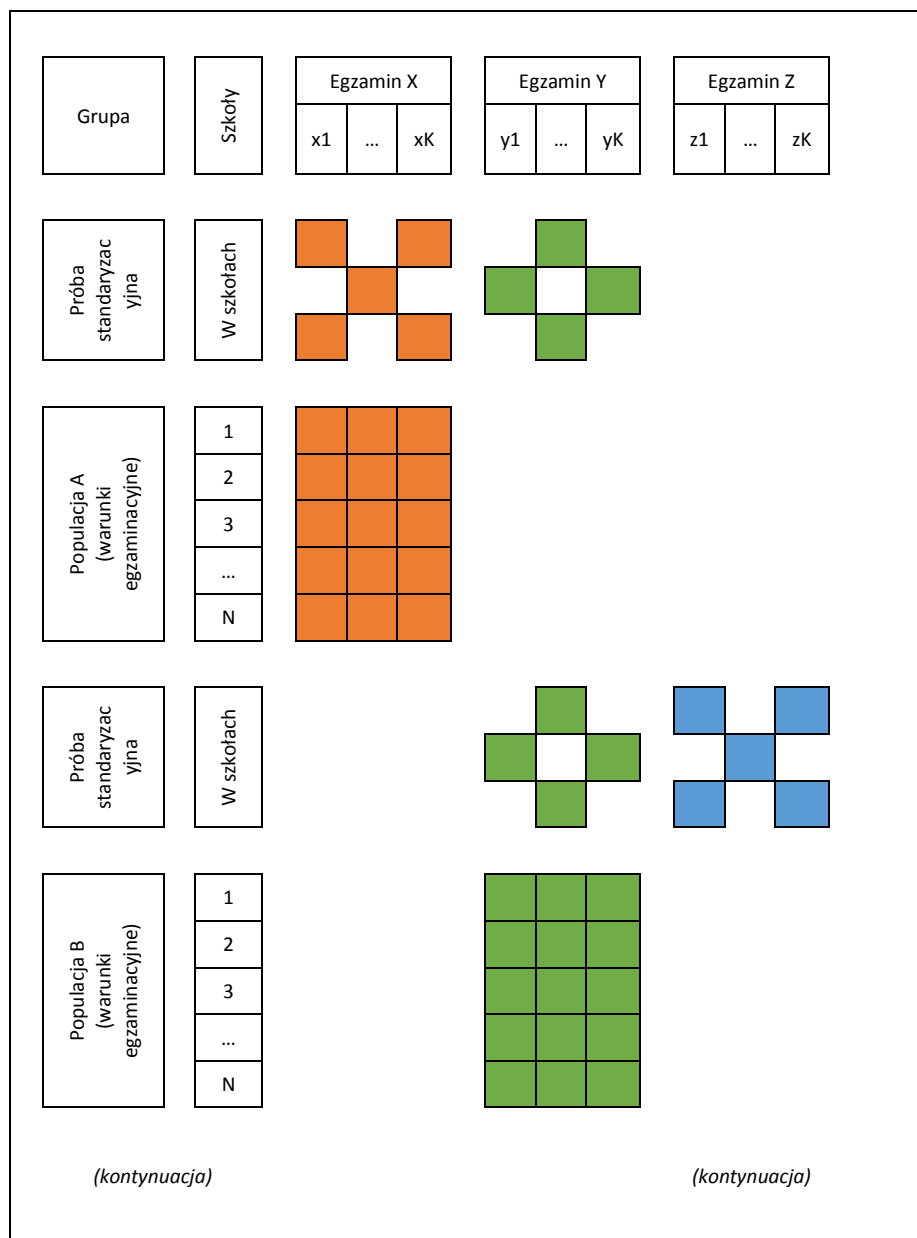
Proponowany wariant wymaga spełnienia następujących wymagań:

- 1) Konieczność wprowadzenia regulacji prawnych umożliwiających włączenie do arkuszy zadań, które nie będą jawne i nie będą wliczane do rezultatu egzaminu (w tym sankcje za nieuprawnione upublicznienie zadań kotwiczących).
- 2) Konieczność przekonania opinii publicznej, że takie rozwiązanie jest korzystne dla zdających, a poniesiony przez nich wysiłek na rozwiązanie dodatkowych zadań jest kompensowany przez realizację zasady wyrównania szans (procesy rekrutacyjne na podstawie wyników różnych egzaminów).
- 3) Wprowadzenie dużej liczby zadań zrównujących do obiegu w pierwszej edycji takiego egzaminu i konieczność corocznego wprowadzania nowych zadań.
- 4) Konieczność stworzenia mechanizmu kontroli związanego z potencjalnym, nieuprawnionym ujawnianiem zadań kotwiczących.
- 5) Wprowadzenie bardziej złożonych niż dotychczas procesów druku i dystrybucji arkuszy egzaminacyjnych zgodnie z planem występowania w nich zadań niejawnych, tak aby jedna szkoła dostawała tylko jeden zestaw zadań dodatkowych w arkuszu egzaminacyjnym.
- 6) Konieczność uwzględnienia zadań kotwiczących podczas składu arkuszy egzaminacyjnych. W celu uniknięcia tworzenia bardzo dużej liczby arkuszy niezależnie, należałoby stworzyć elastyczną strukturę arkusza umożliwiającą wymienianie jego części kotwiczących.

Wymagania opisywane w punktach 4, 5 i 6 byłyby z pewnością łatwiejsze do spełnienia, gdyby egzaminy były rozwiązywane komputerowo, co wydaje się jednocześnie nieuniknionym kierunkiem rozwoju systemu egzaminów zewnętrznych. Przy komputerowym przeprowadzaniu egzaminów problemy związane ze składem, drukiem i dystrybucją arkuszy zostałyby wyeliminowane oraz ujawnienie zadań kotwiczących byłoby znacznie trudniejsze (oczywiście przy założeniu, że baza, w której przechowywano by zadania kotwiczące byłaby należycie zabezpieczona przed ewentualnymi atakami).

10.2.2. Wariant B – jednoczesna standaryzacja zadań z wielu edycji egzaminu

Rozwiązanie polegałoby na przeprowadzaniu w jednym roku standaryzacji arkuszy dla przynajmniej dwóch edycji egzaminu. W tym scenariuszu uczniowie w dodatkowym badaniu rozwiązują zadania, które dopiero mają zostać użyte w przyszłych egzaminach. Warto podkreślić, iż do zrównania potrzebna jest próba uczniów, nie cała populacja. Schemat takiego rozwiązania przedstawia Rysunek 10.2. Można zauważyć, że badania zrównujące prowadzone przez IBE zostały zrealizowane przy pomocy podobnego schematu. Różnica jest jednak taka, że dodatkowe badania zrównujące służące do zebrania informacji o relatywnej łatwości zadań z różnych egzaminów były w badaniu IBE przeprowadzane *ex post*, a tutaj musiałoby to zostać dokonane *ex ante*.



Rysunek 10.2. Wariant B – jednoczesna standaryzacja zadań z wielu edycji egzaminu

W celu oszacowania precyzji szacowania funkcji zrównujących w omawianym wariancie przeprowadzono analogiczne symulacje jak przypadku wariantu A. W symulacjach przyjęto, że próba standaryzacyjna przed administracją egzaminu X jest reprezentatywna dla populacji A oraz liczy 5000 uczniów z 100 szkół (każda szkoła po 50 uczniów). Uczniowie w próbie standaryzacyjnej rozwiązywali dwa arkusze, każdy stworzony z połowy zadań egzaminu X oraz Y, a arkusze rozprowadzono losowo wewnątrz szkół. Wszystkie pozostałe warunki były identyczne jak te opisane dla wariantu A. Uzyskana przy tych warunkach dla wariantu B wielkość błędu oszacowania średniej w populacji Y wyniosła: $SE_B=0,038$, co jest błędem ponad dwukrotnie większym niż dla wariantu A ($SE_A=0,014$).

Zalety takiego rozwiązania:

- 1) Nie jest konieczne zapewnienie reprezentatywności próby. Wszystkie czynniki losowe mogące wpływać na rozwiązanie zadań z różnych egzaminów oddziałują tak samo.

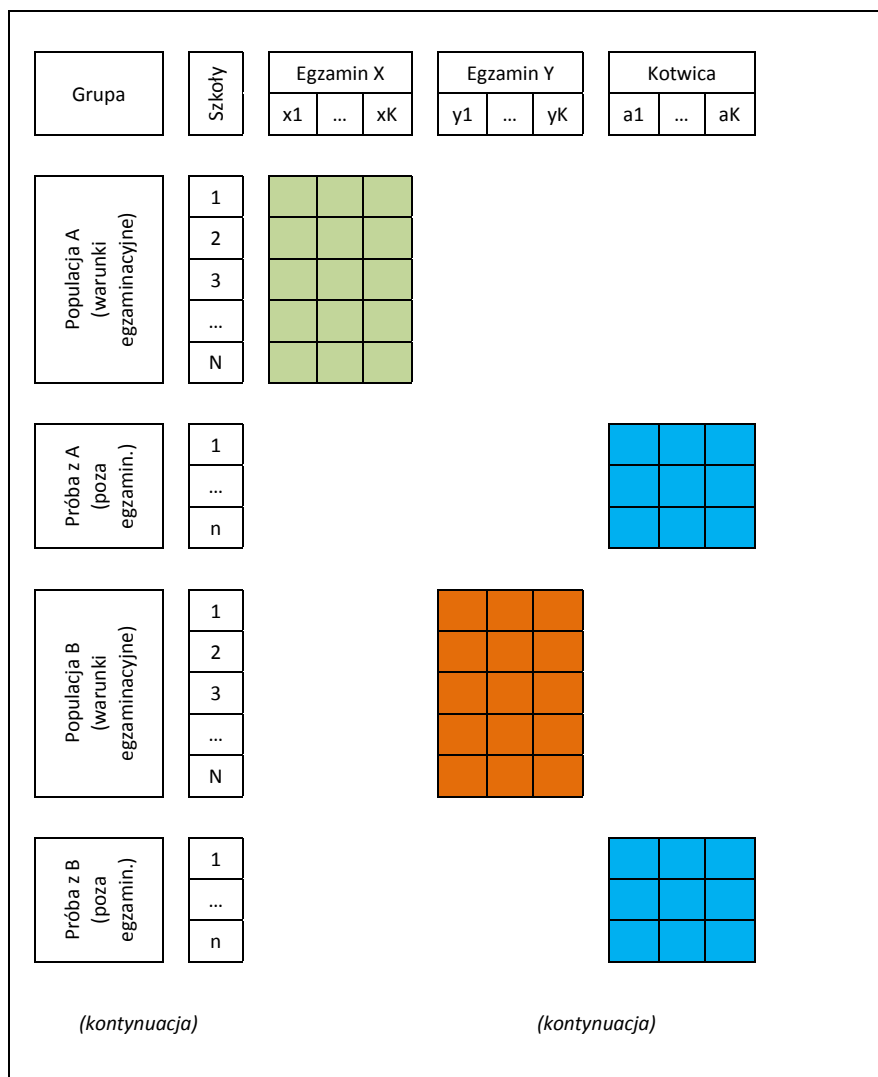
- 2) Schemat ten daje dość dużą moc statystyczną (np. w porównaniu z wariantem C).
- 3) Każde zadanie egzaminacyjne jest standaryzowane. Globalnie przyjęcie takiego schematu nie generuje znaczących dodatkowych kosztów, a polega jedynie na zmianie organizacji standaryzacji.
- 4) Schemat ten umożliwia zrównywanie wszelkich egzaminów i edycji egzaminów, włączając w to sesje poprawkowe, różne poziomy trudności (poziom podstawowy/rozszerzony na maturze) etc.

Wady i problemy związane z wprowadzeniem takiego rozwiązania:

- 1) Wpływ motywacji w tym schemacie jest niekontrolowany i nie zbadany. Zrównanie mogłoby być szczególnie narażone na obciążenie wyników, gdyby wpływ motywacji był niejednorodny dla różnych zadań.
- 2) Po standaryzacji zmiany w treści zadań są niedopuszczalne (w dotychczasowej praktyce zmiany w treści zadań po standaryzacji nie są niczym wyjątkowym).
- 3) Kolejność zadań w teście i inne efekty kontekstu arkusza, mogą wpływać na wyniki zrównywania, co prowadzi do obniżenia ich trafności.
- 4) Konieczność dysponowania większą liczbą zadań do standaryzacji niż to ma miejsce przy obecnych rozwiązaniach przygotowywania arkuszy egzaminacyjnych.

10.2.3. Wariant C – dodatkowe badania na reprezentatywnej próbie

Scenariusz tego rozwiązania polegałby na przeprowadzeniu dodatkowego badania na reprezentatywnej próbie uczniów mających pisać dany egzamin tuż przed prawdziwą datą egzaminu. Schemat takiego rozwiązania przedstawia Rysunek 10.3. W każdym roku reprezentatywna próba uczniów musiałaby rozwiązać takie same zadania (albo przynajmniej część takich samych), które pozwoliłyby określić poziom umiejętności w badanych populacjach i na tej podstawie dokonać zrównania wyników.



Rysunek 10.3. Wariant C – dodatkowe badania na reprezentatywnej próbie

Przy ocenie relatywnej precyzji szacowania funkcji zrównujących w wariantcie C przeprowadzono analogiczne symulacje jak przypadku wariantu A i B. W symulacjach przyjęto, że próby z populacji A oraz z populacji B liczą po 2500 uczniów każda (co łącznie daje liczbę 5000 uczniów – jak w symulacjach dla wariantu B). Dla zasymulowania wyników testu kotwiczącego, wykorzystano parametry IRT oszacowane dla zadań egzaminu maturalnego z matematyki w roku 2011. Wszystkie pozostałe warunki były identyczne jak te opisane dla wariantu A oraz B. Uzyskana przy tych warunkach dla wariantu C wielkość błędu oszacowania średniej w populacji Y wyniosła: $SE_C=0,144$, co jest błędem dziesięciokrotnie większym niż dla wariantu A ($SE_A=0,014$) oraz prawie czterokrotnie większym niż dla wariantu B ($SE_B=0,038$). Przy tak znacznym błędzie standardowym oraz przyjętych warunkach symulacji schemat badania w wariantcie C nie dostarczyłby nawet wystarczających przesłanek do stwierdzenia, że różnica między średnim poziomem umiejętności między populacjami, równa 0,1, jest różna w sposób istotny statystycznie.

Zalety takiego rozwiązania:

- 1) To rozwiązanie jest najłatwiejsze pod względem organizacyjnym z trzech docelowych wariantów.

- 2) Wprowadzenie tego rozwiązania nie wymaga zmian obecnego systemu egzaminacyjnego i może być przeprowadzane poza nim.
- 3) Nie wymaga zmian procedur standaryzacyjnych, organizacji egzaminu ani zmian legislacyjnych. Jest to po prostu dodatkowe badanie, na podstawie którego wyniki z poszczególnych lat mogą być ze sobą związane.

Wady i problemy takiego rozwiązania:

- 1) Rozwiązanie to jest uzależnione od możliwości uzyskania reprezentatywnej informacji o poziomie umiejętności uczniów w danym roku.
 - a) Wymaga to dobrania losowej próby najlepiej z kompletnego i aktualnego operatu (o ile operaty szkół istnieją, to ich kompletność i aktualność nie zawsze jest zagwarantowana).
 - b) Odmowy w udziału w badaniu zarówno szkół, jak i poszczególnych uczniów mogą znacząco obciążać wyniki.
 - c) Nie kontrolujemy tutaj czynników losowych jak: pogoda, ważne wydarzenia społeczne i polityczne odbywające się zaraz przed badaniem i mogące wpływać na jego wyniki.
- 2) Tak jak i w schemacie B istnieje problem zróżnicowanej motywacji.
- 3) Moc wnioskowania statystycznego jest najniższa z wśród wszystkich omawianych schematów, gdyż zrównywanie będzie się opierało na porównaniu rozkładów z dwóch populacji na podstawie tego samego testu. Dodatkowo precyzja takiego oszacowania podlega obniżeniu ze względu na istotną wariancję międzyszkolną. Większa próba (zwiększanie mocy statystycznej) oznacza proporcjonalnie większy koszt oraz większe prawdopodobieństwo ujawnienia treści zadań. Znaczny błąd zrównywania ogranicza możliwość wykorzystania tego schematu do ustalania porównywalnych wyników na poziomie pojedynczych uczniów.
- 4) To rozwiązanie pozwalałoby jedynie na zrównanie części podstawowych egzaminu maturalnego, gdyż tylko one są powszechne. Dla egzaminów poprawkowych oraz egzaminów przeprowadzanych w innych terminach oraz częściach rozszerzonych egzaminów maturalnych mamy do czynienia z nielosowym doбором uczniów do egzaminu i wylosowanie odpowiedniej próby byłoby skrajnie trudne i kosztowne (odpowiedni operat i badanie po kilku uczniów w szkole dla każdego arkusza).

10.3. Podsumowanie

Prezentowane powyżej rozwiązania cechują się zróżnicowanym stopniem wykonalności. Wariant C nie ingeruje w funkcjonowanie systemu egzaminów zewnętrznych, jednak cechuje się bardzo dużym błędem statystycznym, ograniczeniem do zrównywania jedynie podstawowych części egzaminu oraz uzależnieniem jakości wyników od jakości doboru próby. Wariant B, natomiast, nie kontroluje poziomu motywacji uczniów, wymaga stworzenia znacznej liczby zadań do standaryzacji i wiąże się z ryzykiem

ujawnienia treści zadań przed właściwym egzaminem. Zalety wariantu A są bezsprzeczne: duża moc wnioskowania statystycznego, możliwość zrównywania wyników dla wszystkich arkuszy egzaminacyjnych, brak obciążenia wyników zróżnicowanym poziomem motywacji uczniów, standaryzacja zadań oraz zdecydowane niższe koszty niż w przypadku badań niezbędnych do wdrożenia wariantu C lub B. Jednakże jest to wariant najtrudniejszy do realizacji: wymagający zmian prawnych utajniających kotwiczącą część egzaminu⁵⁶ oraz konieczność implementacji mechanizmów kontrolujących potencjalne ujawnianie zadań kotwiczących. Dodatkowo, wariant ten może spotkać się z oporem opinii publicznej.

⁵⁶ Jest to sprzeczne z § 20a Rozporządzenia Ministra Edukacji Narodowej i Sportu z dnia 11 września 2002r., który stanowi, że na prośbę absolwenta, sprawdzone i ocenione zadania egzaminacyjne są udostępniane absolwentowi (MEN, 2002).

Bibliografia

- ACT (2007). *Technical Manual*. Pobrano z: http://www.act.org/aap/pdf/ACT_Technical_Manual.pdf
- Adams, R.J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation*, 31(2–3), 162–172.
- Afrassa, T. i Keeves, J. (1999). Changes in students' mathematics achievement in Australian lower secondary schools over time. *International Education Journal*, 1(1), 1–21.
- Akaike, H. (1973). Information Theory as an Extension of the Maximum Likelihood Principle. W: B.N. Petrov i F. Csaki (red.), *Second International Symposium on Information Theory* (s. 267-281). Budapest: Akademiai Kiado.
- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19, 716-723.
- Allalouf, A. i Ben Shakhar, G. (1998). The effect of coaching on the predictive validity of scholastic aptitude tests. *Journal of Educational Measurement*, 35(1), 31–47.
- Angoff, W.H. (1971). Scales, norms, and equivalent scores, W: L. Thorndike (red.), *Educational Measurement* (wyd. 2, s. 508–600). Washington, DC: American Council on Education.
- Ankudewicz-Blukacz, T. i Małecki, W. (1999). Udział województwa wrocławskiego w modyfikowaniu egzaminów do szkół ponadpodstawowych. W: Niemierko, B., Machowska, B. (red). *Diagnoza edukacyjna. Oczekiwania, problemy, przykłady*. (s.353-362). Legnica: ODESIP.
- 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.
- Baumgartner, H. i Steenkamp, J.B.E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38(2), 143-156.
- Beard, J.J. (2008). *An investigation of vertical scaling with item response theory using a multistage testing framework* (rozprawa doktorska). Iowa: University of Iowa. Pobrano z: <http://ir.uiowa.edu/etd/216/>
- Beller, M. (1994). Psychometric and social issues in admissions to Israeli universities. *Educational Measurement: Issues and Practice*, 13(2), 12–20.
- Béguin, A.A. (2000). *Robustness of equating high-stakes tests* (rozprawa doktorska). Enschede: University of Twente.
- Bland, J.M. i Altman, D.G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2), 135-160.
- Bogdanowicz, M. (1994). *O dysleksji, czyli specyficznych trudnościach w czytaniu i pisaniu*. Warszawa: Wydawnictwa Szkolne i Pedagogiczne.
- 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.
- Camilli, G., Yamamoto, K. i Wang, M. (1993). Scale Shrinkage in Vertical Equating. *Applied Psychological Measurement*, 17(4), 379-388.

- Carlson, J.E. (2011). Statistical Models for Vertical Linking. W: A.A. von Davier (red.), *Statistical Models for Test Equating, Scaling, and Linking* (s. 59-71). New York, NY: Springer.
- CKE (2007). *Informator o egzaminie maturalnym od 2008 roku. Język angielski*. Warszawa: Centralna Komisja Egzaminacyjna.
- CKE (2008). *Informator o egzaminie maturalnym od 2010 roku. Matematyka*. Warszawa: Centralna Komisja Egzaminacyjna.
- CKE (2009). *Informator z aneksem dla uczniów ze specyficznymi trudnościami w uczeniu się. Sprawdzian w klasie szóstej szkoły podstawowej przeprowadzany od roku szkolnego 2009/2010*. Warszawa: Centralna Komisja Egzaminacyjna.
- CKE (2010a) *Informator o egzaminie gimnazjalnym od roku szkolnego 2011/2012*. Warszawa: Centralna Komisja Egzaminacyjna
- CKE (2010b). *Informator o egzaminie maturalnym z języka obcego nowożytnego od roku szkolnego 2011/2012*. Warszawa: Centralna Komisja Egzaminacyjna.
- CKE (2010c). *Osiągnięcia maturzystów w 2010 roku. Sprawozdanie z egzaminu maturalnego w 2010 roku*. Warszawa: Centralna Komisja Egzaminacyjna.
- CKE (2011). *Osiągnięcia maturzystów w 2011 roku. Sprawozdanie z egzaminu maturalnego w 2011 roku*. Warszawa: Centralna Komisja Egzaminacyjna.
- CKE (2012). *Osiągnięcia maturzystów w 2012 roku. Sprawozdanie z egzaminu maturalnego w 2012 roku*. Warszawa: Centralna Komisja Egzaminacyjna.
- CKE (2013). *Osiągnięcia maturzystów w 2013 roku. Sprawozdanie z egzaminu maturalnego w 2013 roku*. Warszawa: Centralna Komisja Egzaminacyjna.
- CKE (2013). *Informator o sprawdzianie od roku szkolnego 2014/2015*. Warszawa: Centralna Komisja Egzaminacyjna.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- von Davier, A.A., Holland, P.W. i Thayer, D.T. (2004). *The Kernel Method of Test Equating*. New York, NY: Springer.
- von Davier, A.A. (2011) (red). *Statistical Models for Test Equating, Scaling, and Linking*. New York, NY: Springer-Verlag.
- von Davier, M. i von Davier, A.A. (2011). A General Model for IRT Scale Linking and Scale Transformations. W: A.A. von Davier (red.), *Statistical Models for Test Equating, Scaling, and Linking* (s. 1-17). New York, NY: Springer-Verlag.
- Dempster, A.P., Laird, N.M. i Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, 1-38.
- Dolata, R. (2008). *Szkoła – segregacje – nierówności*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Dolata, R., Jasińska, A. i Modzelewski, M. (2012). Wykorzystanie krajowych egzaminów jako instrumentu polityki oświatowej na przykładzie procesu różnicowania się gimnazjów w dużych miastach. *Polityka Społeczna*, 1, 41–46.

- Dolata, R., Jakubowski, M., & Pokropek, A. (2013). *Polska oświata w międzynarodowych badaniach umiejętności uczniów PISA OECD. Wyniki, trendy, kontekst i porównywalność*: Wydawnictwa Uniwersytetu Warszawskiego.
- Dolata, R., Hawrot, A., Humenny, G., Jasińska-Maciążek, A., Koniewski, M., Majkut, P. (2014). *Kontekstowy model oceny efektywności nauczania po pierwszym etapie edukacyjnym*. Warszawa: Instytut Badań Edukacyjnych.
- Dorans, N.J. i Holland, P.W. (2000). Population invariance and the equatability of tests: Basic theory and the linear case. *ETS Research Report Series*, 2000(2), i-35.
- EACEA (2010). *Różnice w wynikach nauczania a płeć uczniów. Obecna sytuacja i działania podejmowane w Europie*. Warszawa: Fundacja Rozwoju Systemu Edukacji.
- Ellis, J.L. i van der Woldenberg, A.L. (1993). Local homogeneity in latent trait models. A characterization of the homogenous monotone IRT model. *Psychometrika*, 58 (3), 417–429.
- Gibbons, R.D. i Hedeker, D.R. (1992). Full-information item bi-factor analysis. *Psychometrika*, 57, 423-436.
- Gibbons, R.D., Bock, R.D., Hedeker, D., Weiss, D.J., Segawa, E., Bhaumik, D.K. i Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, 31(1), 4-19.
- Glas C.A.W. i Béguin A.A. (1996). *Appropriateness of IRT Observed-Score*. Enschede: University of Twente, Faculty of Educational Science and Technology.
- Glas C.A.W. i Béguin A.A. (2011). Robustness of IRT Observed-Score Equating. W: A.A. von Davier (red.), *Statistical Models for Test Equating, Scaling, and Linking* (s. 21-42). New York, NY: Springer-Verlag.
- Gruijter, D.N.M. i van der Kamp, L.J. (2005). *Statistical test theory for education and psychology*. Pobrano z: http://irt.com.ne.kr/data/test_theory.pdf
- GUS (2013). *Oświata i wychowanie w roku szkolnym 2012/2013*. Warszawa: Zakład Wydawnictw Statystycznych.
- Hanushek, E.A i Kim, D. (1995). Schooling, labor force quality, and economic growth. *National Bureau of Economic Research Working Paper No. 5399*. Pobrano z: <http://www.nber.org/papers/w5399>
- Harris, D.J. (2007). Practical Issues in Vertical Scaling. W: N.J. Dorans, M. Pommerich i P.W. Holland (red.), *Linking and Aligning Scores and Scales* (s. 233-250). New York, NY: Springer.
- Herbst, M. (2004). Zróżnicowanie jakości kapitału ludzkiego w Polsce. Od czego zależą wyniki edukacyjne? *Studia Regionalne i Lokalne*, 3(17), 93-104.
- Herbst, M. i Herczyński, J. (2005). *School Choice and Student Achievement. Evidence from Poland*. MPRA Paper No. 6138. Pobrano z: <http://mpra.ub.uni-muenchen.de/6138/>
- Herczyński, J. i Herbst, M. (2002). *Pierwsza odsłona: społeczne i terytorialne zróżnicowanie wyników sprawdzianu szóstoklasistów i egzaminu gimnazjalnego przeprowadzonych wiosną 2002 roku. Raport przygotowany na zlecenie Fundacji Klub Obywatelski*. Warszawa: Fundacja „Klub Obywatelski”.
- Herczyński, J. i Sobotka, A. (2014). *Diagnoza zmian w sieci szkół podstawowych i gimnazjów 2007-2012*. Warszawa: Instytut Badań Edukacyjnych.

- Holland, P.W. (2007). A framework and history for score linking. W: N.J. Dorans, M. Pommerich i P.W. Holland (red.), *Linking and aligning scores and scales*. New York, NY: Springer.
- Holland, P.W., Dorans N.J. i Petersen N.S. (2007). Equating test scores. W: C.R. Rao i S. Sinharay (red.), *Handbook of Statistics, Vol. 26. Psychometrics* (s. 169–204). New York, NY: Elsevier.
- Hu, H., Rogers, W.T. i Vukmirovic, Z. (2008). Investigation of IRT-Based Equating Methods in the Presence of Outlier Common Items. *Applied Psychological Measurement*, 32(4), 311-333.
- Hyde, J.S. i Kling, K.C. (2001). Women, motivation, and achievement. *Psychology of Women Quarterly*, 25, 364-378.
- Jamshidian, M. i Jennrich, R. I. (2000). Standard errors for EM estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2), 257-270.
- Jung-Miklaszewska, J. i Rusakowska, D. (1995). *Szkoły społeczne in statu nascendi*. Warszawa: Instytut Badań Edukacyjnych.
- Kancelaria Sejmu (1998). Ustawa o zmianie ustawy o systemie oświaty z dnia 25 lipca 1998 r. (DzU Nr 117 poz. 759).
- Kolen, M.J. (1984). Effectiveness of analytic smoothing in equipercenile equating. *Journal of Educational Statistics*, 9, 25–44.
- Kolen, M.J. i Brennan, R.L. (2004). *Test Equating, Scaling, and Linking. Methods and Practices*. New York, NY: Springer.
- Konarzewski, K. (1996). *Problemy i schematy. Pierwszy rok nauki szkolnej dziecka*. Warszawa: Wydawnictwo "Żak".
- Konarzewski, K. (2012). *TIMSS i PIRLS 2011. Osiągnięcia szkolne polskich trzecioklasistów w perspektywie międzynarodowej*. Warszawa: Centralna Komisja Egzaminacyjna.
- Kondratek, B. i Grudniewska, M. (2013). Test Mantel-Haenshel oraz modelowanie IRT jako narzędzia służące do wykrywania DIF i opisu jego wielkości na przykładzie zadań ocenianych dychotomicznie. *Edukacja*, 2(122), 34-55.
- Kurek, S. (2010). Przestrzenne zróżnicowanie poziomu rozwoju regionalnego w Unii Europejskiej w świetle wybranych mierników. *Prace Komisji Geografii Przemysłu Polskiego Towarzystwa Geograficznego*, 16, 87-104.
- Li, Y. (2011). *Exploring the full-information bifactor model in vertical scaling with construct shift* (rozprawa doktorska). College Park, MD: University of Maryland.
- Li, Y. i Lissitz, R.W. (2000) An Evaluation of the Accuracy of Multidimensional IRT Linking. *Applied Psychological Measurement*, 24(2), 115-138.
- Li, Y. i Lissitz, R.W. (2012). Exploring the full-information bifactor model in vertical scaling with construct shift. *Applied Psychological Measurement*, 36(1), 3-20.
- van der Linden, W.J. (2011). Local observed-score equating. W: A.A. von Davier (red.), *Statistical models for equating, scaling, and linking* (s. 201–223). New York, NY: Springer.
- Liu, J., Moses, T. i Low, A. (2009). Evaluation of the effects of loglinear smoothing models on equating functions in the presence of structured data irregularities. *ETS Research Report No. RR-09-22*. Princeton, NJ: Educational Testing Service.

