



# JESIENNA SZKOŁA EWD 2015



# EDUKACYJNA WARTOŚĆ DODANA

## Jakość danych egzaminacyjnych w kontekście metody EWD



Aleksandra Jasińska-Maciążek  
([a.jasinska@ibe.edu.pl](mailto:a.jasinska@ibe.edu.pl))

Jesienna Szkoła EWD 2015

Warszawa, 1-3.10.2015



KAPITAŁ LUDZKI  
NARODOWA STRATEGIA SPÓJNOŚCI



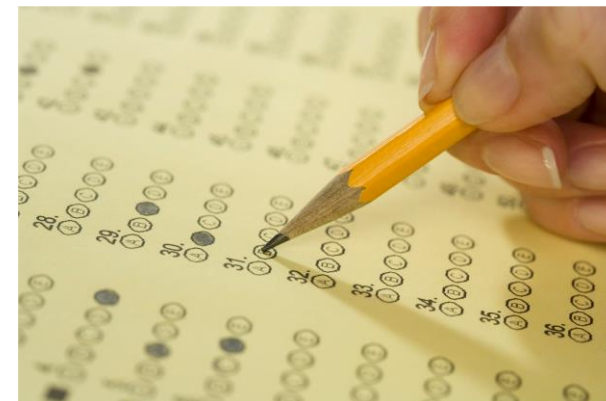
EDUKACYJNA  
WARTOŚĆ  
DODANA

UNIA EUROPEJSKA  
EUROPEJSKI  
FUNDUSZ SPOŁECZNY





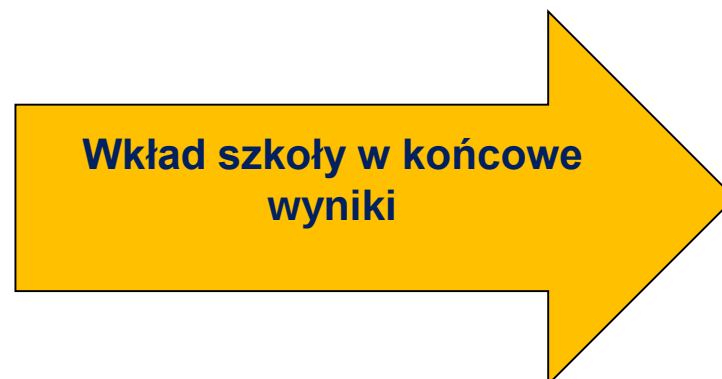
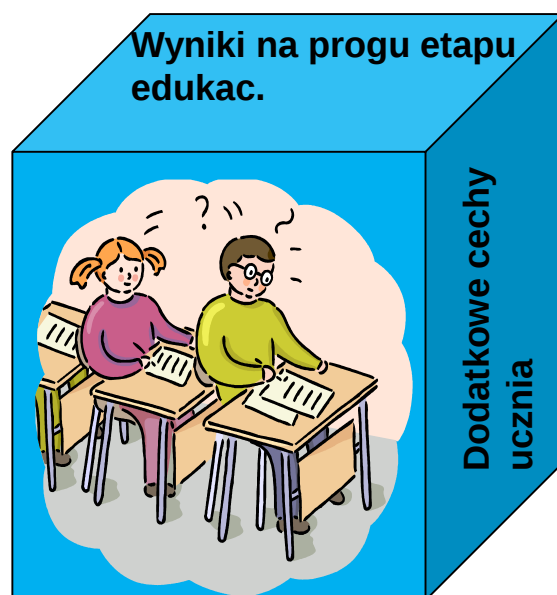
## Dlaczego o jakości danych egzaminacyjnych?



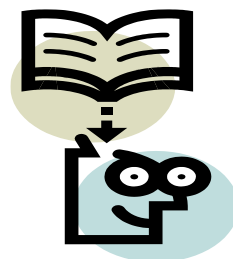
- Wyniki egzaminów zewnętrznych są podstawą liczenia wskaźników edukacyjnej wartości dodanej (EWD).
- Jakość wskaźników EWD zależy więc od tego, co i jak dobrze mierzą egzaminy.
- Jeśli pomiar osiągnięć szkolnych nie będzie wystarczająco dobry, nawet najbardziej wyrafinowane metody statystyczne nie dostarczą wartościowych miar.



# Idea EWD



Nauka w szkole



Efekty pracy



KAPITAŁ LUDZKI  
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



EDUKACYJNA  
WARTOŚĆ  
DODANA

UNIA EUROPEJSKA  
EUROPEJSKI  
FUNDUSZ SPOŁECZNY





## Jakie właściwości muszą mieć egzaminy, by można je było wykorzystać do wyliczenia trafnej miary EWD?

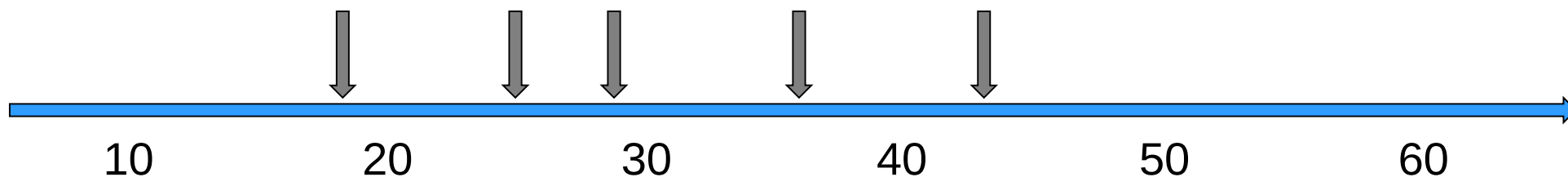
1. Egzaminy muszą być wystarczająco rzetelne, czyli muszą precyzyjnie mierzyć poziom osiągnięć szkolnych.
2. Egzaminy muszą być dobrą miarą tego, w jakim stopniu szkoła zrealizowała najważniejsze cele kształcenia.
3. Testy wykorzystane do szacowania danego wskaźnika EWD powinny mierzyć te same lub podobne umiejętności.
4. Skala, na której prezentowane są wyniki, musi mieć dobrze zdefiniowaną jednostkę pomiaru.



