
Kontrowersje wokół harmonizacji *ex ante* w badaniu PISA

Zbigniew Sawiński

Instytut Badań Edukacyjnych
Instytut Filozofii i Socjologii PAN

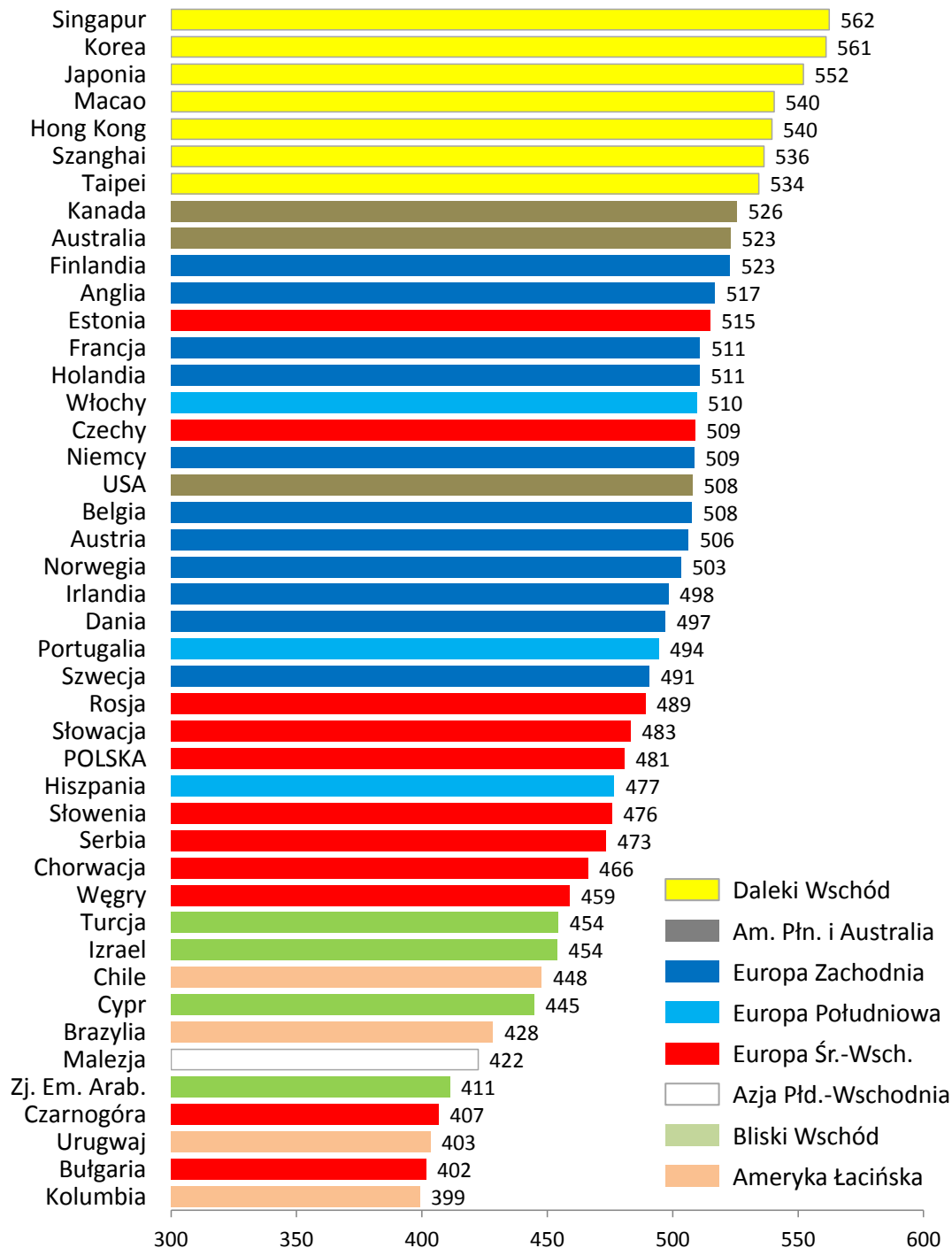
Pomiar PISA = harmonizacja *ex ante*

- Przystępując do badania kraj wyraża przekonanie, że badanie pozwoli trafnie zdiagnozować zjawisko w tym kraju
- Szczegóły badania ustalane są przed jego przeprowadzeniem
- Dbą się o formalną odpowiedniość pytań między krajami
- Pytania są weryfikowane, czy nie wykazują krajowej specyfiki
- Kontroluje się, czy badanie w każdym kraju jest prowadzone w identycznych warunkach