- Liu, J. i Walker, M. (2007). Score linking issues related to test content changes. W: N.J. Dorans, M. Pommerich i P.W. Holland (red.), *Linking and aligning scores and scales* (s. 109- 134). New York, NY: Springer-Verlag.
- Livingston, S.A. (2004). *Equating test stores*. Pobrano z: <http://www.ets.org/Media/Research/pdf/LIVINGSTON.pdf>
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Lord, F.M. i Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Małecki, W. (1999). *Dylematy egzaminacyjne*. Warszawa: MEN.
- MEN (1999a). *Podstawa programowa kształcenia ogólnego dla sześcioletnich szkół podstawowych i gimnazjów. Załącznik nr 1 do rozporządzenia Ministra Edukacji Narodowej z dnia 15 lutego 1999. (DzU z 1999 nr 14 poz. 129) z późniejszymi zmianami.*
- MEN (1999b). *Podstawa programowa kształcenia ogólnego dla szkół ponadpodstawowych. Załącznik nr 2 do rozporządzenia Ministra Edukacji Narodowej z dnia 15 lutego 1999. (DzU z 1999 nr 14 poz. 129) z późniejszymi zmianami.*
- MEN (1999c). *Rozporządzenie Ministra Edukacji Narodowej z dnia 19 kwietnia 1999 r. w sprawie zasad oceniania, klasyfikowania i promowania uczniów i słuchaczy oraz przeprowadzania egzaminów i sprawdzianów w szkołach publicznych (DzU Nr 41, poz.413)*
- MEN (2001a). *Rozporządzenie Ministra Edukacji Narodowej z dnia 21 marca 2001 r. w sprawie warunków i sposobu oceniania, klasyfikowania i promowania uczniów i słuchaczy oraz przeprowadzania egzaminów i sprawdzianów w szkołach publicznych (DzU Nr 29, poz. 323, zm.: DzU Nr 128, poz. 1419, DzU z 2002 r. Nr 46, poz. 433) z późniejszymi zmianami.*
- MEN (2001b). *Rozporządzenie Ministra Edukacji Narodowej z dnia 10 sierpnia 2001 r. w sprawie standardów wymagań będących podstawą przeprowadzania sprawdzianów i egzaminów. Załącznik nr 1. Standardy egzaminacyjne będące podstawą przeprowadzenia sprawdzianu w ostatnim roku nauki w szkole podstawowej. (DzU Nr 92, poz. 1020) z późniejszymi zmianami.*
- MEN (2001c). *Rozporządzenie Ministra Edukacji Narodowej z dnia 6 listopada 2001 r. zmieniające rozporządzenie w sprawie warunków i sposobu oceniania, klasyfikowania i promowania uczniów i słuchaczy oraz przeprowadzania egzaminów i sprawdzianów w szkołach publicznych. (DzU Nr 128, poz. 1419).*
- MEN (2002). *Rozporządzenie Ministra Edukacji Narodowej i Sportu z dnia 11 września 2002 r. zmieniające rozporządzenie w sprawie standardów wymagań będących podstawą przeprowadzania sprawdzianów i egzaminów (Dz U Nr 155, poz. 1289).*
- MEN (2003). *Rozporządzenie Ministra Edukacji Narodowej i Sportu z dnia 10 kwietnia 2003 r. zmieniające rozporządzenie w sprawie standardów wymagań będących podstawą przeprowadzania sprawdzianów i egzaminów (Dz U nr 90, poz. 846).*
- MEN (2006). *Rozporządzenie Ministra Edukacji Narodowej z dnia 8 września 2006 r. zmieniające rozporządzenie w sprawie warunków i sposobu oceniania, klasyfikowania i promowania uczniów*

- i słuchaczy oraz przeprowadzania sprawdzianów i egzaminów w szkołach publicznych (DzU z 2006 r., Nr 164, poz. 1154).
- MEN (2007). *Standardy wymagań będące podstawą przeprowadzania egzaminu w ostatnim roku nauki w gimnazjum*. Załącznik do rozporządzenia Ministra Edukacji Narodowej z dnia 28 sierpnia 2007 r. (Dz.U. z 2007 r. Nr 157, poz. 1102). Pobrano z http://www.gim-nt.com/download/rozporzadzenie_28082007.pdf
- MEN (2008a). *Podstawa programowa kształcenia ogólnego dla gimnazjów i szkół ponadgimnazjalnych*. Pobrano z http://www.bip.men.gov.pl/men_bip/akty_prawne/rozporzadzenie_20081223_zal_4.pdf
- MEN (2008b). Rozporządzenie Ministra Edukacji Narodowej z dnia 25 września 2008 r. zmieniające rozporządzenie w sprawie warunków i sposobu oceniania, klasyfikowania i promowania uczniów i słuchaczy oraz przeprowadzania sprawdzianów i egzaminów w szkołach publicznych (DzU z 2008 r., Nr 178, poz. 1097).
- MEN (2010). Rozporządzenie Ministra Edukacji Narodowej z dnia 17 listopada 2010 r. w sprawie warunków organizowania kształcenia, wychowania i opieki dla dzieci i młodzieży niepełnosprawnych oraz niedostosowanych społecznie w przed szkołach, szkołach i oddziałach ogólnodostępnych lub integracyjnych (Dz.U. nr 228, poz. 1490, z późniejszymi zmianami)
- Messick, S. (1989). Validity. W: R. Linn (red.), *Educational measurement* (wyd. 3, s.13–103). Washington, DC: American Council on Education.
- 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.
- Muthén, L.K. i Muthén, B.O. (1998-2010). *Mplus User's Guide. Sixth Edition*. Los Angeles, CA: Muthén i Muthén.
- Niemierko, B. (2004). Zrównywanie wyników sprawdzianu 2004 do wyników sprawdzianu 2003. W: B. Niemierko i H. Szaleniec (red.), *Standardy wymagań i normy testowe w diagnostyce edukacyjnej* (s. 573-590). Kraków: Polskie Towarzystwo Diagnostyki Edukacyjnej.
- Niemierko, B. (2007). Zrównywanie wyników sprawdzianu 2005 do wyników sprawdzianu 2003 metodą ekwicyntylową. *Egzamin*, 10, 86–104.
- Nijakowska, J.(2009). Hipoteza różnic w kodowaniu językowym – próba wyjaśnienia trudności w uczeniu się języków obcych. W: M. Pawlak, M. Derenowski i B. Wolski (red.), *Problemy współczesnej dydaktyki języków obcych* (s. 41-55). Kalisz: Wydział Pedagogiczno-Artystyczny UAM w Kaliszu.
- 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.
- OECD (2009). *Equally prepared for life? How 15-year-old boys and girls perform in school*. Pobrano z http://www.stemequitypipeline.org/_documents/OECD%20%282009%29%20gender%20diff%20similarities%20in%2015-year-old.pdf
- OECD (2014). *PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science* (Volume I, Revised edition, February 2014). PISA, OECD Publishing.

- 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.
- Patz, R.J. i Junker, B. W. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, 24(2), 146-178.
- Paulhus, D.L. (1991). Measurement and control of response bias. W: J.P. Robinson, P.R. Shaver i L.S. Wrightsman (red.), *Measures of personality and social psychological attitudes* (s. 17-59). San Diego: Academic Press.
- Piekarczyk, M. (2014). *Jak Polacy postrzegają szkoły publiczne i niepubliczne: preferencje dotyczące szkolnictwa w Polsce*. Warszawa: Centrum Badań nad Uprzedzeniami.
- Pokropek, A. (2011a). Matura z języka polskiego. Wybrane problemy psychometryczne. W: B. Niemierko i M.K. Szmigel (red.), *Ewaluacja w edukacji. Koncepcje, metody, perspektywy* (s. 438-450). Kraków: GRUPA TOMAMI.
- Pokropek, A. (2011b). Zrównywanie wyników egzaminów zewnętrznych w kontekście międzynarodowym. W: B. Niemierko i M.K. Szmigel (red.), *Ewaluacja w edukacji. Koncepcje, metody, perspektywy* (s. 524-531). Kraków: GRUPA TOMAMI.
- Pokropek, A. i Kondratek, B. (2012). Zrównywanie wyników testowania. Definicje i przykłady zastosowania. *Edukacja*, 120(4), 52–71.
- Putkiewicz, E. i Wilkomirska, A. (2004). *Szkoły publiczne i niepubliczne. Porównanie środowisk edukacyjnych*. Warszawa: Instytut Spraw Publicznych.
- R Core Team (2014). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Rapp, J. (1999). *Linear and Equipercentile Methods for Equating PET*. Jerusalem: National Institute for Testing and Evaluation. Pobrano z: <https://www.nite.org.il/files/reports/e266.pdf>
- Reckase, M. D. (2009). *Multidimensional Item Response Theory*. New York, NY: Springer-Verlag.
- Reise, S.P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667-696.
- Reise, S.P., Morizot, J. i Hays, R.D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16(1), 19-31.
- RM (2000). Rozporządzenie Rady Ministrów z dnia 13 lipca 2000 r. w sprawie wprowadzenia Nomenklatury Jednostek Terytorialnych do Celów Statystycznych (NTS) (Dz. U. z 2000 r. Nr 58, poz. 685).
- Robert, C.P. i Casella, G. (2004). *Monte Carlo statistical methods*. New York: Springer.
- Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, 63(3), 129-156.
- von Schrader, S. i Ansley, T. (2006). Sex differences in the tendency to omit items on multiple-choice tests: 1980–2000. *Applied Measurement in Education*, 19(1), 41-65.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6, 461-464.
- Skórska, P. i Świst, K. (2014). Wielkość efektu płci w wewnątrzszkolnych i zewnątrzszkolnych wskaźnikach osiągnięć ucznia. W: B. Niemierko i M.K. Szmigel (red.), *Diagnozy edukacyjne: dorobek i nowe zadania* (s. 89-103). Kraków: GRUPA TOMAMI.

- Smolik, M. (2007). Zrównanie wyników sprawdzianu 2004 i sprawdzianu 2005 do wyników sprawdzianu 2003 z wykorzystaniem modelu Rascha. *Egzamin*, 10, 118- 158.
- Stage, C. (2004). *Notes from the Tenth International SweSAT Conference*. Umeå, June 1–3. Pobrano z: http://www.edusci.umu.se/digitalAssets/59/59531_em-no-52.pdf
- Stage, C. i Ögren, G. (2004). *The Swedish Scholastic Assessment Test (SweSAT)*. Umeå: Umeå University, Department of Educational Measurement.
- Szaleniec, H. (2005). Wykorzystanie probabilistycznych modeli zadania testowego do zrównywania wyników sprawdzianu 2003-2005 i budowania banku zadań. W: B. Niemierko i G. Szyling (red.), *Holistyczne i analityczne metody diagnostyki edukacyjnej. Perspektywy informatyczne egzaminów szkolnych* (s.103-123). Gdańsk: Fundacja Rozwoju Uniwersytetu Gdańskiego.
- Szaleniec, H., Grudniewska, M., Kondrątek, B., Kulon, F. i Pokropek, A. (bdw). *Analiza porównawcza wyników egzaminów zewnętrznych – gimnazjum. Raport z badań*. Warszawa: Instytut Badań Edukacyjnych.
- Szaleniec, H., Grudniewska, M., Kondrątek, B., Kulon, F. i Pokropek, A. (2012). Wyniki egzaminu gimnazjalnego 2002–2010 na wspólnej skali. *Edukacja*, 119(3), 9-30.
- Szaleniec, H., Grudniewska, M., Kondrątek, B., Kulon, F., Pokropek, A., Stożek, E. i Żółtak, M. (2013). *Analiza porównawcza wyników egzaminów zewnętrznych – sprawdzian w szóstej klasie szkoły podstawowej i egzamin gimnazjalny*. Warszawa: Instytut Badań Edukacyjnych.
- Szaleniec, H., Kondrątek, B., Kulon, F., Pokropek, A., Skórska, P., Świst, K., Wołodźko, T. i Żółtak, M. (2015). *Efekt egzaminatora w ocenianiu prac maturalnych z języka polskiego i matematyki*. Warszawa: Instytut Badań Edukacyjnych.
- Szmigiel, M.K., Szaleniec, H. (1999). *Badanie kompetencji uczniów klas ósmych, jako diagnoza umiejętności i pilotaż systemu egzaminacyjnego*. W: Niemierko, B., Machowska, B. (red). *Diagnoza edukacyjna. Oczekiwania, problemy, przykłady*. (s.323-341). Legnica.
- 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.
- Sundre, D.L. (2007). *The Student Opinion Scale (SOS): A measure of examinee motivation. Test Manual*. Pobrano z <http://www.jmu.edu/assessment/resources/Overview.html>
- Swerdzewski, P., Finney, S.J. i Harmes, J.C. (2007). *Skippping 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.
- 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.
- Świst, K., Skórska, P. i Pokropek, A. (w druku). Zastosowanie skalowania wertykalnego do oceny przyrostu umiejętności matematycznych polskich uczniów od sprawdzianu do matury. W: A. Pokropek (red.), *Modele cech ukrytych w psychologii, socjologii i badaniach edukacyjnych. Teoria i zastosowania*. Warszawa: Instytut Badań Edukacyjnych.
- Tong, Y. i Kolen, M. (2007). Comparisons of Methodologies and Results in Vertical Scaling for Educational Achievement Tests. *Applied Measurement in Education*, 20(2), 227-253.

- de la Torre, J. (2009). Improving the quality of ability estimates through multidimensional scoring and incorporation of ancillary variables. *Applied Psychological Measurement*, 33(6), 465-485.
- Tyralska-Wojtyczka, E. (2013). Nowy egzamin gimnazjalny z przedmiotów przyrodniczych – lepszy od starego, czy tylko inny?. W: B. Niemierko i M.K. Szmigel (red.), *Polska edukacja w świetle diagnoz prowadzonych z różnych perspektyw badawczych* (s. 432-439). Kraków: gRUPA TOMAMI.
- Wejner, T. (2009). Ocenianie uczniów dyslektycznych w innych krajach. W B. Niemierko i M.K. Szmigel (red.), *Badania międzynarodowe i wzory zagraniczne w diagnostyce edukacyjnej*. (s. 331-341). Kraków: gRUPA TOMAMI.
- Węziak, D. (2007). Metody zrównywania wyników wykorzystywane w skalowaniu Rascha. Propozycja zastosowań w warunkach polskich. *Egzamin*, 10, 75-85.
- Węziak-Białowolska, D. (2011). Skala motywacji testowej – analiza właściwości psychometrycznych. W B. Niemierko, M.K. Szmigel (red.), *Ewaluacja w edukacji: koncepcje, metody, perspektywy : XVII Krajowa Konferencja Diagnostyki Edukacyjnej*. (s. 451-462). Kraków: gRUPA TOMAMI.
- Willingham, W.W. i Cole, N.S. (1997). Gender and fair assessment. Mahwah, NY: Lawrence Erlbaum.
- 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.
- Wołodźko, T., Kondratak, B. i Szaleniec, H. (2014). Zrównanie ekwicyntylowe na tle innych metod zrównywania na przykładzie sprawdzianu i egzaminu gimnazjalnego. *Edukacja*, 3(128), 112–130.
- Wołodźko, T. (2015). *equi: R library for equipercntile equating*. Pobrano z <https://github.com/twołodzko/equi>
- World Economic Forum (2013). *The Global Gender Gap Report*. Pobrano z http://www3.weforum.org/docs/WEF_GenderGap_Report_2013.pdf
- Wright, B.D. i Stone, M. (1979). *Best test design*. Chicago, IL: MESA Press.
- Wright, B.D. i Masters, G.N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.
- Wu. M. (2005). The Role of Plausible Values in Large-Scale Surveys. *Studies in Educational Evaluation*, 31, 114-128.
- Yao, L. i Boughton K. (2009). Multidimensional Linking for Tests with Mixed Item Types. *Journal of Educational Measurement*, 46(2) 177–197.
- Yen, W. M. (1985). Increasing item complexity: A possible cause of scale shrinkage for unidimensional item response theory. *Psychometrika*, 50(4), 399-410.
- Zahorska-Bugaj, M. (1994). Szkoła prywatna czy państwowa. W: R. Siemieńska (red.), *Szkoły niepaństwowe w systemie edukacji w Polsce. Kwartalnik Pedagogiczny*, 1-2. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Zasacka Z. (2014). *Czytelnictwo dzieci i młodzieży*. Warszawa: Instytut Badań Edukacyjnych.
- Zawadzka, E. (2010) *Nauczyciel języków obcych i jego niepełnosprawni uczniowie (z zaburzeniami i dysfunkcjami)*. Kraków: Impuls.

Spis ilustracji

Rysunek 2.1. Cztery etapy studium zrównywania wyników przeprowadzonego w IBE w latach 2011–2014.....	17
Rysunek 2.2. Schemat procesu badawczego w przeprowadzonych badaniach zrównujących	18
Rysunek 4.1. Schemat planu badawczego – sprawdzian w VI klasie szkoły podstawowej	48
Rysunek 5.1. Pochodzenie szkół podstawowych w ostatecznej próbie.....	71
Rysunek 5.2. Pochodzenie gimnazjów w ostatecznej próbie.....	72
Rysunek 5.3. Pochodzenie szkół ponadgimnazjalnych w ostatecznej próbie	72
Rysunek 6.1. Rozkłady wyników obserwowanych ze sprawdzianu w latach 2002-2014. Pełne słupki odnoszą się do niezrównanych wyników egzaminacyjnych, zrównane wyniki przedstawiono za pomocą słupków pustych	85
Rysunek 6.2. Średnie zrównane wyniki sprawdzianu na skali 100;15 wraz z 95% przedziałem ufności	91
Rysunek 6.3. Rozkłady wyników obserwowanych z części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002-2014. Pełne słupki odnoszą się do niezrównanych wyników egzaminacyjnych, zrównane wyniki przedstawiono za pomocą słupków pustych	93
Rysunek 6.4. Rozkłady punktów dla zadań 21, 22 oraz 23 w arkuszu matematyczno-przyrodniczym 2012 (góra) wraz z odpowiadającymi tym zadaniom krzywymi charakterystycznymi (dół)	95
Rysunek 6.5. Średnie zrównane wyniki dla części matematyczno-przyrodniczej egzaminu gimnazjalnego na skali 100;15 wraz z 95% przedziałem ufności	97
Rysunek 6.6. Rozkłady wyników obserwowanych z części humanistycznej egzaminu gimnazjalnego w latach 2002-2014. Pełne słupki odnoszą się do niezrównanych wyników egzaminacyjnych, zrównane wyniki przedstawiono za pomocą słupków pustych	98
Rysunek 6.7. Średnie zrównane wyniki dla części humanistycznej egzaminu gimnazjalnego na skali 100;15 wraz z 95% przedziałem ufności	101
Rysunek 6.8. Rozkłady wyników obserwowanych z matury podstawowej z matematyki w latach 2010–2014. Pełne słupki odnoszą się do niezrównanych wyników egzaminacyjnych, zrównane wyniki przedstawiono za pomocą słupków pustych	102
Rysunek 6.9. Średnie zrównane wyniki dla egzaminu maturalnego z matematyki na skali 100;15 wraz z 95% przedziałem ufności.....	105
Rysunek 6.10. Rozkłady wyników obserwowanych z matury podstawowej z języka polskiego w latach 2010–2014. Pełne słupki odnoszą się do niezrównanych wyników egzaminacyjnych, zrównane wyniki przedstawiono za pomocą słupków pustych	110
Rysunek 6.11. Średnie zrównane wyniki dla egzaminu maturalnego z języka polskiego na skali 100;15 wraz z 95% przedziałem ufności	112
Rysunek 6.12. Rozkłady wyników obserwowanych z matury podstawowej z języka angielskiego w latach 2010–2014. Pełne słupki odnoszą się do niezrównanych wyników egzaminacyjnych, zrównane wyniki przedstawiono za pomocą słupków pustych	113
Rysunek 6.13. Średnie zrównane wyniki dla egzaminu maturalnego z języka angielskiego na skali 100;15 wraz z 95% przedziałem ufności	115
Rysunek 6.14. Wyniki zrównania ekwicyntylowego sprawdzianu w zestawieniu z wynikami zrównania opartego o IRT.....	117
Rysunek 6.15. Wyniki zrównania ekwicyntylowego części humanistycznej egzaminu gimnazjalnego w zestawieniu z wynikami zrównania opartego o IRT.....	117
Rysunek 6.16. Wyniki zrównania części matematyczno-przyrodniczej egzaminu gimnazjalnego w zestawieniu z wynikami zrównania opartego o IRT.....	118

Rysunek 6.17. Wyniki zrównania egzaminu maturalnego z języka polskiego w zestawieniu z wynikami zrównania opartego o IRT	118
Rysunek 6.18. Wyniki zrównania egzaminu maturalnego z matematyki w zestawieniu z wynikami zrównania opartego o IRT	119
Rysunek 6.19. Wyniki zrównania egzaminu maturalnego z języka angielskiego w zestawieniu z wynikami zrównania opartego o IRT	119
Rysunek 6.20. Model podwójnego czynnika dla dwóch etapów edukacyjnych	123
Rysunek 6.21. Model podwójnego czynnika dla trzech etapów edukacyjnych	124
Rysunek 7.1. Porównywalne wyniki sprawdzianu w latach 2002–2013	143
Rysunek 7.2. Porównywalne wyniki części humanistycznej egzaminu gimnazjalnego w latach 2002–2013	145
Rysunek 7.3. Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013	146
Rysunek 7.4. Porównywalne wyniki egzaminu maturalnego z matematyki w latach 2010–2013	147
Rysunek 7.5. Porównywalne wyniki egzaminu maturalnego z języka angielskiego w latach 2010–2013	148
Rysunek 7.6. Porównywalne wyniki sprawdzianu w latach 2002–2013 w podziale na regiony wg NTS	149
Rysunek 7.7. Porównywalne wyniki części humanistycznej egzaminu gimnazjalnego w latach 2002–2013 w podziale na regiony wg NTS	150
Rysunek 7.8. Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013 w podziale na regiony wg NTS	151
Rysunek 7.9. Porównywalne wyniki egzaminu maturalnego z matematyki w latach 2010–2013 w podziale na regiony wg NTS	152
Rysunek 7.10. Porównywalne wyniki egzaminu maturalnego z języka angielskiego w latach 2010–2013 w podziale na regiony wg NTS	153
Rysunek 7.11. Wybieralność języka angielskiego na maturze (procent uczniów) w latach 2010–2013 w podziale na regiony	154
Rysunek 7.12. Porównywalne wyniki sprawdzianu w latach 2002–2013 w podziale na płeć uczniów	155
Rysunek 7.13. Porównywalne wyniki części humanistycznej egzaminu gimnazjalnego w latach 2002–2013 w podziale na płeć uczniów	156
Rysunek 7.14. Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013 w podziale na płeć uczniów	157
Rysunek 7.15. Porównywalne wyniki egzaminu maturalnego z matematyki w latach 2010–2013 w podziale na płeć uczniów	158
Rysunek 7.16. Porównywalne wyniki egzaminu maturalnego z języka angielskiego w latach 2010–2013 w podziale na płeć uczniów	159
Rysunek 7.17. Porównywalne wyniki sprawdzianu w latach 2002–2013 w podziale na grupy bez dysleksji rozwojowej i z dysleksją rozwojową	160
Rysunek 7.18. Porównywalne wyniki części humanistycznej egzaminu gimnazjalnego w latach 2002–2013 w podziale na grupy uczniów bez dysleksji rozwojowej i z dysleksją rozwojową	161
Rysunek 7.19. Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013 w podziale na grupy uczniów bez dysleksji rozwojowej i z dysleksją rozwojową ..	162
Rysunek 7.20. Porównywalne wyniki egzaminu maturalnego z matematyki w latach 2010–2013 w podziale na grupy uczniów bez dysleksji rozwojowej i z dysleksją rozwojową	163
Rysunek 7.21. Porównywalne wyniki egzaminu maturalnego z języka angielskiego w latach 2010–2013 w podziale na grupy uczniów bez dysleksji rozwojowej i z dysleksją rozwojową	164
Rysunek 7.22. Porównywalne wyniki sprawdzianu w latach 2002–2013 w podziale na typ szkoły (niepubliczna vs publiczna)	165

Rysunek 7.23. Różnica porównywalnych wyników sprawdzianu oraz odsetek uczniów szkół niepublicznych w latach 2002–2013.....	166
Rysunek 7.24. Porównywalne wyniki części humanistycznej egzaminu gimnazjalnego w latach 2002–2013 w podziale na typ szkoły (niepubliczna vs publiczna)	167
Rysunek 7.25. Różnica porównywalnych wyników części humanistycznej egzaminu gimnazjalnego oraz odsetek uczniów szkół niepublicznych w latach 2002–2013.....	168
Rysunek 7.26. Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013 w podziale na typ szkoły (niepubliczna vs publiczna)	169
Rysunek 7.27. Różnica porównywalnych wyników części matematyczno-przyrodniczej egzaminu gimnazjalnego oraz odsetek uczniów szkół niepublicznych w latach 2002–2013	169
Rysunek 7.28. Porównywalne wyniki egzaminu maturalnego z matematyki w latach 2010–2013 w podziale na rodzaje szkół	171
Rysunek 7.29. Porównywalne wyniki egzaminu maturalnego z języka angielskiego w latach 2010–2013 w podziale na rodzaje szkół.....	172
Rysunek 7.30. Porównywalne wyniki sprawdzianu w latach 2002–2013 w podziale na rodzaj gminy, w której znajduje się szkoła.....	173
Rysunek 7.31. Porównywalne wyniki części humanistycznej egzaminu gimnazjalnego w latach 2002–2013 w podziale na rodzaj gminy, w której znajduje się szkoła	174
Rysunek 7.32. Porównywalne wyniki części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013 w podziale na rodzaj gminy, w której znajduje się szkoła	175
Rysunek 8.1. Widok serwisu http://pwe.ibe.edu.pl	177
Rysunek 8.2. Relacja między wykresem skrzynkowym a rozkładem wyników na podstawie przykładowego rozkładu wyników w powiecie.....	181
Rysunek 8.3. Średnie PWE (z 95% przedziałami ufności) z lat 2003-2010 dla trzech wybranych województw oraz Polski dla części matematyczno-przyrodniczej egzaminu gimnazjalnego	182
Rysunek 8.4. Średnie PWE (z 95% przedziałami ufności) dla dwóch przykładowych szkół podstawowych oraz gminy, w której się znajdują.....	183
Rysunek 8.5. Kwartyle oraz minimalne i maksymalne wyniki nieodstające PWE matury z języka angielskiego dla dwóch wybranych liceów ogólnokształcących oraz gminy, w której się znajdują	184
Rysunek 9.1. Schemat przepływu danych w bazie (grupy danych w bazie oznaczone prostokątami)	186
Rysunek 10.1. Wariant A – włączenie do egzaminu niejawnych zadań kotwiczących	196
Rysunek 10.2. Wariant B – jednoczesna standaryzacja zadań z wielu edycji egzaminu	198
Rysunek 10.3. Wariant C – dodatkowe badania na reprezentatywnej próbie	200
Rysunek A5.1. Widok zadania na stronie WWW stanowiącej interfejs dla banku zadań – w dolnej części widoczna tabela z parametrami psychometrycznymi zadania (objaśnienia oznaczeń parametrów w tekście)	284

Spis tabel

Tabela 3.1. Motyw przewodni sprawdzianu w latach 2002–2009	21
Tabela 3.2. Przydział punktów dla poszczególnych standardów wymagań egzaminacyjnych na sprawdzianie	22
Tabela 3.3. Wybrane własności psychometryczne sprawdzianów z lat 2002–2014.....	24
Tabela 3.4. Liczba zadań określonego formatu i czas trwania dla poszczególnych części sprawdzianu od 2015 roku.....	26
Tabela 3.5. Przydział punktów dla poszczególnych standardów wymagań egzaminacyjnych dla egzaminu gimnazjalnego w części humanistycznej	26
Tabela 3.6. Przydział punktów dla poszczególnych standardów wymagań egzaminacyjnych dla egzaminu gimnazjalnego w części matematyczno-przyrodniczej	27
Tabela 3.7. Wybrane własności psychometryczne egzaminu gimnazjalnego w części humanistycznej z lat 2002–2014	29
Tabela 3.8. Wybrane własności psychometryczne egzaminu gimnazjalnego w części matematyczno-przyrodniczej z lat 2002–2014.....	30
Tabela 3.9. Wybrane własności psychometryczne egzaminu maturalnego z języka polskiego na poziomie podstawowym z lat 2010–2014.....	34
Tabela 3.10. Standardy wymagań egzaminacyjnych – matura z matematyki	36
Tabela 3.11. Wybrane własności psychometryczne egzaminu maturalnego z matematyki na poziomie podstawowym z lat 2010–2014	37
Tabela 3.12. Wybieralność języka angielskiego na egzaminie maturalnym w latach 2010–2014.....	37
Tabela 3.13. Wybrane własności psychometryczne egzaminu maturalnego z języka angielskiego na poziomie podstawowym z lat 2010–2014.....	39
Tabela 4.1. Liczebność prób w poszczególnych latach badania w podziale na typy szkół oraz egzaminów.....	46
Tabela 4.2. Liczba zadań i ich źródła w zeszytach testowych wykorzystanych w kolejnych edycjach badań zrównujących (sprawdzian)	53
Tabela 4.3. Liczba zadań i ich źródła w zeszytach testowych wykorzystanych w kolejnych edycjach badań zrównujących (egzamin gimnazjalny – część humanistyczna)	54
Tabela 4.4. Liczba zadań i ich źródła w zeszytach testowych wykorzystanych w kolejnych edycjach badań zrównujących (egzamin gimnazjalny – część matematyczno-przyrodnicza)	55
Tabela 4.5. Liczba zadań i ich źródła w zeszytach testowych wykorzystanych w kolejnych edycjach badań zrównujących (matura – język polski).....	56
Tabela 4.6. Liczba zadań i ich źródła w zeszytach testowych wykorzystanych w kolejnych edycjach badań zrównujących (matura – matematyka)	56
Tabela 4.7. Liczba zadań i ich źródła w zeszytach testowych wykorzystanych w kolejnych edycjach badań zrównujących (matura – język angielski).....	57
Tabela 4.8. Schemat planu NEAT	60
Tabela 4.9. Schematyczny podział metod zrównywania wyników testowych	61
Tabela 4.10. Zmienne wykorzystane przy warunkowaniu podczas generowania PV w rozbiciu na egzaminy	67
Tabela 5.1. Stopy realizacji badania w podziale na lata i typy szkół (egzaminu).....	74
Tabela 5.2. Interpretacja wskaźnika zgodności kappa (Bland i Altman, 1999).....	76
Tabela 5.3. Zgodność kodowania zadań otwartych – zrównywanie sprawdzianu 2012.....	77
Tabela 5.4. Zgodność kodowania zadań otwartych – zrównywanie sprawdzianu 2013.....	78
Tabela 5.5. Zgodność kodowania zadań otwartych – zrównywanie sprawdzianu 2014.....	79
Tabela 5.6. Zgodność kodowania zadań otwartych – zrównywanie egzaminu gimnazjalnego 2012... 80	