# Test powinien być wystarczająco rzetelny – powinien mierzyć precyzyjnie

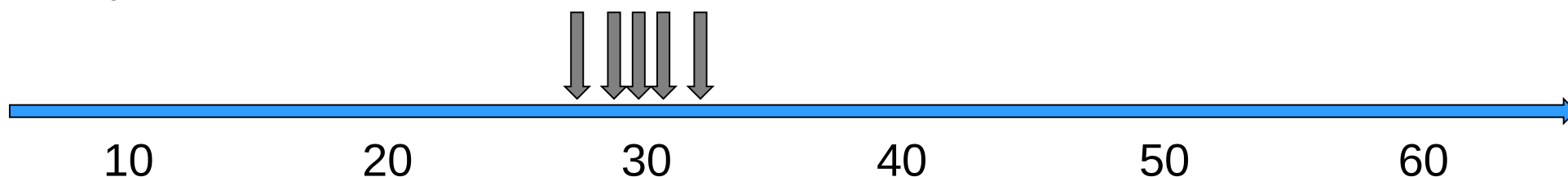
Test o niskiej rzetelności

rzetelność 0,6



Test o wysokiej rzetelności

rzetelność 0,9



KAPITAŁ LUDZKI  
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



EDUKACYJNA  
WARTOŚĆ  
DODANA

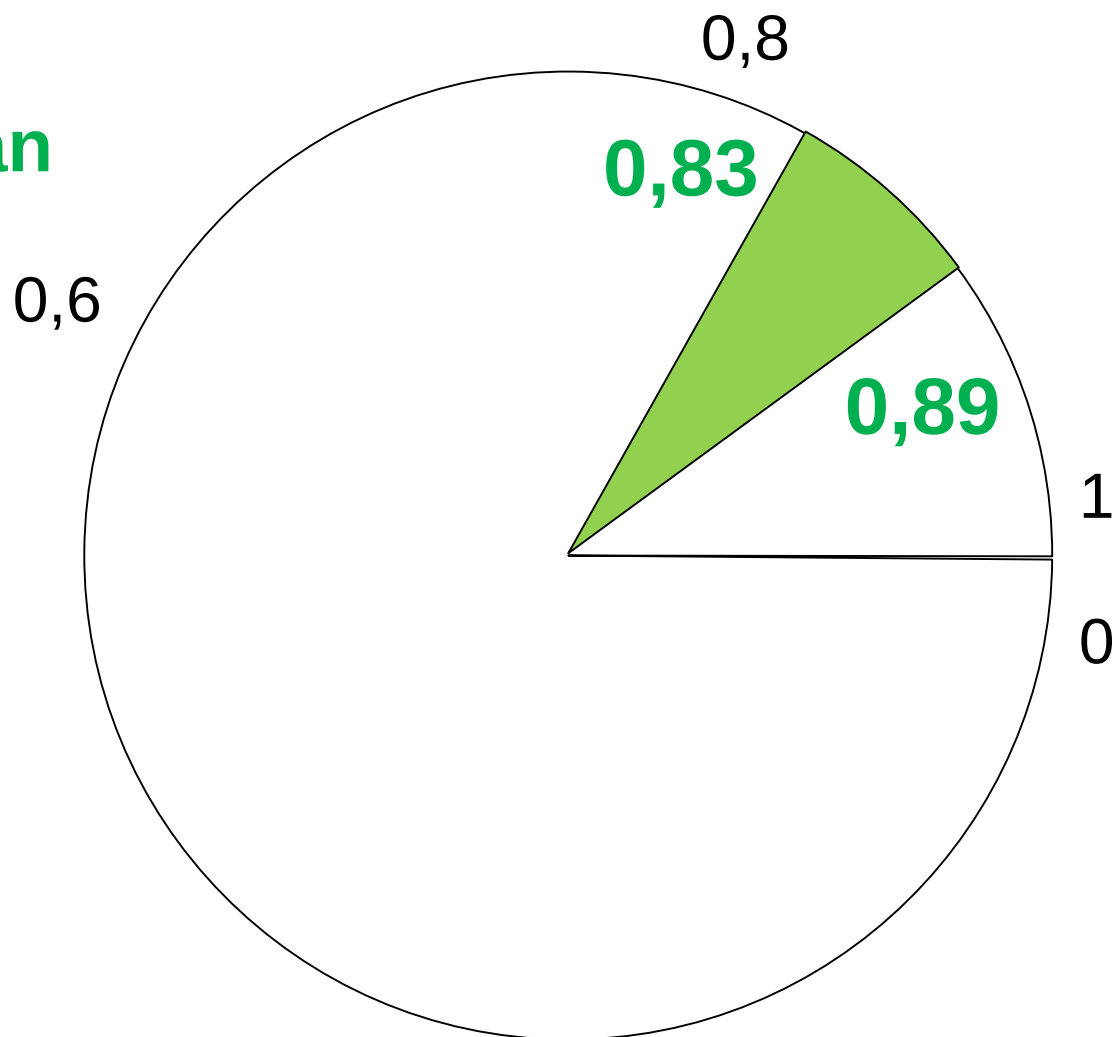
UNIA EUROPEJSKA  
EUROPEJSKI  
FUNDUSZ SPOŁECZNY





# Rzetelność egzaminów

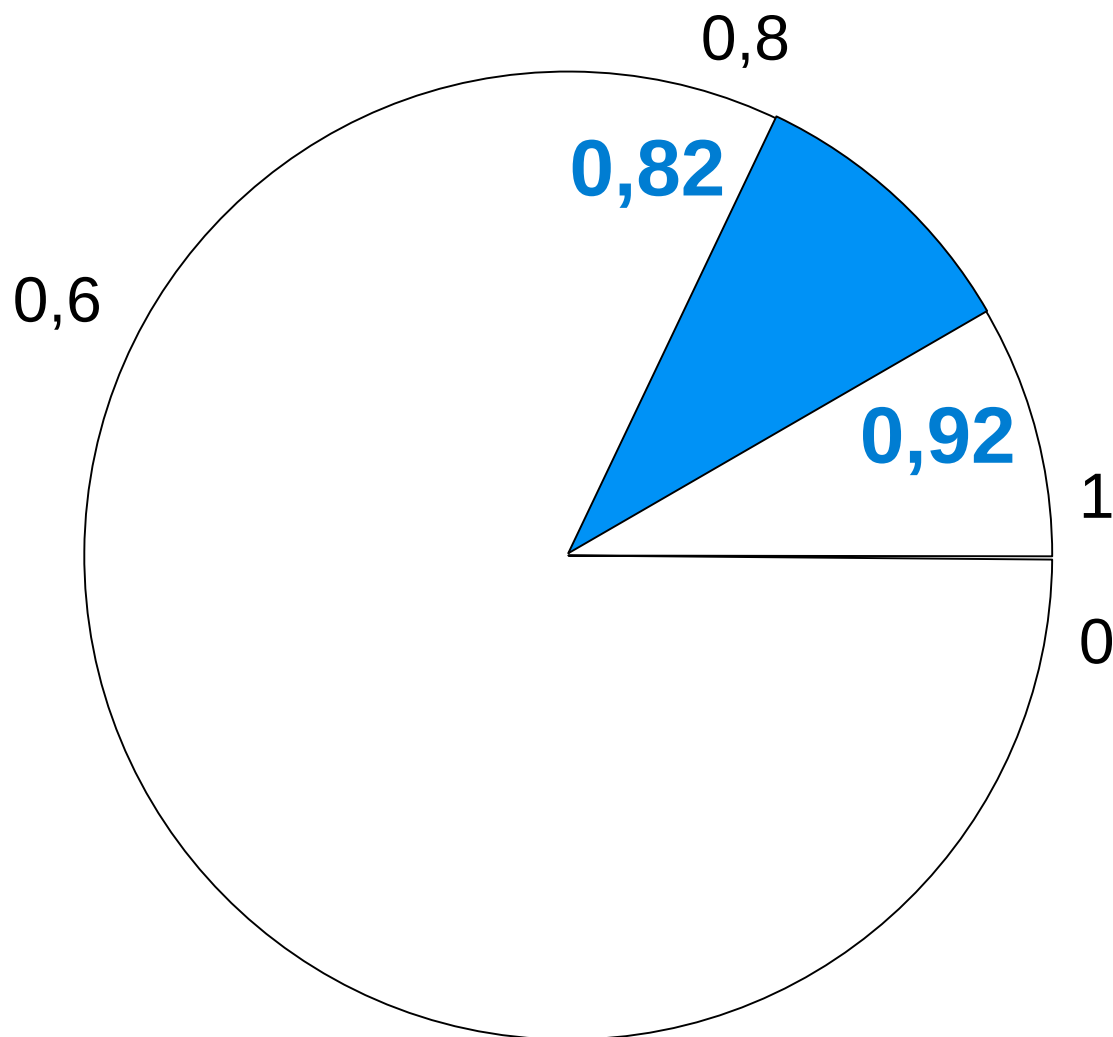
Sprawdzian





## Rzetelność egzaminów

**Egzamin  
gimnazjalny  
- części  
główne**



KAPITAŁ LUDZKI  
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



EDUKACYJNA  
WARTOŚĆ  
DODANA

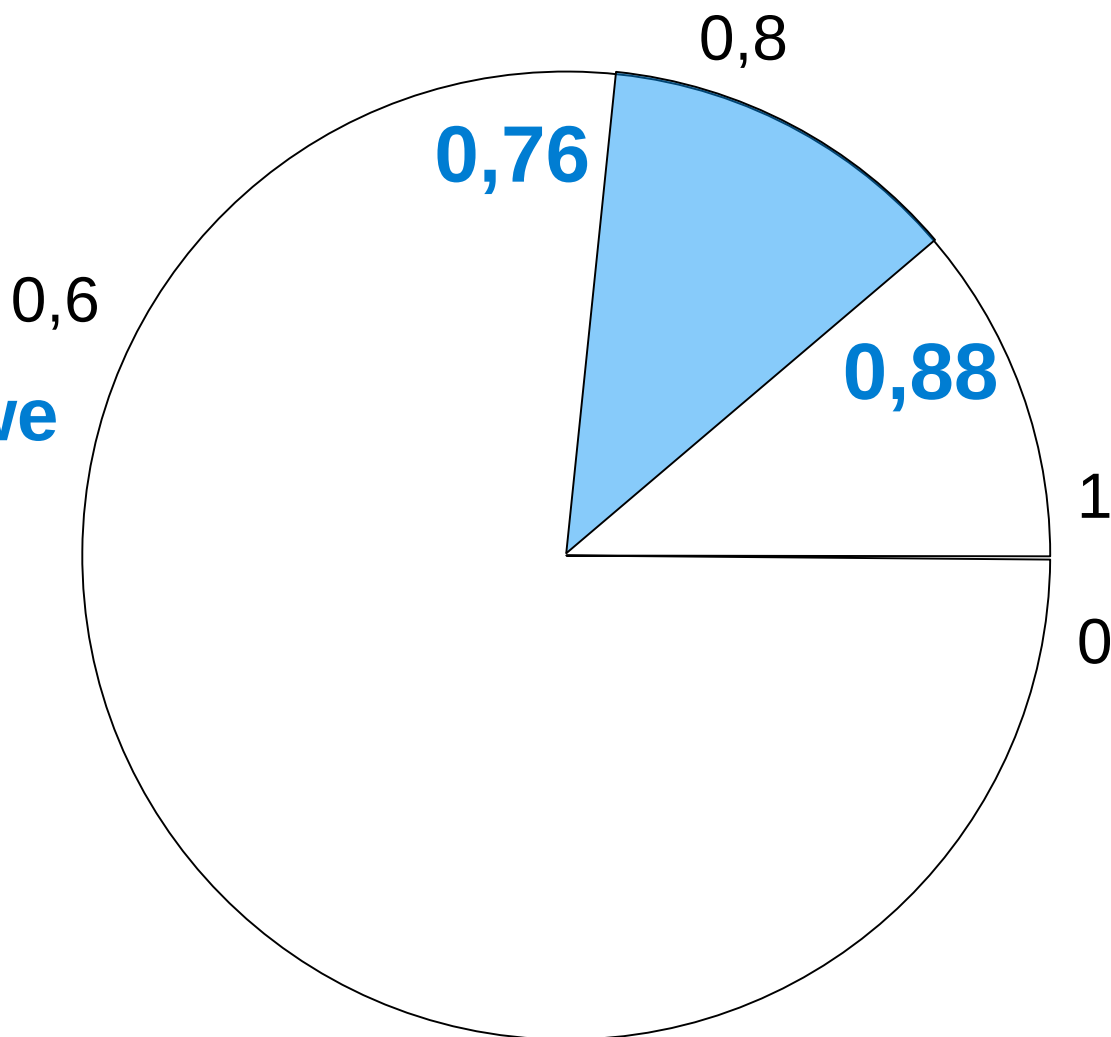
UNIA EUROPEJSKA  
EUROPEJSKI  
FUNDUSZ SPOŁECZNY





## Rzetelność egzaminów

**Egzamin  
gimnazjalny  
- części  
przedmiotowe**



KAPITAŁ LUDZKI  
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



EDUKACYJNA  
WARTOŚĆ  
DODANA

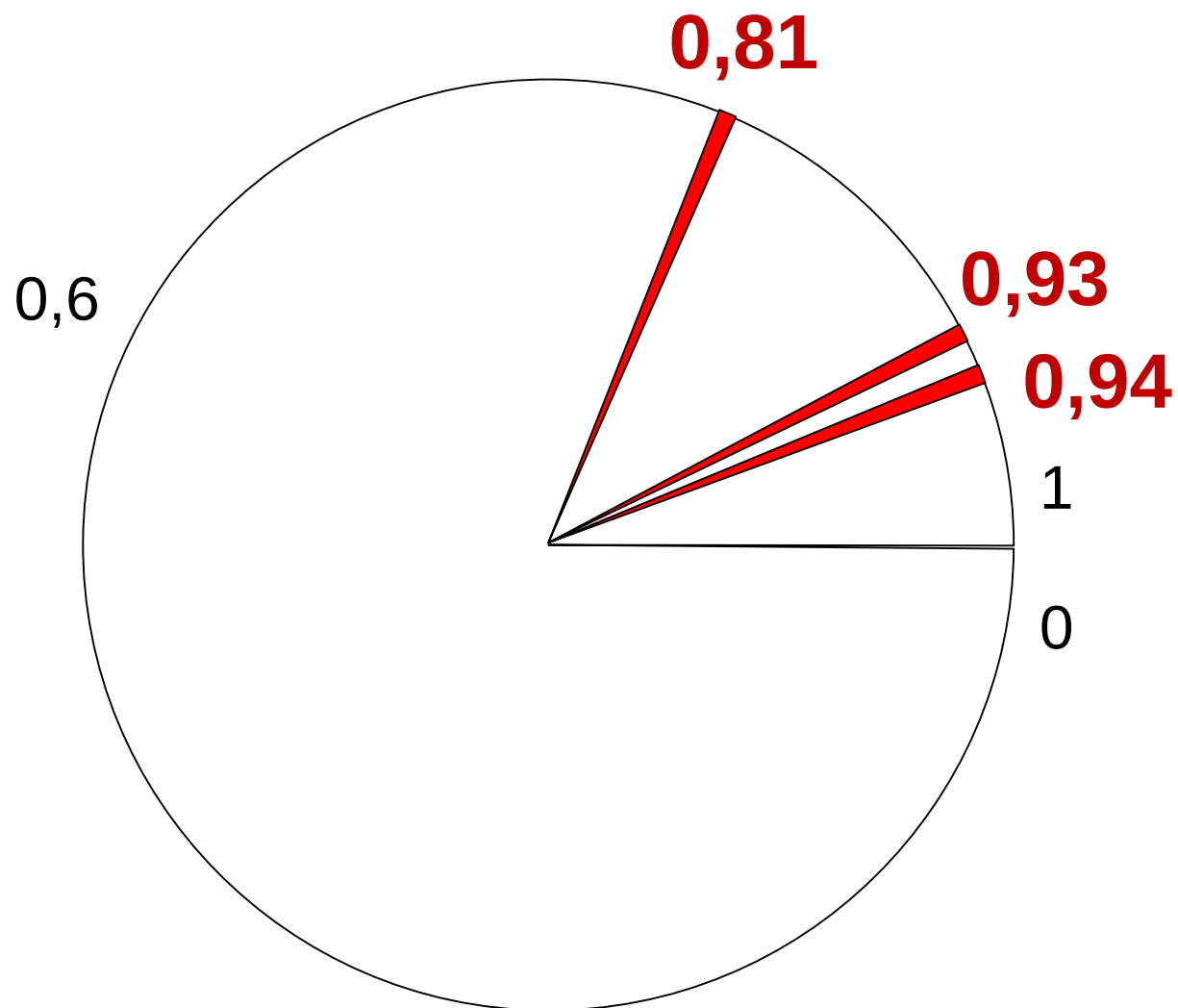
UNIA EUROPEJSKA  
EUROPEJSKI  
FUNDUSZ SPOŁECZNY





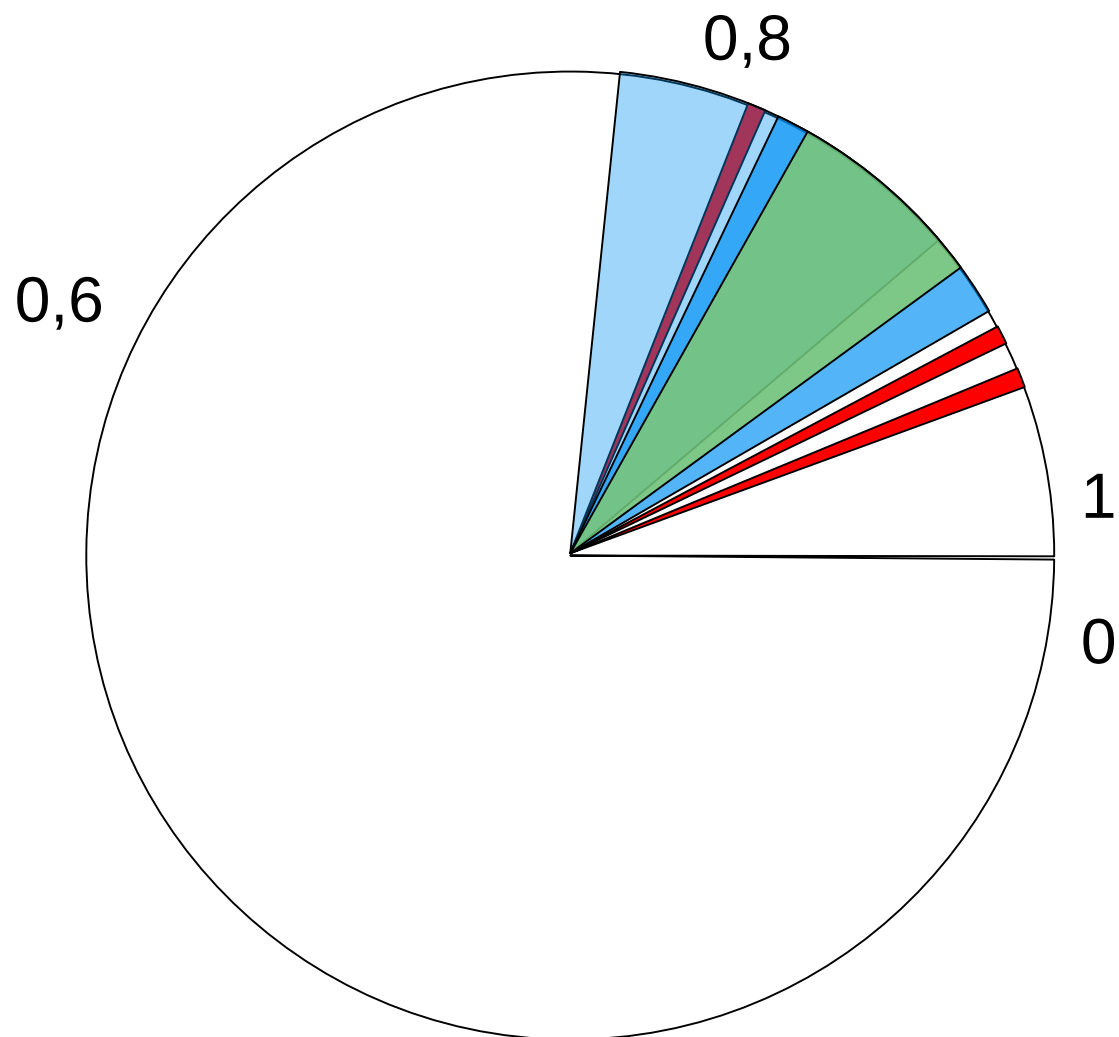
# Rzetelność egzaminów

**Matura**





# Rzetelność egzaminów





## **Egzaminy muszą być dobrą miarą tego, w jakim stopniu szkoła zrealizowała najważniejsze cele kształcenia.**

- Testy konstruowane w celu pomiaru realizacji wymagań ogólnych i szczegółowych zawartych w podstawie programowej kształcenia ogólnego.
- Na ile testy, na podstawie których liczone są wskaźniki EWD, pozwalają mierzyć coś więcej niż tylko skuteczność rozwiązywania danego zestawu zadań?
- Na ile wyniki egzaminów są powiązane z szerzej ujętymi efektami kształcenia i czy dzięki temu mogą stanowić podstawę bardziej kompleksowej oceny pracy szkoły?