PISA 2012

Wyniki komputerowego testu
z rozwiązywania problemów

44 kraje



PISA 2012: wyjaśnienie wyników przez zmienne makrostrukturalne (regresja liniowa MNK)

Matematyka (test papierowy)	zmienne	25 krajów		44 kraje	
		β	ΔR^2	β	ΔR^2
	PKB	-0,38	4,2	0,51	17,9
	ESCS	0,88	13,1	-0,26	2,7
	NSP	0,11	0,6	-0,50	10,7
	NED	0,10	0,7	0,23	4,6
	R^2	25,7		44,2	

Rozwiązywanie problemów (test komputerowy)	zmienne	25 krajów		44 kraje	
		β	ΔR^2	β	ΔR^2
	PKB	0,31	2,9	0,40	9,1
	ESCS	0,23	0,9	-0,08	0,2
	NSP	-0,09	0,4	-0,33	4,2
	NED	-0,23	3,5	-0,22	3,0
	R^2	47,7		51,6	

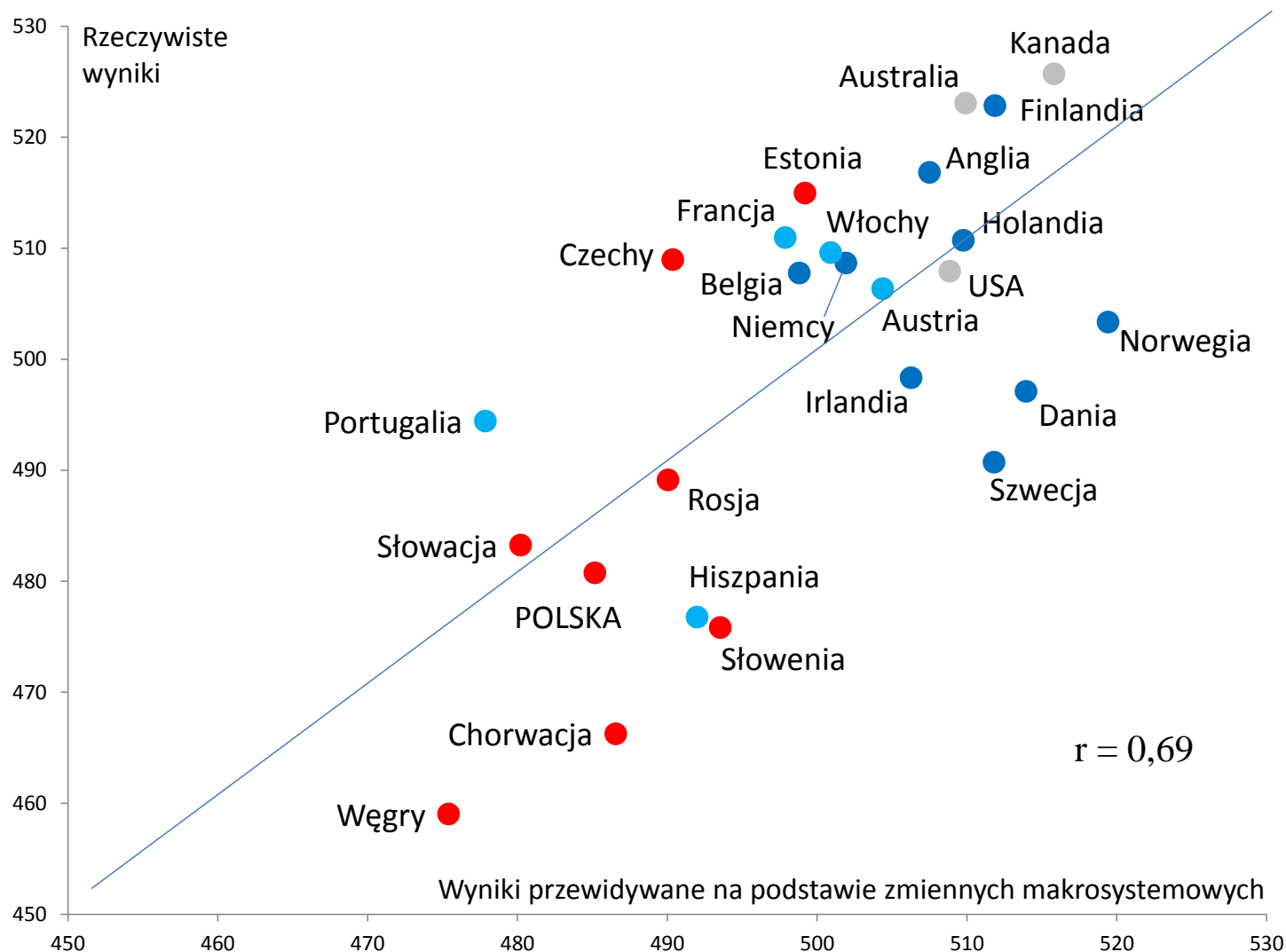
PKB – roczny krajowy produkt brutto na mieszkańca;