Tabela 5.7. Zgodność kodowania zadań otwartych – zrównywanie egzaminu gimnazjalnego 2013 ...	81
Tabela 5.8. Zgodność kodowania zadań otwartych – zrównywanie egzaminu gimnazjalnego 2014 ...	81
Tabela 6.1. Tablica przeliczeniowa obserwowanych wyników sprawdzianu na wyniki obserwowane w roku 2012.....	87
Tabela 6.2. Średnia oraz odchylenie standardowe zmiennej umiejętności dla sprawdzianu w latach 2002–2013, wyniki zrównane na skali 100; 15 zakotwiczone w roku 2012	90
Tabela 6.3. Średnia oraz odchylenie standardowe zmiennej umiejętności dla części matematyczno-przyrodniczej egzaminu gimnazjalnego w latach 2002–2013, wyniki zrównane na skali 100; 15 zakotwiczone w roku 2012	96
Tabela 6.4. Średnia oraz odchylenie standardowe zmiennej umiejętności dla części humanistycznej egzaminu gimnazjalnego w latach 2002–2013, wyniki zrównane na skali 100; 15 zakotwiczone w roku 2012.....	100
Tabela 6.5. Zdawalność egzaminu maturalnego z matematyki wyliczona na niezrównanych wynikach egzaminacyjnych oraz na wynikach wyrażonych na skali egzaminu z roku 2012	103
Tabela 6.6. Średnia oraz odchylenie standardowe zmiennej umiejętności dla egzaminu maturalnego z matematyki w latach 2010–2013, wyniki zrównane na skali 100; 15 zakotwiczone w roku 2012	104
Tabela 6.7. Wyniki matury z języka polskiego w zależności od wybranego tematu*	107
Tabela 6.8. Zdawalność egzaminu maturalnego z języka polskiego wyliczona na niezrównanych wynikach egzaminacyjnych oraz na wynikach wyrażonych na skali egzaminu z roku 2012	111
Tabela 6.9. Średnia oraz odchylenie standardowe zmiennej umiejętności dla egzaminu maturalnego z języka polskiego w latach 2010–2013, wyniki zrównane na skali 100; 15 zakotwiczone w roku 2012	111
Tabela 6.10. Zdawalność egzaminu maturalnego z języka angielskiego wyliczona na niezrównanych wynikach egzaminacyjnych oraz na wynikach wyrażonych na skali egzaminu z roku 2012	114
Tabela 6.11. Średnia oraz odchylenie standardowe zmiennej umiejętności dla egzaminu maturalnego z języka angielskiego w latach 2010–2013, wyniki zrównane na skali 100; 15 zakotwiczone w roku 2012	114
Tabela 6.12. Schemat kotwiczenia zadań dla języka polskiego w latach 2012–2014	121
Tabela 6.13. Schemat kotwiczenia zadań dla matematyki w latach 2012–2014	121
Tabela 6.14. Oszacowania średnich oraz wariancji dla przyrostu umiejętności uczniów z języka polskiego dla modelu podwójnego czynnika w latach 2012–2014.....	126
Tabela 6.15. Oszacowania średniego przyrostu umiejętności, jego zróżnicowania dla języka polskiego w latach 2012–2014	126
Tabela 6.16. Oszacowania średnich oraz wariancji dla przyrostu umiejętności uczniów z matematyki w przypadku modelu podwójnego czynnika w latach 2012–2014	127
Tabela 6.17. Oszacowania średniego przyrostu umiejętności oraz jego zróżnicowania dla matematyki w latach 2012–2014	127
Tabela 6.18. Współczynniki regresji i procent wyjaśnionej wariancji dla motywacji testowej w badaniach zrównujących 2011-2013.....	130
Tabela 6.19. Współczynniki regresji i procent wyjaśnionej wariancji dla motywacji testowej w badaniach zrównujących 2014	131
Tabela 6.20. Różnice w motywacji między zeszytami testowymi	132
Tabela 6.21. Różnice w motywacji między zeszytami testowymi (wymiar ważności i wysiłku)	133
Tabela 6.22. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, jak nauczyciel ocenia przygotowanie uczniów do sprawdzianu	134
Tabela 6.23. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel uważa, że uczniom zależy na jak najlepszym rozwiązaniu testu w badaniu zrównującym (sprawdzian)	135
Tabela 6.24. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel uważa, że udziału w badaniu zrównującym może mieć wpływ na wyniki sprawdzianu	135

Tabela 6.25. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel wykorzystywał arkusze z poprzednich edycji sprawdzianu podczas zajęć z uczniami (sprawdzian)	136
Tabela 6.26. Procentowy rozkład odpowiedzi na pytanie o to, czy nauczyciele uważają, że trudność arkuszy sprawdzianu jest porównywalna między latami	136
Tabela 6.27. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, jak nauczyciel ocenia przygotowanie uczniów do egzaminu gimnazjalnego	136
Tabela 6.28. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel uważa, że uczniom zależy na jak najlepszym rozwiązaniu testu w badaniu zrównującym	137
Tabela 6.29. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel uważa, że udziału w badaniu zrównującym mógł mieć wpływ na wyniki egzaminu gimnazjalnego	137
Tabela 6.30. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel wykorzystywał arkusze z poprzednich edycji egzaminu gimnazjalnego podczas zajęć z uczniami	137
Tabela 6.31. Procentowy rozkład odpowiedzi na pytanie o to, czy nauczyciele uważają, że trudność arkuszy egzaminu gimnazjalnego jest porównywalna między latami	138
Tabela 6.32. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, jak nauczyciel ocenia przygotowanie uczniów do matury	138
Tabela 6.33. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel uważa, że uczniom zależy na jak najlepszym rozwiązaniu testu w badaniu zrównującym (matura)	139
Tabela 6.34. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel uważa, że udziału w badaniu zrównującym mógł mieć wpływ na wyniki matury	139
Tabela 6.35. Procentowy rozkład odpowiedzi na pytanie dotyczące tego, czy nauczyciel wykorzystywał arkusze z poprzednich edycji matury podczas zajęć z uczniami	140
Tabela 6.36. Procentowy rozkład odpowiedzi na pytanie o to, czy nauczyciele uważają, że trudność arkuszy maturalnych jest porównywalna między latami	140
Tabela 7.1. Regiony i województwa wchodzące w ich skład (wg NTS)	149
Tabela A5.1. Opisy zadań pobrane za pomocą opisanego w aneksie 5 skryptu	287
Tabela A5.2. Parametry IRT zadań pobrane za pomocą opisanego w aneksie 5 skryptu	289
Tabela A5.3. Parametry KKT zadań pobrane za pomocą opisanego w aneksie 5 skryptu	289

Aneksy

Aneks 1. Tabele przeliczeniowe

1.1. Sprawdzian

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012											
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
0	1	0	1	0	1	0	0	1	1	0	0	1
1	1	1	2	1	2	1	1	1	1	1	1	1
2	2	1	3	2	3	2	1	2	2	2	2	2
3	2	2	4	2	4	3	2	3	3	3	3	3
4	3	2	4	3	5	3	3	4	4	4	4	4
5	3	3	5	4	6	4	4	5	5	5	5	5
6	4	4	6	4	7	5	5	6	6	6	6	6
7	4	4	7	5	7	6	5	7	7	6	7	7
8	5	5	8	6	8	6	6	8	8	7	8	8
9	5	5	8	6	9	7	7	9	9	8	9	9
10	6	6	9	7	10	8	8	10	10	9	10	10
11	6	6	10	8	11	9	9	11	11	10	11	11
12	7	7	11	8	12	9	9	12	12	11	12	12
13	8	7	11	9	12	10	10	13	13	12	13	13
14	8	8	12	10	13	11	11	14	13	12	14	14
15	9	9	13	10	14	12	12	15	14	13	15	15
16	9	9	14	11	15	12	12	16	15	14	16	16
17	10	10	14	12	16	13	13	17	16	15	17	17
18	11	11	15	13	17	14	14	18	17	16	18	18
19	11	12	16	13	17	15	15	19	18	17	19	19
20	12	13	17	14	18	15	15	20	19	18	20	19
21	13	13	18	15	19	16	16	21	20	19	21	20
22	14	14	19	15	20	17	17	22	21	20	22	21
23	14	15	20	16	21	18	18	23	21	21	23	22
24	15	16	21	17	22	19	19	24	22	22	24	23
25	16	17	22	18	23	20	20	25	23	23	25	24
26	17	18	23	18	23	21	21	26	24	24	26	25
27	18	19	24	19	24	22	22	27	25	25	27	26
28	19	20	25	20	25	23	23	28	26	26	28	27
29	20	21	26	21	26	24	25	29	27	27	29	27
30	21	22	27	22	27	25	26	30	28	28	30	28
31	22	24	28	23	28	26	27	31	29	29	31	29
32	24	25	29	24	29	27	28	32	30	30	32	30
33	25	27	30	26	30	28	30	32	31	31	33	31
34	26	28	31	27	31	30	31	33	32	32	34	32
35	28	30	32	29	32	31	32	34	34	34	35	33
36	30	31	33	30	33	32	34	35	35	35	36	34
37	31	33	34	32	34	34	35	36	36	36	37	36
38	33	35	35	34	35	35	36	37	37	38	38	37

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012											
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
39	36	37	37	36	37	37	38	38	39	39	39	38
40	38	40	39	39	40	40	40	40	40	40	40	40

1.2. Egzamin gimnazjalny w części humanistycznej

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012											
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
0		6	6	4	6	4	6	4	6	4	0	
1	6	6	7	6	6	6	6	6	7	6		
2	7	7	8	8	7	7	8	6	8	8		0
3	9	8	9	9	9	8	9	8	9	9		
4	10	9	11	10	10	10	10	9	10	11	4	
5	11	10	12	11	11	10	12	10	12	12		
6	12	11	13	12	12	12	13	11	13	13	6	6
7	13	12	14	13	13	13	14	12	14	15	7	6
8	14	13	16	14	14	14	15	13	15	16	8	7
9	16	14	17	15	15	15	16	14	16	18	9	8
10	17	15	18	16	15	16	17	15	17	19	10	9
11	18	16	19	17	16	17	18	16	18	20	11	10
12	19	17	21	18	17	18	19	17	19	22	12	11
13	20	18	22	19	18	19	21	18	20	23	13	12
14	21	19	23	20	20	19	22	19	21	24	14	13
15	22	20	24	20	21	20	23	20	22	26	15	14
16	24	21	26	21	22	21	24	21	23	27	16	15
17	25	22	27	22	23	22	25	22	24	28	17	16
18	26	24	28	23	24	23	26	23	25	30	18	17
19	27	25	30	24	25	24	27	24	26	31	19	18
20	29	26	31	25	26	25	28	25	27	32	20	20
21	30	27	32	26	27	26	29	26	28	34	21	21
22	31	28	34	27	28	27	30	27	29	35	22	22
23	33	29	35	27	29	28	31	28	31	36	23	23
24	34	30	36	28	30	29	32	29	32	38	24	24
25	35	31	38	29	32	30	33	30	33	39	25	25
26	36	32	39	30	33	31	35	31	34	40	26	27
27	37	33	40	32	34	32	36	32	35	42	27	28
28	39	34	42	33	35	33	37	33	37	43	28	29
29	40	35	43	34	37	34	38	34	38	45	29	30
30	41	37	44	35	38	35	39	36	39	46	30	31
31	43	38	46	36	39	37	40	37	41	47	31	33
32	44	39	47	37	41	38	41	38	42	48	32	34
33	45	40	48	39	42	39	42	40	44	50	33	35
34	47	42	49	40	43	40	43	41	45	51	34	36
35	48	43	51	41	45	42	44	42	47	52	35	38
36	49	45	52	43	46	43	46	44	48	53	36	39
37	51	46	53	44	48	45	47	45	50	54	37	40
38	52	48	54	46	49	46	48	47	51	55	38	41

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012											
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
39	53	49	55	48	50	48	50	48	52	56	39	43
40	54	50	56	49	52	49	51	50	54	57	40	44
41	56	52	57	51	53	51	52	51	55	58	41	45
42	57	53	58	52	54	52	54	52	56	59	42	46
43	58	55	59	54	56	54	55	54	57	60	43	47
44	59	56	60	55	57	55	56	55	59	61	44	48
45	60	57	61	57	58	57	57	56	60	61	45	49
46	61	59	62	59	59	58	59	58	61	62	46	50
47	63	60	63	60	60	60	60	59	62	63	47	51
48	65	62	65	62	61	61	61	60	62	64	48	52
49	65	63	65	64	62	63	62	61	63	65	49	53
50	65	65	65	65	64	65	65	63	65	65	50	54
51											51	55
52											52	55
53											53	56
54											54	57
55											55	58
56											56	59
57											57	59
58											58	60
59											59	61
60											60	62
61											61	63
62											62	64
63											63	65
64											64	65
65											65	65

1.3. Egzamin gimnazjalny w części matematyczno-przyrodniczej

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012											
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
0	3	4	3	3	3	3	3	3	4	3		
1		5	4	4	4	3	4	3	5	4		
2	4	6	6	5	6	4	5	4	6	5	2	2
3	5	7	7	6	7	5	6	5	6	6	3	3
4	6	8	8	7	8	6	6	6	7	6	4	3
5	7	9	9	8	9	7	7	7	8	7	5	4
6	8	10	10	9	10	7	8	8	9	8	6	5
7	9	11	11	10	11	8	9	9	10	9	7	6
8	10	12	12	11	12	9	10	10	11	10	8	6
9	10	14	13	12	13	10	11	11	12	11	9	7
10	11	15	14	13	14	11	12	12	13	12	10	8
11	12	16	15	14	15	13	13	13	14	13	11	9
12	13	16	16	15	16	14	14	14	15	14	12	10

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012											
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
13	14	17	17	16	17	15	15	15	16	15	13	10
14	15	18	18	17	18	16	15	16	17	16	14	11
15	16	19	19	18	18	17	16	17	18	17	15	12
16	16	20	19	19	19	18	17	17	18	18	16	13
17	17	21	20	19	20	19	18	18	19	19	17	14
18	18	22	21	20	21	20	19	19	20	19	18	15
19	19	22	22	21	22	21	20	20	21	20	19	16
20	20	23	23	22	23	22	21	21	22	21	20	16
21	21	24	24	23	24	23	22	22	23	22	21	17
22	22	25	25	25	25	24	23	23	24	23	22	18
23	23	26	25	26	26	25	24	24	25	24	23	19
24	24	27	26	27	28	26	24	25	26	25	24	20
25	26	28	27	28	29	27	25	26	27	26	25	21
26	27	29	28	29	30	28	26	27	29	28	26	22
27	28	30	29	31	31	30	27	28	30	29	27	23
28	30	30	30	32	32	31	29	29	31	30	28	23
29	31	31	31	33	34	32	30	30	32	31	29	24
30	32	32	33	35	35	33	31	31	33	33	30	25
31	34	34	34	36	36	34	32	32	35	34	31	26
32	35	35	35	37	37	35	33	33	36	35	32	27
33	37	36	36	39	39	37	34	34	37	37	33	28
34	38	37	37	40	40	38	35	36	38	38	34	29
35	40	38	38	41	41	39	36	37	39	39	35	31
36	41	39	39	42	42	40	38	38	41	41	36	32
37	42	41	40	44	43	41	39	39	42	42	37	33
38	44	42	41	45	44	42	40	41	43	43	38	34
39	45	43	42	46	45	43	41	42	44	45	39	35
40	46	44	44	47	46	45	42	43	45	46	40	36
41	47	45	45	48	47	46	44	44	46	47	41	38
42	49	46	46	49	48	47	45	46	47	48	42	39
43	50	47	47	50	49	48	46	47	48	49	43	40
44	51	48	48	51	50	49	47	48	49	50	44	41
45	52	50	49	52	51	50	48	49	50	51	45	42
46	54	51	51	53	52	51	50	50	51	52	46	44
47	56	52	52	56	53	53	51	52	52	53	47	45
48	56	54	54	56	56	55	52	53	53	54	48	46
49	56	56	56	56	56	56	54	56	54	54	49	47
50	56	56	56	56	56	56	56	56	56	56	50	48
51											51	49
52											52	50
53											53	52
54											54	53
55											55	55
56											56	56
57												56

1.4. Matura z j. angielskiego

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012			
	2010	2011	2012	2013
0	0	0	0	0
1	3			
2	3	4		0
2,5	3			
3	4	4	3	0
3,5	4			
4	5	5	4	3
4,5	5	5		
5	6	6	5	4
5,5	6	6	5,5	5
6	7	6	6	5
6,5	7	7	6,5	6
7	8	7	7	6
7,5	8	7,5	7,5	7
8	9	8	8	7
8,5	9	8	8,5	8
9	10	9	9	8
9,5	10,5	9	9,5	9
10	11	10	10	9
10,5	11,5	10	10,5	10
11	12	11	11	10
11,5	12,5	11	11,5	11
12	13	11,5	12	11
12,5	14	12	12,5	12
13	14,5	12	13	12
13,5	15	13	13,5	12,5
14	15,5	13	14	13
14,5	16	13,5	14,5	14
15	16,5	14	15	15
15,5	17	15	15,5	15
16	17,5	15	16	16
16,5	18	15,5	16,5	16,5
17	18,5	16	17	17
17,5	19	16	17,5	17,5
18	20	16,5	18	18
18,5	20,5	17	18,5	18,5
19	21	17,5	19	19
19,5	21,5	18	19,5	20
20	22	18,5	20	20,5
20,5	22,5	19	20,5	21
21	23	19	21	21,5
21,5	23,5	19,5	21,5	22
22	24	20	22	23

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012			
	2010	2011	2012	2013
26	28,5	23,5	26	27,5
26,5	29	24	26,5	28
27	29,5	24,5	27	28,5
27,5	30	25	27,5	29
28	30,5	25	28	29,5
28,5	31	25,5	28,5	30
29	31,5	26	29	31
29,5	32	26,5	29,5	31,5
30	32,5	27	30	32
30,5	33	27,5	30,5	32,5
31	33,5	28	31	33
31,5	34	28	31,5	33,5
32	34	28,5	32	34
32,5	35	29	32,5	34,5
33	35	29,5	33	35
33,5	36	30	33,5	35,5
34	36	30,5	34	36
34,5	36,5	31	34,5	36,5
35	37	31,5	35	37
35,5	37,5	32	35,5	37
36	38	32,5	36	38
36,5	38,5	33	36,5	38
37	39	33,5	37	38,5
37,5	39	34	37,5	39
38	39,5	34,5	38	39,5
38,5	40	35	38,5	40
39	40,5	35,5	39	40,5
39,5	41	36	39,5	41
40	41	37	40	41
40,5	42	37	40,5	42
41	42	38	41	42
41,5	42,5	38,5	41,5	42,5
42	43	39	42	43
42,5	43	39,5	42,5	43,5
43	44	40	43	44
43,5	44	41	43,5	44
44	44,5	41,5	44	45
44,5	45	42	44,5	45
45	45	42,5	45	46
45,5	46	43	45,5	46
46	46	44	46	46,5
46,5	46,5	44,5	46,5	47
47	47	45	47	48

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012			
	2010	2011	2012	2013
22,5	24,5	20,5	22,5	23,5
23	25	21	23	24
23,5	25,5	21,5	23,5	24,5
24	26	22	24	25
24,5	26,5	22	24,5	26
25	27	22,5	25	26,5
25,5	28	23	25,5	27

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012			
	2010	2011	2012	2013
47,5	47	46	47,5	48
48	48	46,5	48	49
48,5	48	47	48,5	49
49	49	48	49	50
49,5	49	49	49,5	
50	50	50	50	

1.5. Matura z matematyki

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012			
	2010	2011	2012	2013
0	0	2	0	0
1	1	2	1	1
2	2	3	2	2
3	3	3	3	3
4	3	4	4	4
5	4	5	5	5
6	5	6	6	6
7	6	7	7	7
8	6	8	8	8
9	7	9	9	9
10	8	11	10	10
11	8	12	11	10
12	9	13	12	11
13	10	15	13	12
14	11	16	14	13
15	12	18	15	15
16	13	20	16	16
17	14	21	17	17
18	15	23	18	18
19	16	24	19	19
20	17	26	20	20
21	18	27	21	21
22	19	28	22	22
23	20	30	23	24
24	22	31	24	25
25	23	32	25	26

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012			
	2010	2011	2012	2013
26	24	33	26	27
27	26	34	27	28
28	27	36	28	29
29	28	37	29	30
30	29	38	30	32
31	31	39	31	33
32	32	39	32	34
33	33	40	33	35
34	35	41	34	36
35	36	42	35	37
36	37	43	36	38
37	39	43	37	39
38	40	44	38	40
39	41	44	39	41
40	43	45	40	42
41	44	45	41	42
42	45	46	42	43
43	46	46	43	44
44	46	47	44	45
45	47	47	45	46
46	48	48	46	46
47	49	49	47	47
48	49	49	48	48
49	50	50	49	49
50	50	50	50	50

1.6. Matura z j. polskiego

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012			
	2010	2011	2012	2013
0	0	5	0	4
1	1	7	1	6
2	2	7	2	7
3	3	7	3	8
4	3	8	4	9
5	4	9	5	10
6	5	10	6	11
7	6	12	7	12
8	8	13	8	13
9	9	15	9	14
10	10	16	10	15
11	11	17	11	16
12	12	21	12	17
13	14	21	13	21
14	15	21	14	21
15	16	22	15	21
16	18	23	16	22
17	21	23	17	22
18	21	24	18	22
19	21	24	19	23
20	21	24	19	23
21	22	26	21	24
22	24	28	22	26
23	25	29	23	27
24	26	30	24	28
25	28	31	25	29
26	29	32	26	30
27	30	33	27	31
28	31	34	28	31
29	32	34	29	32
30	33	35	30	33
31	34	36	31	34
32	35	37	32	35
33	36	38	33	36
34	37	38	34	36
35	38	39	35	37

Wynik z egzaminu	Przeliczenie wyników na skalę egzaminu z roku 2012			
	2010	2011	2012	2013
36	39	40	36	38
37	40	41	37	39
38	41	42	38	40
39	42	43	39	41
40	43	43	40	42
41	44	44	41	42
42	45	45	42	43
43	46	46	43	44
44	47	47	44	45
45	48	48	45	46
46	49	48	46	47
47	50	49	47	48
48	51	50	48	49
49	52	51	49	49
50	53	52	50	50
51	54	53	51	51
52	55	53	52	52
53	56	54	53	53
54	57	55	54	54
55	58	56	55	54
56	59	57	56	55
57	60	57	57	56
58	61	58	58	57
59	62	59	59	58
60	63	60	60	59
61	64	61	61	60
62	65	62	62	60
63	66	62	63	61
64	67	63	64	62
65	68	64	65	63
66	69	65	66	64
67	70	66	67	65
68	70	67	68	66
69	70	68	69	67
70	70	70	70	70

Aneks 2. Plany zrównania wyników egzaminów

2.1. Sprawdzian 2012-2014

Źródło	Populacja	Próba	N	Zbiory zadań egzaminacyjnych														Zbiory dodatkowych zadań kotwiczących			
				T_{02}^A	T_{03}^A	T_{04}^A	T_{05}^A	T_{06}^A	T_{07}^A	T_{08}^A	T_{09}^A	T_{10}^A	T_{11}^A	T_{12}	T_{13}	T_{14}	C^1	C^2	C^3	C^4	
Dane egzaminacyjne	$\mathcal{P}_{02}$		$5,4 \cdot 10^5$	■																	
	$\mathcal{P}_{03}$		$5,3 \cdot 10^5$		■																
	$\mathcal{P}_{04}$		$5,2 \cdot 10^5$			■															
	$\mathcal{P}_{05}$		$4,9 \cdot 10^5$				■														
	$\mathcal{P}_{06}$		$4,6 \cdot 10^5$					■													
	$\mathcal{P}_{07}$		$4,4 \cdot 10^5$						■												
	$\mathcal{P}_{08}$		$4,1 \cdot 10^5$							■											
	$\mathcal{P}_{09}$		$4,0 \cdot 10^5$								■										
	$\mathcal{P}_{10}$		$3,8 \cdot 10^5$									■									
	$\mathcal{P}_{11}$		$3,7 \cdot 10^5$										■								
	$\mathcal{P}_{12}$		$3,6 \cdot 10^5$	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	
	$\mathcal{P}_{13}$		$3,5 \cdot 10^5$			■								■	■		■	■	■	■	
	$\mathcal{P}_{14}$		$3,4 \cdot 10^5$												■	■	■	■		■	
	Dane z sesji zrównujących	$\mathcal{P}_{12}$	S_{12}^1	800	■	■									■			■	■		
$\mathcal{P}_{12}$		S_{12}^2	800		■	■								■				■		■	
$\mathcal{P}_{12}$		S_{12}^3	800			■	■							■			■			■	
$\mathcal{P}_{12}$		S_{12}^4	800				■	■						■			■		■		
$\mathcal{P}_{12}$		S_{12}^5	800					■	■					■				■	■		
$\mathcal{P}_{12}$		S_{12}^6	800						■	■				■				■		■	
$\mathcal{P}_{12}$		S_{12}^7	800							■	■			■			■			■	
$\mathcal{P}_{12}$		S_{12}^8	800								■	■		■			■		■		
$\mathcal{P}_{12}$		S_{12}^9	800									■	■	■				■	■		
$\mathcal{P}_{12}$		S_{12}^{10}	800		■								■	■				■		■	
$\mathcal{P}_{12}$		S_{12}^{11}	800						■				■	■			■			■	
$\mathcal{P}_{13}$		S_{13}	800			■									■		■		■	■	
$\mathcal{P}_{13}$		S_{13}	800											■	■		■	■		■	
$\mathcal{P}_{14}$		S_{14}	1600												■	■	■	■		■	

Legenda:

□ uczniowie nie rozwiązują żadnego zadania

■ uczniowie rozwiązują wszystkie zadania

■ uczniowie rozwiązują wszystkie zadania na sesji innego typu

■ część uczniów rozwiązuje wszystkie zadania na sesji innego typu

▣ uczniowie rozwiązują podzbiór zadań

▣ część uczniów rozwiązuje podzbiór zadań na sesji innego typu

2.2. Gimnazjum 2011–2014

Źródło	Populacja	Próba	N	Zbiory zadań egzaminacyjnych														Zbiory dodatkowych zadań kotwiczących										
				T_{02}	T_{03}	T_{04}	T_{05}	T_{06}	T_{07}	T_{08}	T_{09}	T_{10}	T_{11}	T_{12}	T_{13}	T_{14}	C^1	C^2	C^3	C^4	C^5	C^6	C^7	C^8	C^9	C^{10}	C^{11}	
Dane egzaminacyjne	$\mathcal{P}_{02}$		$5 \cdot 10^5$	■																								
	$\mathcal{P}_{03}$		$5 \cdot 10^5$		■																							
	$\mathcal{P}_{04}$		$5 \cdot 10^5$			■																						
	$\mathcal{P}_{05}$		$5 \cdot 10^5$				■																					
	$\mathcal{P}_{06}$		$5 \cdot 10^5$					■																				
	$\mathcal{P}_{07}$		$5 \cdot 10^5$						■																			
	$\mathcal{P}_{08}$		$5 \cdot 10^5$							■																		
	$\mathcal{P}_{09}$		$5 \cdot 10^5$								■																	
	$\mathcal{P}_{10}$		$4 \cdot 10^5$									■																
	$\mathcal{P}_{11}$		$4 \cdot 10^5$	■	■	■	■	■	■	■	■	■	■	■			■	■	■	■	■	■	■	■	■	■		
	$\mathcal{P}_{12}$		$4 \cdot 10^5$									■		■			■	■	■		■		■	■				
	$\mathcal{P}_{13}$		$4 \cdot 10^5$				■					■		■	■		■	■		■			■					
	$\mathcal{P}_{14}$		$4 \cdot 10^5$											■	■		■	■										
	Dane z sesji zrównujących	$\mathcal{P}_{11}$	S_{11}^1	800	■	■							■					■								■		
$\mathcal{P}_{11}$		S_{11}^2	800		■	■						■					■					■						
$\mathcal{P}_{11}$		S_{11}^3	800			■	■					■									■	■						
$\mathcal{P}_{11}$		S_{11}^4	800				■	■				■						■			■							
$\mathcal{P}_{11}$		S_{11}^5	800					■	■			■				■		■										
$\mathcal{P}_{11}$		S_{11}^6	800						■	■			■				■							■				
$\mathcal{P}_{11}$		S_{11}^7	800							■	■		■						■					■				
$\mathcal{P}_{11}$		S_{11}^8	800								■	■	■						■	■								
$\mathcal{P}_{11}$		S_{11}^9	800		■				■				■							■						■		
$\mathcal{P}_{11}$		S_{11}^{10}	800					■				■	■										■		■			
$\mathcal{P}_{11}$		S_{11}^{11}	800		■							■	■										■		■			
$\mathcal{P}_{12}$		S_{12}	1600								■	■		■			■	■	■		■		■	■				
$\mathcal{P}_{13}$		S_{13}	1600				■					■		■	■		■	■		■			■					
$\mathcal{P}_{14}$		S_{14}	1600											■	■		■	■										

Legenda:

□ uczniowie nie rozwiązują żadnego zadania

■ uczniowie rozwiązują wszystkie zadania

■ uczniowie rozwiązują wszystkie zadania na sesji innego typu

■ część uczniów rozwiązuje wszystkie zadania na sesji innego typu

▣ uczniowie rozwiązują podzbiór zadań

▣ część uczniów rozwiązuje podzbiór zadań na sesji innego typu

2.3. Matura z języka angielskiego 2014

Źródło	Populacja	Próba	N	Zbiory zadań egzaminacyjnych										Zbiory dodatkowych zadań kotwiczących
				P_{10}	D_{10}	P_{11}	D_{11}	P_{12}	D_{12}	P_{13}	D_{13}	P_{14}	D_{14}	C^1
Dane egzaminacyjne	P_{10}		$2,6 \cdot 10^5$	■	■									
	P_{11}		$2,6 \cdot 10^5$			■	■							
	P_{12}		$2,6 \cdot 10^5$					■	■					
	P_{13}		$2,6 \cdot 10^5$							■	■			
	P_{14}		$2,6 \cdot 10^5$	■	■	■	■	■	■	■	■	■	■	■
Dane z sesji zrównujących	P_{14}	S_{14}^1	600	■			■					■	■	■
	P_{14}	S_{14}^2	600	■	■							■	■	■
	P_{14}	S_{14}^3	600	■						■		■	■	■
	P_{14}	S_{14}^4	600	■		■						■	■	■
	P_{14}	S_{14}^5	600			■			■			■	■	■
	P_{14}	S_{14}^6	600			■	■					■	■	■
	P_{14}	S_{14}^7	600			■		■				■	■	■
	P_{14}	S_{14}^8	600					■			■	■	■	■
	P_{14}	S_{14}^9	600					■	■			■	■	■
	P_{14}	S_{14}^{10}	600					■		■		■	■	■
	P_{14}	S_{14}^{11}	600							■	■	■	■	■
	P_{14}	S_{14}^{12}	600		■					■		■	■	■

Legenda:

□ uczniowie nie rozwiązują żadnego zadania

■ uczniowie rozwiązują wszystkie zadania

■ część uczniów rozwiązuje wszystkie zadania

■ uczniowie rozwiązują wszystkie zadania na sesji innego typu

■ część uczniów rozwiązuje wszystkie zadania na sesji innego typu

■ uczniowie rozwiązują podzbiór zadań

■ część uczniów rozwiązuje podzbiór zadań na sesji innego typu

2.4. Matura z matematyki 2013-2014

Źródło	Populacja	Próba	N	Zbiory zadań egzaminacyjnych										Zbiory dodatkowych zadań kotwiczących	
				P_{10}	D_{10}	P_{11}	D_{11}	P_{12}	D_{12}	P_{13}	D_{13}	P_{14}	D_{14}	C^1	C^2
Dane egzaminacyjne	P_{10}		$3,2 \cdot 10^5$	■	■										
	P_{11}		$3,2 \cdot 10^5$			■	■								
	P_{12}		$3,0 \cdot 10^5$					■	■						
	P_{13}		$3,0 \cdot 10^5$	■	■	■	■	■	■	■	■			■	
	P_{14}		$2,9 \cdot 10^5$					■	■	■	■	■	■		■
Dane z sesji zrównujących	2013	P_{13}	S_{13}^1	600	■		■			■	■			■	
		P_{13}	S_{13}^2	600	■	■		■		■	■				
		P_{13}	S_{13}^3	600		■	■		■	■	■				
		P_{13}	S_{13}^4	600			■	■		■	■				
		P_{13}	S_{13}^5	600				■	■	■	■			■	
		P_{13}	S_{13}^6	600	■			■	■	■	■				
		P_{13}	S_{13}^7	600		■			■	■	■			■	
	2014	P_{14}	S_{14}^1	600				■	■			■	■		■
		P_{14}	S_{14}^2	600				■		■		■	■		■
		P_{14}	S_{14}^3	600						■	■	■	■		■
		P_{14}	S_{14}^4	600					■	■		■	■		■

Legenda:

-  uczniowie nie rozwiązują żadnego zadania
-  uczniowie rozwiązują wszystkie zadania
-  część uczniów rozwiązuje wszystkie zadania
-  uczniowie rozwiązują wszystkie zadania na sesji innego typu
-  część uczniów rozwiązuje wszystkie zadania na sesji innego typu
-  uczniowie rozwiązują podzbiór zadań
-  część uczniów rozwiązuje podzbiór zadań na sesji innego typu