## Zewnętrzne kryteria trafności egzaminów gimnazjalnych

| Kryterium                               | Zależność                                                                            | Uwagi                                                                                    |
|-----------------------------------------|--------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| oceny szkolne                           |    | Silniejsze związki między tymi samymi przedmiotami.                                      |
| motywacja niska                         |    | Gr. odniesienia: motywacja o przeciętnej sile                                            |
| motywacja wysoka                        |    |                                                                                          |
| motywacja zewnętrzna                    |    |                                                                                          |
| motywacja wewnętrzna                    |    |                                                                                          |
| pozytywne nastawienie do nauki i szkoły |   | Podobne efekty dla każdego egzaminu.                                                     |
| wiara we własne możliwości uczenia się  |  |                                                                                          |
| bezradność intelektualna                |  | Silniejsze związki między tymi samymi przedmiotami.                                      |
| inteligencja                            |  | Silniejsza zależność z GMP niż GH. Wielkość efektu porównywalna z wynikami innych badań. |
| wyniki testów PISA                      |  | Silniejsze związki między tymi samymi dziedzinami.                                       |



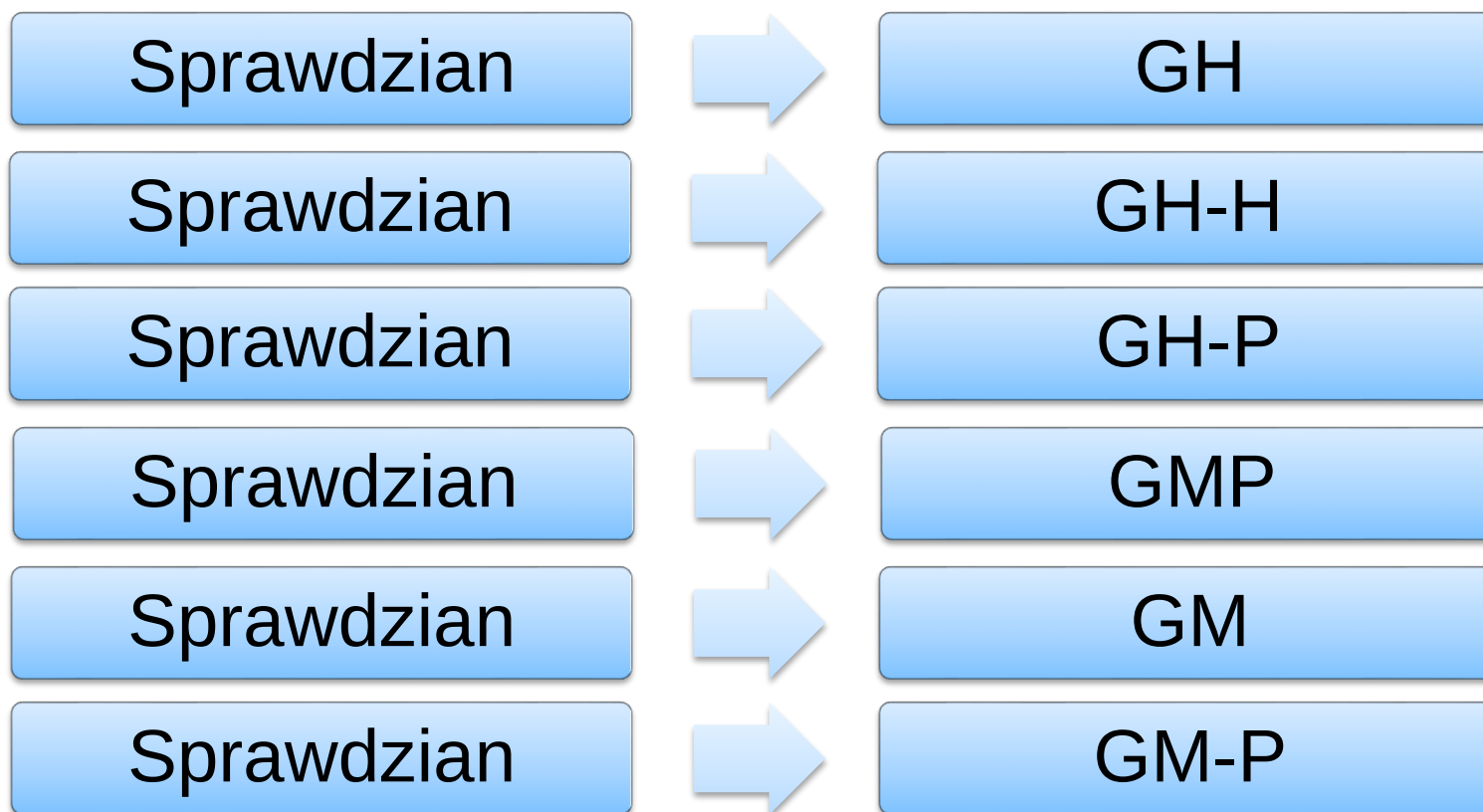
## Zewnętrzne kryteria trafności egzaminu maturalnego

| Kryterium               | Zależność                                                                          | Uwagi                                                                                                                                                                                   |
|-------------------------|------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| inteligencja            |  | Silniejsza zależność dla matematyki i wyników z przedmiotów matematyczno-przyrodniczych niż humanistycznych i z języka polskiego. Wielkość efektu porównywalna z wynikami innych badań. |
| wyniki testów PISA      |  | Silniejsze związki między odpowiadającymi sobie dziedzinami.                                                                                                                            |
| nastawienie do czytania |  | Związki tylko z wynikami z języka polskiego i przedmiotów humanistycznych.                                                                                                              |



## Testy wykorzystane do szacowania danego wskaźnika EWD powinny mierzyć te same lub podobne umiejętności.

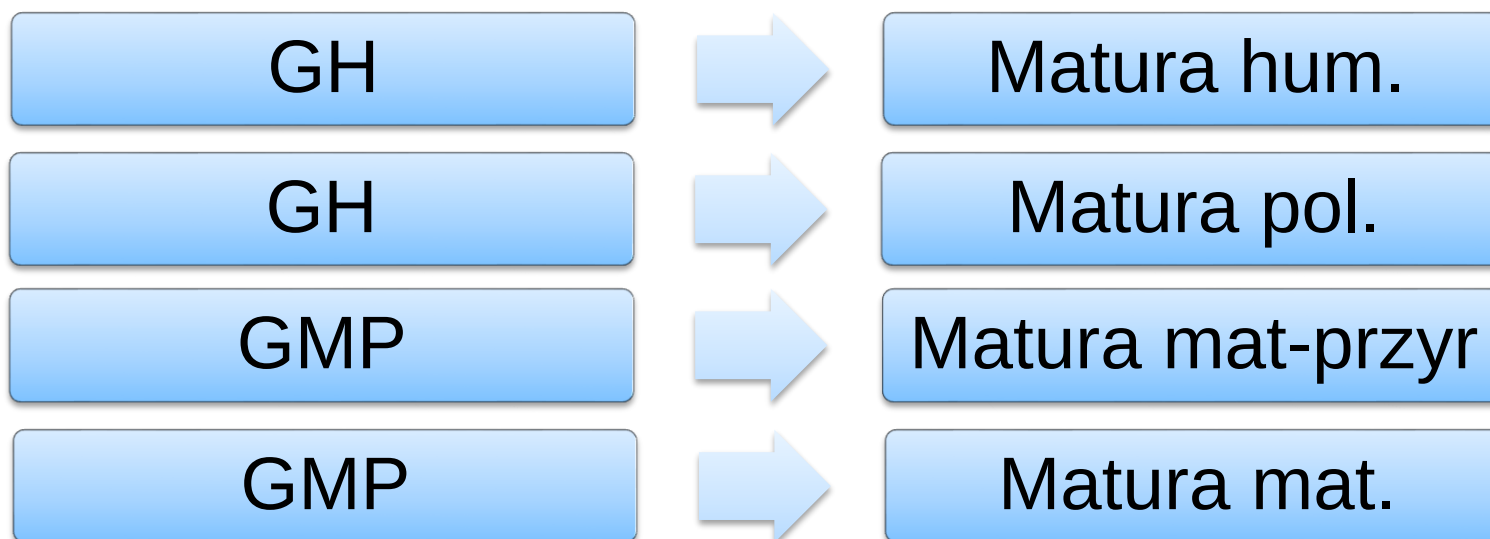
- Im dokładniej oszacujemy wynik przewidywany (na podstawie cech uczniów niezależnych od szkoły), tym miara EWD będzie lepiej odzwierciedlała rzeczywisty wkład szkoły.
- Najlepszą zmienną pozwalającą przewidzieć wynik na egzaminie z danego przedmiotu wydaje się być uprzedni poziom umiejętności z tego przedmiotu.





## Testy wykorzystane do szacowania danego wskaźnika EWD powinny mierzyć te same lub podobne umiejętności.

- Im dokładniej oszacujemy wynik przewidywany (na podstawie cech uczniów niezależnych od szkoły), tym miara EWD będzie lepiej odzwierciedlała rzeczywisty wkład szkoły.
- Najlepszą zmienną pozwalającą przewidzieć wynik na egzaminie z danego przedmiotu wydaje się być uprzedni poziom umiejętności z tego przedmiotu.





# Podobieństwo mierzonych umiejętności

Korelacje między wynikami „na wejściu” i „na wyjściu”

Sprawdzian 2009 – Egzamin gimnazjalny 2012

Egzamin gimnazjalny 2008 – Matura 2011

|      | SPR   | M (hum) | M (pol) | M (mp) | M (mat) |
|------|-------|---------|---------|--------|---------|
| GMP  | 0,782 |         |         | 0,842  | 0,822   |
| GM-M | 0,755 |         |         |        |         |
| GM-P | 0,719 |         |         |        |         |
| GH   | 0,769 | 0,591   | 0,547   |        |         |
| GH-P | 0,720 |         |         |        |         |
| GH-H | 0,707 |         |         |        |         |



**Skala, na której prezentowane są wyniki, musi mieć dobrze zdefiniowaną jednostkę pomiaru.**

Czyli jaka musi być skala wyników?

- ☐ Umożliwiająca porównywanie między latami
- ☐ O wystarczającej precyzji pomiaru
- ☐ O interwałowym charakterze
- ☐ Posiadająca interpretację ilościową



Czy skala wyników surowych (suma punktów, % punktów) spełnia te wymagania?