2.5. Matura z polskiego 2014

Źródło	Populacja	Próba	Pod-próba	N	Zbiory zadań egzaminacyjnych																		C^1	
					$P_{10}^{\otimes}$	P_{10}^I	P_{10}^{II}	D_{10}	$P_{11}^{\otimes}$	P_{11}^I	P_{11}^{II}	D_{11}	$P_{12}^{\otimes}$	P_{12}^I	P_{12}^{II}	D_{12}	$P_{13}^{\otimes}$	P_{13}^I	P_{13}^{II}	$D_{11}^{\otimes}$	D_{13}^I	D_{13}^{II}		P_{14}
Dane egzaminacyjne	$\mathcal{P}_{10}$			$3,2 \cdot 10^5$	■	■	■	■																
	$\mathcal{P}_{11}$			$3,2 \cdot 10^5$					■	■	■	■												
	$\mathcal{P}_{12}$			$3,0 \cdot 10^5$									■	■	■	■								
	$\mathcal{P}_{13}$			$3,0 \cdot 10^5$												■	■	■	■	■	■			
	$\mathcal{P}_{14}$			$2,9 \cdot 10^5$	■	■	■		■	■	■		■	■	■		■	■	■	■	■	■	■	■
Dane z sesji zrównujących	$\mathcal{P}_{14}$	S_{14}^1	$S_{14}^{1.I}$	377	■										■							■	■	■
	$\mathcal{P}_{14}$	S_{14}^1	$S_{14}^{1.II}$	229	■	■																■	■	■
	$\mathcal{P}_{14}$	S_{14}^2	$S_{14}^{2.I}$	484					■									■					■	■
	$\mathcal{P}_{14}$	S_{14}^2	$S_{14}^{2.II}$	149					■	■												■	■	■
	$\mathcal{P}_{14}$	S_{14}^3	$S_{14}^{3.I}$	581									■								■	■	■	■
	$\mathcal{P}_{14}$	S_{14}^3	$S_{14}^{3.II}$	82									■	■								■	■	■
	$\mathcal{P}_{14}$	S_{14}^4	$S_{14}^{4.I}$	245												■	■					■	■	■
	$\mathcal{P}_{14}$	S_{14}^4	$S_{14}^{4.II}$	442							■					■						■	■	■
	$\mathcal{P}_{14}$	S_{14}^5	$S_{14}^{5.I}$	113														■	■			■	■	■
	$\mathcal{P}_{14}$	S_{14}^5	$S_{14}^{5.II}$	511			■											■				■	■	■
	$\mathcal{P}_{14}$	S_{14}^6	$S_{14}^{6.I}$	482	■						■											■	■	■
	$\mathcal{P}_{14}$	S_{14}^6	$S_{14}^{6.II}$	195	■					■												■	■	■
	$\mathcal{P}_{14}$	S_{14}^7	$S_{14}^{7.I}$	600					■						■							■	■	■
	$\mathcal{P}_{14}$	S_{14}^7	$S_{14}^{7.II}$	54					■					■								■	■	■
	$\mathcal{P}_{14}$	S_{14}^8	$S_{14}^{8.I}$	522									■					■				■	■	■
	$\mathcal{P}_{14}$	S_{14}^8	$S_{14}^{8.II}$	180									■				■					■	■	■
	$\mathcal{P}_{14}$	S_{14}^9	$S_{14}^{9.I}$	474												■					■	■	■	■
	$\mathcal{P}_{14}$	S_{14}^9	$S_{14}^{9.II}$	154											■					■		■	■	■
	$\mathcal{P}_{14}$	S_{14}^{10}	$S_{14}^{10.I}$	388			■											■				■	■	■
	$\mathcal{P}_{14}$	S_{14}^{10}	$S_{14}^{10.II}$	224		■												■				■	■	■

Legenda:

□ uczniowie nie rozwiązują żadnego zadania

■ uczniowie rozwiązują wszystkie zadania

■ część uczniów rozwiązuje wszystkie zadania

■ część uczniów rozwiązuje wszystkie zadania na sesji innego typu

■ uczniowie rozwiązują podzbiór zadań

■ część uczniów rozwiązuje podzbiór zadań na sesji innego typu

■ uczniowie rozwiązują podzbiór zadań na sesji innego typu

■ część uczniów rozwiązuje podzbiór zadań

Aneks 3. Zadania wchodzące w skład zeszytów testowych

3.1. Sprawdzian

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty													
					2012											2013		2014
					1	2	3	4	5	6	7	8	9	10	11	1	2	1
sprawdzian	2002	S-A1-021	02	1091	x													
sprawdzian	2002	S-A1-021	03	1092	x													
sprawdzian	2002	S-A1-021	05	1094	x													
sprawdzian	2002	S-A1-021	06	1095	x													
sprawdzian	2002	S-A1-021	08	1097	x													
sprawdzian	2002	S-A1-021	09	1098	x													
sprawdzian	2002	S-A1-021	10	1099	x													
sprawdzian	2002	S-A1-021	11	1100	x													
sprawdzian	2002	S-A1-021	18	1107	x													
sprawdzian	2002	S-A1-021	20	1109	x													
sprawdzian	2002	S-A1-021	22	1111	x													
sprawdzian	2002	S-A1-021	23	1112	x													
sprawdzian	2003	S-A1-032	02	1124	x	x								x				
sprawdzian	2003	S-A1-032	03	1125	x	x								x				
sprawdzian	2003	S-A1-032	05	1127	x	x								x				
sprawdzian	2003	S-A1-032	09	1131	x	x								x				
sprawdzian	2003	S-A1-032	10	1132	x	x								x				
sprawdzian	2003	S-A1-032	12	1134	x	x								x				
sprawdzian	2003	S-A1-032	15	1137	x	x								x				
sprawdzian	2003	S-A1-032	18	1140	x	x								x				
sprawdzian	2003	S-A1-032	19	1141	x	x								x				
sprawdzian	2003	S-A1-032	22	1144	x	x								x				
sprawdzian	2003	S-A1-032	25	1160	x	x								x				

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty													
					2012											2013		2014
					1	2	3	4	5	6	7	8	9	10	11	1	2	1
sprawdzian	2004	S-A1-04	04	1165		x	x										x	
sprawdzian	2004	S-A1-04	05	1166		x	x										x	
sprawdzian	2004	S-A1-04	06	1167		x	x										x	
sprawdzian	2004	S-A1-04	07	1168		x	x											
sprawdzian	2004	S-A1-04	13	1174		x	x										x	
sprawdzian	2004	S-A1-04	15	1176		x	x										x	
sprawdzian	2004	S-A1-04	16	1177		x	x										x	
sprawdzian	2004	S-A1-04	17	1178		x	x										x	
sprawdzian	2004	S-A1-04	18	1179		x	x										x	
sprawdzian	2004	S-A1-04	23	1193		x	x										x	
sprawdzian	2004	S-A1-04	24	1195		x	x										x	
sprawdzian	2005	S-A1-052	02	1202			x	x										
sprawdzian	2005	S-A1-052	03	1203			x	x										
sprawdzian	2005	S-A1-052	05	1205			x	x										
sprawdzian	2005	S-A1-052	06	1206			x	x										
sprawdzian	2005	S-A1-052	07	1207			x	x										
sprawdzian	2005	S-A1-052	10	1210			x	x										
sprawdzian	2005	S-A1-052	11	1211			x	x										
sprawdzian	2005	S-A1-052	13	1213			x	x										
sprawdzian	2005	S-A1-052	15	1215			x	x										
sprawdzian	2005	S-A1-052	20	1220			x	x										
sprawdzian	2005	S-A1-052	25	1231			x	x										
sprawdzian	2006	S-A1-062	05	1242				x	x									
sprawdzian	2006	S-A1-062	06	1243				x	x									
sprawdzian	2006	S-A1-062	07	1244				x	x									
sprawdzian	2006	S-A1-062	08	1245				x	x									
sprawdzian	2006	S-A1-062	09	1246				x	x									

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty													
					2012											2013		2014
					1	2	3	4	5	6	7	8	9	10	11	1	2	1
sprawdzian	2006	S-A1-062	10	1247				x	x									
sprawdzian	2006	S-A1-062	12	1249				x	x									
sprawdzian	2006	S-A1-062	13	1250				x	x									
sprawdzian	2006	S-A1-062	14	1251				x	x									
sprawdzian	2006	S-A1-062	15	1252				x	x									
sprawdzian	2006	S-A1-062	16	1253				x	x									
sprawdzian	2006	S-A1-062	18	1255				x	x									
sprawdzian	2006	S-A1-062	20	1257				x	x									
sprawdzian	2006	S-A1-062	22	1264				x	x									
sprawdzian	2006	S-A1-062	23	1265				x	x									
sprawdzian	2007	S-1-072	01	1276					x	x					x			
sprawdzian	2007	S-1-072	02	1277					x	x					x			
sprawdzian	2007	S-1-072	03	1278					x	x					x			
sprawdzian	2007	S-1-072	06	1281					x	x					x			
sprawdzian	2007	S-1-072	08	1283					x	x					x			
sprawdzian	2007	S-1-072	12	1287					x	x					x			
sprawdzian	2007	S-1-072	15	1290					x	x					x			
sprawdzian	2007	S-1-072	17	1292					x	x					x			
sprawdzian	2007	S-1-072	18	1293					x	x					x			
sprawdzian	2007	S-1-072	19	1294					x	x					x			
sprawdzian	2007	S-1-072	22	1299					x	x					x			
sprawdzian	2007	S-1-072	23	1303					x	x					x			
sprawdzian	2008	S-1-082	04	1315						x	x							
sprawdzian	2008	S-1-082	05	1316						x	x							
sprawdzian	2008	S-1-082	07	1318						x	x							
sprawdzian	2008	S-1-082	11	1322						x	x							
sprawdzian	2008	S-1-082	12	1323						x	x							

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty													
					2012											2013		2014
					1	2	3	4	5	6	7	8	9	10	11	1	2	1
sprawdzian	2008	S-1-082	16	1327						x	x							
sprawdzian	2008	S-1-082	17	1328						x	x							
sprawdzian	2008	S-1-082	18	1329						x	x							
sprawdzian	2008	S-1-082	19	1330						x	x							
sprawdzian	2008	S-1-082	23	1336						x	x							
sprawdzian	2008	S-1-082	24	1340						x	x							
sprawdzian	2009	S-1-092	01	1347							x	x						
sprawdzian	2009	S-1-092	02	1348							x	x						
sprawdzian	2009	S-1-092	04	1350							x	x						
sprawdzian	2009	S-1-092	05	1351							x	x						
sprawdzian	2009	S-1-092	07	1353							x	x						
sprawdzian	2009	S-1-092	11	1357							x	x						
sprawdzian	2009	S-1-092	15	1361							x	x						
sprawdzian	2009	S-1-092	16	1362							x	x						
sprawdzian	2009	S-1-092	17	1363							x	x						
sprawdzian	2009	S-1-092	19	1365							x	x						
sprawdzian	2009	S-1-092	20	1366							x	x						
sprawdzian	2009	S-1-092	22_1	1368							x	x						
sprawdzian	2009	S-1-092	22_2	1369							x	x						
sprawdzian	2009	S-1-092	23	1370							x	x						
sprawdzian	2009	S-1-092	23	1370							x	x						
sprawdzian	2009	S-1-092	23	1370							x	x						
sprawdzian	2010	S-1-102	01	1382								x	x					
sprawdzian	2010	S-1-102	02	1383								x	x					
sprawdzian	2010	S-1-102	03	1384								x	x					
sprawdzian	2010	S-1-102	05	1386								x	x					
sprawdzian	2010	S-1-102	06	1387								x	x					

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty													
					2012											2013		2014
					1	2	3	4	5	6	7	8	9	10	11	1	2	1
sprawdzian	2012	S-1-122	11	1456												x		x
sprawdzian	2012	S-1-122	12	1457												x		x
sprawdzian	2012	S-1-122	13	1458												x	x	x
sprawdzian	2012	S-1-122	15	1460												x	x	x
sprawdzian	2012	S-1-122	16	1461												x	x	x
sprawdzian	2012	S-1-122	18	1463												x	x	x
sprawdzian	2012	S-1-122	20	1465												x	x	x
sprawdzian	2012	S-1-122	21	1466												x	x	x
sprawdzian	2012	S-1-122	25	1470													x	
sprawdzian	2013	S-1-132	01	3991														x
sprawdzian	2013	S-1-132	02	3992														x
sprawdzian	2013	S-1-132	03	3993														x
sprawdzian	2013	S-1-132	05	3995														x
sprawdzian	2013	S-1-132	07	3997														x
sprawdzian	2013	S-1-132	08	3998														x
sprawdzian	2013	S-1-132	13	4003														x
sprawdzian	2013	S-1-132	14	4004														x
sprawdzian	2013	S-1-132	18	4008														x
sprawdzian	2013	S-1-132	19	4009														x
sprawdzian	2013	S-1-132	20	4010														x
sprawdzian	2013	S-1-132	21	4011														x
sprawdzian	2013	S-1-132	23	4013														x
sprawdzian	2013	S-1-132	25	4015														x
arkusz niepub.		ht1	25	1506	x	x			x	x			x	x		x		
arkusz niepub.		ht1	29	1507	x	x			x	x			x	x		x		
arkusz niepub.		ht2	26	1515				x	x			x	x				x	
arkusz niepub.		ht2	27	1510		x	x			x	x			x	x	x	x	x

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty													
					2012											2013		2014
					1	2	3	4	5	6	7	8	9	10	11	1	2	1
arkusz niepub.		ht2	30	1511		x	x			x	x			x	x	x	x	x
arkusz niepub.		ht2	33	1516				x	x			x	x					
arkusz niepub.		ht2	35	1504	x		x	x			x	x			x			x
arkusz niepub.		ht2	37	1505	x		x	x			x	x			x			x
arkusz niepub.		mc1	29	1508		x	x			x	x			x	x		x	
arkusz niepub.		mc1	30	1503	x	x			x	x			x	x				
arkusz niepub.		mc1	36	1514				x	x			x	x					
arkusz niepub.		mc2	26	1502	x	x			x	x			x	x		x		x
arkusz niepub.		mc2	40	1501	x		x	x			x	x			x	x	x	x
arkusz niepub.		mcs	23	1509		x	x			x	x			x	x	x	x	x

3.2. Egzamin gimnazjalny – część humanistyczna

[illegible]

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część humanistyczna											historia	j.polski	historia	j.polski	historia	j.polski
																1	1	1	1	1	1
gimn.	2002	GH-A1-021	25	92	x																
gimn.	2002	GH-A1-021	26	93	x																
gimn.	2002	GH-A1-021	27	94	x																
gimn.	2002	GH-A1-021	28	96	x																
gimn.	2003	GH-A1-031	10	124	x	x								x		x					
gimn.	2003	GH-A1-031	12	126	x	x								x		x					
gimn.	2003	GH-A1-031	14	128	x	x								x		x					
gimn.	2003	GH-A1-031	22	136	x	x								x		x					
gimn.	2003	GH-A1-031	23	137	x	x								x		x					
gimn.	2003	GH-A1-031	24	138	x	x								x		x					
gimn.	2003	GH-A1-031	25	139	x	x								x		x					
gimn.	2003	GH-A1-031	26	140	x	x								x		x					
gimn.	2003	GH-A1-031	27	141	x	x								x		x					
gimn.	2003	GH-A1-031	28	142	x	x								x		x					
gimn.	2003	GH-A1-031	29	143	x	x								x		x					
gimn.	2003	GH-A1-031	30	144	x	x								x		x					
gimn.	2004	GH-A1-042	13	175		x	x														
gimn.	2004	GH-A1-042	15	177		x	x														
gimn.	2004	GH-A1-042	19	181		x	x														
gimn.	2004	GH-A1-042	20	182		x	x														
gimn.	2004	GH-A1-042	21	183		x	x														
gimn.	2004	GH-A1-042	22	184		x	x														
gimn.	2004	GH-A1-042	23	185		x	x														

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część humanistyczna											historia	j.polski	historia	j.polski	historia	j.polski
gimn.	2004	GH-A1-042	24	186		x	x														
gimn.	2004	GH-A1-042	25	187		x	x														
gimn.	2004	GH-A1-042	26	188		x	x														
gimn.	2004	GH-A1-042	28	190		x	x														
gimn.	2004	GH-A1-042	29	191		x	x														
gimn.	2004	GH-A1-042	30	192		x	x														
gimn.	2005	GH-A1-052	11	221			x	x										x			
gimn.	2005	GH-A1-052	12	222			x	x													
gimn.	2005	GH-A1-052	14	224			x	x										x			
gimn.	2005	GH-A1-052	16	226			x	x													
gimn.	2005	GH-A1-052	18	228			x	x													
gimn.	2005	GH-A1-052	20	230			x	x													
gimn.	2005	GH-A1-052	22	232			x	x													
gimn.	2005	GH-A1-052	23	233			x	x													
gimn.	2005	GH-A1-052	26	236			x	x													
gimn.	2005	GH-A1-052	27	237			x	x													
gimn.	2005	GH-A1-052	28	238			x	x													
gimn.	2005	GH-A1-052	31	241			x	x													
gimn.	2005	GH-A1-052	32	242			x	x													
gimn.	2006	GH-A1-062	8	266				x	x					x	x						
gimn.	2006	GH-A1-062	9	267				x	x					x	x						
gimn.	2006	GH-A1-062	11	269				x	x					x	x						
gimn.	2006	GH-A1-062	16	274				x	x					x	x						

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część humanistyczna											historia	j.polski	historia	j.polski	historia	j.polski
1	2	3	4	5	6	7	8	9	10	11	1	1	1	1	1	1					
gimn.	2006	GH-A1-062	17	275				x	x				x	x							
gimn.	2006	GH-A1-062	18	276				x	x				x	x							
gimn.	2006	GH-A1-062	20	278				x	x				x	x							
gimn.	2006	GH-A1-062	23	284				x	x				x	x							
gimn.	2006	GH-A1-062	24	286				x	x				x	x							
gimn.	2006	GH-A1-062	25	288				x	x				x	x							
gimn.	2007	GH-A1-072	14	319					x	x											
gimn.	2007	GH-A1-072	15	320					x	x											
gimn.	2007	GH-A1-072	16	321					x	x											
gimn.	2007	GH-A1-072	17	322					x	x											
gimn.	2007	GH-A1-072	22	327					x	x											
gimn.	2007	GH-A1-072	23	328					x	x											
gimn.	2007	GH-A1-072	25	330					x	x											
gimn.	2007	GH-A1-072	26	331					x	x											
gimn.	2007	GH-A1-072	28	333					x	x											
gimn.	2007	GH-A1-072	29	334					x	x											
gimn.	2008	GH-A1-082	7	358						x	x					x					
gimn.	2008	GH-A1-082	8	359						x	x					x					
gimn.	2008	GH-A1-082	9	360						x	x					x					
gimn.	2008	GH-A1-082	11	362						x	x					x					
gimn.	2008	GH-A1-082	13	364						x	x					x					
gimn.	2008	GH-A1-082	14	365						x	x					x					
gimn.	2008	GH-A1-082	22	373						x	x										

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część humanistyczna											historia	j.polski	historia	j.polski	historia	j.polski
gimn.	2008	GH-A1-082	23	374						x	x										
gimn.	2008	GH-A1-082	26	377						x	x										
gimn.	2008	GH-A1-082	27	378						x	x										
gimn.	2008	GH-A1-082	28	379						x	x										
gimn.	2008	GH-A1-082	29	380						x	x										
gimn.	2008	GH-A1-082	30	381						x	x										
gimn.	2008	GH-A1-082	31	382						x	x										
gimn.	2009	GH-A1-092	10	408							x	x									
gimn.	2009	GH-A1-092	11	409												x					
gimn.	2009	GH-A1-092	12	410							x	x									
gimn.	2009	GH-A1-092	13	411							x	x									
gimn.	2009	GH-A1-092	14	412												x					
gimn.	2009	GH-A1-092	16	414							x	x									
gimn.	2009	GH-A1-092	17	415							x	x				x		x			
gimn.	2009	GH-A1-092	18	416							x	x				x		x			
gimn.	2009	GH-A1-092	19	417							x	x				x		x			
gimn.	2009	GH-A1-092	22	420							x	x									
gimn.	2009	GH-A1-092	24	422							x	x									
gimn.	2009	GH-A1-092	26	424							x	x									
gimn.	2009	GH-A1-092	28	428							x	x									
gimn.	2010	GH-A1-102	7	451								x		x	x						
gimn.	2010	GH-A1-102	8	452								x		x	x						
gimn.	2010	GH-A1-102	9	453								x		x	x						

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																	
					2011											2012		2013		2014		
					część humanistyczna											historia	j.polski	historia	j.polski	historia	j.polski	
																						1
gimn.	2010	GH-A1-102	10	454									x		x	x						
gimn.	2010	GH-A1-102	11	455									x		x	x						
gimn.	2010	GH-A1-102	12	456									x		x	x						
gimn.	2010	GH-A1-102	18	462									x		x	x						
gimn.	2010	GH-A1-102	19	463									x		x	x						
gimn.	2010	GH-A1-102	20	464									x		x	x						
gimn.	2010	GH-A1-102	26	474									x		x	x						
gimn.	2010	GH-A1-102	27	475									x		x	x						
gimn.	2010	GH-A1-102	28	476									x		x	x						
gimn.	2011	GH-A1-112	7	496													x		x			
gimn.	2011	GH-A1-112	8	497														x		x		
gimn.	2011	GH-A1-112	9	498														x		x		
gimn.	2011	GH-A1-112	10	499														x		x		
gimn.	2011	GH-A1-112	11	500													x		x			
gimn.	2011	GH-A1-112	12	501														x		x		
gimn.	2011	GH-A1-112	14	503													x					
gimn.	2011	GH-A1-112	15	504													x		x			
gimn.	2011	GH-A1-112	16	505													x		x			
gimn.	2011	GH-A1-112	18	507														x		x		
gimn.	2011	GH-A1-112	19	508														x		x		
gimn.	2011	GH-A1-112	20	509														x		x		
gimn.	2011	GH-A1-112	27	516														x		x		
gimn.	2012	GH-HA1-122	3	37																	x	

[illegible]

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część humanistyczna											historia	j.polski	historia	j.polski	historia	j.polski
																1	2	3	4	5	6
gimn.	2012	GH-PA1-122	16	541															x		x
gimn.	2012	GH-PA1-122	17	542															x		x
gimn.	2012	GH-PA1-122	19	544															x		x
gimn.	2012	GH-PA1-122	20	545															x		x
gimn.	2012	GH-PA1-122	21	546															x		x
gimn.	2013	GH-HA1-132	2	4069																x	
gimn.	2013	GH-HA1-132	4	4071																x	
gimn.	2013	GH-HA1-132	7	4074																x	
gimn.	2013	GH-HA1-132	8	4075																x	
gimn.	2013	GH-HA1-132	9	4076																x	
gimn.	2013	GH-HA1-132	12	4081																x	
gimn.	2013	GH-HA1-132	13	4082																x	
gimn.	2013	GH-HA1-132	14	4083																x	
gimn.	2013	GH-HA1-132	19_1	4090																x	
gimn.	2013	GH-HA1-132	19_2	4091																x	
gimn.	2013	GH-HA1-132	19_3	4092																x	
gimn.	2013	GH-HA1-132	20_1	4093																x	
gimn.	2013	GH-HA1-132	20_2	4094																x	
gimn.	2013	GH-HA1-132	21	4095																x	
gimn.	2013	GH-PA1-132	1	4152																	x
gimn.	2013	GH-PA1-132	3	4154																	x
gimn.	2013	GH-PA1-132	5	4156																	x
gimn.	2013	GH-PA1-132	7	4158																	x

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część humanistyczna											historia	j.polski	historia	j.polski	historia	j.polski
gimn.	2013	GH-PA1-132	9	4160																	x
gimn.	2013	GH-PA1-132	10	4161																	x
gimn.	2013	GH-PA1-132	11	4162																	x
gimn.	2013	GH-PA1-132	14	4165																	x
gimn.	2013	GH-PA1-132	15	4166																	x
gimn.	2013	GH-PA1-132	17	4168																	x
gimn.	2013	GH-PA1-132	18	4169																	x
gimn.	2013	GH-PA1-132	19	4170																	x
gimn.	2013	GH-PA1-132	21	4172																	x
arkusz niepub.		ht1	2	1527	x	x											x		x		
arkusz niepub.		ht1	3	1528	x	x											x		x		
arkusz niepub.		ht1	4	1531	x	x											x		x		
arkusz niepub.		ht1	5	1532	x	x											x		x		
arkusz niepub.		ht1	6	1529	x	x											x		x		
arkusz niepub.		ht1	7	1530	x	x											x		x		
arkusz niepub.		ht1	8	1609	x	x															
arkusz niepub.		ht1	9	1610	x	x															
arkusz niepub.		ht1	10	1617		x	x														
arkusz niepub.		ht1	11	1540		x	x											x			
arkusz niepub.		ht1	12	1541		x	x											x			
arkusz niepub.		ht1	13	1542		x	x											x			
arkusz niepub.		ht1	14	1544		x	x											x			
arkusz niepub.		ht1	15	1545		x	x											x			

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																	
					2011											2012		2013		2014		
					część humanistyczna											historia	j.polski	historia	j.polski	historia	j.polski	
																1	2	3	4	5	6	7
arkusz niepub.		ht1	16	1546		x	x										x					
arkusz niepub.		ht1	17	1555								x	x				x		x			
arkusz niepub.		ht1	18	1556								x	x				x		x			
arkusz niepub.		ht1	20	1558								x	x				x					
arkusz niepub.		ht1	24	1557								x	x				x		x			
arkusz niepub.		ht1	25	1506			x	x								x		x		x		
arkusz niepub.		ht1	26	1525			x	x								x		x		x		
arkusz niepub.		ht1	27	1627			x	x														
arkusz niepub.		ht1	28	1628			x	x														
arkusz niepub.		ht1	29	1507			x	x														
arkusz niepub.		ht1	30	1630			x	x														
arkusz niepub.		ht1	31	1631			x	x														
arkusz niepub.		ht1	32	1632			x	x														
arkusz niepub.		ht1	33	1633			x	x														
arkusz niepub.		ht1	34	1548					x	x							x		x			x
arkusz niepub.		ht1	40	1634				x	x													
arkusz niepub.		ht1	41	1672								x	x									
arkusz niepub.		ht1	42	1549								x	x				x					
arkusz niepub.		ht1	43	1674								x	x									
arkusz niepub.		ht1	44	1675								x	x									
arkusz niepub.		ht1	45	1611	x										x							
arkusz niepub.		ht2	1	1538				x	x								x					
arkusz niepub.		ht2	2	1543				x	x								x					

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część humanistyczna											historia	j.polski	historia	j.polski	historia	j.polski
																1	1	1	1	1	1
arkusz niepub.		ht2	3	1539				x	x								x				
arkusz niepub.		ht2	4	1547				x	x								x				
arkusz niepub.		ht2	5	1639				x	x												
arkusz niepub.		ht2	6	1640				x	x												
arkusz niepub.		ht2	8	1641				x	x												
arkusz niepub.		ht2	9	1642					x	x											
arkusz niepub.		ht2	10	1559					x	x							x		x		
arkusz niepub.		ht2	12	1644					x	x											
arkusz niepub.		ht2	13	1560					x	x							x		x		
arkusz niepub.		ht2	14	1561					x	x							x		x		
arkusz niepub.		ht2	15	1562					x	x							x		x		
arkusz niepub.		ht2	16	1563					x	x							x		x		
arkusz niepub.		ht2	17	1649					x	x											
arkusz niepub.		ht2	19	1526						x	x					x		x			
arkusz niepub.		ht2	20	1652						x	x										
arkusz niepub.		ht2	21	1653						x	x										
arkusz niepub.		ht2	23	1654						x	x										
arkusz niepub.		ht2	24	1655						x	x										
arkusz niepub.		ht2	25	1656						x	x										
arkusz niepub.		ht2	26	1515							x	x					x		x		x
arkusz niepub.		ht2	27	1510							x	x					x		x		x
arkusz niepub.		ht2	29	1537							x	x					x		x		
arkusz niepub.		ht2	30	1511							x	x					x		x		x

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część humanistyczna											historia	j.polski	historia	j.polski	historia	j.polski
																1	1	1	1	1	1
arkusz niepub.		ht2	31	1664							x	x									
arkusz niepub.		ht2	32	1665							x	x									
arkusz niepub.		ht2	33	1516							x	x					x				
arkusz niepub.		ht2	34	1667							x	x									
arkusz niepub.		ht2	35	1504										x	x		x		x		x
arkusz niepub.		ht2	36	1551										x	x		x		x		
arkusz niepub.		ht2	37	1505										x	x		x		x		x
arkusz niepub.		ht2	39	1684										x	x						
arkusz niepub.		ht2	40	1552										x	x		x		x		x
arkusz niepub.		ht2	41	1553										x	x		x		x		
arkusz niepub.		ht2	45	1657							x	x									
arkusz niepub.		ht2	46	1659							x	x									
arkusz niepub.		ht2	47	1676										x	x						

3.3. Egzamin gimnazjalny – część matematyczno-przyrodnicza

Egzamin	Rok egz.	Kod arkusza	Nr. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część matematyczno-przyrodnicza											matem.	przyroda	matem.	przyroda	matem.	przyroda
					1	2	3	4	5	6	7	8	9	10	11	1	1	1	1	1	1
gimn.	2002	GM-A1-021	3	555	x									x							
gimn.	2002	GM-A1-021	4	556	x									x							
gimn.	2002	GM-A1-021	5	557	x									x							
gimn.	2002	GM-A1-021	9	561	x									x							
gimn.	2002	GM-A1-021	10	562	x									x							
gimn.	2002	GM-A1-021	16	568	x									x							
gimn.	2002	GM-A1-021	19	571	x									x							
gimn.	2002	GM-A1-021	20	572	x									x							
gimn.	2002	GM-A1-021	22	574	x									x							
gimn.	2002	GM-A1-021	25	577	x									x							
gimn.	2002	GM-A1-021	26	578	x									x							
gimn.	2002	GM-A1-021	32	592	x									x							
gimn.	2003	GM-A1-031	2	604	x	x							x		x						
gimn.	2003	GM-A1-031	3	605	x	x							x		x						
gimn.	2003	GM-A1-031	5	607	x	x							x		x						
gimn.	2003	GM-A1-031	6	608	x	x							x		x						
gimn.	2003	GM-A1-031	8	610	x	x							x		x						
gimn.	2003	GM-A1-031	9	611	x	x							x		x						
gimn.	2003	GM-A1-031	11	613	x	x							x		x						
gimn.	2003	GM-A1-031	13	615	x	x							x		x						
gimn.	2003	GM-A1-031	14	616	x	x							x		x						
gimn.	2003	GM-A1-031	18	620	x	x							x		x						

Egzamin	Rok egz.	Kod arkusza	Nr. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część matematyczno-przyrodnicza											matem.	przyroda	matem.	przyroda	matem.	przyroda
					1	2	3	4	5	6	7	8	9	10	11	1	1	1	1	1	1
gimn.	2003	GM-A1-031	21	623	x	x							x		x						
gimn.	2003	GM-A1-031	25	627	x	x							x		x						
gimn.	2003	GM-A1-031	33	646	x	x							x		x						
gimn.	2004	GM-A1-042	1	653		x	x														
gimn.	2004	GM-A1-042	3	655		x	x														
gimn.	2004	GM-A1-042	4	656		x	x														
gimn.	2004	GM-A1-042	5	657		x	x														
gimn.	2004	GM-A1-042	6	658		x	x														
gimn.	2004	GM-A1-042	10	662		x	x														
gimn.	2004	GM-A1-042	12	664		x	x														
gimn.	2004	GM-A1-042	18	670		x	x														
gimn.	2004	GM-A1-042	22	674		x	x														
gimn.	2004	GM-A1-042	23	675		x	x														
gimn.	2004	GM-A1-042	24	676		x	x														
gimn.	2004	GM-A1-042	34	696		x	x														
gimn.	2005	GM-A1-052	8	708			x	x													
gimn.	2005	GM-A1-052	9	709			x	x													
gimn.	2005	GM-A1-052	12	712			x	x													
gimn.	2005	GM-A1-052	13	713			x	x													
gimn.	2005	GM-A1-052	14	714			x	x													
gimn.	2005	GM-A1-052	17	717			x	x													
gimn.	2005	GM-A1-052	18	718			x	x													
gimn.	2005	GM-A1-052	19	719			x	x													