# Pomiar wzrostu za pomocą kwestionariusza

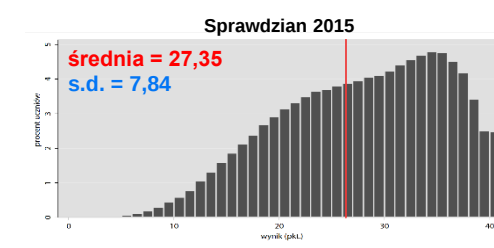
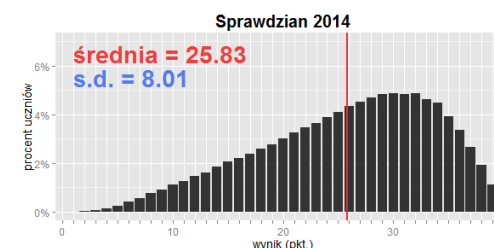
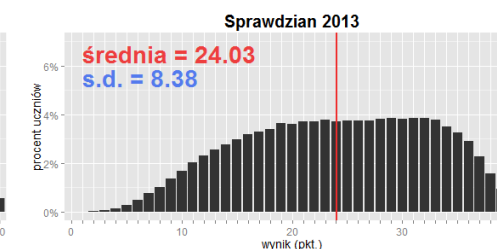
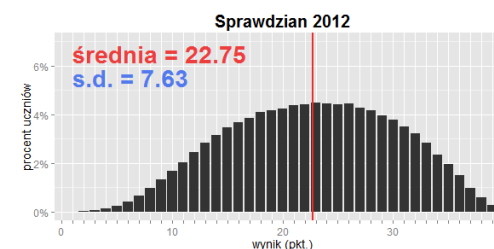
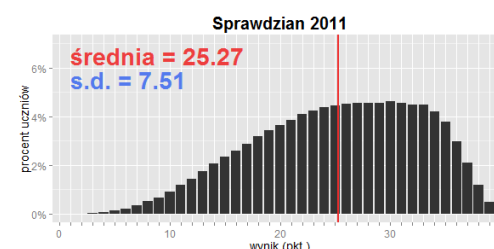
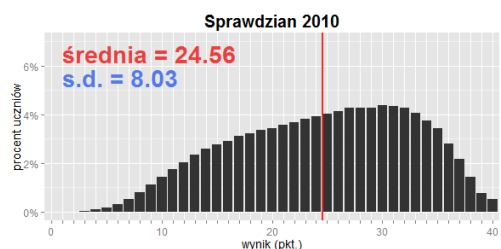
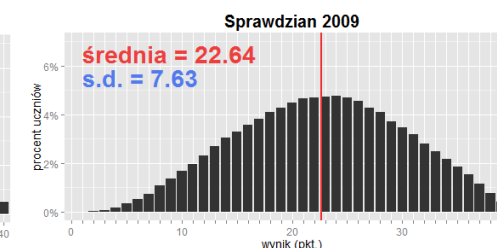
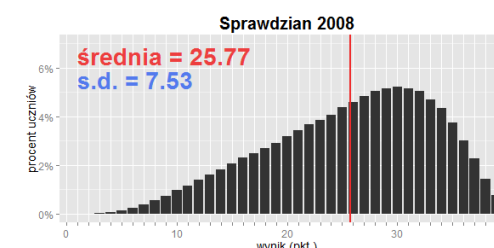
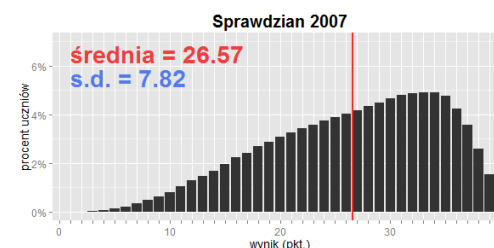
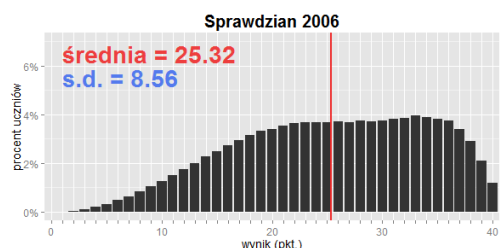
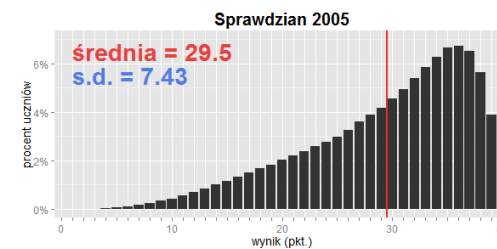
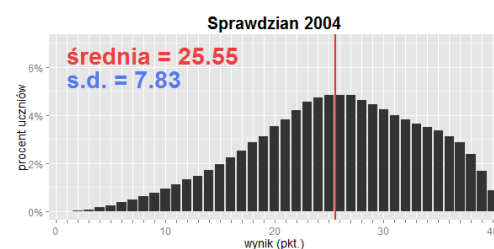
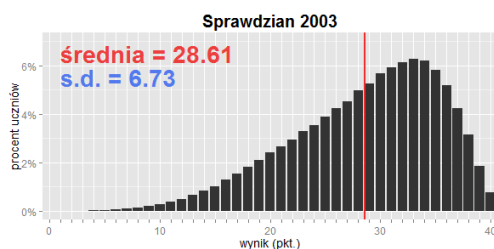
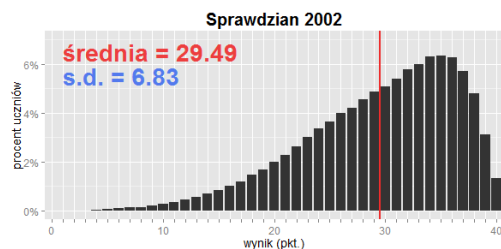
| Jak często zdarzają/zdarzały Ci się następujące sytuacje?                    | bardzo rzadko | rzadko | często | bardzo często |
|------------------------------------------------------------------------------|---------------|--------|--------|---------------|
| 1. Uderzam się w głowę o framugę drzwi.                                      | 1             | 2      | 3      | 4             |
| 2. Na zdjęciach klasowych zawsze byłem ustawiany w pierwszym rzędzie.        | 1             | 2      | 3      | 4             |
| 3. Kiedy leżę w łóżku, marzną mi stopy.                                      | 1             | 2      | 3      | 4             |
| 4. Wchodząc po schodach przeskakuję o dwa schodki.                           | 1             | 2      | 3      | 4             |
| 5. W szkole chętnie byłem wybierany do drużyny, kiedy graliśmy w koszykówkę. | 1             | 2      | 3      | 4             |
| 6. W samolotach mam wystarczająco dużo miejsca, by siedzieć komfortowo.      | 1             | 2      | 3      | 4             |
| 7. Kiedy rozmawiam ze znajomymi, spoglądam na nich z góry.                   | 1             | 2      | 3      | 4             |
| 8. itd...                                                                    | 1             | 2      | 3      | 4             |

Źródło: Cess Glas





# Wyniki Sprawdzianu w klasie szóstej 2002-2015

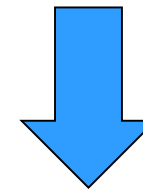
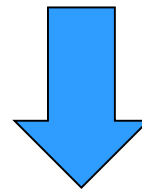
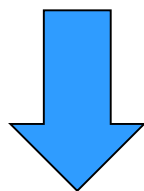
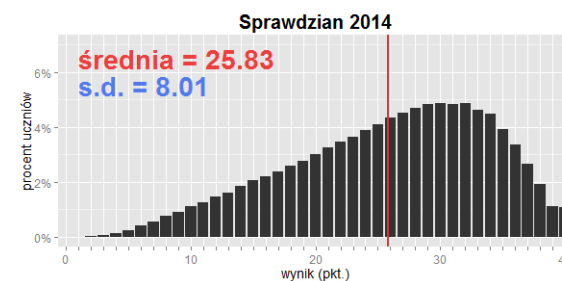
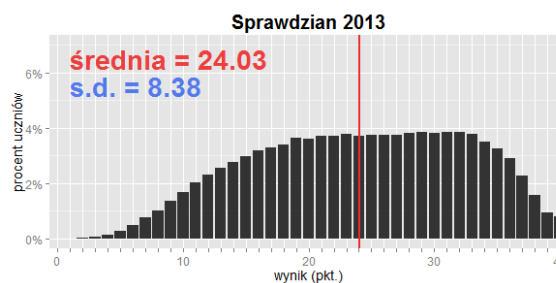
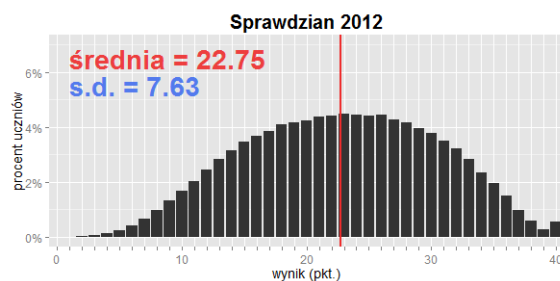




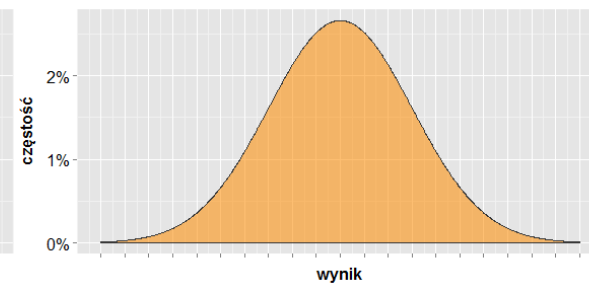
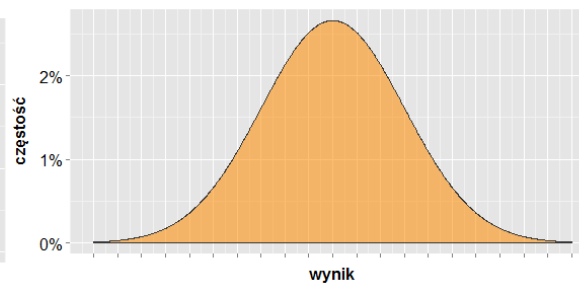
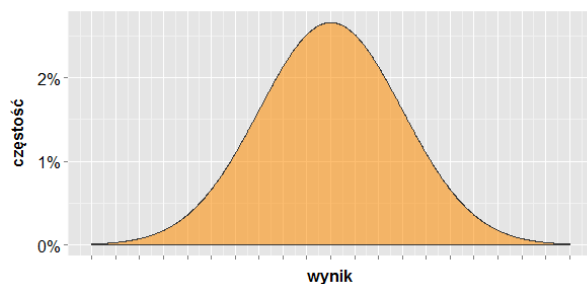
# Zakładany kształt rozkładu wyników testu...

... a faktyczny rozkład wyników.