[illegible]

Egzamin	Rok egz.	Kod arkusza	Nr. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część matematyczno-przyrodnicza											matem.	przyroda	matem.	przyroda	matem.	przyroda
1	2	3	4	5	6	7	8	9	10	11	1	1	1	1	1	1					
gimn.	2007	GM-A1-072	12	811					x	x											
gimn.	2007	GM-A1-072	14	813					x	x											
gimn.	2007	GM-A1-072	15	814					x	x											
gimn.	2007	GM-A1-072	16	815					x	x											
gimn.	2007	GM-A1-072	19	818					x	x											
gimn.	2007	GM-A1-072	22	821					x	x											
gimn.	2007	GM-A1-072	24	823					x	x											
gimn.	2007	GM-A1-072	32	837					x	x											
gimn.	2008	GM-A1-082	1	848						x	x										
gimn.	2008	GM-A1-082	2	849						x	x										
gimn.	2008	GM-A1-082	7	854						x	x										
gimn.	2008	GM-A1-082	11	858						x	x										
gimn.	2008	GM-A1-082	12	859						x	x										
gimn.	2008	GM-A1-082	14	861						x	x										
gimn.	2008	GM-A1-082	15	862						x	x										
gimn.	2008	GM-A1-082	16	863						x	x										
gimn.	2008	GM-A1-082	17	864						x	x										
gimn.	2008	GM-A1-082	20	867						x	x										
gimn.	2008	GM-A1-082	21	868						x	x										
gimn.	2008	GM-A1-082	25	872						x	x										
gimn.	2008	GM-A1-082	33	892						x	x										
gimn.	2009	GM-A1-092	5	901							x	x									
gimn.	2009	GM-A1-092	6	902							x	x									

Egzamin	Rok egz.	Kod arkusza	Nr. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część matematyczno-przyrodnicza											matem.	przyroda	matem.	przyroda	matem.	przyroda
1	2	3	4	5	6	7	8	9	10	11	1	1	1	1	1	1					
gimn.	2009	GM-A1-092	9	905							x	x									
gimn.	2009	GM-A1-092	10	906							x	x									
gimn.	2009	GM-A1-092	13	909							x	x									
gimn.	2009	GM-A1-092	14	910							x	x									
gimn.	2009	GM-A1-092	18	914							x	x									
gimn.	2009	GM-A1-092	19	915							x	x									
gimn.	2009	GM-A1-092	20	916							x	x									
gimn.	2009	GM-A1-092	24	920							x	x									
gimn.	2009	GM-A1-092	25	921							x	x									
gimn.	2009	GM-A1-092	27	924							x	x									
gimn.	2009	GM-A1-092	33	935							x	x									
gimn.	2010	GM-A1-102	4	950								x		x	x						
gimn.	2010	GM-A1-102	6	952								x		x	x						
gimn.	2010	GM-A1-102	7	953								x		x	x						
gimn.	2010	GM-A1-102	8	954								x		x	x						
gimn.	2010	GM-A1-102	11	957								x		x	x						
gimn.	2010	GM-A1-102	12	958								x		x	x						
gimn.	2010	GM-A1-102	16	962								x		x	x						
gimn.	2010	GM-A1-102	18	964								x		x	x						
gimn.	2010	GM-A1-102	19	965								x		x	x						
gimn.	2010	GM-A1-102	25	971								x		x	x						
gimn.	2010	GM-A1-102	29	981								x		x	x						
gimn.	2010	GM-A1-102	32	987								x		x	x						

Egzamin	Rok egz.	Kod arkusza	Nr. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część matematyczno-przyrodnicza											matem.	przyroda	matem.	przyroda	matem.	przyroda
					1	2	3	4	5	6	7	8	9	10	11	1	1	1	1	1	1
gimn.	2011	GM-A1-112	1	995												x		x			
gimn.	2011	GM-A1-112	2	996												x		x			
gimn.	2011	GM-A1-112	3	997												x		x			
gimn.	2011	GM-A1-112	5	999												x		x			
gimn.	2011	GM-A1-112	6	1000												x		x			
gimn.	2011	GM-A1-112	7	1001													x				
gimn.	2011	GM-A1-112	9	1003													x		x		
gimn.	2011	GM-A1-112	11	1005													x		x		
gimn.	2011	GM-A1-112	12	1006													x		x		
gimn.	2011	GM-A1-112	13	1007													x		x		
gimn.	2011	GM-A1-112	14	1008													x		x		
gimn.	2011	GM-A1-112	18	1012													x		x		
gimn.	2011	GM-A1-112	21	1015													x		x		
gimn.	2011	GM-A1-112	24	1018												x		x			
gimn.	2011	GM-A1-112	25	1019												x		x			
gimn.	2012	GM-MA1-122	2	1042														x		x	
gimn.	2012	GM-MA1-122	3	1043														x		x	
gimn.	2012	GM-MA1-122	4	1044														x		x	
gimn.	2012	GM-MA1-122	7	1047														x		x	
gimn.	2012	GM-MA1-122	10	1050														x		x	
gimn.	2012	GM-MA1-122	12	1052														x		x	
gimn.	2012	GM-MA1-122	13	1053														x		x	
gimn.	2012	GM-MA1-122	14	1054														x		x	

[illegible]

Egzamin	Rok egz.	Kod arkusza	Nr. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część matematyczno-przyrodnicza											matem.	przyroda	matem.	przyroda	matem.	przyroda
					1	2	3	4	5	6	7	8	9	10	11	1	1	1	1	1	1
gimn.	2013	GM-MA1-132	22	4150																x	
gimn.	2013	GM-PA1-132	2	4103																	x
gimn.	2013	GM-PA1-132	3	4104																	x
gimn.	2013	GM-PA1-132	8	4109																	x
gimn.	2013	GM-PA1-132	9	4110																	x
gimn.	2013	GM-PA1-132	13	4115																	x
gimn.	2013	GM-PA1-132	14	4116																	x
gimn.	2013	GM-PA1-132	18	4121																	x
gimn.	2013	GM-PA1-132	21	4125																	x
gimn.	2013	GM-PA1-132	24	4128																	x
arkusz niepub.		mc1	2	1576							x	x					x				
arkusz niepub.		mc1	3	1761									x	x							
arkusz niepub.		mc1	7	1688	x	x															
arkusz niepub.		mc1	8	1752							x										
arkusz niepub.		mc1	10	1760								x									
arkusz niepub.		mc1	11	1764									x	x							
arkusz niepub.		mc1	12	1692	x	x															
arkusz niepub.		mc1	16	1753							x	x									
arkusz niepub.		mc1	17	1583									x	x		x		x			
arkusz niepub.		mc1	20	1565									x	x			x		x		x
arkusz niepub.		mc1	21	1754							x	x									
arkusz niepub.		mc1	22	1566	x	x											x		x		x
arkusz niepub.		mc1	29	1508							x	x				x		x			

Egzamin	Rok egz.	Kod arkusza	Nr. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część matematyczno-przyrodnicza											matem.	przyroda	matem.	przyroda	matem.	przyroda
					1	2	3	4	5	6	7	8	9	10	11	1	1	1	1	1	1
arkusz niepub.		mc1	30	1503	x	x										x		x			
arkusz niepub.		mc1	31	1586									x	x		x					
arkusz niepub.		mc1	34	1596	x	x										x		x			
arkusz niepub.		mc1	35	1756							x	x									
arkusz niepub.		mc1	36	1514							x	x									
arkusz niepub.		mc1	42	1770									x	x							
arkusz niepub.		mc2	1	1704		x	x														
arkusz niepub.		mc2	2	1726				x	x												
arkusz niepub.		mc2	3	1581										x	x		x		x		x
arkusz niepub.		mc2	4	1740						x	x										
arkusz niepub.		mc2	8	1582						x	x						x		x		
arkusz niepub.		mc2	13	1729				x	x												
arkusz niepub.		mc2	14	1579										x	x		x		x		
arkusz niepub.		mc2	16	1708		x	x														
arkusz niepub.		mc2	17	1567				x	x								x		x		
arkusz niepub.		mc2	18	1568						x	x						x		x		
arkusz niepub.		mc2	19	1592										x	x	x		x			
arkusz niepub.		mc2	24	1569										x	x		x		x		
arkusz niepub.		mc2	25	1580						x	x						x		x		
arkusz niepub.		mc2	26	1502		x	x										x		x		x
arkusz niepub.		mc2	28	1734				x	x												
arkusz niepub.		mc2	29	1587				x	x								x		x		
arkusz niepub.		mc2	30	1588										x	x	x		x			

Egzamin	Rok egz.	Kod arkusza	Nr. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część matematyczno-przyrodnicza											matem.	przyroda	matem.	przyroda	matem.	przyroda
					1	2	3	4	5	6	7	8	9	10	11	1	1	1	1	1	1
arkusz niepub.		mc2	34	1712		x	x														
arkusz niepub.		mc2	36	1715		x	x														
arkusz niepub.		mc2	37	1744						x	x										
arkusz niepub.		mc2	38	1736				x	x												
arkusz niepub.		mc2	39	1779										x	x						
arkusz niepub.		mc2	40	1501		x	x									x		x		x	
arkusz niepub.		mc2	41	1595						x	x					x		x		x	
arkusz niepub.		mc2	87	1718			x														
arkusz niepub.		mc2	97	1707		x															
arkusz niepub.		mcs	2	1572	x				x	x					x		x		x		
arkusz niepub.		mcs	3	1574			x	x				x	x				x		x		x
arkusz niepub.		mcs	4	1690	x				x	x					x						
arkusz niepub.		mcs	5	1691	x				x	x					x						
arkusz niepub.		mcs	9	1575			x	x				x	x				x		x		
arkusz niepub.		mcs	11	1564			x	x				x	x				x		x		x
arkusz niepub.		mcs	12	1570			x	x				x	x				x		x		x
arkusz niepub.		mcs	13	1571	x				x	x					x		x		x		
arkusz niepub.		mcs	18	1589			x	x				x	x			x		x			
arkusz niepub.		mcs	21	1590											x	x		x			
arkusz niepub.		mcs	23	1509											x	x		x		x	
arkusz niepub.		mcs	24	1573			x	x				x	x				x		x		
arkusz niepub.		mcs	25	1594	x				x	x					x	x		x			
arkusz niepub.		mcs	26	1724			x	x				x	x								

Egzamin	Rok egz.	Kod arkusza	Nr. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty																
					2011											2012		2013		2014	
					część matematyczno-przyrodnicza											matem.	przyroda	matem.	przyroda	matem.	przyroda
					1	2	3	4	5	6	7	8	9	10	11	1	1	1	1	1	1
arkusz niepub.		mcs	27	1697	x				x	x					x						

3.4. Matura z języka angielskiego

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty											
					2014											
					1	2	3	4	5	6	7	8	9	10	11	12
gimnazjalny	2012	GJA-A1-20122	4	4193	x	x	x	x	x	x	x	x	x	x	x	x
gimnazjalny	2012	GJA-A1-20122	5	4197	x	x	x	x	x	x	x	x	x	x	x	x
gimnazjalny	2013	GJA-A1-20132	3	4224	x	x	x	x	x	x	x	x	x	x	x	x
matura	2010	MJA-P1A1P-102	2	6123	x			x	x				x			
matura	2010	MJA-P1A1P-102	06_1	6147	x			x	x				x			
matura	2010	MJA-P1A1P-102	06_2	6148	x			x	x				x			
matura	2010	MJA-P1A1P-102	06_3	6149	x			x	x				x			
matura	2010	MJA-P1A1P-102	06_4	6150	x			x	x				x			
matura	2010	MJA-P1A1P-102	06_5	6151	x			x	x				x			
matura	2010	MJA-P1A1P-102	06_6	6152	x			x	x				x			
matura	2010	MJA-P1A1P-102	7	6153	x			x	x				x			
matura	2011	MJA-P1A1P-112	03_1	6175	x	x				x				x		
matura	2011	MJA-P1A1P-112	03_2	6176	x	x				x				x		
matura	2011	MJA-P1A1P-112	03_3	6177	x	x				x				x		

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty											
					2014											
					1	2	3	4	5	6	7	8	9	10	11	12
matura	2011	MJA-P1A1P-112	03_4	6178	x	x				x				x		
matura	2011	MJA-P1A1P-112	03_5	6179	x	x				x				x		
matura	2011	MJA-P1A1P-112	06_1	6193	x	x				x				x		
matura	2011	MJA-P1A1P-112	06_2	6194	x	x				x				x		
matura	2011	MJA-P1A1P-112	06_3	6195	x	x				x				x		
matura	2011	MJA-P1A1P-112	06_4	6196	x	x				x				x		
matura	2011	MJA-P1A1P-112	06_5	6197	x	x				x				x		
matura	2011	MJA-P1A1P-112	06_6	6198	x	x				x				x		
matura	2011	MJA-P1A1P-112	06_7	6199	x	x				x				x		
matura	2011	MJA-P1A1P-112	8	6205	x	x				x				x		
matura	2012	MJA-P1A1P-122	03_1	6222		x	x				x				x	
matura	2012	MJA-P1A1P-122	03_2	6223		x	x				x				x	
matura	2012	MJA-P1A1P-122	03_3	6224		x	x				x				x	
matura	2012	MJA-P1A1P-122	03_4	6225		x	x				x				x	
matura	2012	MJA-P1A1P-122	03_5	6226		x	x				x				x	
matura	2012	MJA-P1A1P-122	06_1	6240		x	x				x				x	
matura	2012	MJA-P1A1P-122	06_2	6241		x	x				x				x	
matura	2012	MJA-P1A1P-122	06_3	6242		x	x				x				x	
matura	2012	MJA-P1A1P-122	06_4	6243		x	x				x				x	
matura	2012	MJA-P1A1P-122	06_5	6244		x	x				x				x	
matura	2012	MJA-P1A1P-122	06_6	6245		x	x				x				x	
matura	2012	MJA-P1A1P-122	06_7	6246		x	x				x				x	
matura	2012	MJA-P1A1P-122	7	6247		x	x				x				x	
matura	2013	MJA-P1A1P-132	2	6264			x	x				x				x
matura	2013	MJA-P1A1P-132	06_1	6287			x	x				x				x

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty											
					2014											
					1	2	3	4	5	6	7	8	9	10	11	12
matura	2013	MJA-P1A1P-132	06_2	6288			x	x				x				x
matura	2013	MJA-P1A1P-132	06_3	6289			x	x				x				x
matura	2013	MJA-P1A1P-132	06_4	6290			x	x				x				x
matura	2013	MJA-P1A1P-132	06_5	6291			x	x				x				x
matura	2013	MJA-P1A1P-132	06_6	6292			x	x				x				x
matura	2013	MJA-P1A1P-132	06_7	6293			x	x				x				x
matura	2013	MJA-P1A1P-132	7	6294			x	x				x				x
matura poprawkowa	2010	MJA-P1A1P-104	03_1	6904					x							x
matura poprawkowa	2010	MJA-P1A1P-104	03_2	6905					x							x
matura poprawkowa	2010	MJA-P1A1P-104	03_3	6906					x							x
matura poprawkowa	2010	MJA-P1A1P-104	03_4	6907					x							x
matura poprawkowa	2010	MJA-P1A1P-104	03_5	6908					x							x
matura poprawkowa	2010	MJA-P1A1P-104	4	6909					x							x
matura poprawkowa	2010	MJA-P1A1P-104	8	6934					x							x
matura poprawkowa	2011	MJA-P1A1P-114	1	6941						x			x			
matura poprawkowa	2011	MJA-P1A1P-114	4	6956						x			x			
matura poprawkowa	2011	MJA-P1A1P-114	7	6976						x			x			
matura poprawkowa	2012	MJA-P1A1P-124	2	6993							x			x		
matura poprawkowa	2012	MJA-P1A1P-124	4	7003							x			x		
matura poprawkowa	2012	MJA-P1A1P-124	7	7023							x			x		
matura poprawkowa	2013	MJA-P1A1P-134	2	7211								x			x	
matura poprawkowa	2013	MJA-P1A1P-134	4	7221								x			x	
matura poprawkowa	2013	MJA-P1A1P-134	8	7246								x			x	

3.5. Matura z języka matematyki

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty											
					2013							2014				
					1	2	3	4	5	6	7	1	2	3	4	
gimnazjalny	2011	GM-A1-112	5	999	x				x		x					
gimnazjalny	2011	GM-A1-112	6	1000	x				x		x					
gimnazjalny	2011	GM-A1-112	24	1018	x				x		x					
gimnazjalny	2011	GM-A1-112	25	1019							x					
gimnazjalny	2012	GM-MA1-122	12	1052								x	x	x	x	
gimnazjalny	2012	GM-MA1-122	22	1062								x	x	x	x	
gimnazjalny	2012	GM-MA1-122	23	1063								x	x	x	x	
gimnazjalny	2013	GM-MA1-132	5	4133								x	x	x	x	
gimnazjalny	2013	GM-MA1-132	7	4135								x	x	x	x	
gimnazjalny	2013	GM-MA1-132	15	4143								x	x	x	x	
gimnazjalny	2013	GM-MA1-132	18	4146								x	x	x	x	
gimnazjalny	2013	GM-MA1-132	20	4148								x	x	x	x	
gimnazjalny	2013	GM-MA1-132	22	4150								x	x	x	x	
matura	2010	MMA-P1A1P-102	2	1828		x										
matura	2010	MMA-P1A1P-102	3	1829		x										
matura	2010	MMA-P1A1P-102	5	1831	x	x				x						
matura	2010	MMA-P1A1P-102	7	1833	x	x				x						
matura	2010	MMA-P1A1P-102	9	1835	x	x				x						
matura	2010	MMA-P1A1P-102	10	1836		x										
matura	2010	MMA-P1A1P-102	11	1837	x	x				x						
matura	2010	MMA-P1A1P-102	13	1839		x										
matura	2010	MMA-P1A1P-102	14	1840						x						
matura	2010	MMA-P1A1P-102	17	1843	x	x				x						
matura	2010	MMA-P1A1P-102	18	1844	x	x				x						
matura	2010	MMA-P1A1P-102	19	1845						x						

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty											
					2013							2014				
					1	2	3	4	5	6	7	1	2	3	4	
matura	2011	MMA-P1A1P-112	33	1938			x	x								
matura	2012	MMA-P1A1P-122	1	1984			x									
matura	2012	MMA-P1A1P-122	2	1985						x		x			x	
matura	2012	MMA-P1A1P-122	6	1989			x		x	x						
matura	2012	MMA-P1A1P-122	7	1990								x			x	
matura	2012	MMA-P1A1P-122	8	1991			x		x	x						
matura	2012	MMA-P1A1P-122	9	1992			x		x	x						
matura	2012	MMA-P1A1P-122	10	1993			x									
matura	2012	MMA-P1A1P-122	11	1994					x			x			x	
matura	2012	MMA-P1A1P-122	14	1997								x			x	
matura	2012	MMA-P1A1P-122	15	1998			x		x	x						
matura	2012	MMA-P1A1P-122	16	1999						x						
matura	2012	MMA-P1A1P-122	18	2001			x		x	x		x			x	
matura	2012	MMA-P1A1P-122	22	2005						x						
matura	2012	MMA-P1A1P-122	23	2006								x			x	
matura	2012	MMA-P1A1P-122	24	2007			x		x	x						
matura	2012	MMA-P1A1P-122	26	2009								x			x	
matura	2012	MMA-P1A1P-122	27	2010					x	x						
matura	2012	MMA-P1A1P-122	29	2012			x		x	x		x			x	
matura	2012	MMA-P1A1P-122	30	2013			x		x							
matura	2012	MMA-P1A1P-122	31	2014								x			x	
matura	2012	MMA-P1A1P-122	32	2015			x			x						
matura	2012	MMA-P1A1P-122	34	2017					x			x			x	
matura	2013	MMA-P1A1P-132	1	4022									x	x	x	
matura	2013	MMA-P1A1P-132	6	4027									x	x	x	
matura	2013	MMA-P1A1P-132	7	4028									x	x	x	
matura	2013	MMA-P1A1P-132	8	4029									x	x	x	

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty											
					2013							2014				
					1	2	3	4	5	6	7	1	2	3	4	
matura	2013	MMA-P1A1P-132	18	4039									x	x	x	
matura	2013	MMA-P1A1P-132	20	4041									x	x	x	
matura	2013	MMA-P1A1P-132	22	4043									x	x	x	
matura	2013	MMA-P1A1P-132	27	4048									x	x	x	
matura	2013	MMA-P1A1P-132	29	4050									x	x	x	
matura	2013	MMA-P1A1P-132	31	4052									x	x	x	
matura	2013	MMA-P1A1P-132	33	4054									x	x	x	
matura poprawkowa	2010	MMA-P1A1P-104	3	1874		x	x				x					
matura poprawkowa	2010	MMA-P1A1P-104	4	1875		x	x				x					
matura poprawkowa	2010	MMA-P1A1P-104	7	1878		x	x				x					
matura poprawkowa	2010	MMA-P1A1P-104	9	1880		x	x				x					
matura poprawkowa	2010	MMA-P1A1P-104	13	1884		x	x				x					
matura poprawkowa	2010	MMA-P1A1P-104	15	1886		x	x				x					
matura poprawkowa	2010	MMA-P1A1P-104	23	1894		x	x				x					
matura poprawkowa	2010	MMA-P1A1P-104	27	1898		x	x				x					
matura poprawkowa	2010	MMA-P1A1P-104	30	1901		x					x					
matura poprawkowa	2010	MMA-P1A1P-104	31	1902		x	x				x					
matura poprawkowa	2010	MMA-P1A1P-104	32	1903							x					
matura poprawkowa	2010	MMA-P1A1P-104	34	1905		x	x									
matura poprawkowa	2011	MMA-P1A1P-114	6	1956		x		x	x							
matura poprawkowa	2011	MMA-P1A1P-114	8	1958		x		x	x							
matura poprawkowa	2011	MMA-P1A1P-114	12	1962		x		x	x							
matura poprawkowa	2011	MMA-P1A1P-114	21	1971		x		x	x							
matura poprawkowa	2011	MMA-P1A1P-114	22	1972		x		x	x							
matura poprawkowa	2011	MMA-P1A1P-114	24	1974					x							
matura poprawkowa	2011	MMA-P1A1P-114	25	1975		x		x	x							
matura poprawkowa	2011	MMA-P1A1P-114	28	1978		x		x								

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty											
					2013							2014				
					1	2	3	4	5	6	7	1	2	3	4	
matura poprawkowa	2011	MMA-P1A1P-114	29	1979		x		x	x							
matura poprawkowa	2011	MMA-P1A1P-114	32	1982		x		x	x							
matura poprawkowa	2012	MMA-P1A1P-124	2	2030				x		x	x					
matura poprawkowa	2012	MMA-P1A1P-124	4	2032								x	x			
matura poprawkowa	2012	MMA-P1A1P-124	5	2033				x		x	x	x	x			
matura poprawkowa	2012	MMA-P1A1P-124	6	2034				x		x	x					
matura poprawkowa	2012	MMA-P1A1P-124	10	2038								x	x			
matura poprawkowa	2012	MMA-P1A1P-124	13	2041								x	x			
matura poprawkowa	2012	MMA-P1A1P-124	15	2043				x		x	x					
matura poprawkowa	2012	MMA-P1A1P-124	17	2045								x	x			
matura poprawkowa	2012	MMA-P1A1P-124	20	2048				x		x	x					
matura poprawkowa	2012	MMA-P1A1P-124	21	2049				x		x	x					
matura poprawkowa	2012	MMA-P1A1P-124	22	2050				x		x	x					
matura poprawkowa	2012	MMA-P1A1P-124	23	2051								x	x			
matura poprawkowa	2012	MMA-P1A1P-124	24	2052							x					
matura poprawkowa	2012	MMA-P1A1P-124	25	2053								x	x			
matura poprawkowa	2012	MMA-P1A1P-124	26	2054				x		x	x					
matura poprawkowa	2012	MMA-P1A1P-124	27	2055								x	x			
matura poprawkowa	2012	MMA-P1A1P-124	28	2056				x		x	x					
matura poprawkowa	2012	MMA-P1A1P-124	29	2057								x	x			
matura poprawkowa	2012	MMA-P1A1P-124	31	2059								x	x			
matura poprawkowa	2012	MMA-P1A1P-124	33	2061								x	x			
matura poprawkowa	2012	MMA-P1A1P-124	34	2062				x		x	x					
matura poprawkowa	2013	MMA-P1A1P-134	1	7140										x		
matura poprawkowa	2013	MMA-P1A1P-134	6	7145										x		
matura poprawkowa	2013	MMA-P1A1P-134	10	7149										x		
matura poprawkowa	2013	MMA-P1A1P-134	12	7151										x		

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty									
					2014									
					1	2	3	4	5	6	7	8	9	10
matura	2010	MPO-P1A1P-102	1	6512	x					x				
matura	2010	MPO-P1A1P-102	2	6513	x					x				
matura	2010	MPO-P1A1P-102	3	6514	x					x				
matura	2010	MPO-P1A1P-102	4	6515	x					x				
matura	2010	MPO-P1A1P-102	5	6516	x					x				
matura	2010	MPO-P1A1P-102	6	6517	x					x				
matura	2010	MPO-P1A1P-102	7	6518	x					x				
matura	2010	MPO-P1A1P-102	12	6523	x					x				
matura	2010	MPO-P1A1P-102	13	6524	x					x				
matura	2010	MPO-P1A1P-102	14	6525	x					x				
matura	2010	MPO-P1A1P-102	I	7255	x									x
matura	2010	MPO-P1A1P-102	II	7261					x					x
matura	2011	MPO-P1A1P-112	1	6532		x					x			
matura	2011	MPO-P1A1P-112	2	6533		x					x			
matura	2011	MPO-P1A1P-112	4	6535		x					x			
matura	2011	MPO-P1A1P-112	5	6536		x					x			
matura	2011	MPO-P1A1P-112	6	6537		x					x			
matura	2011	MPO-P1A1P-112	7	6538		x					x			
matura	2011	MPO-P1A1P-112	10	6541		x					x			
matura	2011	MPO-P1A1P-112	11	6542		x					x			
matura	2011	MPO-P1A1P-112	12	6543		x					x			
matura	2011	MPO-P1A1P-112	13	6544		x					x			
matura	2011	MPO-P1A1P-112	I	7267		x				x				
matura	2011	MPO-P1A1P-112	II	7273				x		x				
matura	2012	MPO-P1A1P-122	1	6551								x		
matura	2012	MPO-P1A1P-122	2	6552			x							
matura	2012	MPO-P1A1P-122	3	6553			x					x		

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty									
					2014									
					1	2	3	4	5	6	7	8	9	10
matura	2012	MPO-P1A1P-122	4	6554			x					x		
matura	2012	MPO-P1A1P-122	5	6555			x					x		
matura	2012	MPO-P1A1P-122	6	6556			x					x		
matura	2012	MPO-P1A1P-122	7	6557			x					x		
matura	2012	MPO-P1A1P-122	10	6560			x					x		
matura	2012	MPO-P1A1P-122	11	6561			x					x		
matura	2012	MPO-P1A1P-122	12	6562			x					x		
matura	2012	MPO-P1A1P-122	13	6563			x					x		
matura	2012	MPO-P1A1P-122	I	6565			x				x			
matura	2012	MPO-P1A1P-122	II	6571	x						x			
matura	2013	MPO-P1A1P-132	1	6583				x					x	
matura	2013	MPO-P1A1P-132	4	6586				x					x	
matura	2013	MPO-P1A1P-132	5	6587				x					x	
matura	2013	MPO-P1A1P-132	6	6588				x					x	
matura	2013	MPO-P1A1P-132	7	6589				x					x	
matura	2013	MPO-P1A1P-132	8	6590				x					x	
matura	2013	MPO-P1A1P-132	10	6592				x					x	
matura	2013	MPO-P1A1P-132	12	6594				x					x	
matura	2013	MPO-P1A1P-132	14	6596				x					x	
matura	2013	MPO-P1A1P-132	I	6597				x				x		
matura	2013	MPO-P1A1P-132	II	6603		x						x		
matura poprawkowa	2013	MPO-P1A1P-134	1	7174					x					x
matura poprawkowa	2013	MPO-P1A1P-134	4	7177					x					x
matura poprawkowa	2013	MPO-P1A1P-134	5	7178					x					x
matura poprawkowa	2013	MPO-P1A1P-134	6	7179					x					x
matura poprawkowa	2013	MPO-P1A1P-134	7	7180					x					x
matura poprawkowa	2013	MPO-P1A1P-134	8	7181					x					x

Egzamin	Rok egz.	Kod arkusza	Nr zad. w arkuszu	Nr zad. w bazie	Edycje zrównania i zeszyty									
					2014									
					1	2	3	4	5	6	7	8	9	10
matura poprawkowa	2013	MPO-P1A1P-134	10	7183					x					x
matura poprawkowa	2013	MPO-P1A1P-134	12	7185					x					x
matura poprawkowa	2013	MPO-P1A1P-134	13	7186					x					x
matura poprawkowa	2013	MPO-P1A1P-134	14	7187					x					x
matura poprawkowa	2013	MPO-P1A1P-134	I	7188					x				x	
matura poprawkowa	2013	MPO-P1A1P-134	II	7194			x						x	

Aneks 4. Kwestionariusze ankiet

4.1. Skala Motywacji Testowej⁵⁷ – ankiet dla ucznia



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



entuzjaści
edukacji

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Kwestionariusz ankiety ucznia

Kod szkoły (wypełnia ankieter)	
Kod ucznia (wypełnia ankieter)	

Chcielibyśmy zadać Ci jeszcze kilka pytań dotyczących testu, który właśnie wypełniłeś(-aś).

Przeczytaj uważnie poniższe pytania i udziel na nie odpowiedzi.

1. W jakim stopniu zgadzasz się lub nie zgadzasz z następującymi stwierdzeniami?

Przy każdym stwierdzeniu zaznacz jedną odpowiedź na podanej skali od 1 do 5.

Stwierdzenia:	Zdecydowanie się nie zgadzam	Raczej się nie zgadzam	Ani się zgadzam, ani nie zgadzam	Raczej się zgadzam	Zdecydowanie się zgadzam
1. Byłem(-am) dobrze przygotowany(-a) do tego testu.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
2. Starałbym(-łabym) się bardziej, gdyby ten test był właściwym sprawdzianem szóstoklasisty.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
3. Uczyłem(-am) się kilka dni specjalnie do rozwiązywanego dzisiaj testu.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
4. Część zadań była nudna, nie rozwiązywałem(-am) ich wcale.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
5. Jestem dobrze przygotowany(-a) do sprawdzianu.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
6. Spotkałem(-am) się z zadaniami z tego testu lub z zadaniami bardzo podobnymi.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
7. Przygotowując się do sprawdzianu rozwiązywałem(-am) arkusze testowe z poprzednich lat.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
8. Dzisiaj nie czułem(-am) się na siłach, aby dobrze rozwiązać ten test.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
9. Gdyby ten test był prawdziwym sprawdzianem szóstoklasisty, byłbym(-abym) lepiej przygotowany(-a).	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅

⁵⁷ Przedstawiamy w aneksie dwie wersje kwestionariusza ankiety ucznia, gdyż w zależności od etapu edukacyjnego ankiety nieznacznie się od siebie różniły. Aneks zawiera przykładową ankietę dla ucznia piszącego sprawdzian i ucznia piszącego egzamin maturalny.