wyniki surowe



wyniki na skali standardowej



KAPITAŁ LUDZKI  
NARODOWA STRATEGIA SPÓJNOŚCI



EDUKACYJNA  
WARTOŚĆ  
DODANA

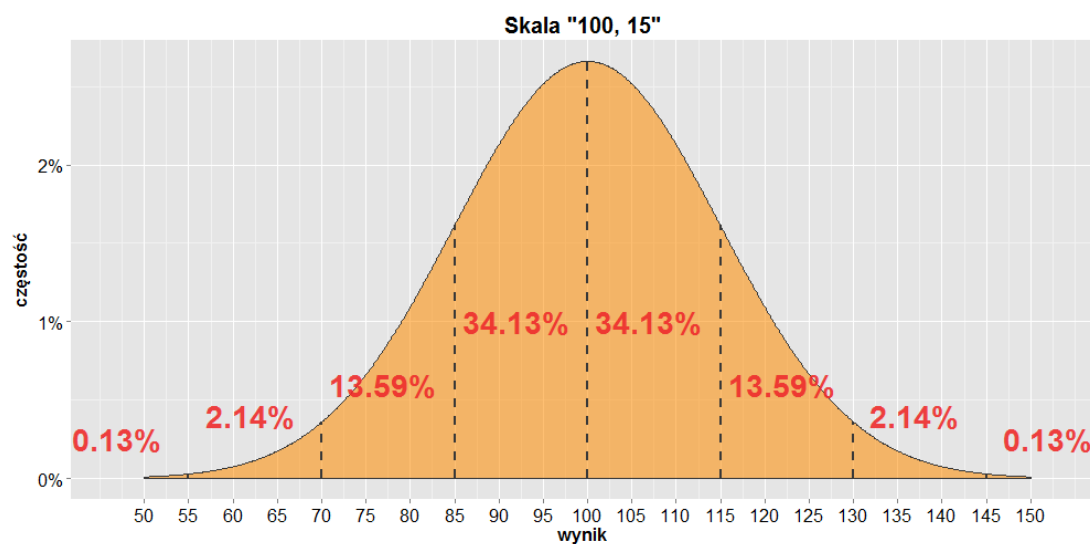
UNIA EUROPEJSKA  
EUROPEJSKI  
FUNDUSZ SPOŁECZNY



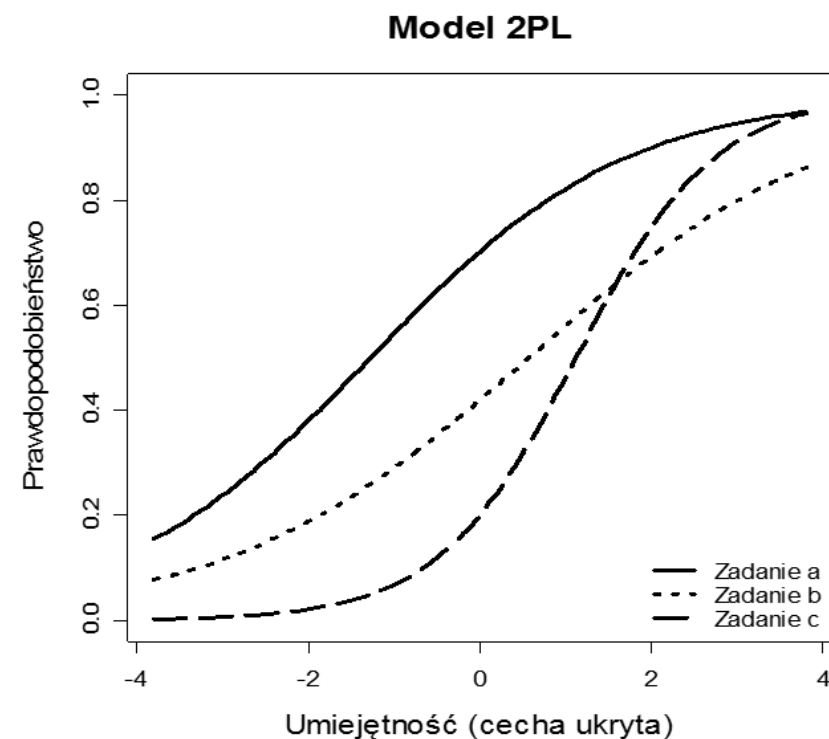


# Dwie wykorzystywane przez nas metody skalowania

## Normalizacja

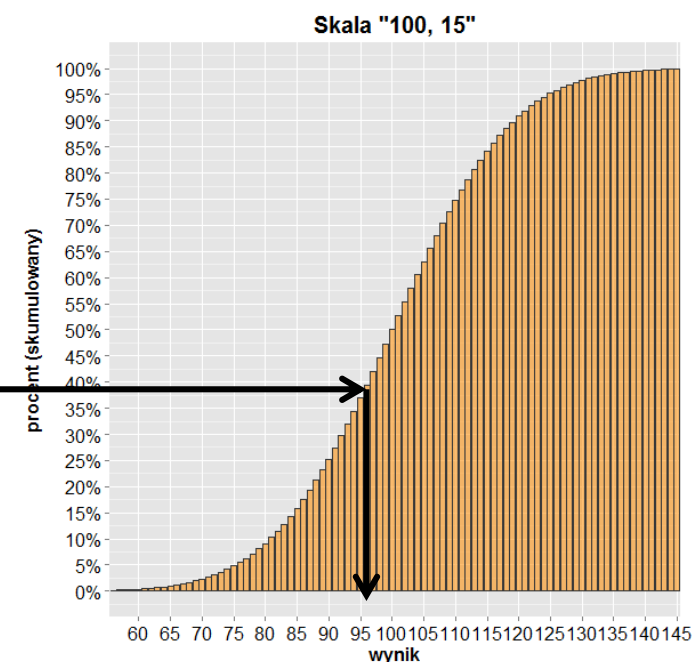
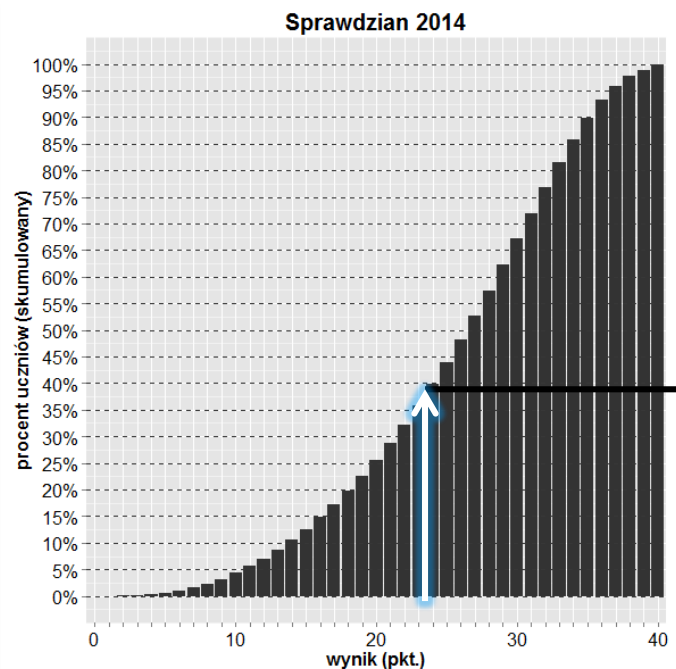
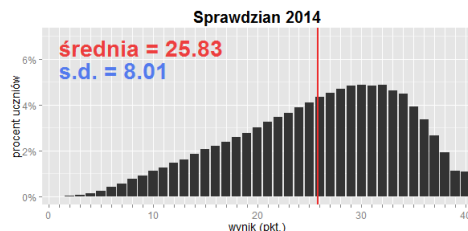


## Modele IRT





# Normalizacja wyników



Bierzemy rozkład  
wyników surowych  
(procentowy)

Przekształcamy  
go na rozkład  
skumulowany

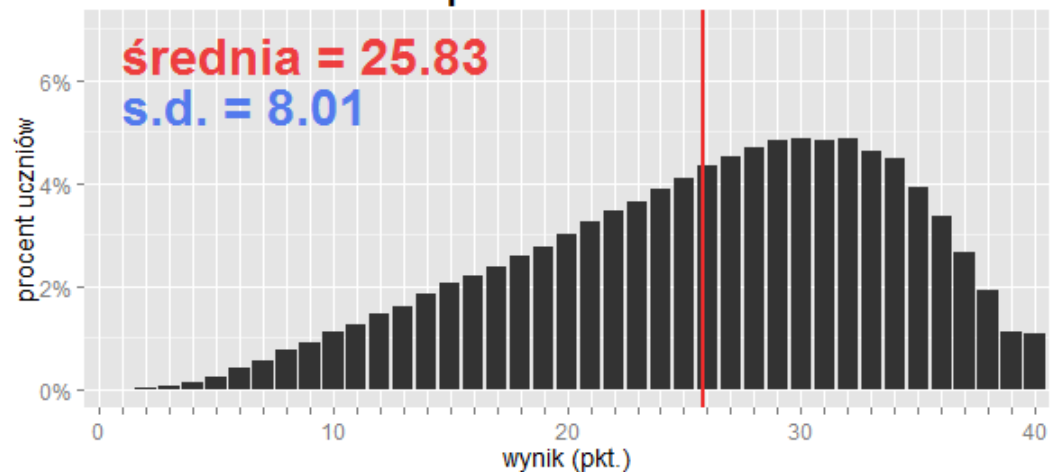
Porównujemy z rozkładem  
skumulowanym dla  
rozkładu normalnego skali  
100/15