Stwierdzenia:	Zdecydowa nie się zgadzam	Raczej się nie zgadzam	Ani się zgadzam, ani nie zgadzam	Raczej się zgadzam	Zdecydowa nie się zgadzam
10. Rozwiązywałem(-am) przede wszystkim pytania zamknięte, na resztę nie miałem(-am) ochoty.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
11. Do sprawdzianu zostało jeszcze tyle czasu, że można się dużo nauczyć.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
12. Do tej pory nie uczyłem(-am) się do sprawdzianu, dopiero zamierzam to zrobić.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
13. Zadania z tego testu rozwiązałem(-am) najlepiej jak potrafiłem(-am).	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
14. Dzięki temu testowi sprawdzę swoje przygotowanie do sprawdzianu.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅

2. Jak trudny był, według Ciebie, dzisiejszy test?

Zaznacz jedną odpowiedź.

1. Bardzo trudny.
2. Raczej trudny.
3. Ani trudny, ani łatwy.
4. Raczej łatwy.
5. Bardzo łatwy.

3. Czy, według Ciebie, czas przeznaczony na rozwiązanie wszystkich zadań w dzisiejszym teście był wystarczający?

Zaznacz jedną odpowiedź.

1. Zdecydowanie tak.
2. Raczej tak.
3. Raczej nie.
4. Zdecydowanie nie.

4. Czy w ostatnich 3 miesiącach rozwiązywałeś(-aś) na komputerze (w Internecie lub za pomocą programu komputerowego) testy sprawdzające wiedzę szkolną z zakresu obejmującego rozwiązywany dzisiaj test?

Zaznacz jedną odpowiedź.

TAK NIE

Bardzo dziękujemy za udzielenie odpowiedzi!



Kwestionariusz ankiety ucznia

Kod szkoły (wypełnia ankieter)	
Kod ucznia (wypełnia ankieter)	

Chcielibyśmy zadać Ci jeszcze kilka pytań dotyczących testu, który właśnie wypełniłeś(-aś).

Przeczytaj uważnie poniższe pytania i udziel na nie odpowiedzi.

1. W jakim stopniu zgadzasz się lub nie zgadzasz z następującymi stwierdzeniami?

Przy każdym stwierdzeniu zaznacz jedną odpowiedź na podanej skali od 1 do 5.

Stwierdzenia:	Zdecydowa nie się nie zgadzam	Raczej się nie zgadzam	Ani się zgadzam, ani nie zgadzam	Raczej się zgadzam	Zdecydowa nie się zgadzam
1. Byłem(-am) dobrze przygotowany(-a) do tego testu.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
2. Starałbym(-łabym) się bardziej, gdyby ten test był właściwym egzaminem maturalnym.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
3. Uczyłem(-am) się kilka dni specjalnie do rozwiązywanego dzisiaj testu.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
4. Jestem dobrze przygotowany(-a) do egzaminu maturalnego z matematyki.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
5. Część zadań była nudna, nie rozwiązywałem(-am) ich wcale.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
6. Spotkałem(-am) się z zadaniami z tego testu lub z zadaniami bardzo podobnymi.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
7. Przygotowując się do egzaminu rozwiązywałem(-am) arkusze egzaminacyjne z poprzednich lat.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
8. Dzisiaj nie czułem(-am) się na siłach, aby dobrze rozwiązać ten test.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
9. Gdyby ten test był prawdziwym egzaminem maturalnym, byłbym(-abym) lepiej przygotowany(-a).	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅

Stwierdzenia:	Zdecydowa nie się nie zgadzam	Raczej się nie zgadzam	Ani się zgadzam, ani nie zgadzam	Raczej się zgadzam	Zdecydowa nie się zgadzam
10. Rozwiązywałem(-am) przede wszystkim pytania zamknięte, na resztę nie miałem(-am) ochoty.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
11. Do matury zostało jeszcze tyle czasu, że można się dużo nauczyć.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
12. Do tej pory nie uczyłem(-am) się do egzaminu z matematyki, dopiero zamierzam to zrobić.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
13. Zadania z tego testu rozwiązałem(-am) najlepiej jak potrafiłem(-am).	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
14. Dzięki temu testowi sprawdzę swoje przygotowanie do egzaminu z matematyki.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
15. Zależy mi, aby uzyskać jak najlepszy wynik z egzaminu maturalnego z matematyki.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅

2. Jak trudny był, według Ciebie, dzisiejszy test?

Zaznacz jedną odpowiedź.

- 6. Bardzo trudny.
- 7. Raczej trudny.
- 8. Ani trudny, ani łatwy.
- 9. Raczej łatwy.
- 10. Bardzo łatwy.

3. Czy, według Ciebie, czas przeznaczony na rozwiązanie wszystkich zadań w dzisiejszym teście był wystarczający?

Zaznacz jedną odpowiedź.

- 5. Zdecydowanie tak.
- 6. Raczej tak.
- 7. Raczej nie.
- 8. Zdecydowanie nie.

4. Czy w ostatnich 3 miesiącach rozwiązywałeś(-aś) na komputerze (w Internecie lub za pomocą programu komputerowego) testy sprawdzające wiedzę szkolną z zakresu obejmującego rozwiązywany dzisiaj test?

Zaznacz jedną odpowiedź.

TAK NIE

5. Czy po ukończeniu szkoły maturalnej zamierzasz kontynuować naukę na studiach wyższych?

Zaznacz jedną odpowiedź.

TAK

NIE → *przejdź do pytania 7*

6. Do której z poniższych grup zaliczysz kierunek studiów, który najbardziej chcesz studiować?

Zaznacz jedną odpowiedź.

☐₁ Kierunki ekonomiczne

☐₅ Prawo i administracja

☐₂ Kierunki medyczne

☐₆ Kierunki biologiczne i przyrodnicze

☐₃ Kierunki humanistyczne

☐₇ Kierunki ścisłe

☐₄ Kierunki społeczne

☐₈ Kierunki wychowania fizycznego

7. Czy w Twojej szkole organizowane były dodatkowe kursy przygotowujące do egzaminu maturalnego z matematyki?

Zaznacz jedną odpowiedź.

TAK

NIE

8. Czy poza szkołą brałeś(-aś) udział w innych kursach, prywatnych lekcjach, korepetycjach przygotowujących do egzaminu maturalnego z matematyki?

Zaznacz jedną odpowiedź.

TAK

NIE

Bardzo dziękujemy za udzielenie odpowiedzi!

4.2. Skala Opinii Uczniów (polskie tłumaczenie Student Opinion Scale, SOS; Sundre, 2007) – ankieta dla ucznia

Chcielibyśmy zadać Ci jeszcze kilka pytań dotyczących testu, który właśnie wypełniłeś(-aś).

Przeczytaj uważnie poniższe pytania i udziel na nie odpowiedzi.

W jakim stopniu zgadzasz się lub nie zgadzasz z następującymi stwierdzeniami?

Przy każdym stwierdzeniu zaznacz jedną odpowiedź na podanej skali od 1 do 5.

Stwierdzenia:	Zdecydowa nie się nie zgadzam	Raczej się nie zgadzam	Ani się zgadzam, ani nie zgadzam	Raczej się zgadzam	Zdecydo wanie się zgadzam
1. Ważne było dla mnie uzyskanie dobrego wyniku w teście.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
2. Silnie zaangażowałem(-am) się w rozwiązywanie testu.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
3. Nie ciekawi mnie, jak mi poszło w teście w porównaniu do innych osób.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
4. Nie przejmuję się wynikiem, jaki uzyskam w teście.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
5. Ten test był dla mnie ważny.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
6. Dałem(-am) z siebie wszystko rozwiązując test.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
7. Mogłem(-am) się bardziej przyłożyć do rozwiązywania testu.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
8. Chciałbym(-ałabym) wiedzieć, jak mi poszło w teście.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
9. Nie poświęciłem(-am) całej swojej uwagi testowi.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅
10. Pracowałem(-am) nad rozwiązaniem testu aż do końca.	☐ ₁	☐ ₂	☐ ₃	☐ ₄	☐ ₅

Bardzo dziękujemy za udzielenie odpowiedzi!

4.3. Ankiety dla nauczyciela (szkoła podstawowa)



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



entuzjaści
edukacji

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Kwestionariusz ankiety nauczyciela: SZKOŁA PODSTAWOWA SN

WYPEŁNIA ANKIETER KOD NAUCZYCIELA			ANKIETER: NAKLEJ KOD SZKOŁY
Przedmioty nauczyciela <i>Proszę zaznaczyć WSZYSTKIE przedmioty, których naucza Pan/i w BADANEJ KLASIE. Możliwe wiele przedmiotów.</i>			Przedmiot ankiety <i>Proszę zaznaczyć JEDEN przedmiot, którego dotyczy ta ankieta. Ankieter: W przypadku, gdy jeden nauczyciel naucza wielu przedmiotów w BADANEJ KLASIE należy przeprowadzić oddzielną ankietę za każdy przedmiot.</i>
☐ 01 J. polski ☐ 09 Historia i społeczeństwo ☐ 02 Matematyka ☐ 10 Przyroda			☐ 01 J. polski ☐ 09 Historia i społeczeństwo ☐ 02 Matematyka ☐ 10 Przyroda

Chcielibyśmy zadać Panu(-i) kilka pytań dotyczących sprawdzianu i przygotowania do niego uczniów. Proszę przeczytać uważnie wszystkie pytania i udzielić na nie odpowiedzi.

1. Z jakich testów komercyjnych korzystał(-a) Pan(-i) podczas zajęć z uczniami?

Proszę przy każdym teście zaznaczyć wszystkie właściwe odpowiedzi.

	W poprzednich latach szkolnych	W tym roku szkolnym	W badanej klasie	W ogóle nie korzystałem(-am)
1. Operon	☐ ₁	☐ ₁	☐ ₁	☐ ₁
2. Diagnozy WSiP	☐ ₁	☐ ₁	☐ ₁	☐ ₁
3. IBK (Wałbrzych)	☐ ₁	☐ ₁	☐ ₁	☐ ₁
4. Nowa Era	☐ ₁	☐ ₁	☐ ₁	☐ ₁
5. Gdańskie Wydawnictwo Oświatowe	☐ ₁	☐ ₁	☐ ₁	☐ ₁
6. Inne, jakie?	☐ ₁	☐ ₁	☐ ₁	☐ ₁

2. Proszę ocenić, jak dobrze, Pana(-i) zdaniem, badani uczniowie przygotowani są do sprawdzianu z przedmiotu, którego Pan(-i) uczy.

Proszę zaznaczyć jedną odpowiedź.

1. Bardzo dobrze.
2. Raczej dobrze.
3. Raczej źle.
4. Bardzo źle.

3. **Proszę ocenić, jak bardzo, Pana(-i) zdaniem, badanym uczniom zależy na jak najlepszym rozwiązaniu testu w prowadzonym przez IBE badaniu.**

Proszę zaznaczyć jedną odpowiedź.

1. Bardzo zależy.
2. Trochę zależy.
3. Niewiele zależy.
4. Wcale nie zależy.

4. **Czy zamierza Pan(-i) użyć wyników testu z tego badania w ocenianiu uczniów?**

Proszę zaznaczyć jedną odpowiedź.

1. Tak 2. Nie → proszę przejść do pytania 6

5. **Czy poinformował(-a) Pan(-i) o tym uczniów przed badaniem?**

Proszę zaznaczyć jedną odpowiedź.

1. Tak 2. Nie

6. **Czy, Pana(-i) zdaniem, udział w prowadzonym przez IBE badaniu spowoduje, że uczniowie biorący w nim udział uzyskają wyższy wynik ze sprawdzianu niż pozostali uczniowie?**

Proszę zaznaczyć jedną odpowiedź.

1. Zdecydowanie tak.
2. Raczej tak.
3. Ani tak, ani nie.
4. Raczej nie.
5. Zdecydowanie nie.

7. **Czy używał(-a) Pan(-i) arkuszy z poprzednich edycji sprawdzianu podczas zajęć z uczniami badanej klasy?**

Proszę zaznaczyć jedną odpowiedź.

1. Tak 2. Nie 3. Nie pamiętam

8. **Kiedy zaczął(-ęła) Pan(-i) uczyć badaną klasę?**

Proszę zaznaczyć jedną odpowiedź.

1. Od początku klasy czwartej.
2. W trakcie jej nauki w klasie czwartej.
3. W trakcie jej nauki w klasie piątej.
4. Dopiero w tym roku.

9. **Czy uczestniczył(-a) Pan(-i) kiedykolwiek w ocenianiu sprawdzianu?**

Proszę zaznaczyć jedną odpowiedź.

1. Tak 2. Nie → proszę przejść do pytania 11

10. **Czy udział w sesjach egzaminacyjnych w roli egzaminatora pomaga Panu(-i) lepiej przygotować uczniów do sprawdzianu?**

Proszę zaznaczyć jedną odpowiedź.

1. Zdecydowanie tak.
2. Raczej tak.
3. Ani tak, ani nie.
4. Raczej nie.
5. Zdecydowanie nie.

11. Czy zmiana formuły sprawdzianu od 2015 roku będzie, Pana(-i) zdaniem, miała istotny wpływ na wyniki egzaminacyjne?

Proszę zaznaczyć jedną odpowiedź.

1. Zdecydowanie tak.
2. Raczej tak.
3. Ani tak, ani nie.
4. Raczej nie.
5. Zdecydowanie nie.
6. Nie wiem jakie zmiany zostaną wprowadzone do sprawdzianu w 2015 roku.

12. Jaki jest Pana(-i) stopień awansu zawodowego?

Proszę zaznaczyć jedną odpowiedź.

1. Profesor oświaty.
2. Nauczyciel dyplomowany.
3. Nauczyciel mianowany.
4. Nauczyciel kontraktowy.
5. Brak stopnia awansu.

13. Czy uważa Pan(-i), że trudność sprawdzianu jest z roku na rok porównywalna?

Proszę zaznaczyć jedną odpowiedź.

1. Tak 2. Nie 3. Nie wiem

Bardzo dziękujemy za udzielenie odpowiedzi!

4.4. Ankiety dla nauczyciela (gimnazjum)



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



entuzjaści
edukacji

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Kwestionariusz ankiety nauczyciela: GIMNAZJUM GN

WYPEŁNIA ANKIETER		ANKIETER: NAKLEJ KOD SZKOŁY
KOD NAUCZYCIELA:		
Przedmioty nauczyciela <i>Proszę zaznaczyć WSZYSTKIE przedmioty, których uczy Pan/i w BADANEJ KLASIE. Możliwe wiele przedmiotów.</i>		Przedmiot ankiety <i>Proszę zaznaczyć JEDEN przedmiot, którego dotyczy ta ankieta.</i> <i>Ankieter: W przypadku, gdy jeden nauczyciel naucza wielu przedmiotów w BADANEJ KLASIE należy przeprowadzić oddzielną ankietę za każdy przedmiot.</i>
☐ 01 J. polski ☐ 03 Historia ☐ 04 WOS ☐ 02 Matematyka	☐ 05 Biologia ☐ 06 Geografia ☐ 07 Fizyka ☐ 08 Chemia	☐ 01 J. polski ☐ 03 Historia ☐ 04 WOS ☐ 02 Matematyka
		☐ 05 Biologia ☐ 06 Geografia ☐ 07 Fizyka ☐ 08 Chemia

Chcielibyśmy zadać Panu(-i) kilka pytań dotyczących egzaminu gimnazjalnego i przygotowania do niego uczniów. Proszę przeczytać uważnie wszystkie pytania i udzielić na nie odpowiedzi.

1. Z jakich testów komercyjnych korzystał(-a) Pan(-i) podczas zajęć z uczniami?

Proszę przy każdym teście zaznaczyć wszystkie właściwe odpowiedzi.

	W poprzednich latach szkolnych	W tym roku szkolnym	W badanej klasie	W ogóle nie korzystałem(-a m)
1. Operon	☐ ₁	☐ ₁	☐ ₁	☐ ₁
2. Diagnozy WSiP	☐ ₁	☐ ₁	☐ ₁	☐ ₁
3. IBK (Wałbrzych)	☐ ₁	☐ ₁	☐ ₁	☐ ₁
4. Nowa Era	☐ ₁	☐ ₁	☐ ₁	☐ ₁
5. Gdańskie Wydawnictwo Oświatowe	☐ ₁	☐ ₁	☐ ₁	☐ ₁
6. Inne, jakie?	☐ ₁	☐ ₁	☐ ₁	☐ ₁

2. Czy korzystał(-a) Pan(-i) z „Bazy materiałów dydaktycznych” na stronie internetowej IBE?

Proszę zaznaczyć jedną odpowiedź.

1. Tak 2. Nie

3. **Czy, Pana(-i) zdaniem, udział w prowadzonym przez IBE badaniu spowoduje, że uczniowie biorący w nim udział uzyskają wyższy wynik z egzaminu gimnazjalnego niż pozostali uczniowie?**

Proszę zaznaczyć jedną odpowiedź.

1. Zdecydowanie tak.
2. Raczej tak.
3. Ani tak, ani nie.
4. Raczej nie.
5. Zdecydowanie nie.

4. **Czy używał(-a) Pan(-i) arkuszy z poprzednich edycji egzaminu gimnazjalnego podczas zajęć z uczniami badanej klasy?**

Proszę zaznaczyć jedną odpowiedź.

1. Tak 2. Nie 3. Nie pamiętam

5. **Proszę ocenić, jak dobrze, Pana(-i) zdaniem, badani uczniowie przygotowani są do egzaminu gimnazjalnego z przedmiotu, którego Pan(-i) uczy.**

Proszę zaznaczyć jedną odpowiedź.

1. Bardzo dobrze.
2. Raczej dobrze.
3. Raczej źle.
4. Bardzo źle.

6. **Proszę ocenić, jak bardzo, Pana(-i) zdaniem, badanym uczniom zależy na jak najlepszym rozwiązaniu testu w prowadzonym przez IBE badaniu.**

Proszę zaznaczyć jedną odpowiedź.

1. Bardzo zależy.
2. Trochę zależy.
3. Niewiele zależy.
4. Wcale nie zależy.

7. **Czy zamierza Pan(-i) użyć wyników testu z tego badania w ocenianiu uczniów?**

Proszę zaznaczyć jedną odpowiedź.

1. Tak 2. Nie → proszę przejść do pytania 9

8. **Czy poinformował(-a) Pan(-i) o tym uczniów przed badaniem?**

Proszę zaznaczyć jedną odpowiedź.

1. Tak 2. Nie

9. **Kiedy zaczął(-ęła) Pan(-i) uczyć badaną klasę?**

Proszę zaznaczyć jedną odpowiedź.

1. Uczę ją od początku jej istnienia.
2. W trakcie jej pierwszego roku nauki.
3. W trakcie jej drugiego roku nauki.
4. Dopiero w tym roku.

10. **Czy uczestniczył(-a) Pan(-i) kiedykolwiek w ocenianiu egzaminu gimnazjalnego?**

Proszę zaznaczyć jedną odpowiedź.

1. Tak 2. Nie → proszę przejść do pytania 12

11. Czy udział w sesjach egzaminacyjnych w roli egzaminatora pomaga Panu(-i) lepiej przygotować uczniów do egzaminu?

Proszę zaznaczyć jedną odpowiedź.

1. Zdecydowanie tak.
2. Raczej tak.
3. Ani tak, ani nie.
4. Raczej nie.
5. Zdecydowanie nie.

12. Czy Pana(-i) zdaniem wprowadzona w 2012 zmiana formuły egzaminu gimnazjalnego wpłynęła w sposób istotny na wyniki egzaminacyjne uczniów?

Proszę zaznaczyć jedną odpowiedź.

1. Zdecydowanie tak.
2. Raczej tak.
3. Ani tak, ani nie.
4. Raczej nie.
5. Zdecydowanie nie.

13. Jaki jest Pana(-i) stopień awansu zawodowego?

Proszę zaznaczyć jedną odpowiedź.

1. Profesor oświaty.
2. Nauczyciel dyplomowany.
3. Nauczyciel mianowany.
4. Nauczyciel kontraktowy.
5. Brak stopnia awansu.

14. Czy uważa Pan(-i) że trudność egzaminu (z przedmiotu, którego Pan(-i) uczy) jest z roku na rok porównywalna?

Proszę zaznaczyć jedną odpowiedź.

1. Tak 2. Nie 3. Nie wiem

Bardzo dziękujemy za udzielenie odpowiedzi!

4.5. Przykładowa ankieta nauczyciela (szkoła ponadgimnazjalna)



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



entuzjaści
edukacji

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Kwestionariusz ankiety nauczyciela:

SZKOŁA PONADGIMNAZJALNA
MATEMATYKA **PN2**

WYPEŁNIA ANKIETER
KOD NAUCZYCIELA

Chcielibyśmy zadać Panu(-i) kilka pytań dotyczących egzaminu maturalnego z matematyki i przygotowania do niego uczniów. Proszę przeczytać uważnie wszystkie pytania i udzielić na nie odpowiedzi.

1. Z jakich testów komercyjnych korzystał(-a) Pan(-i) podczas zajęć z uczniami?

Proszę przy każdym teście zaznaczyć wszystkie właściwe odpowiedzi.

	W poprzednich latach szkolnych	W tym roku szkolnym	W badanej klasie	W ogóle nie korzystałem(-am)
1. Operon	☐ ₁	☐ ₁	☐ ₁	☐ ₁
2. WSiP	☐ ₁	☐ ₁	☐ ₁	☐ ₁
3. Gdańskie Wydawnictwo Oświatowe	☐ ₁	☐ ₁	☐ ₁	☐ ₁
4. Inne, jakie?	☐ ₁	☐ ₁	☐ ₁	☐ ₁
.....				
.....				

2. Czy, Pana(-i) zdaniem, udział w prowadzonym przez IBE badaniu spowoduje, że uczniowie biorący w nim udział uzyskają wyższy wynik z egzaminu maturalnego z matematyki niż pozostali uczniowie?

Proszę zaznaczyć jedną odpowiedź.

1. Zdecydowanie tak.
2. Raczej tak.
3. Ani tak, ani nie.
4. Raczej nie.
5. Zdecydowanie nie.

3. Czy używał(-a) Pan(-i) arkuszy z poprzednich edycji egzaminu maturalnego z matematyki podczas zajęć z uczniami badanej klasy?

Proszę zaznaczyć jedną odpowiedź.

1. Tak
2. Nie
3. Nie pamiętam

4. Proszę ocenić, jak dobrze, Pana(-i) zdaniem, badani uczniowie przygotowani są do egzaminu maturalnego z matematyki.

Proszę zaznaczyć jedną odpowiedź.

1. Bardzo dobrze.
2. Raczej dobrze.
3. Raczej źle.
4. Bardzo źle.

5. Czy w Pana(-i) szkole organizowane są dodatkowe zajęcia przygotowujące uczniów do egzaminu maturalnego z matematyki?

Proszę zaznaczyć jedną odpowiedź.

1. Tak
2. Nie

6. Proszę ocenić, jak bardzo, Pana(-i) zdaniem, badanym uczniom zależy na jak najlepszym rozwiązaniu testu w prowadzonym przez IBE badaniu.

Proszę zaznaczyć jedną odpowiedź.

1. Bardzo zależy.
2. Trochę zależy.
3. Niewiele zależy.
4. Wcale nie zależy.

7. Czy zamierza Pan(-i) użyć wyników testu z tego badania w ocenianiu uczniów?

Proszę zaznaczyć jedną odpowiedź.

1. Tak

2. Nie → proszę przejść do pytania 9

8. Czy poinformował(-a) Pan(-i) o tym uczniów przed badaniem?

Proszę zaznaczyć jedną odpowiedź.

1. Tak

2. Nie

9. Kiedy zaczął(-ęła) Pan(-i) uczyć badaną klasę?

Proszę zaznaczyć jedną odpowiedź.

1. Uczę ją od początku jej istnienia.
2. W trakcie jej pierwszego roku nauki.
3. W trakcie jej drugiego roku nauki.
4. W trakcie jej trzeciego roku nauki.
5. [tylko szkoły czteroletnie] W trakcie jej czwartego roku nauki.

10. Czy uczestniczył(-a) Pan(-i) kiedykolwiek w ocenianiu egzaminu maturalnego z matematyki?

Proszę zaznaczyć jedną odpowiedź.

1. Tak

2. Nie → proszę przejść do pytania 12

11. Czy udział w sesjach egzaminacyjnych w roli egzaminatora pomaga Panu(-i) lepiej przygotować uczniów do egzaminu?

Proszę zaznaczyć jedną odpowiedź.

1. Zdecydowanie tak.
2. Raczej tak.
3. Ani tak, ani nie.
4. Raczej nie.
5. Zdecydowanie nie.

12. Jaka jest tygodniowa liczba godzin nauki matematyki w badanej klasie?

Proszę zaznaczyć jedną odpowiedź.

- | | |
|---------------|-----------------------|
| 1. 1 godzina. | 5. 5 godzin. |
| 2. 2 godziny. | 6. 6 godzin. |
| 3. 3 godziny. | 7. 7 godzin. |
| 4. 4 godziny. | 8. 8 godzin i więcej. |

13. Jaki jest Pana(-i) stopień awansu zawodowego?

Proszę zaznaczyć jedną odpowiedź.

1. Profesor oświaty.
2. Nauczyciel dyplomowany.
3. Nauczyciel mianowany.
4. Nauczyciel kontraktowy.
5. Brak stopnia awansu.

14. Czy uważa Pan(-i), że trudność egzaminu maturalnego z matematyki jest z roku na rok porównywalna?

Proszę zaznaczyć jedną odpowiedź.

1. Tak

2. Nie

3. Nie wiem

Bardzo dziękujemy za udzielenie odpowiedzi!

Aneks 5. Psychometryczne własności zadań egzaminacyjnych

Obliczone w trakcie badań parametry psychometryczne zadań egzaminacyjnych zapisane zostały i są upubliczniane za pośrednictwem bazy danych IBE (zob. Rozdział 9). Dostęp do tych informacji możliwy jest na kilka sposobów, z czego dwa najważniejsze to:

- Dostęp do bazy danych IBE za pomocą pakietu ZPD dla programu statystycznego R.
- Strona WWW umożliwiająca przeglądanie banku zadań stanowiącego element bazy IBE (<http://zpd.ibe.edu.pl/doku.php?id=bazatestypytania>) – w widoku zadania wyświetlane są także jego parametry psychometryczne w poszczególnych testach (Rysunek A5.1).

bazapytania

[Powrót do wyszukiwarki pytań](#)

Pytanie 1451 - sprawdzian;;2012;06

Kryteria oceny

sposób oceny	numer kryterium oceny	id kryterium oceny	liczba punktów
nie dotyczy		1451	1

Treść pytania i wiązki pytań

id pytania	opis	treść	tekst dla wiązki
1446	sprawdzian;;2012;01	treść pytania (docx)	tekst dla wiązki (docx)
1447	sprawdzian;;2012;02	treść pytania (docx)	
1448	sprawdzian;;2012;03	treść pytania (docx)	
1449	sprawdzian;;2012;04	treść pytania (docx)	
1450	sprawdzian;;2012;05	treść pytania (docx)	
1451	sprawdzian;;2012;06	treść pytania (docx)	
1452	sprawdzian;;2012;07	treść pytania (docx)	

Testy

id testu	opis testu	data testu	ewd	wariant oceny	poprawna odpowiedź
652	sprawdzian 2012 - arkusz S-A1-122	2012-04-03	nie	nie dotyczy	A
653	sprawdzian 2012 - arkusz S-B1-122	2012-04-03	nie	nie dotyczy	D
654	sprawdzian 2012 - arkusz S-A1-122	2012-04-03	tak	nie dotyczy	A
655	sprawdzian 2012 - arkusz S-B1-122	2012-04-03	tak	nie dotyczy	D
1269	sprawdzian 2012 - arkusz S-7-122	2012-04-03	tak	nie dotyczy	

Skale i skalowania

id skali	nazwa skali	opis skali	kryteria	model	estymacja	parametry									
						ar	mean	rir	rit	rsk	sd	trudność	a	b1	c
145	ktt;652	ktt;S-A1-122;cke	n.d.	KTT	nie dotyczy	0.84	0.30	0.21	0.27	0.46	0.46	0.70			
146	ktt;653	ktt;S-B1-122;cke	n.d.	KTT	nie dotyczy	0.84	0.23	0.22	0.27	0.42	0.42	0.77			
788	paou;s;2012		n.d.	3PL	MML (IRT.ADO)								1.09	1.88	0.13

Rysunek A5.1. Widok zadania na stronie WWW stanowiącej interfejs dla banku zadań – w dolnej części widoczna tabela z parametrami psychometrycznymi zadania (objaśnienia oznaczeń parametrów w tekście)

W obydwu przypadkach dostępne parametry psychometryczne zadań to:

- Parametry KTT (Klasycznej Teorii Testu):

- *ar* (Alpha Rest coefficient) – wskaźnik rzetelności testu, jaki charakteryzowałby test po usunięciu danego zadania;
 - *mean* – średni wynik punktowy uzyskany za dane zadanie;
 - *rir* – współczynnik mocy różnicującej, który jest obliczony na podstawie korelacji danego zadania z wynikiem testu pozbawionego danego zadania;
 - *rit* – współczynnik mocy różnicującej, który jest obliczony na podstawie korelacji danego zadania z wynikiem testu;
 - *rsk* (relative standard deviation) – wartość względnego odchylenia standardowego obliczana jest poprzez podzielenie odchylenia standardowego przez maksimum punktów, jakie uczeń mógł uzyskać za dane zadanie;
 - *sd* – odchylenie standardowe wyników danego zadania;
 - *trudność* – 1 mniej iloraz średniego wyniku punktowego za dane zadanie przez maksymalną liczbę punktów, jakie można było uzyskać w zadaniu.
- Parametry IRT (Item Response Theory)
 - *a* – parametr dyskryminacji zadania;
 - *b*, *b1*, *b2*, itd. – parametr trudności zadania (w wypadku zadań o wielu poziomach wykonania parametry trudności kolejnych poziomów wykonania);
 - *c* – parametr zgadywania zadania.

Dla zilustrowania drugiej z opisanych metod poniżej zamieszczony został przykładowy skrypt programu statystycznego R (wraz z wynikiem działania) pobierający parametry IRT oraz KTT zadań. W analogiczny sposób, posiłkując się w razie potrzeby dokumentacją i przykładami dostępnymi na stronach internetowych <https://github.com/zoziak/ZPD/wiki/> oraz <http://zpd.ibe.edu.pl>, pobrać można parametry psychometryczne zadań dla pozostałych egzaminów. Tabela A5.1 zawiera zwrócone przez skrypt informacje o zadaniach, Tabela A5.2 – parametry IRT zadań, a Tabela A5.3 – parametry KTT.