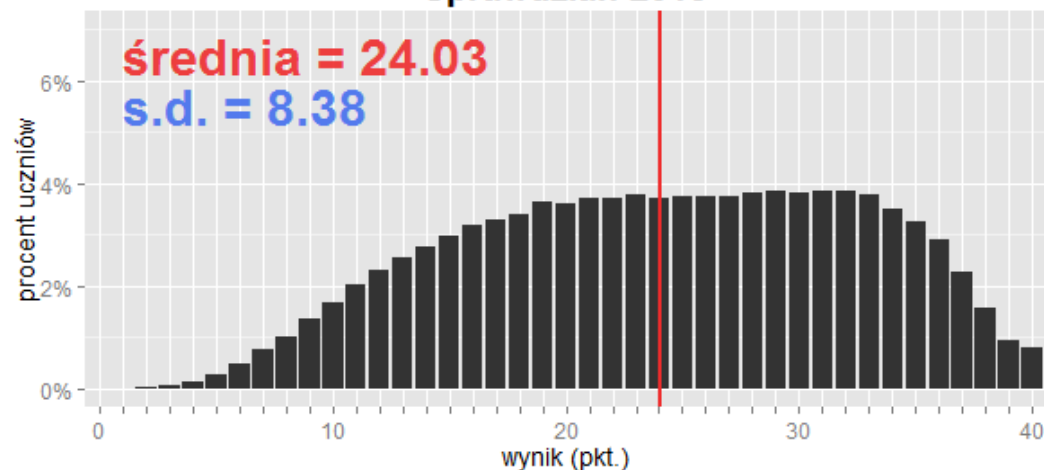


# Normalizacja wyników - przykłady

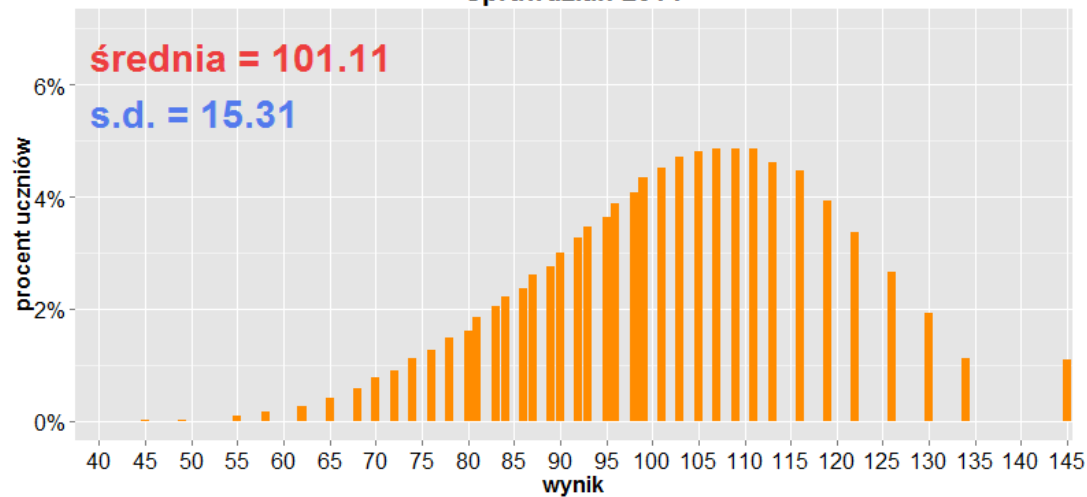
Sprawdzian 2014



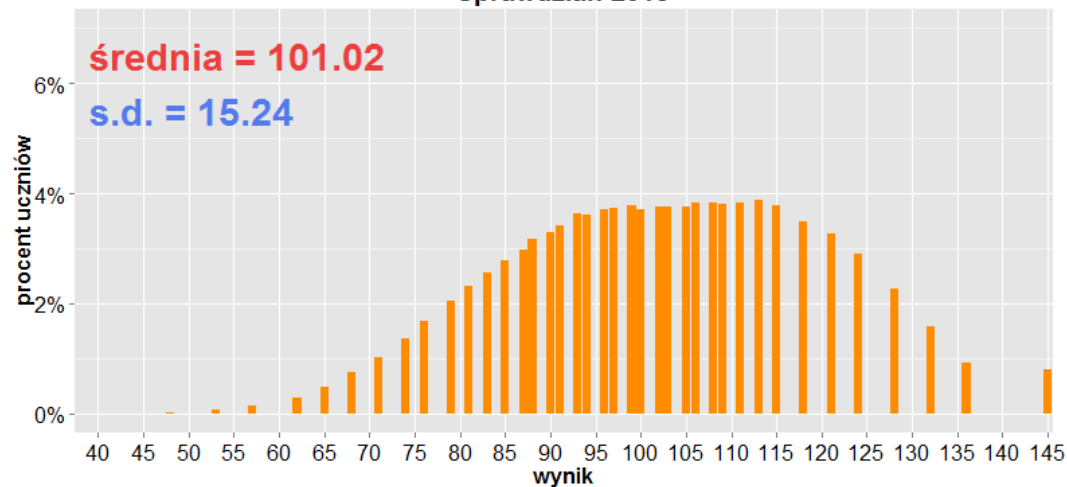
Sprawdzian 2013



Sprawdzian 2014



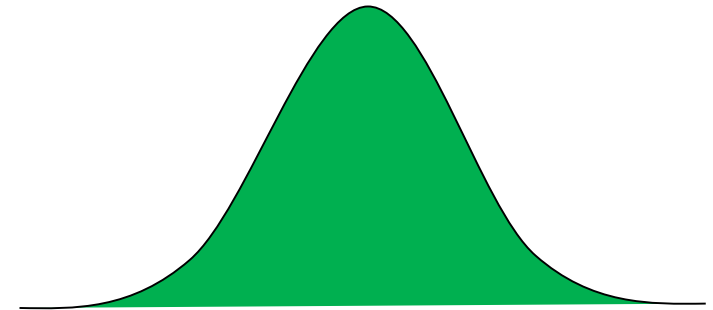
Sprawdzian 2013



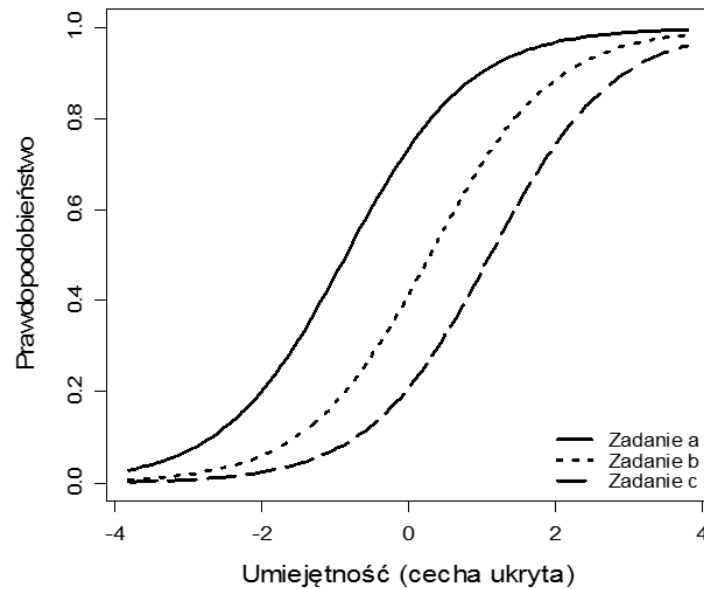


# Modele IRT

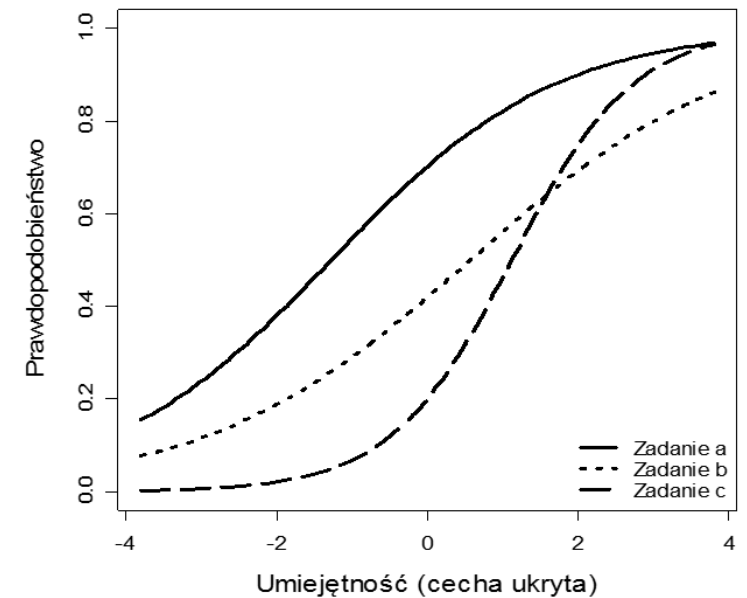
| Id  | Zad 1 | Zad 2 | Zad 3 | Zad 4 | Zad 5 | Zad 6 | Zad 7 | Zad 8 | Zad 9 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 001 | 1     | 1     | 0     | 0     | 1     | 1     | 1     | 1     | 1     |
| 002 | 1     | 0     | 0     | 1     | 0     | 1     | 2     | 3     | 1     |
| 003 | 1     | 1     | 1     | 1     | 0     | 0     | 1     | 3     | 1     |
| 004 | 0     | 0     | 1     | 1     | 0     | 1     | 0     | 1     | 1     |
| 005 | 0     | 0     | 1     | 1     | 0     | 0     | 2     | 0     | 0     |
| 006 | 1     | 1     | 0     | 0     | 1     | 2     | 2     | 4     | 2     |
| 007 | 1     | 1     | 1     | 1     | 1     | 1     | 2     | 4     | 1     |



Model 1PL (Rasch)



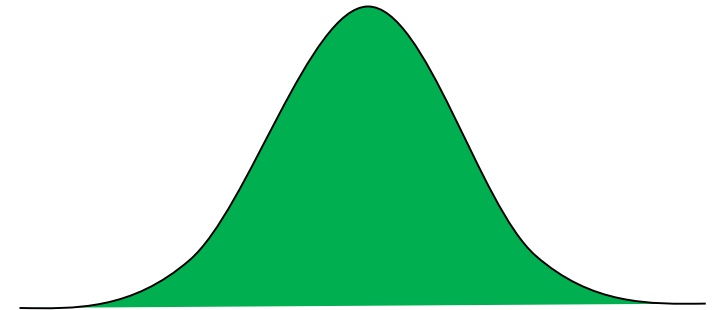
Model 2PL



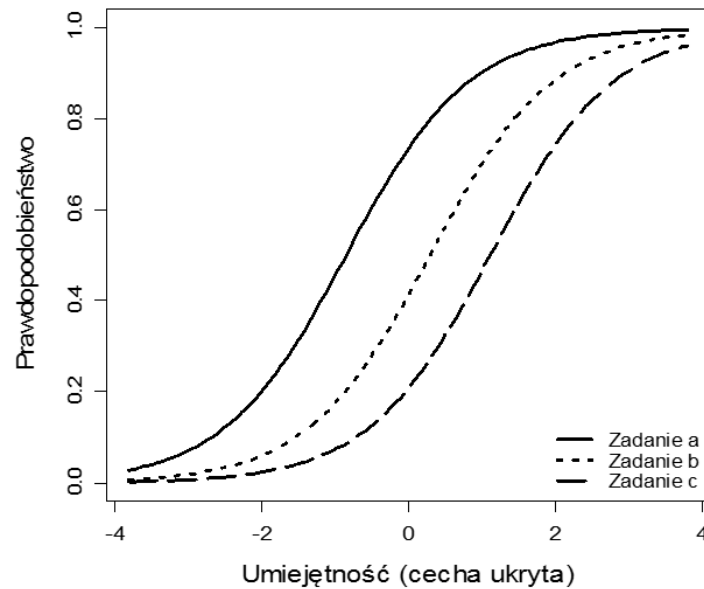


# Modele IRT

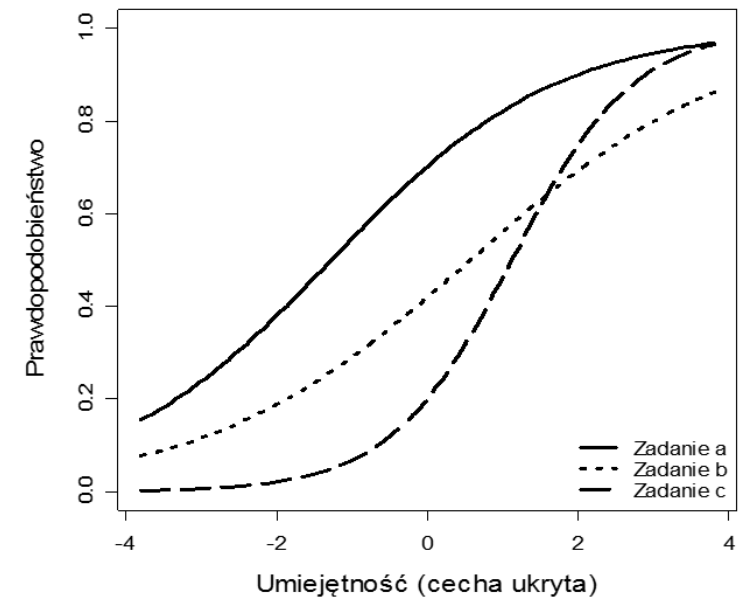
| Id  | Zad 1 | Zad 2 | Zad 3 | Zad 4 | Zad 5 | Zad 6 | Zad 7 | Zad 8 | Zad 9 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 001 | 1     | 1     | 0     | 0     | 1     | 1     |       |       |       |
| 002 | 1     | 0     | 0     | 1     | 0     | 1     | 2     | 3     | 1     |
| 003 |       |       |       | 1     | 0     | 0     | 1     | 3     | 1     |
| 004 |       |       |       | 1     | 0     | 1     | 0     | 1     | 1     |
| 005 | 0     | 0     | 1     | 1     | 0     | 0     | 2     | 0     | 0     |
| 006 | 1     | 1     | 0     | 0     | 1     | 2     |       |       |       |
| 007 | 1     | 1     | 1     | 1     | 1     | 1     |       |       |       |



Model 1PL (Rasch)



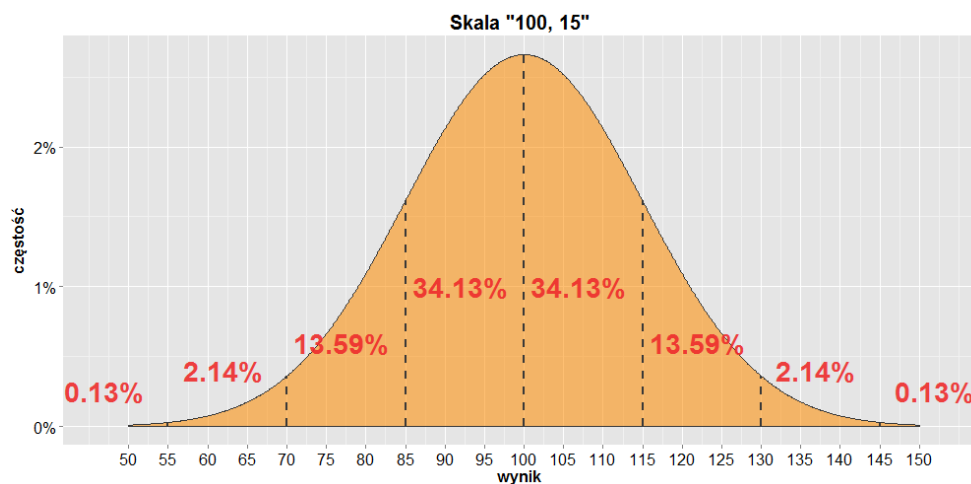
Model 2PL



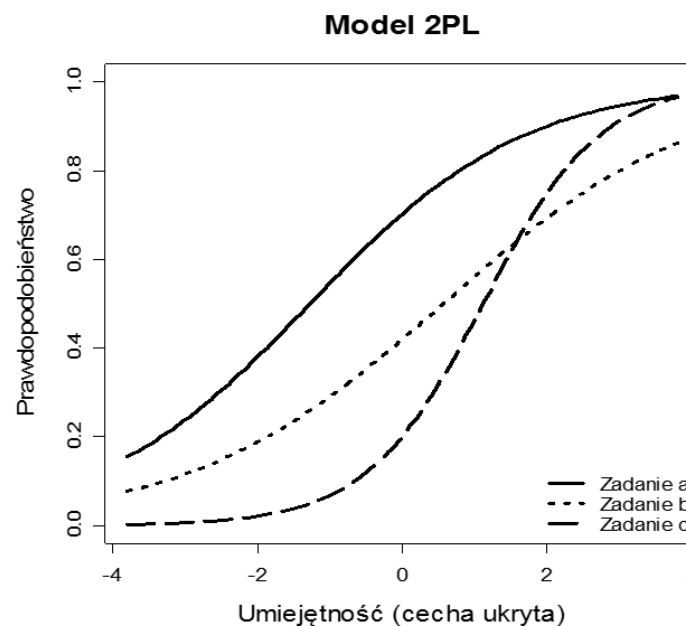


# Dwie wykorzystywane przez nas metody skalowania

## Normalizacja



## Modele IRT

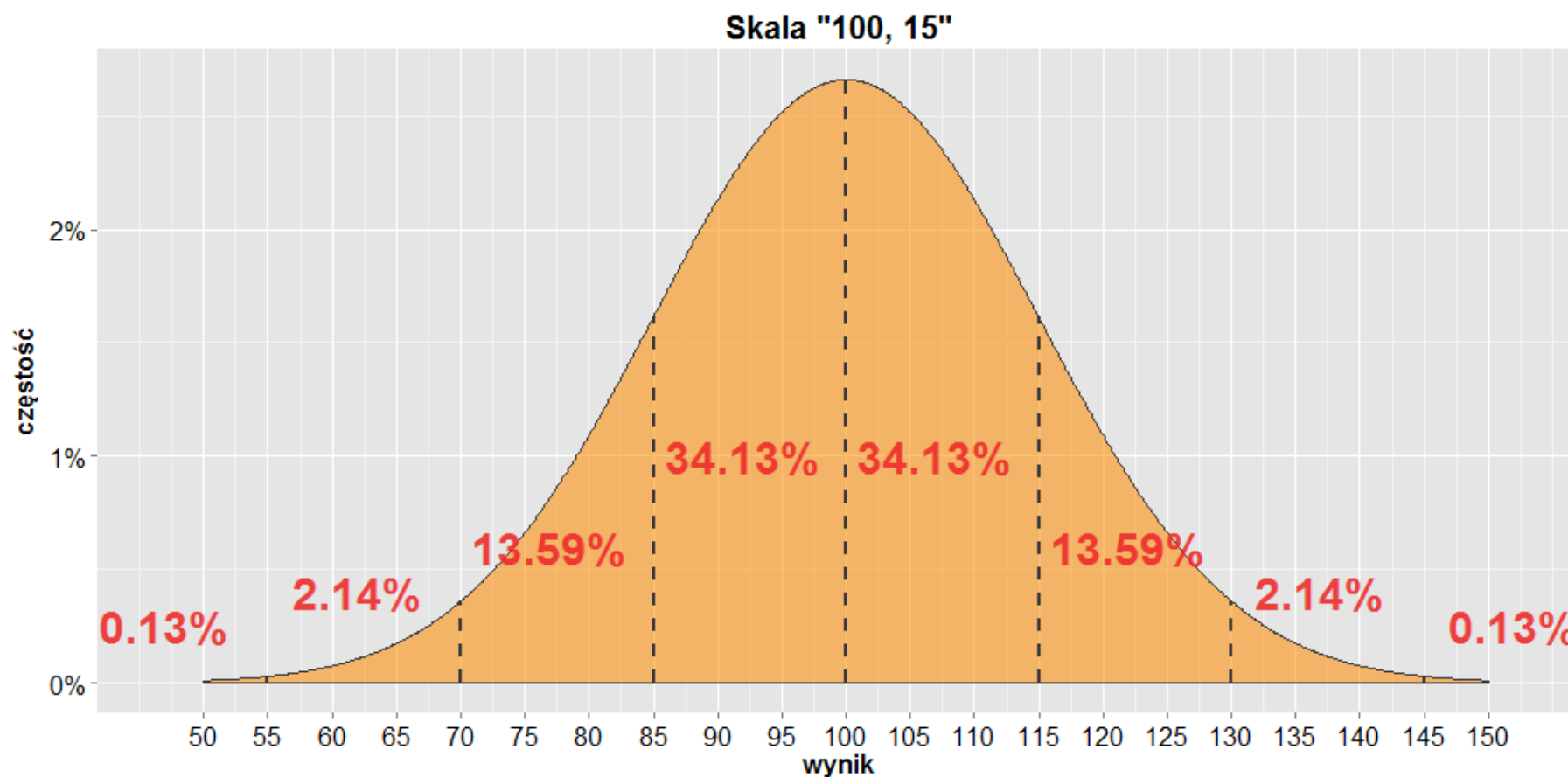


Kalkulator EWD dla gimnazjów

wskaźniki trzyletnie (2PL),  
maturalny Kalkulator EWD (1PL),  
Kalkulator EWD dla SP (1 PL)



# Skala standardowa



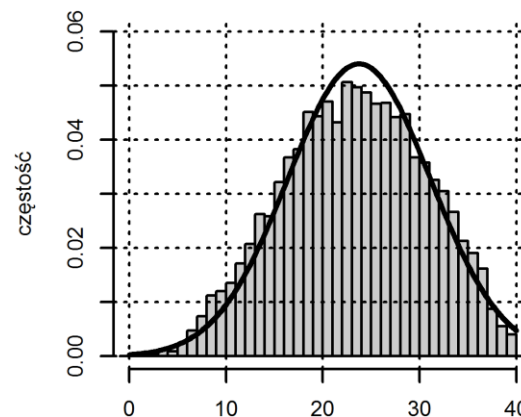
100 – zawsze wynik średni w populacji zdających w danym roku

Jednostka – odchylenie standardowe (równe 15)