Skrypt programu statystycznego R pobierający z bazy danych IBE informacje o parametrach psychometrycznych sprawdzianu szóstoklasisty w 2012 roku:

```
# Instalacja pakietu ZPD
install.packages('devtools', 'reshape2')
devtools::install_github('zoziak/ZPD')

# Załadowanie potrzebnych pakietów i nawiązanie połączenia z bazą IBE
library(ZPD)
library(reshape2)
src = polacz()

# Wyszukanie testów ze sprawdzianu z 2012 roku z danych CKE
testy = pobierz_testy(src) %>%
```

```

filter(
  rodzaj_egzaminu == 'sprawdzian',
  czesc_egzaminu == '',
  rok == 2012,
  czy_egzamin == TRUE,
  dane_ewd == FALSE
) %>%
select(id_testu, arkusz)

# Wyszukanie skal zrównujących (parametry IRT) oraz skal KTT dla tych testów
skaleIRT = pobierz_skale(src, czyKtt = TRUE) %>%
  filter(rodzaj_skali == 'zrównywanie') %>%
  semi_join(testy)
skaleKTT = pobierz_skale(src, czyKtt = TRUE) %>%
  filter(rodzaj_skali == 'ktt') %>%
  semi_join(testy)

# Pobranie przydatnych informacji o zadaniach (kryteriach oceny)
kryteriaOceny = pobierz_kryteria_oceny(src) %>%
  select(kryterium, numer_pytania, numer_kryterium, l_punktow, schemat_odp, standard,
standard_szcz, opis_standardu) %>%
  distinct()

# Pobranie parametrów zadań (kryteriów oceny) dla wyszukanych skal
# i połączenie ich z informacjami o zadaniach (kryteriach oceny)
parametryIRT = pobierz_parametry(src) %>%
  semi_join(skaleIRT) %>%
  inner_join(kryteriaOceny) %>%
  arrange(as.numeric(numer_pytania), numer_kryterium) %>%
  collect()
parametryKTT = pobierz_parametry(src) %>%
  inner_join(skaleKTT %>% select(id_skali, id_testu)) %>%
  inner_join(testy) %>%
  inner_join(kryteriaOceny) %>%
  collect()

# Wyświetlenie wyników:
# 1. Informacje o zadaniach (kryteriach oceny)
parametryIRT %>%
  select(kryterium, numer_pytania, numer_kryterium, l_punktow, schemat_odp, standard,
standard_szcz, opis_standardu) %>%
  distinct() %>%
  as.data.frame()
# 2. Parametry IRT
dcast(parametryIRT, kryterium + numer_pytania + numer_kryterium ~ parametr, value.var
= 'wartosc')
# 3. Parametry KTT
dcast(parametryKTT, kryterium + numer_pytania + numer_kryterium + arkusz ~ parametr,
value.var = 'wartosc') %>%
  arrange(as.numeric(numer_pytania), numer_kryterium)

```

Tabela A5.1. Opisy zadań pobrane za pomocą opisanego w aneksie 5 skryptu

Kryterium	Numer pyt.	Numer kryt.	L. pkt.	Schemat odp.	Standard ⁵⁸	Standard szcz. ⁵⁸	Czynność ucznia ⁵⁸
k_1446	1		1	A..D	sprawdzian Czytanie	sprawdzian 1.1	wyszukuje informację w tekście
k_1447	2		1	A..D	sprawdzian Czytanie	sprawdzian 1.1	porównuje informacje z różnych tekstów kultury
k_1448	3		1	A..D	sprawdzian Czytanie	sprawdzian 1.1	wnioskuje na podstawie przesłanek w tekście
k_1449	4		1	A..D	sprawdzian Rozumowanie	sprawdzian 3.1	porządkuje chronologicznie zdarzenia
k_1450	5		1	A..D	sprawdzian Czytanie	sprawdzian 1.1	odczytuje z kontekstu znaczenie użytego w tekście sformułowania
k_1451	6		1	A..D	sprawdzian Czytanie	sprawdzian 1.2	na podstawie użytych środków stylistycznych odczytuje intencję autora
k_1452	7		1	A..D	sprawdzian Czytanie	sprawdzian 1.1	wyszukuje informację w tekście
k_1453	8		1	A..D	sprawdzian Korzystanie z informacji	sprawdzian 4.2	na podstawie informacji z oferty handlowej ustala minimalną liczbę uczestników wycieczki
k_1454	9		1	A..D	sprawdzian Korzystanie z informacji	sprawdzian 4.2	na podstawie informacji z oferty handlowej ustala warunki uzyskania najniższej ceny
k_1455	10		1	A..D	sprawdzian Korzystanie z informacji	sprawdzian 4.2	na podstawie informacji z oferty handlowej ustala cenę usługi
k_1456	11		1	A..D	sprawdzian Czytanie	sprawdzian 1.2	rozpoznaje funkcję stylistyczną porównania
k_1457	12		1	A..D	sprawdzian Czytanie	sprawdzian 1.1	rozpoznaje osobę mówiącą w wierszu
k_1458	13		1	A..D	sprawdzian Czytanie	sprawdzian 1.2	określa funkcję stylistyczną wyrazów dźwiękonaśladowczych
k_1459	14		1	A..D	sprawdzian Czytanie	sprawdzian 1.1	odczytuje z kontekstu znaczenie wyrazu użytego w tekście

⁵⁸ Standardy wymagań egzaminacyjnych obowiązujące dla egzaminu gimnazjalnego do 2011 roku, a dla pozostałych egzaminów do 2014 roku (włącznie).

Kryterium	Numer pyt.	Numer kryt.	L. pkt.	Schemat odp.	Standard ⁵⁸	Standard szcz. ⁵⁸	Czynność ucznia ⁵⁸
k_1460	15		1	A..D	sprawdzian Wykorzystywanie wiedzy w praktyce	sprawdzian 5.3	oblicza czas trwania filmu
k_1461	16		1	A..D	sprawdzian Rozumowanie	sprawdzian 3.6	wyznacza liczbę spełniającą warunki zadania
k_1462	17		1	A..D	sprawdzian Rozumowanie	sprawdzian 3.1	nazywa państwa leżące nad Bałtykiem
k_1463	18		1	A..D	sprawdzian Wykorzystywanie wiedzy w praktyce	sprawdzian 5.5	wyznacza sumę długości odcinków
k_1464	19		1	A..D	sprawdzian Rozumowanie	sprawdzian 3.9	podaje średnicę koła spełniającego warunek określony w zadaniu
k_1465	20		1	A..D	sprawdzian Rozumowanie	sprawdzian 3.5	wskazuje ilustrację graficzną sytuacji opisanej w zadaniu
k_1466	21		4		sprawdzian Wykorzystywanie wiedzy w praktyce	sprawdzian 5.3	wyznacza ceny towarów
k_1467	22		1		sprawdzian Rozumowanie	sprawdzian 3.6	wskazuje osiemnastą część kwadratu
k_1468	23		2		sprawdzian Wykorzystywanie wiedzy w praktyce	sprawdzian 5.3	dzieli przedział czasu na równe części
k_1469	24		2		sprawdzian Rozumowanie	sprawdzian 3.6	ustala liczbę brył o wskazanej własności
p_181	25	1-2	3				
k_1472	25	3	1		sprawdzian Pisanie	sprawdzian 2.1	pisze w funkcjonalnym stylu z dbałością o dobór słownictwa
p_182	26	1-2	4				
p_183	26	3-5	3				

Tabela A5.2. Parametry IRT zadań pobrane za pomocą opisanego w aneksie 5 skryptu

Kryterium	Numer pyt.	Numer kryt.	a	b1	b2	b3	b4	c
k_1446	1		0,6115	-1,8979				
k_1447	2		1,0748	-1,8187				
k_1448	3		1,4112	0,5589				0,144
k_1449	4		0,9660	-1,7771				
k_1450	5		1,3644	-0,5418				0,1711
k_1451	6		1,0867	1,8753				0,1306
k_1452	7		1,2939	-1,2368				0,3074
k_1453	8		1,8864	0,3355				0,2585
k_1454	9		1,5071	-0,1738				0,1870
k_1455	10		1,3509	-1,3853				0,0540
k_1456	11		1,0956	0,0511				0,3185
k_1457	12		1,4082	-1,6721				
k_1458	13		1,6165	-0,2184				0,0767
k_1459	14		1,1300	1,8117				0,1722
k_1460	15		1,5786	0,5893				0,0851
k_1461	16		1,5790	0,3917				0,1831
k_1462	17		0,9932	0,9252				0,2033
k_1463	18		1,9655	0,1826				0,2332
k_1464	19		1,5076	0,7163				0,2060
k_1465	20		1,4753	0,2225				0,1389
k_1466	21		1,8008	-0,5477	-0,3910	0,0992	0,5716	
k_1467	22		1,7537	-0,5439				
k_1468	23		1,4847	0,4197	0,9292			
k_1469	24		1,7514	1,3545				
k_1472	25	3	0,6343	-3,7058				
p_181	25	1-2	0,7783	-3,5617	-1,8066	-0,7337		
p_182	26	1-2	0,8654	-3,7549	-1,4775	1,2547	2,4514	
p_183	26	3-5	1,1019	-0,9293	0,2608	1,6093		

Tabela A5.3. Parametry KKT zadań pobrane za pomocą opisanego w aneksie 5 skryptu

Kryterium	Numer pyt.	Numer kryt.	Arkusz	ar	mean	rir	rit	rsk	sd	trudn.
k_1446	1		S-A1-122	0,843	0,742	0,228	0,281	0,438	0,438	0,258
k_1446	1		S-B1-122	0,836	0,749	0,228	0,282	0,433	0,433	0,251
k_1447	2		S-A1-122	0,842	0,827	0,313	0,357	0,378	0,378	0,173
k_1447	2		S-B1-122	0,834	0,842	0,314	0,357	0,365	0,365	0,158
k_1448	3		S-A1-122	0,840	0,462	0,378	0,432	0,499	0,499	0,538
k_1448	3		S-B1-122	0,833	0,443	0,369	0,425	0,497	0,497	0,557
k_1449	4		S-A1-122	0,842	0,811	0,302	0,348	0,392	0,392	0,189
k_1449	4		S-B1-122	0,835	0,812	0,300	0,347	0,391	0,391	0,188
k_1450	5		S-A1-122	0,839	0,692	0,407	0,456	0,462	0,462	0,308
k_1450	5		S-B1-122	0,832	0,703	0,383	0,434	0,457	0,457	0,297

Kryterium	Numer pyt.	Numer kryt.	Arkusz	ar	mean	rir	rit	rsk	sd	trudn.
k_1451	6		S-A1-122	0,844	0,301	0,214	0,270	0,459	0,459	0,699
k_1451	6		S-B1-122	0,836	0,232	0,222	0,275	0,422	0,422	0,768
k_1452	7		S-A1-122	0,842	0,853	0,313	0,354	0,355	0,355	0,147
k_1452	7		S-B1-122	0,834	0,834	0,316	0,360	0,372	0,372	0,166
k_1453	8		S-A1-122	0,839	0,568	0,385	0,440	0,495	0,495	0,432
k_1453	8		S-B1-122	0,832	0,548	0,393	0,448	0,498	0,498	0,452
k_1454	9		S-A1-122	0,839	0,632	0,415	0,466	0,482	0,482	0,368
k_1454	9		S-B1-122	0,832	0,633	0,404	0,457	0,482	0,482	0,367
k_1455	10		S-A1-122	0,840	0,817	0,373	0,416	0,387	0,387	0,183
k_1455	10		S-B1-122	0,833	0,815	0,373	0,417	0,389	0,389	0,185
k_1456	11		S-A1-122	0,842	0,675	0,290	0,345	0,468	0,468	0,325
k_1456	11		S-B1-122	0,835	0,629	0,285	0,343	0,483	0,483	0,371
k_1457	12		S-A1-122	0,841	0,847	0,365	0,405	0,360	0,360	0,153
k_1457	12		S-B1-122	0,834	0,862	0,360	0,400	0,344	0,344	0,138
k_1458	13		S-A1-122	0,837	0,623	0,474	0,522	0,485	0,485	0,377
k_1458	13		S-B1-122	0,829	0,567	0,486	0,535	0,495	0,495	0,433
k_1459	14		S-A1-122	0,843	0,283	0,222	0,277	0,451	0,451	0,717
k_1459	14		S-B1-122	0,837	0,323	0,188	0,248	0,468	0,468	0,677
k_1460	15		S-A1-122	0,838	0,403	0,427	0,478	0,491	0,491	0,597
k_1460	15		S-B1-122	0,831	0,398	0,420	0,472	0,489	0,489	0,602
k_1461	16		S-A1-122	0,839	0,516	0,393	0,447	0,500	0,500	0,484
k_1461	16		S-B1-122	0,832	0,496	0,399	0,453	0,500	0,500	0,504
k_1462	17		S-A1-122	0,842	0,432	0,301	0,358	0,495	0,495	0,568
k_1462	17		S-B1-122	0,836	0,481	0,226	0,288	0,500	0,500	0,519
k_1463	18		S-A1-122	0,838	0,573	0,431	0,483	0,495	0,495	0,427
k_1463	18		S-B1-122	0,831	0,579	0,421	0,474	0,494	0,494	0,421
k_1464	19		S-A1-122	0,841	0,457	0,340	0,397	0,498	0,498	0,543
k_1464	19		S-B1-122	0,833	0,459	0,335	0,393	0,498	0,498	0,541
k_1465	20		S-A1-122	0,839	0,522	0,412	0,465	0,500	0,500	0,478
k_1465	20		S-B1-122	0,832	0,519	0,403	0,458	0,500	0,500	0,481
k_1466	21		S-A1-122	0,845	2,091	0,592	0,726	0,427	1,706	0,477
k_1466	21		S-B1-122	0,837	2,078	0,584	0,723	0,426	1,705	0,481
k_1467	22		S-A1-122	0,836	0,653	0,518	0,563	0,476	0,476	0,347
k_1467	22		S-B1-122	0,829	0,654	0,515	0,560	0,476	0,476	0,346
k_1468	23		S-A1-122	0,835	0,663	0,500	0,583	0,438	0,875	0,668
k_1468	23		S-B1-122	0,828	0,652	0,483	0,569	0,435	0,870	0,674
k_1469	24		S-A1-122	0,839	0,710	0,387	0,467	0,370	0,740	0,645
k_1469	24		S-B1-122	0,833	0,700	0,364	0,446	0,366	0,732	0,650
k_1470	25	1	S-A1-122	0,842	0,789	0,294	0,342	0,408	0,408	0,211
k_1470	25	1	S-B1-122	0,835	0,789	0,290	0,339	0,408	0,408	0,211
k_1471	25	2	S-A1-122	0,840	1,543	0,365	0,438	0,329	0,659	0,229
k_1471	25	2	S-B1-122	0,833	1,540	0,361	0,435	0,329	0,658	0,230
k_1472	25	3	S-A1-122	0,844	0,900	0,180	0,218	0,299	0,299	0,100
k_1472	25	3	S-B1-122	0,837	0,899	0,183	0,221	0,301	0,301	0,101
k_1473	26	1	S-A1-122	0,839	1,861	0,404	0,484	0,251	0,754	0,380
k_1473	26	1	S-B1-122	0,832	1,854	0,384	0,466	0,248	0,744	0,382

Kryterium	Numer pyt.	Numer kryt.	Arkusz	ar	mean	rir	rit	rsk	sd	trudn.
k_1474	26	2	S-A1-122	0,841	0,257	0,342	0,392	0,437	0,437	0,743
k_1474	26	2	S-B1-122	0,834	0,250	0,323	0,374	0,433	0,433	0,750
k_1475	26	3	S-A1-122	0,839	0,420	0,402	0,455	0,494	0,494	0,580
k_1475	26	3	S-B1-122	0,832	0,417	0,389	0,443	0,493	0,493	0,583
k_1476	26	4	S-A1-122	0,839	0,526	0,402	0,455	0,499	0,499	0,474
k_1476	26	4	S-B1-122	0,832	0,523	0,391	0,446	0,499	0,499	0,477
k_1477	26	5	S-A1-122	0,840	0,391	0,359	0,413	0,488	0,488	0,609
k_1477	26	5	S-B1-122	0,833	0,384	0,341	0,397	0,486	0,486	0,616

Aneks 6. Zgodność kodowania zadań otwartych

6.1. Egzamin gimnazjalny – część humanistyczna 2011

	Numer zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd std.	z
			obserwowana	oczekiwana			
Z1	1	91	97,75%	55,31%	0,95	0,11	8,96
	2	92	100,00%	52,78%	1,00	0,11	9,43
	3	93	97,75%	50,49%	0,95	0,11	9,02
	4	94	92,13%	39,43%	0,87	0,08	10,87
	5	96	84,27%	24,97%	0,79	0,06	13,94
	6	136	98,88%	50,40%	0,98	0,11	9,22
	7	137	97,75%	51,80%	0,95	0,11	9,00
	8	138	96,63%	71,23%	0,88	0,11	8,34
	9	139	94,38%	50,09%	0,89	0,11	8,37
	10	140	98,88%	64,54%	0,97	0,11	9,14
	11	141	94,38%	50,17%	0,89	0,11	8,39
	12	142	96,63%	69,79%	0,89	0,11	8,39
	13	143	96,63%	61,13%	0,91	0,11	8,62
	14	144	87,64%	19,40%	0,85	0,05	16,80
	15	1609	94,38%	50,40%	0,89	0,11	8,37
	16	1610	97,75%	78,31%	0,90	0,11	8,50
	17	1611	88,76%	14,27%	0,87	0,04	21,10
Z2	1	142	87,78%	60,22%	0,69	0,10	6,64
	2	143	94,44%	80,96%	0,71	0,10	7,02
	3	136	94,44%	50,00%	0,89	0,11	8,45
	4	137	96,67%	50,44%	0,93	0,11	8,87
	5	138	96,67%	63,63%	0,91	0,11	8,62
	6	139	96,67%	52,72%	0,93	0,11	8,82
	7	140	94,44%	58,30%	0,87	0,10	8,30
	8	141	92,22%	50,74%	0,84	0,11	7,99
	9	144	90,00%	19,81%	0,88	0,05	17,33
	10	187	94,44%	58,44%	0,87	0,11	8,22
	11	190	98,89%	57,56%	0,97	0,11	9,24
	12	191	93,33%	55,46%	0,85	0,10	8,11
	13	192	82,22%	23,86%	0,77	0,06	13,72
	14	1609	98,89%	53,26%	0,98	0,11	9,26
	15	1610	97,78%	70,74%	0,92	0,11	8,79
	16	1544	98,89%	50,30%	0,98	0,11	9,28
	17	1545	93,33%	50,22%	0,87	0,11	8,22
	18	1546	96,67%	39,04%	0,95	0,08	11,73
Z3	1	1200	100,00%	52,38%	1,00	0,11	9,33
	2	1211	98,85%	50,42%	0,98	0,11	9,11
	3	1201	100,00%	52,91%	1,00	0,11	9,33

	4	1212	97,70%	50,80%	0,95	0,11	8,89
	5	1202	95,40%	55,53%	0,90	0,11	8,37
	6	1213	97,70%	72,97%	0,92	0,11	8,57
	7	1204	97,70%	51,12%	0,95	0,11	8,89
	8	1215	97,70%	60,05%	0,94	0,11	8,79
	9	1205	70,11%	27,63%	0,59	0,06	9,52
	10	1199	94,25%	50,42%	0,88	0,11	8,25
	11	1210	95,40%	50,51%	0,91	0,11	8,47
	12	1143	94,25%	61,59%	0,85	0,11	7,96
	13	1158	97,70%	71,46%	0,92	0,11	8,58
	14	1144	95,40%	68,53%	0,85	0,11	7,99
	15	1159	85,06%	28,27%	0,79	0,06	12,54
	16	1149	97,70%	52,38%	0,95	0,11	8,88
	17	1164	96,55%	51,24%	0,93	0,11	8,69
	18	1151	97,70%	37,19%	0,96	0,08	12,38
	19	1168	98,85%	55,94%	0,97	0,11	9,09
	20	1145	94,25%	51,24%	0,88	0,11	8,25
	21	1160	94,25%	58,50%	0,86	0,11	8,06
	22	1150	94,25%	50,23%	0,88	0,11	8,25
	23	1165	98,85%	50,10%	0,98	0,11	9,11
Z4	1	232	93,48%	50,07%	0,87	0,10	8,35
	2	233	95,65%	50,21%	0,91	0,10	8,75
	3	236	92,39%	49,86%	0,85	0,10	8,23
	4	237	96,74%	50,24%	0,93	0,10	8,98
	5	238	95,65%	50,19%	0,91	0,10	8,76
	6	241	100,00%	75,73%	1,00	0,10	9,59
	7	242	90,22%	23,33%	0,87	0,05	16,02
	8	286	90,22%	37,42%	0,84	0,08	11,22
	9	288	85,87%	23,71%	0,81	0,06	14,41
	10	1507	93,48%	52,84%	0,86	0,10	8,27
	11	1630	94,57%	50,43%	0,89	0,10	8,56
	12	1631	93,48%	53,90%	0,86	0,10	8,27
	13	1632	96,74%	50,28%	0,93	0,10	8,96
	14	1633	98,91%	53,69%	0,98	0,10	9,37
	15	1634	100,00%	78,95%	1,00	0,10	9,59
	16	1547	100,00%	51,51%	1,00	0,10	9,59
	17	1639	97,83%	51,16%	0,96	0,10	9,16
	18	1640	95,65%	65,88%	0,87	0,10	8,44
	19	1641	98,91%	50,14%	0,98	0,10	9,38
Z5	1	284	94,94%	41,00%	0,91	0,09	10,43
	2	286	94,94%	38,58%	0,92	0,09	10,76
	3	288	93,67%	31,98%	0,91	0,07	12,73
	4	327	92,41%	62,06%	0,80	0,11	7,17
	5	328	96,20%	51,15%	0,92	0,11	8,20
	6	330	93,67%	51,15%	0,87	0,11	7,74

	7	331	93,67%	51,56%	0,87	0,11	7,73
	8	333	97,47%	35,19%	0,96	0,08	11,85
	9	334	87,34%	35,88%	0,80	0,08	10,61
	10	1547	96,20%	59,25%	0,91	0,11	8,06
	11	1640	92,41%	60,94%	0,81	0,11	7,18
	12	1641	96,20%	50,02%	0,92	0,11	8,22
	13	1634	100,00%	67,70%	1,00	0,11	8,89
	14	1561	100,00%	50,39%	1,00	0,11	8,89
	15	1562	98,73%	55,41%	0,97	0,11	8,64
	16	1563	100,00%	52,89%	1,00	0,11	8,89
	17	1649	100,00%	58,72%	1,00	0,11	8,89
	18	1548	92,41%	49,88%	0,85	0,11	7,58
Z6	1	327	97,83%	52,34%	0,95	0,10	9,16
	2	328	94,57%	51,65%	0,89	0,10	8,53
	3	330	91,30%	49,63%	0,83	0,10	8,09
	4	331	92,39%	51,23%	0,84	0,10	8,25
	5	333	97,83%	33,93%	0,97	0,07	13,06
	6	334	88,04%	22,22%	0,85	0,05	15,77
	7	1561	96,74%	52,45%	0,93	0,10	9,09
	8	1562	98,91%	49,47%	0,98	0,10	9,54
	9	1563	97,83%	52,34%	0,95	0,10	9,16
	10	373	94,57%	51,03%	0,89	0,10	8,69
	11	374	97,83%	53,38%	0,95	0,10	9,15
	12	377	96,74%	51,65%	0,93	0,10	8,97
	13	378	96,74%	50,50%	0,93	0,10	9,12
	14	379	95,65%	49,56%	0,91	0,10	8,91
	15	380	88,04%	49,88%	0,76	0,10	7,63
	16	381	97,83%	56,47%	0,95	0,10	9,29
	17	382	91,30%	21,09%	0,89	0,05	16,79
	18	1649	95,65%	54,58%	0,90	0,10	8,84
	19	1548	95,65%	50,12%	0,91	0,10	8,79
	20	1653	96,74%	50,95%	0,93	0,10	8,97
	21	1654	98,91%	69,19%	0,96	0,10	9,26
	22	1655	89,13%	49,62%	0,78	0,10	7,70
	23	1656	92,39%	54,68%	0,83	0,10	8,10
Z7	1	373	95,45%	50,10%	0,91	0,11	8,53
	2	374	92,05%	52,69%	0,83	0,11	7,86
	3	377	98,86%	50,00%	0,98	0,11	9,17
	4	378	90,91%	49,97%	0,82	0,11	7,68
	5	379	92,05%	49,90%	0,84	0,11	7,94
	6	380	90,91%	50,00%	0,82	0,11	7,68
	7	381	92,05%	60,54%	0,80	0,10	7,65
	8	382	75,00%	21,27%	0,68	0,05	12,56
	9	420	94,32%	52,27%	0,88	0,11	8,29
	10	422	98,86%	51,45%	0,98	0,11	9,16

	11	424	87,50%	50,93%	0,75	0,11	7,04
	12	428	64,77%	16,81%	0,58	0,05	12,40
	13	1653	98,86%	51,45%	0,98	0,11	9,16
	14	1654	96,59%	61,93%	0,91	0,11	8,54
	15	1655	100,00%	67,46%	1,00	0,11	9,38
	16	1656	94,32%	53,41%	0,88	0,11	8,24
	17	1667	100,00%	53,72%	1,00	0,11	9,38
Z8	1	420	98,97%	52,12%	0,98	0,10	9,64
	2	422	100,00%	51,54%	1,00	0,10	9,85
	3	424	95,88%	51,54%	0,91	0,10	9,01
	4	428	70,10%	17,05%	0,64	0,05	14,22
	5	1602	94,85%	35,80%	0,92	0,07	12,57
	6	474	94,85%	53,54%	0,89	0,10	8,78
	7	475	93,81%	49,92%	0,88	0,10	8,66
	8	476	83,51%	21,08%	0,79	0,05	15,39
	9	1667	100,00%	58,93%	1,00	0,10	9,85
	10	1557	100,00%	51,20%	1,00	0,10	9,85
Z9	1	142	93,75%	75,31%	0,75	0,11	6,76
	2	143	93,75%	54,81%	0,86	0,11	7,73
	3	136	97,50%	50,28%	0,95	0,11	8,49
	4	137	95,00%	50,78%	0,90	0,11	8,04
	5	138	100,00%	76,28%	1,00	0,11	8,94
	6	139	98,75%	50,37%	0,97	0,11	8,72
	7	140	95,00%	60,09%	0,87	0,11	7,84
	8	141	96,25%	50,63%	0,92	0,11	8,27
	9	144	77,50%	18,42%	0,72	0,05	13,84
	10	284	93,75%	38,98%	0,90	0,09	10,53
	11	286	88,75%	33,94%	0,83	0,08	10,49
	12	288	81,25%	27,63%	0,74	0,07	11,24
	13	1557	97,50%	51,97%	0,95	0,11	8,49
	14	1676	12,50%	12,67%	0,00	0,04	-0,05
Z10	1	284	95,06%	37,75%	0,92	0,08	11,10
	2	286	80,25%	35,47%	0,69	0,08	8,87
	3	288	79,01%	40,88%	0,65	0,07	8,91
	4	1602	66,67%	35,86%	0,48	0,08	6,06
	5	474	93,83%	53,62%	0,87	0,11	7,83
	6	475	88,89%	49,84%	0,78	0,11	7,06
	7	476	76,54%	20,03%	0,71	0,06	12,75
	8	1505	91,36%	79,29%	0,58	0,11	5,25
	9	1676	54,32%	11,16%	0,49	0,04	12,53
Z11	1	142	95,29%	65,29%	0,86	0,11	7,97
	2	143	94,12%	67,24%	0,82	0,11	7,61
	3	136	94,12%	50,44%	0,88	0,11	8,13
	4	137	89,41%	51,43%	0,78	0,11	7,32

	5	138	98,82%	64,64%	0,97	0,11	8,92
	6	139	97,65%	50,15%	0,95	0,11	8,79
	7	140	94,12%	62,20%	0,84	0,11	7,79
	8	141	98,82%	50,02%	0,98	0,11	9,01
	9	144	75,29%	20,11%	0,69	0,05	13,05
	10	1602	75,29%	37,84%	0,60	0,08	7,57
	11	474	85,88%	51,31%	0,71	0,11	6,62
	12	475	91,76%	51,90%	0,83	0,11	7,76
	13	476	77,65%	23,86%	0,71	0,06	12,02
	14	1505	94,12%	65,94%	0,83	0,11	7,63
	15	1611	76,47%	15,54%	0,72	0,04	16,30

6.2. Egzamin gimnazjalny – część matematyczno-przyrodnicza 2011

	Numer zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd std.	z
			obserwowana	oczekiwana			
Z1	1	592	92,77%	35,97%	0,89	0,08	11,12
	2	646	83,13%	24,46%	0,78	0,06	13,58
	3	1697	95,18%	35,90%	0,92	0,08	11,96
	4	1596	83,13%	29,31%	0,76	0,06	12,35
Z2	1	646	85,71%	26,70%	0,81	0,06	12,83
	2	696	85,71%	31,34%	0,79	0,06	12,50
	3	1596	89,61%	37,00%	0,84	0,07	12,64
	4	1501	92,21%	41,81%	0,87	0,08	10,41
	5	1715	92,21%	58,96%	0,81	0,09	9,40
Z3	1	696	83,33%	22,77%	0,78	0,05	14,48
	2	739	91,67%	34,27%	0,87	0,08	11,23
	3	741	82,14%	26,50%	0,76	0,06	12,30
	4	1501	91,67%	38,90%	0,86	0,08	10,83
	5	1715	97,62%	66,50%	0,93	0,08	11,14
	6	1724	91,67%	42,22%	0,86	0,09	9,63
Z4	1	739	90,59%	35,97%	0,85	0,08	10,80
	2	741	90,59%	30,03%	0,87	0,07	13,26
	3	784	84,71%	25,36%	0,80	0,06	13,04
	4	1724	87,06%	45,30%	0,76	0,08	9,27
	5	1736	96,47%	56,58%	0,92	0,10	9,56
Z5	1	784	92,39%	24,30%	0,90	0,06	15,40
	2	837	90,22%	31,38%	0,86	0,07	12,74
	3	1736	84,78%	41,22%	0,74	0,07	10,18
	4	1697	94,57%	38,34%	0,91	0,08	11,84
Z6	1	837	89,36%	30,98%	0,85	0,06	13,31
	2	892	94,68%	26,86%	0,93	0,06	15,32
	3	1697	97,87%	32,50%	0,97	0,07	14,12
	4	1744	97,87%	66,50%	0,94	0,08	11,77

	5	1595	97,87%	43,52%	0,96	0,09	11,32
Z7	1	924	97,53%	43,76%	0,96	0,09	10,29
	2	892	86,42%	27,27%	0,81	0,06	12,96
	3	935	90,12%	41,03%	0,83	0,07	12,02
	4	1744	95,06%	47,23%	0,91	0,08	10,67
	5	1595	90,12%	35,18%	0,85	0,08	11,15
	6	1756	91,36%	39,23%	0,86	0,08	10,47
	7	1514	96,30%	45,48%	0,93	0,09	9,94
Z8	1	924	91,57%	39,66%	0,86	0,08	10,14
	2	935	83,13%	40,54%	0,72	0,07	10,29
	3	981	90,36%	37,73%	0,85	0,08	10,37
	4	987	79,52%	33,24%	0,69	0,06	10,70
	5	1756	91,57%	45,01%	0,85	0,09	9,34
	6	1514	93,98%	44,22%	0,89	0,09	9,71
	7	1724	93,98%	55,15%	0,87	0,09	9,72
Z9	1	1770	92,41%	58,15%	0,82	0,09	9,12
	2	784	88,61%	23,19%	0,85	0,06	14,21
	3	646	84,81%	20,99%	0,81	0,06	14,62
	4	1724	89,87%	39,59%	0,83	0,08	9,84
Z10	1	1779	82,46%	30,61%	0,75	0,06	12,77
	2	1770	83,33%	53,42%	0,64	0,07	9,48
	3	578	78,07%	27,28%	0,70	0,06	12,67
	4	592	89,47%	36,08%	0,84	0,07	12,45
	5	981	80,70%	40,14%	0,68	0,07	9,61
	6	987	74,56%	25,01%	0,66	0,05	12,94
Z11	1	646	80,52%	20,80%	0,75	0,06	13,45
	2	981	84,42%	45,72%	0,71	0,09	8,36
	3	987	83,12%	26,77%	0,77	0,06	12,56
	4	1697	89,61%	37,34%	0,83	0,08	10,34
	5	1779	84,42%	29,82%	0,78	0,07	11,02

6.3. Matura z matematyki 2013

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd std.	Z
		obserwowana	oczekiwana			
1	1929	98,55%	37,30%	0,98	0,09	10,95
2	1853	96,97%	42,29%	0,95	0,09	10,48
3	1856	98,15%	53,91%	0,96	0,10	9,40
4	1932	100,00%	44,27%	1,00	0,11	9,28
5	1501	98,41%	39,76%	0,97	0,09	10,63
6	1934	98,44%	96,90%	0,50	0,06	8,00
7	1858	100,00%	28,85%	1,00	0,08	13,02
8	1982	93,33%	25,06%	0,91	0,07	12,82
9	1596	100,00%	25,98%	1,00	0,07	13,86