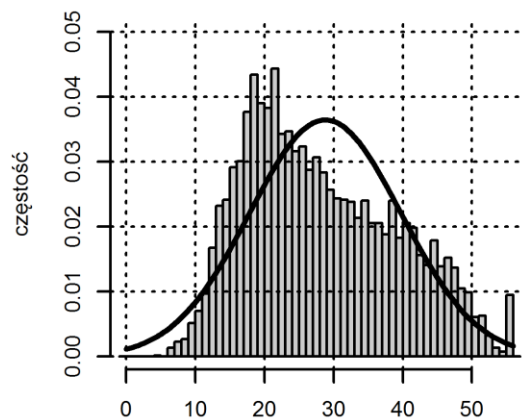


# Efekty skalowania: wyniki surowe...

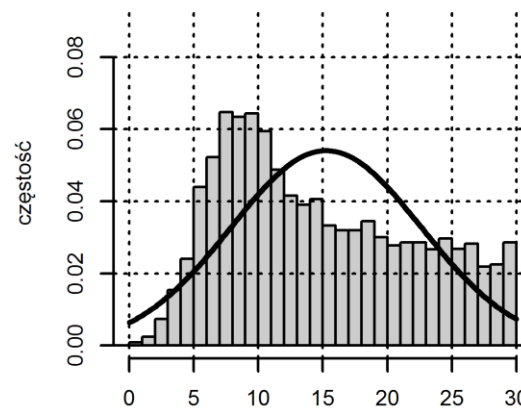
**Sprawdzian**



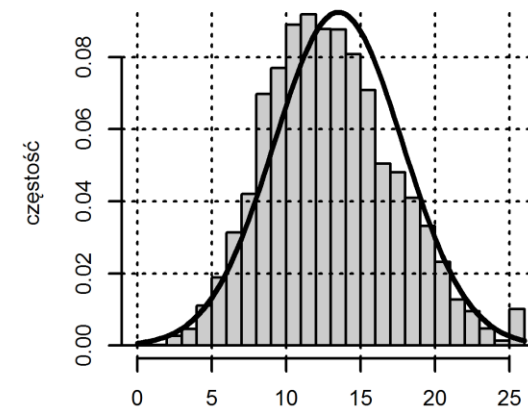
**Egz. gimn.  
część mat.-przr.**



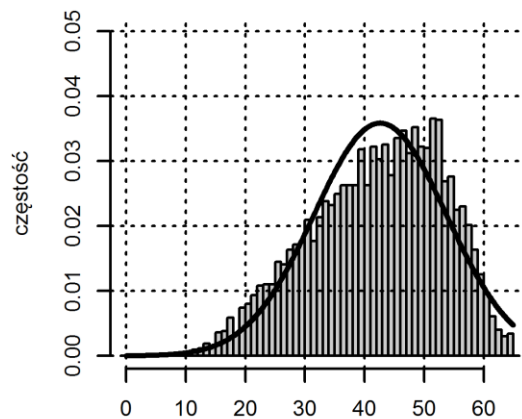
**Egz. gimn.  
test z matematyki**



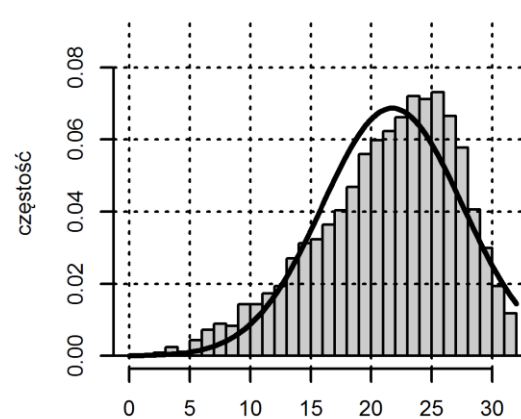
**Egz. gimn.  
test z przedm. przyr.**



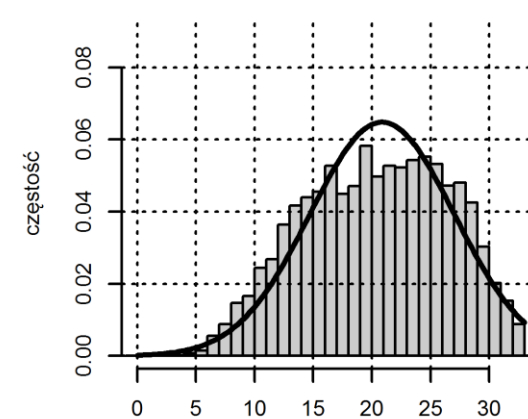
**Egz. gimn.  
część humanistyczna**



**Egz. gimn.  
test z j. polskiego**



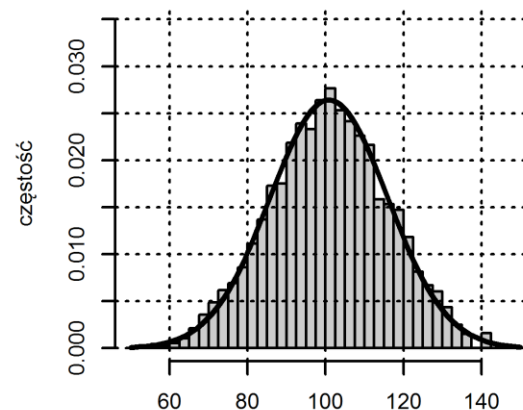
**Egz. gimn.  
test z historii i WOS**



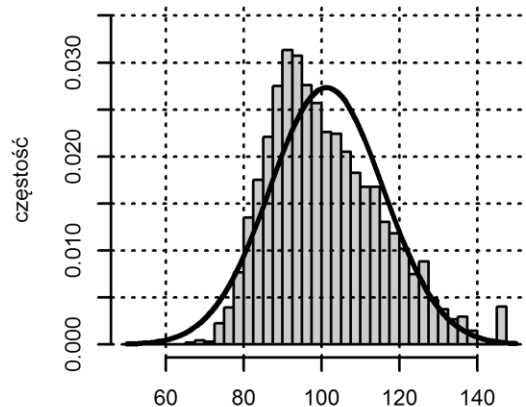


# Efekty skalowania: ... a wyskalowane.

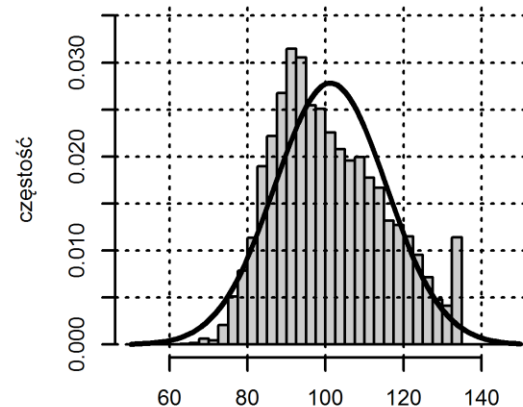
**Sprawdzian**



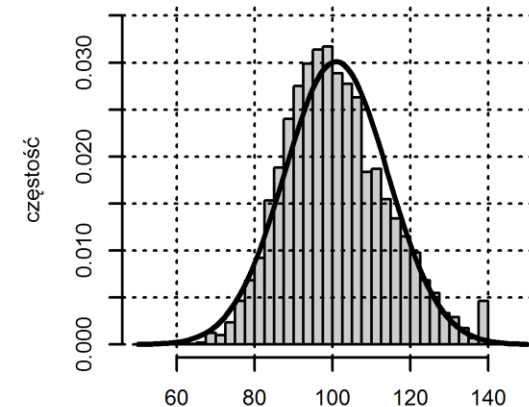
**Egz. gimn.  
część mat.-przyr.**



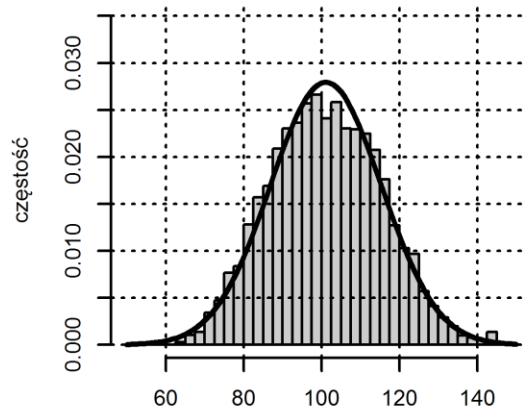
**Egz. gimn.  
test z matematyki**



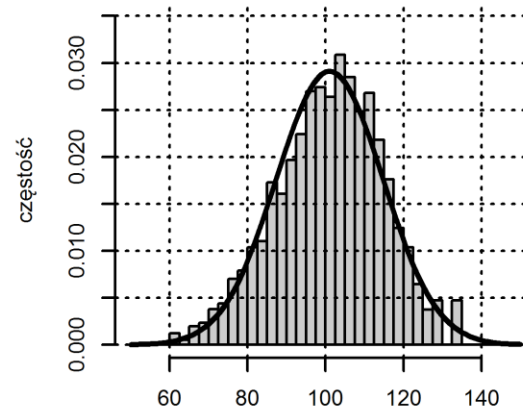
**Egz. gimn.  
test z przedm. przyr.**



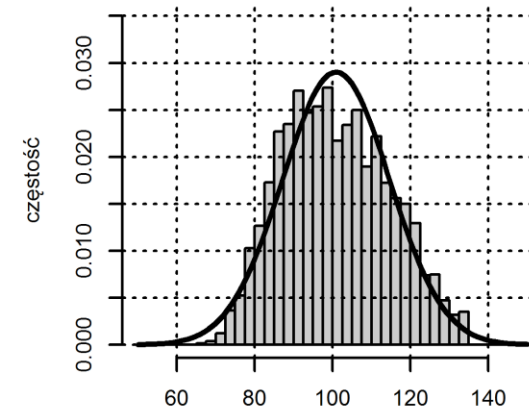
**Egz. gimn.  
część humanistyczna**



**Egz. gimn.  
test z j. polskiego**



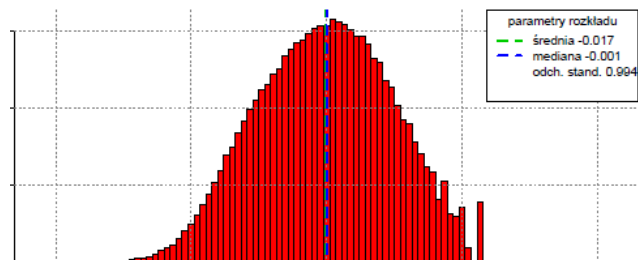
**Egz. gimn.  
test z historii i WOS**



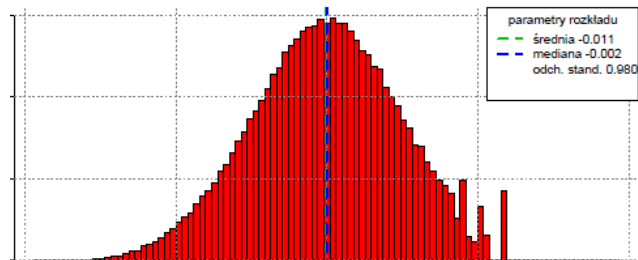


# Sprawdzian (2003-2012) – wyniki wyskalowane

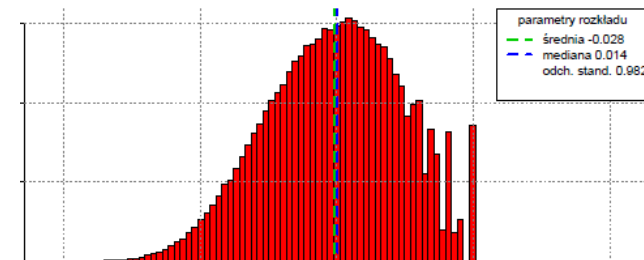
Sprawdzian 2003  
rozkład oszacowań (n = 483\*145)



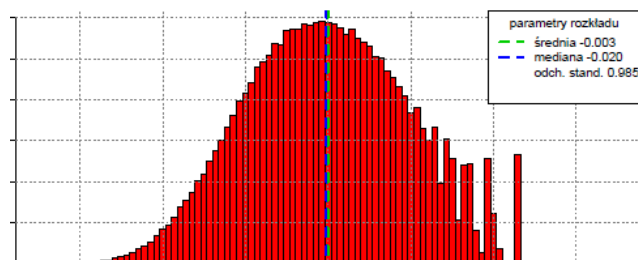
Sprawdzian 2004  
rozkład oszacowań (n = 466\*390)



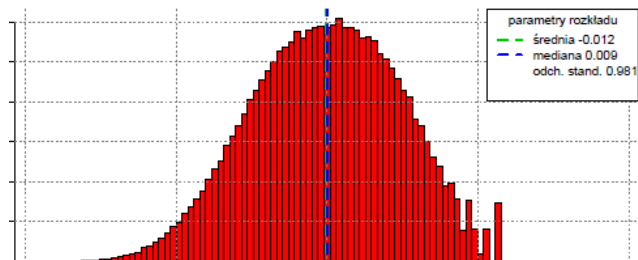
Sprawdzian 2005  
rozkład oszacowań (n = 459\*160)



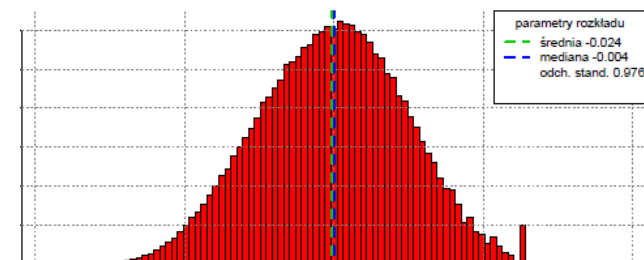
Sprawdzian 2006  
rozkład oszacowań (n = 440\*482)



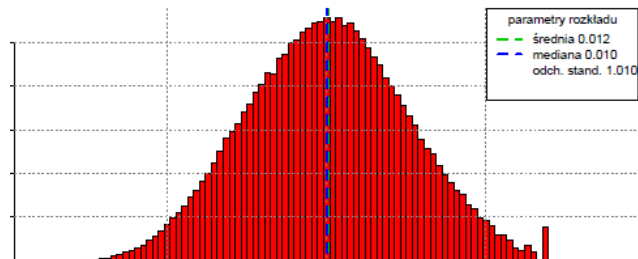
Sprawdzian 2007  
rozkład oszacowań (n = 410\*844)



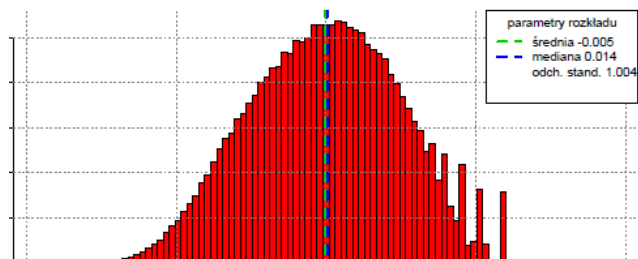
Sprawdzian 2008  
rozkład oszacowań (n = 394\*569)



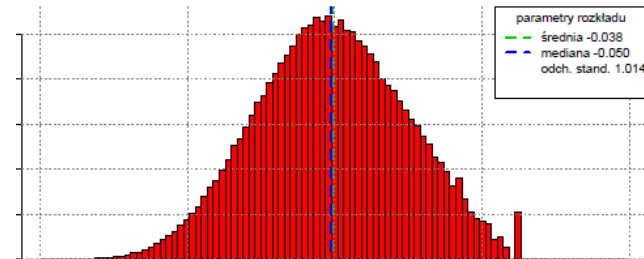
Sprawdzian 2009  
rozkład oszacowań (n = 383\*207)



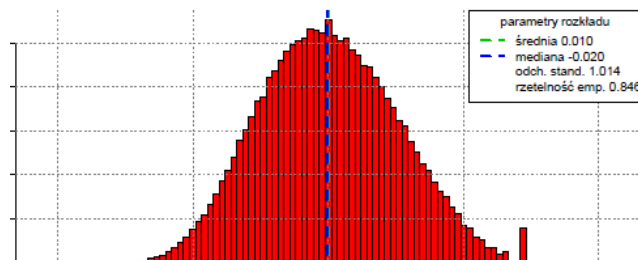
Sprawdzian 2010  
rozkład oszacowań (n = 366\*216)



Sprawdzian 2011  
rozkład oszacowań (n = 346\*787)



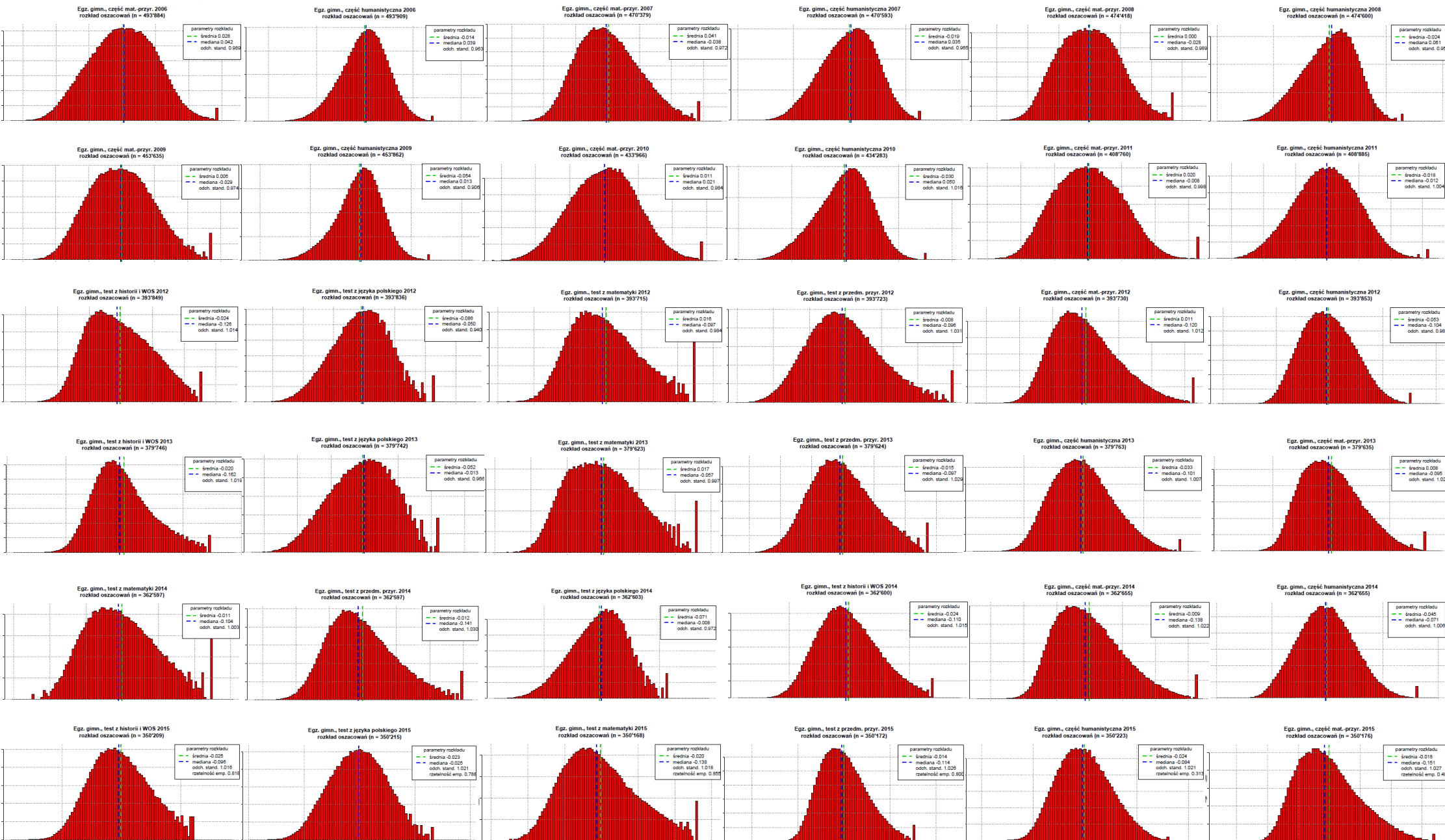
Sprawdzian 2012  
rozkład oszacowań (n = 324\*976)





# Egzamin gimnazjalny (2005-2015)

## – wyniki wyskalowane





## Skala - podsumowanie



Wyniki egzaminacyjne na skali standardowej 100/15:

- ☐ Mają znany rozkład, zgodny z rozkładem teoretycznym
- ☐ Mogą być sensownie porównywane między latami
- ☐ Są sensownie interpretowane na tle grupy odniesienia
- ☐ Posiadają ustaloną, interpretowalną jednostkę
- ☐ Mają wystarczająco dużo wartości, by zapewnić odpowiednią precyzję pomiaru
- ☐ Są znacznie bliższe spełnienia założenia o interwałowym charakterze skali niż wyniki surowe (skalowanie nie naprawi jednak wad narzędzi)



## Podsumowanie

- **Egzaminy zewnętrzne wykorzystane do konstrukcji gimnazjalnych i maturalnych wskaźników EWD mają wystarczająco dobre właściwości, by mogły stanowić podstawę wyliczania trafnej miary EWD.**
- Rzetelność testów egzaminacyjnych jest wystarczająca dla tworzenia miar charakteryzujących szkoły.
- Związki z zewnętrznymi kryteriami przemawiają za tym, że egzaminy są dobrą miarą skuteczności realizacji szerzej rozumianych celów nauczania.
- Wyniki na progu szkoły w wystarczającym stopniu pozwalają przewidywać wyniki na egzaminie końcowym, czyli mogą być wartościową podstawą wyliczania wskaźników EWD.
- Wyniki egzaminów wykorzystywane do liczenia wskaźników EWD są przedstawione na skali o dobrze zdefiniowanej jednostce pomiaru.



# EDUKACYJNA WARTOŚĆ DODANA



Dziękuję za uwagę.

Źródło: <http://www.stepitupqueen.com>

**Instytut Badań Edukacyjnych**

ul. Górczewska 8, 01-180 Warszawa

tel.: (22) 241 71 00, e-mail: [ewd@ewd.edu.pl](mailto:ewd@ewd.edu.pl)



KAPITAŁ LUDZKI  
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



EDUKACYJNA  
WARTOŚĆ  
DODANA

UNIA EUROPEJSKA  
EUROPEJSKI  
FUNDUSZ SPOŁECZNY