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd std.	Z
		obserwowana	oczekiwana			
10	1852	0,00%	23,67%	-0,31	0,04	-7,21
11	1898	98,55%	36,04%	0,98	0,09	11,20
12	1978	96,77%	92,17%	0,59	0,10	6,04
13	1975	86,44%	48,81%	0,74	0,10	7,27
14	1902	100,00%	83,48%	1,00	0,12	8,67
15	1901	100,00%	52,75%	1,00	0,10	9,55
16	1979	100,00%	51,66%	1,00	0,12	8,03
17	1982	93,10%	31,60%	0,90	0,08	10,71
18	1905	76,92%	34,84%	0,65	0,08	8,12
19	1858	94,20%	28,21%	0,92	0,07	12,80
20	1929	100,00%	39,19%	1,00	0,09	11,38
21	1898	98,55%	39,00%	0,98	0,09	11,02
22	1931	100,00%	34,73%	1,00	0,08	11,97
23	1902	0,00%	0,77%	-0,01	0,01	-0,72
24	2012	98,41%	73,49%	0,94	0,10	9,07
25	2013	0,00%	0,41%	0,00	0,01	-0,56
26	2015	21,74%	6,97%	0,16	0,02	6,46
27	1905	72,58%	24,53%	0,64	0,06	10,16
28	1938	95,65%	45,09%	0,92	0,11	8,52
29	2054	97,37%	42,21%	0,95	0,09	10,36
30	1979	100,00%	45,95%	1,00	0,11	9,04
31	1975	94,83%	59,51%	0,87	0,11	7,89
32	1933	96,43%	43,11%	0,94	0,11	8,47
33	1978	93,44%	84,63%	0,57	0,10	5,59
34	2056	97,22%	55,61%	0,94	0,09	10,25
35	1935	95,24%	34,44%	0,93	0,09	10,34
36	1982	87,88%	35,67%	0,81	0,07	11,62
37	2062	34,43%	15,96%	0,22	0,05	4,66
38	1938	41,86%	24,45%	0,23	0,07	3,31
39	1974	0,00%	8,54%	-0,09	0,02	-3,82
40	1975	96,88%	59,77%	0,92	0,10	9,37
41	2010	98,28%	58,59%	0,96	0,11	8,65
42	2012	98,39%	62,59%	0,96	0,11	8,75
43	1501	97,01%	53,71%	0,94	0,09	10,17
44	2013	100,00%	76,16%	1,00	0,12	8,50
45	1979	100,00%	52,47%	1,00	0,12	8,62
46	1596	95,59%	23,75%	0,94	0,07	14,26
47	2017	98,48%	37,74%	0,98	0,07	13,90
48	1982	93,10%	40,73%	0,88	0,09	10,12
49	2054	96,30%	39,52%	0,94	0,08	11,31
50	1853	91,25%	49,16%	0,83	0,08	9,86
51	1854	96,55%	83,83%	0,79	0,10	7,87

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd std.	Z
		obserwowana	oczekiwana			
52	2010	0,00%	7,03%	-0,08	0,02	-3,23
53	2012	98,59%	65,42%	0,96	0,09	10,56
54	2056	0,00%	13,24%	-0,15	0,02	-6,59
55	2015	0,00%	7,61%	-0,08	0,02	-5,38
56	1858	57,89%	20,53%	0,47	0,05	10,44
57	2062	87,88%	30,00%	0,83	0,07	12,08
58	2054	98,41%	34,09%	0,98	0,09	10,88
59	1898	95,16%	33,48%	0,93	0,09	10,32
60	1901	98,33%	47,36%	0,97	0,11	8,57
61	1902	0,00%	2,25%	-0,02	0,01	-3,04
62	1501	96,67%	52,64%	0,93	0,10	9,59
63	2056	0,00%	15,84%	-0,19	0,03	-6,68
64	1596	0,00%	6,19%	-0,07	0,03	-2,33
65	1903	86,67%	32,17%	0,80	0,08	10,41
66	2062	57,89%	34,35%	0,36	0,05	6,64

6.4. Matura z matematyki 2014

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		Kappa	Błąd std.	Z
		Obserwowana	Oczekiwana			
1	2009	99,35%	57,84%	0,98	0,04	25,53
2	2012	100,00%	58,21%	1,00	0,04	24,85
3	2055	99,67%	58,24%	0,99	0,04	24,54
4	2057	99,67%	57,88%	0,99	0,04	25,58
5	4150	99,67%	57,53%	0,99	0,04	26,84
6	2059	100,00%	57,87%	1,00	0,04	25,84
7	2014	99,67%	57,26%	0,99	0,04	27,92
8	1062	98,37%	57,62%	0,96	0,04	25,72
9	2061	99,02%	56,92%	0,98	0,03	29,07
10	1063	99,67%	57,07%	0,99	0,03	28,85
11	2017	99,02%	57,38%	0,98	0,04	27,12
12	4052	99,67%	59,82%	0,99	0,04	24,81
13	4048	99,67%	59,42%	0,99	0,04	26,23
14	2055	99,02%	60,11%	0,98	0,04	23,48
15	2057	99,35%	59,59%	0,98	0,04	25,34
16	4150	97,71%	59,96%	0,94	0,04	23,17
17	2059	99,35%	59,61%	0,98	0,04	25,25
18	4050	99,35%	59,51%	0,98	0,04	25,66
19	1063	98,04%	59,42%	0,95	0,04	25,20
20	2061	97,71%	58,78%	0,94	0,03	27,75

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		Kappa	Błąd std.	Z
		Obserwowana	Oczekiwana			
21	4054	96,73%	58,75%	0,92	0,03	27,27
22	1062	97,39%	59,44%	0,94	0,04	24,67
23	4052	100,00%	58,16%	1,00	0,04	24,93
24	4048	99,02%	57,39%	0,98	0,04	26,95
25	7165	98,69%	57,53%	0,97	0,04	26,20
26	4150	99,67%	57,99%	0,99	0,04	25,26
27	7168	99,02%	57,78%	0,98	0,04	25,52
28	7170	99,67%	58,37%	0,99	0,04	24,24
29	4050	99,35%	57,43%	0,98	0,04	26,99
30	1062	98,04%	57,86%	0,95	0,04	24,74
31	1063	98,37%	57,12%	0,96	0,03	27,73
32	4054	99,35%	56,92%	0,98	0,03	29,32
33	7171	99,02%	56,77%	0,98	0,03	29,83
34	2009	99,35%	58,70%	0,98	0,04	25,49
35	2012	99,67%	59,13%	0,99	0,04	24,38
36	4052	99,67%	58,78%	0,99	0,04	25,43
37	4048	100,00%	58,51%	1,00	0,04	26,62
38	4150	99,02%	58,44%	0,98	0,04	26,23
39	4050	99,67%	58,41%	0,99	0,04	26,72
40	2014	99,67%	58,09%	0,99	0,04	28,10
41	1063	99,02%	57,93%	0,98	0,03	28,53
42	2017	99,67%	58,35%	0,99	0,04	27,11
43	4054	99,35%	57,85%	0,98	0,03	29,12
44	1062	99,67%	58,45%	0,99	0,04	26,63

6.5. Matura z języka angielskiego 2014

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		Kappa	Błąd stand.	Z
		Obserwowana	Oczekiwana			
1	6153	99,77%	83,99%	0,99	0,03	37,15
2	6154	100,00%	84,02%	1,00	0,03	36,62
3	6155	99,77%	83,99%	0,99	0,03	37,14
4	6156	99,43%	83,88%	0,96	0,02	40,21
5	6157	100,00%	83,92%	1,00	0,03	40,05
6	6205	99,54%	83,91%	0,97	0,02	39,34
7	6206	99,43%	83,88%	0,96	0,02	40,28
8	6207	99,77%	83,96%	0,99	0,03	37,95
9	6208	99,43%	83,83%	0,96	0,02	42,28
10	6209	99,77%	83,84%	0,99	0,02	42,82
11	6210	99,89%	83,85%	0,99	0,02	42,53

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		Kappa	Błąd stand.	Z
		Obserwowana	Oczekiwana			
12	6211	99,77%	83,84%	0,99	0,02	43,00
13	6205	98,74%	84,09%	0,92	0,02	38,13
14	6206	98,52%	84,05%	0,91	0,02	38,95
15	6207	98,86%	84,19%	0,93	0,03	34,83
16	6208	97,83%	84,02%	0,86	0,02	38,29
17	6209	99,77%	84,06%	0,99	0,02	42,00
18	6210	98,74%	84,02%	0,92	0,02	40,95
19	6211	99,89%	84,06%	0,99	0,02	42,39
20	6247	99,09%	84,13%	0,94	0,03	37,41
21	6248	98,29%	84,16%	0,89	0,03	34,33
22	6249	99,43%	84,15%	0,96	0,03	37,31
23	6250	99,09%	84,18%	0,94	0,03	35,67
24	6251	99,54%	84,13%	0,97	0,03	38,52
25	6247	99,54%	83,53%	0,97	0,03	38,47
26	6248	98,63%	83,56%	0,92	0,03	35,46
27	6249	99,09%	83,55%	0,94	0,03	36,94
28	6250	99,43%	83,66%	0,97	0,03	34,45
29	6251	99,43%	83,53%	0,97	0,03	38,18
30	6294	99,77%	83,61%	0,99	0,03	36,68
31	6295	99,54%	83,61%	0,97	0,03	36,16
32	6296	99,54%	83,50%	0,97	0,02	39,67
33	6297	99,66%	83,51%	0,98	0,02	39,42
34	6298	99,77%	83,49%	0,99	0,02	40,46
35	6153	99,54%	85,18%	0,97	0,03	36,56
36	6154	99,89%	85,32%	0,99	0,03	33,14
37	6155	99,43%	85,27%	0,96	0,03	33,55
38	6156	99,09%	85,13%	0,94	0,03	37,48
39	6157	99,77%	85,13%	0,98	0,03	39,10
40	6294	99,20%	85,25%	0,95	0,03	33,38
41	6295	99,66%	85,27%	0,98	0,03	34,07
42	6296	99,43%	85,22%	0,96	0,03	35,16
43	6297	99,54%	85,18%	0,97	0,03	36,71
44	6298	99,54%	85,14%	0,97	0,03	38,39
45	6153	98,86%	84,23%	0,93	0,03	33,78
46	6154	99,54%	84,32%	0,97	0,03	32,92
47	6155	99,43%	84,26%	0,96	0,03	34,27
48	6156	99,66%	84,16%	0,98	0,03	37,73
49	6157	99,66%	84,16%	0,98	0,03	37,70
50	6934	98,52%	84,11%	0,91	0,02	36,69
51	6935	98,97%	84,11%	0,94	0,02	37,68
52	6936	98,97%	84,12%	0,94	0,02	37,49
53	6937	98,52%	84,13%	0,91	0,03	36,02

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		Kappa	Błąd stand.	Z
		Obserwowana	Oczekiwana			
54	6938	99,89%	84,13%	0,99	0,03	39,22
55	6939	99,20%	84,06%	0,95	0,02	40,28
56	6940	99,32%	84,06%	0,96	0,02	40,68
57	6205	99,43%	84,35%	0,96	0,03	37,49
58	6206	99,09%	84,34%	0,94	0,03	37,14
59	6207	98,86%	84,44%	0,93	0,03	33,44
60	6208	98,86%	84,24%	0,93	0,02	40,55
61	6209	99,66%	84,30%	0,98	0,02	39,79
62	6210	99,54%	84,26%	0,97	0,02	41,45
63	6211	99,89%	84,26%	0,99	0,02	42,46
64	6976	99,89%	84,49%	0,99	0,03	34,36
65	6977	99,77%	84,43%	0,99	0,03	35,58
66	6978	99,66%	84,46%	0,98	0,03	34,53
67	6979	99,66%	84,40%	0,98	0,03	36,35
68	6980	99,89%	84,36%	0,99	0,03	38,15
69	6247	99,32%	85,35%	0,95	0,03	37,25
70	6248	98,97%	85,40%	0,93	0,03	34,47
71	6249	99,20%	85,34%	0,95	0,03	37,52
72	6250	99,09%	85,39%	0,94	0,03	35,13
73	6251	99,20%	85,34%	0,95	0,03	37,39
74	7023	99,43%	85,36%	0,96	0,03	37,32
75	7024	99,32%	85,35%	0,95	0,03	37,31
76	7025	99,32%	85,36%	0,95	0,03	37,10
77	7026	99,09%	85,36%	0,94	0,03	36,49
78	7027	99,20%	85,33%	0,95	0,02	37,97
79	6294	99,54%	84,22%	0,97	0,03	35,59
80	6295	99,32%	84,29%	0,96	0,03	33,21
81	6296	99,54%	84,16%	0,97	0,03	37,35
82	6297	99,32%	84,19%	0,96	0,03	35,86
83	6298	99,54%	84,14%	0,97	0,03	38,07
84	7246	98,52%	84,15%	0,91	0,03	35,43
85	7247	98,74%	84,11%	0,92	0,02	37,09
86	7248	98,74%	84,17%	0,92	0,03	35,15
87	7249	98,17%	84,09%	0,89	0,02	36,43
88	7250	99,54%	84,06%	0,97	0,02	41,33
89	7251	99,09%	84,10%	0,94	0,02	38,43
90	7252	99,66%	84,04%	0,98	0,02	42,35
91	6153	99,20%	85,38%	0,95	0,03	35,85
92	6154	99,54%	85,45%	0,97	0,03	34,43
93	6155	99,43%	85,37%	0,96	0,03	36,69
94	6156	99,66%	85,32%	0,98	0,02	39,73
95	6157	99,77%	85,32%	0,98	0,02	40,08

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		Kappa	Błąd stand.	Z
		Obserwowana	Oczekiwana			
96	6976	99,54%	85,43%	0,97	0,03	34,82
97	6977	99,66%	85,38%	0,98	0,03	36,92
98	6978	99,77%	85,45%	0,98	0,03	34,99
99	6979	99,32%	85,34%	0,95	0,03	37,85
100	6980	100,00%	85,34%	1,00	0,03	39,51
101	6205	98,74%	83,75%	0,92	0,03	36,07
102	6206	98,97%	83,70%	0,94	0,02	38,06
103	6207	98,97%	83,77%	0,94	0,03	35,98
104	6208	98,63%	83,62%	0,92	0,02	40,53
105	6209	98,63%	83,66%	0,92	0,02	38,83
106	6210	98,86%	83,63%	0,93	0,02	40,56
107	6211	99,09%	83,64%	0,94	0,02	40,92
108	7023	98,86%	83,74%	0,93	0,03	36,39
109	7024	98,86%	83,76%	0,93	0,03	35,97
110	7025	99,32%	83,75%	0,96	0,03	37,26
111	7026	99,20%	83,74%	0,95	0,03	37,36
112	7027	99,32%	83,71%	0,96	0,02	38,70
113	6247	98,86%	84,35%	0,93	0,03	36,06
114	6248	98,74%	84,42%	0,92	0,03	33,62
115	6249	98,63%	84,39%	0,91	0,03	34,11
116	6250	99,54%	84,45%	0,97	0,03	34,54
117	6251	98,97%	84,35%	0,93	0,03	36,29
118	7246	99,54%	84,36%	0,97	0,03	37,52
119	7247	99,32%	84,34%	0,96	0,03	37,46
120	7248	98,52%	84,36%	0,91	0,03	34,75
121	7249	98,86%	84,36%	0,93	0,03	35,81
122	7250	99,89%	84,29%	0,99	0,02	41,03
123	7251	99,66%	84,27%	0,98	0,02	41,35
124	7252	100,00%	84,25%	1,00	0,02	43,31
125	6247	98,97%	83,99%	0,94	0,03	35,28
126	6248	98,86%	83,88%	0,93	0,02	38,81
127	6249	98,97%	83,94%	0,94	0,03	36,82
128	6250	98,86%	83,87%	0,93	0,02	38,99
129	6251	98,86%	83,88%	0,93	0,02	38,52
130	7246	99,09%	83,86%	0,94	0,02	39,91
131	7247	99,66%	83,85%	0,98	0,02	41,96
132	7248	99,32%	84,02%	0,96	0,03	35,11
133	7249	99,43%	84,06%	0,96	0,03	34,26
134	7250	99,32%	83,93%	0,96	0,03	37,81
135	7251	99,54%	83,94%	0,97	0,03	38,15
136	7252	99,66%	83,91%	0,98	0,02	39,50

6.6. Matura z języka polskiego 2014

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd stand.	Z
		Obserwowana	Oczekiwana			
1	6512	99,23%	82,66%	0,96	0,03	31,05
2	6513	99,61%	82,50%	0,98	0,03	35,35
3	6514	99,10%	82,47%	0,95	0,03	34,99
4	6515	98,71%	82,39%	0,93	0,03	36,66
5	6516	98,97%	82,48%	0,94	0,03	34,37
6	6517	99,48%	82,59%	0,97	0,03	32,95
7	6518	99,48%	82,61%	0,97	0,03	32,46
8	6523	99,23%	82,52%	0,96	0,03	34,09
9	6524	98,84%	82,34%	0,93	0,02	38,37
10	6525	98,45%	82,34%	0,91	0,02	37,68
11	7255	99,36%	93,17%	0,91	0,02	46,47
12	7256	99,23%	93,20%	0,89	0,02	37,00
13	7257	99,23%	93,20%	0,89	0,02	35,82
14	7258	99,36%	93,19%	0,91	0,02	39,09
15	7259	99,36%	93,19%	0,91	0,02	39,23
16	7260	99,61%	93,27%	0,94	0,03	27,47
17	6571	98,45%	88,52%	0,87	0,02	44,85
18	6572	99,36%	88,63%	0,94	0,03	36,56
19	6573	99,36%	88,65%	0,94	0,03	35,06
20	6574	99,10%	88,60%	0,92	0,02	38,16
21	6575	99,23%	88,60%	0,93	0,02	38,70
22	6576	99,87%	88,84%	0,99	0,04	27,78
23	6532	99,87%	82,09%	0,99	0,03	35,17
24	6533	99,87%	82,12%	0,99	0,03	34,45
25	6535	99,87%	81,91%	0,99	0,02	40,33
26	6536	99,87%	82,17%	0,99	0,03	33,35
27	6537	99,61%	81,92%	0,98	0,02	39,23
28	6538	99,61%	81,91%	0,98	0,02	39,48
29	6541	99,74%	81,86%	0,99	0,02	41,61
30	6542	100,00%	82,10%	1,00	0,03	35,16
31	6543	99,74%	82,14%	0,99	0,03	33,74
32	6544	99,87%	93,92%	0,98	0,02	48,47
33	7267	99,74%	93,95%	0,96	0,03	36,54
34	7268	100,00%	93,96%	1,00	0,03	35,82
35	7269	99,87%	93,93%	0,98	0,02	41,20
36	7270	99,74%	93,94%	0,96	0,02	39,96
37	7271	100,00%	93,98%	1,00	0,03	30,54
38	7272	99,87%	87,33%	0,99	0,02	50,44
39	6603	99,74%	87,54%	0,98	0,03	33,38
40	6604	100,00%	87,54%	1,00	0,03	33,98

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd stand.	Z
		Obserwowana	Oczekiwana			
41	6605	99,48%	87,42%	0,96	0,02	39,16
42	6606	99,87%	87,42%	0,99	0,02	40,12
43	6607	100,00%	87,72%	1,00	0,04	27,86
44	6552	99,87%	81,23%	0,99	0,03	34,74
45	6553	99,74%	81,21%	0,99	0,03	35,03
46	6554	99,74%	81,28%	0,99	0,03	33,61
47	6555	99,10%	81,18%	0,95	0,03	34,47
48	6556	98,84%	80,94%	0,94	0,02	40,45
49	6557	99,87%	81,22%	0,99	0,03	35,13
50	6560	100,00%	81,36%	1,00	0,03	32,64
51	6561	98,97%	81,01%	0,95	0,02	38,44
52	6562	100,00%	81,05%	1,00	0,03	39,36
53	6563	99,87%	81,12%	0,99	0,03	37,34
54	6565	99,87%	97,19%	0,95	0,02	42,96
55	6566	100,00%	97,19%	1,00	0,02	41,47
56	6567	100,00%	97,19%	1,00	0,02	40,91
57	6568	99,87%	97,19%	0,95	0,03	38,03
58	6569	100,00%	97,19%	1,00	0,03	39,36
59	6570	100,00%	97,21%	1,00	0,04	27,86
60	7194	99,61%	83,30%	0,98	0,02	50,62
61	7195	99,61%	83,49%	0,98	0,02	39,56
62	7196	100,00%	83,60%	1,00	0,03	36,49
63	7197	99,87%	83,45%	0,99	0,02	42,00
64	7198	99,87%	83,47%	0,99	0,02	41,13
65	7199	100,00%	83,95%	1,00	0,03	28,87
66	6583	99,74%	82,70%	0,99	0,03	31,28
67	6586	99,61%	82,33%	0,98	0,02	40,85
68	6587	99,61%	82,38%	0,98	0,03	38,95
69	6588	99,74%	82,39%	0,99	0,03	38,71
70	6589	99,23%	82,39%	0,96	0,03	37,64
71	6590	99,87%	82,56%	0,99	0,03	34,36
72	6592	99,61%	82,57%	0,98	0,03	33,59
73	6594	99,87%	82,53%	0,99	0,03	35,02
74	6596	99,61%	82,46%	0,98	0,03	36,43
75	6597	99,87%	93,42%	0,98	0,02	48,17
76	6598	100,00%	93,45%	1,00	0,03	38,82
77	6599	99,87%	93,45%	0,98	0,03	38,86
78	6600	99,87%	93,44%	0,98	0,02	41,47
79	6601	100,00%	93,44%	1,00	0,02	42,02
80	6602	100,00%	93,52%	1,00	0,03	28,69
81	7273	98,84%	88,28%	0,90	0,02	46,64
82	7274	99,87%	88,43%	0,99	0,03	35,38

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd stand.	Z
		Obserwowana	Oczekiwana			
83	7275	99,61%	88,43%	0,97	0,03	34,74
84	7276	99,36%	88,39%	0,94	0,03	36,71
85	7277	99,87%	88,37%	0,99	0,02	40,22
86	7278	99,87%	88,61%	0,99	0,04	27,78
87	7174	98,84%	81,73%	0,94	0,03	36,61
88	7177	100,00%	82,19%	1,00	0,03	29,72
89	7178	100,00%	81,86%	1,00	0,03	35,66
90	7179	99,61%	81,86%	0,98	0,03	34,78
91	7180	99,48%	81,99%	0,97	0,03	32,05
92	7181	99,36%	81,74%	0,96	0,03	37,51
93	7183	98,84%	81,75%	0,94	0,03	35,92
94	7185	99,36%	81,78%	0,96	0,03	36,32
95	7186	99,48%	81,85%	0,97	0,03	34,98
96	7187	99,23%	81,76%	0,96	0,03	36,74
97	7188	99,87%	96,68%	0,96	0,02	47,57
98	7189	99,74%	96,69%	0,92	0,02	37,03
99	7190	99,87%	96,69%	0,96	0,03	37,90
100	7191	99,87%	96,68%	0,96	0,02	42,27
101	7192	99,87%	96,69%	0,96	0,02	38,77
102	7193	100,00%	96,71%	1,00	0,04	27,86
103	7261	98,97%	84,47%	0,93	0,02	48,84
104	7262	99,61%	84,69%	0,97	0,03	37,15
105	7263	99,10%	84,68%	0,94	0,03	36,24
106	7264	99,74%	84,63%	0,98	0,02	40,12
107	7265	99,23%	84,63%	0,95	0,02	38,88
108	7266	99,74%	85,06%	0,98	0,04	27,74
109	6512	100,00%	80,94%	1,00	0,03	32,64
110	6513	99,48%	80,81%	0,97	0,03	33,84
111	6514	99,74%	80,73%	0,99	0,03	35,87
112	6515	99,87%	80,63%	0,99	0,03	38,57
113	6516	100,00%	80,84%	1,00	0,03	34,30
114	6517	99,87%	80,88%	0,99	0,03	33,36
115	6518	100,00%	80,99%	1,00	0,03	31,78
116	6523	99,87%	80,79%	0,99	0,03	34,99
117	6524	99,74%	80,57%	0,99	0,02	40,14
118	6525	99,74%	80,57%	0,99	0,02	40,11
119	7267	99,87%	93,17%	0,98	0,02	49,82
120	7268	99,87%	93,21%	0,98	0,03	36,76
121	7269	99,87%	93,25%	0,98	0,03	30,21
122	7270	99,74%	93,21%	0,96	0,03	37,08
123	7271	99,87%	93,21%	0,98	0,03	38,05
124	7272	100,00%	93,28%	1,00	0,04	27,86

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd stand.	Z
		Obserwowana	Oczekiwana			
125	7273	99,23%	86,61%	0,94	0,02	48,26
126	7274	99,61%	86,79%	0,97	0,03	35,49
127	7275	99,48%	86,87%	0,96	0,03	31,90
128	7276	99,36%	86,78%	0,95	0,03	35,37
129	7277	99,36%	86,72%	0,95	0,02	38,36
130	7278	100,00%	87,03%	1,00	0,04	28,27
131	6532	99,48%	80,72%	0,97	0,03	35,63
132	6533	99,87%	80,81%	0,99	0,03	34,50
133	6535	99,87%	80,53%	0,99	0,02	41,70
134	6536	99,87%	80,90%	0,99	0,03	33,03
135	6537	99,48%	80,67%	0,97	0,03	37,01
136	6538	99,36%	80,56%	0,97	0,02	39,50
137	6541	99,10%	80,51%	0,95	0,02	40,41
138	6542	98,84%	80,66%	0,94	0,03	35,89
139	6543	99,74%	80,74%	0,99	0,03	35,86
140	6544	100,00%	97,70%	1,00	0,02	45,95
141	6565	100,00%	97,70%	1,00	0,02	40,26
142	6566	100,00%	97,70%	1,00	0,03	38,11
143	6567	100,00%	97,70%	1,00	0,02	44,83
144	6568	100,00%	97,70%	1,00	0,02	42,82
145	6569	100,00%	97,70%	1,00	0,03	32,41
146	6570	98,32%	82,38%	0,90	0,02	46,79
147	6571	99,74%	82,78%	0,99	0,03	34,11
148	6572	99,87%	82,87%	0,99	0,03	32,53
149	6573	99,87%	82,64%	0,99	0,03	38,45
150	6574	99,74%	82,58%	0,99	0,02	40,17
151	6575	100,00%	83,17%	1,00	0,04	27,86
152	6551	100,00%	80,46%	1,00	0,03	33,38
153	6553	100,00%	80,43%	1,00	0,03	34,00
154	6554	99,61%	80,32%	0,98	0,03	35,39
155	6555	99,74%	80,28%	0,99	0,03	36,39
156	6556	100,00%	80,04%	1,00	0,02	43,27
157	6557	99,74%	80,26%	0,99	0,03	36,88
158	6560	100,00%	80,47%	1,00	0,03	33,27
159	6561	99,74%	80,11%	0,99	0,02	40,61
160	6562	100,00%	80,14%	1,00	0,02	40,14
161	6563	100,00%	80,28%	1,00	0,03	36,79
162	6597	99,87%	93,67%	0,98	0,02	46,94
163	6598	99,87%	93,70%	0,98	0,03	37,42
164	6599	99,74%	93,71%	0,96	0,03	34,69
165	6600	99,74%	93,70%	0,96	0,03	36,72
166	6601	100,00%	93,70%	1,00	0,03	38,38

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd stand.	Z
		Obserwowana	Oczekiwana			
167	6602	100,00%	93,76%	1,00	0,04	27,86
168	6603	99,74%	85,65%	0,98	0,02	50,99
169	6604	99,74%	85,79%	0,98	0,02	40,01
170	6605	100,00%	85,88%	1,00	0,03	36,22
171	6606	99,48%	85,77%	0,96	0,02	40,28
172	6607	99,87%	85,76%	0,99	0,02	41,83
173	6608	100,00%	86,15%	1,00	0,04	28,25
174	6583	99,87%	81,44%	0,99	0,03	31,14
175	6586	98,71%	80,98%	0,93	0,02	38,88
176	6587	98,71%	80,99%	0,93	0,02	38,52
177	6588	99,36%	81,03%	0,97	0,02	38,81
178	6589	98,84%	81,02%	0,94	0,02	37,79
179	6590	99,10%	81,20%	0,95	0,03	33,94
180	6592	99,74%	81,20%	0,99	0,03	35,35
181	6594	98,84%	81,13%	0,94	0,03	34,95
182	6596	99,61%	81,09%	0,98	0,03	37,60
183	7188	99,10%	94,92%	0,82	0,02	41,72
184	7189	99,61%	94,93%	0,92	0,02	39,27
185	7190	99,74%	94,93%	0,95	0,02	39,42
186	7191	99,36%	94,93%	0,87	0,02	37,36
187	7192	99,74%	94,93%	0,95	0,02	41,94
188	7193	99,74%	94,96%	0,95	0,03	29,53
189	7194	98,20%	85,41%	0,88	0,02	46,17
190	7195	99,36%	85,64%	0,96	0,03	35,08
191	7196	99,87%	85,61%	0,99	0,03	37,35
192	7197	99,48%	85,57%	0,96	0,03	38,32
193	7198	99,74%	85,54%	0,98	0,02	40,88
194	7199	99,87%	85,94%	0,99	0,04	27,79
195	7174	99,74%	81,71%	0,99	0,03	39,27
196	7177	100,00%	82,05%	1,00	0,03	31,85
197	7178	100,00%	81,86%	1,00	0,03	35,66
198	7179	100,00%	81,83%	1,00	0,03	36,32
199	7180	99,87%	81,92%	0,99	0,03	34,02
200	7181	99,87%	81,87%	0,99	0,03	35,28
201	7183	99,48%	81,75%	0,97	0,03	37,44
202	7185	99,61%	81,80%	0,98	0,03	36,24
203	7186	100,00%	81,81%	1,00	0,03	36,88
204	7187	99,61%	81,71%	0,98	0,03	38,75
205	7255	99,74%	94,42%	0,95	0,02	48,62
206	7256	99,87%	94,44%	0,98	0,02	39,33
207	7257	99,87%	94,46%	0,98	0,03	33,05
208	7258	99,74%	94,44%	0,95	0,02	38,53

Numer zadania	Numer kryterium/zadania w bazie	Zgodność		kappa	Błąd stand.	Z
		Obserwowana	Oczekiwana			
209	7259	100,00%	94,44%	1,00	0,03	39,25
210	7260	99,87%	94,48%	0,98	0,03	28,65
211	7261	98,97%	86,61%	0,92	0,02	46,85
212	7262	100,00%	86,82%	1,00	0,03	35,17
213	7263	99,48%	86,81%	0,96	0,03	34,22
214	7264	99,74%	86,70%	0,98	0,02	41,34
215	7265	99,36%	86,72%	0,95	0,02	38,92
216	7266	99,10%	86,97%	0,93	0,03	27,80

Instytut Badań Edukacyjnych

Głównym zadaniem Instytutu jest prowadzenie badań, analiz i prac przydatnych w rozwoju polityki i praktyki edukacyjnej.

Instytut zatrudnia ponad 150 badaczy zajmujących się edukacją – pedagogów, socjologów, psychologów, ekonomistów, politologów i przedstawicieli innych dyscyplin naukowych – wybitnych specjalistów w swoich dziedzinach, o różnorodnych doświadczeniach zawodowych, które obejmują, oprócz badań naukowych, także pracę dydaktyczną, doświadczenie w administracji publicznej czy działalność w organizacjach pozarządowych.

Instytut w Polsce uczestniczy w realizacji międzynarodowych projektów badawczych w tym *PIAAC*, *PISA*, *TALIS*, *ESLC*, *SHARE*, *TIMSS* i *PIRLS* oraz projektów systemowych współfinansowanych przez Unię Europejską ze środków Europejskiego Funduszu Społecznego.