ESCS – ekonomiczno-społeczno-kulturowy status rodziców (wykształcenie i zawód ojca i matki, dochód i wyposażenie gospodarstwa, liczba książek, warunki do nauki, jakie dziecko ma w domu)

NSP – poziom nierówności społecznych (różnica średnich ESCS w najwyższej i najniższej grupie kwartyłowej);

NED – poziom nierówności edukacyjnych (różnica w wynikach z matematyki lub z rozwiązywania problemów między najwyższą a najniższą grupą kwartyłową według ESCS).

Estymowane i faktyczne wyniki w dziedzinie rozwiązywania problemów (PISA 2012, test komputerowy, 25 krajów)



Analiza niejednakowego funkcjonowania zadań testu: DIF (*differential item functioning*)

$$\text{logit}(\pi_{ki}) = \beta_0 + \beta_1 \hat{\theta} + \beta_2 G;$$

π_{ki} prawdopodobieństwo uzyskania k punktów w i – tym zadaniu

$\hat{\theta}$ poziom umiejętności ucznia w dwuparametrycznym modelu IRT

G przynależność do grupy (potencjalne źródło obciążenia)

Trzy kategorie obciążeń:

A β_2 nieistotny na poziomie 0,05

B β_2 istotny lub przekroczona wielkość statystyki testu

C β_2 istotny oraz przekroczona wielkość statystyki testu

Zestawienie DIF dla testu komputerowego i papierowego: specyfika Polski wobec krajów OECD

item testu	komp.	papier
cm002q01		B+
cm002q02		
cm004q01t		
cm004q02t		B+
cm004q03		
cm005q01		B+
cm005q02	B+	C+
cm005q03		
cm005q04		
cm006q01		
cm006q02		
cm006q03		B+
cm008q01		
cm008q02		
cm011q01t		
cm011q02	C+	C+
cm011q03		
cm014q01		
cm014q02		
cm014q03t		B+
cm015q01		
cm015q02d		B+
cm015q03d		C+
cm016q01		
cm020q01	B-	
cm020q02		
cm020q03		
cm020q04		
cm025q01	B+	B+
cm025q02t		
cm028q03	B-	
cm035q01		
cm035q02	C-	
cm035q03		
cm035q04	C-	
cm036q01t		
cm036q02	C+	C+
cm036q03		
cm038q03t	B-	
cm038q05	C+	C+
cm038q06		B+

Matematyka (41 pytań)

Test komputerowy:

5 x B

5 x C

Test papierowy:

8 x B

5 x C

Czytanie (19 pytań)

Test komputerowy:

2 x B

0 x C

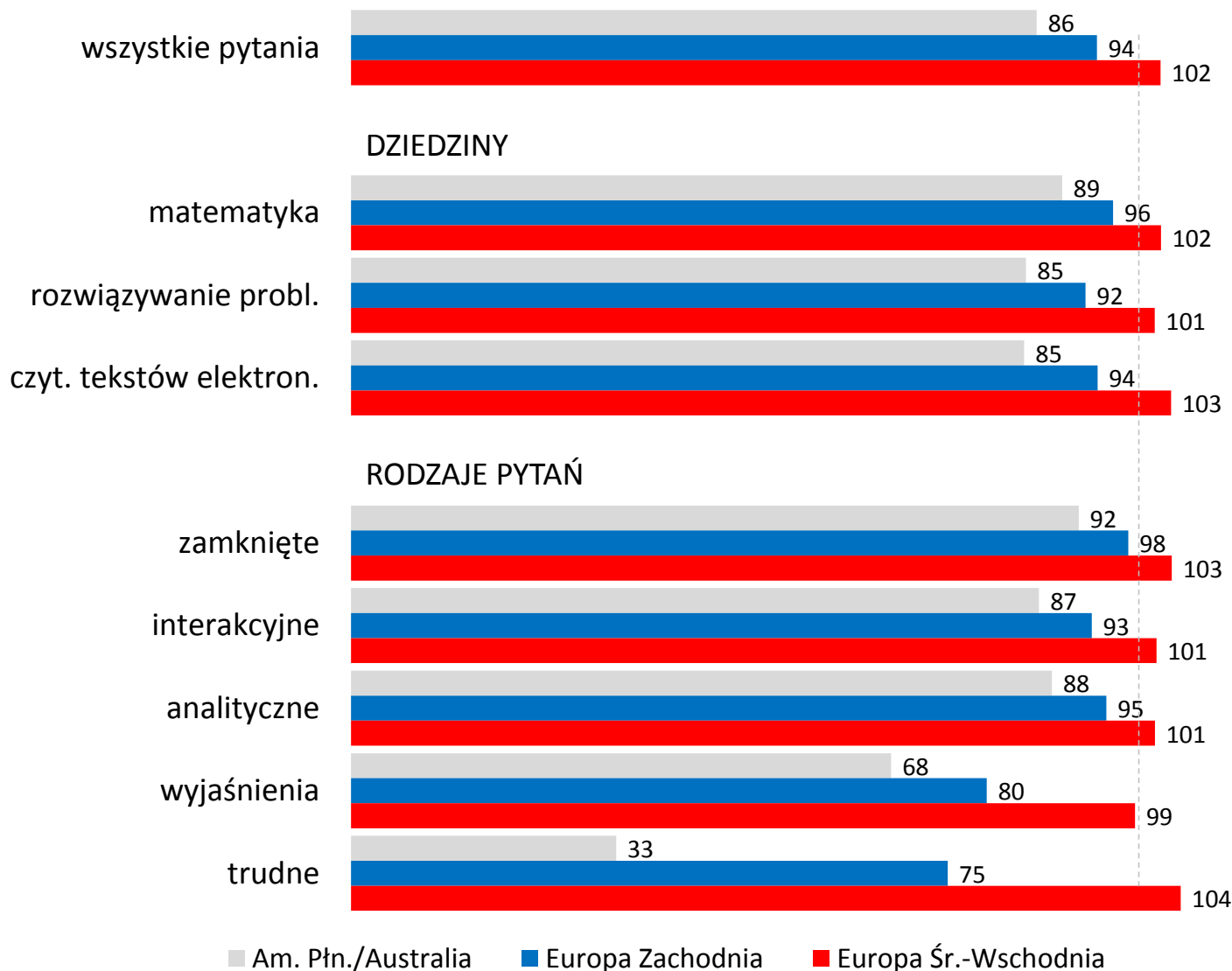
Test papierowy:

9 x B

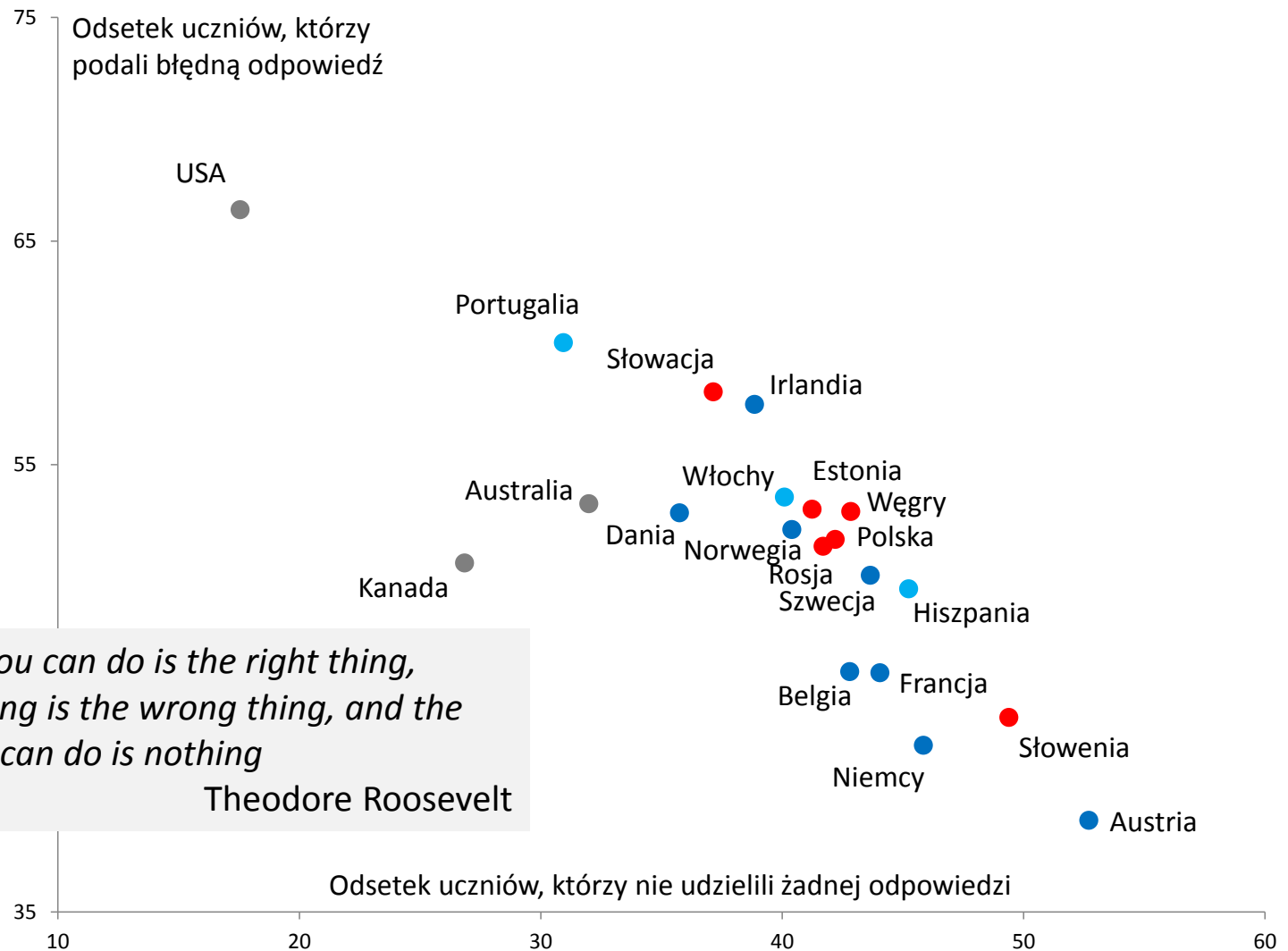
1 x C

itemy testu	komp.	ppier
cr002q01	B-	
cr002q03		B+
cr002q05		
cr011q01t		B+
cr011q01bt		
cr013q01		
cr013q04		B+
cr013q07		
cr014q01		B+
cr014q06		
cr014q07		
cr014q11		B+
cr017q01	B-	
cr017q04		B+
cr017q07		B+
cr021q01		
cr021q04		B+
cr021q05		B+
cr021q08		C+

PISA 2012 (test komputerowy): Indeksy wyników uczniów z Polski wobec wyników uczniów w trzech grupach krajów



PISA 2012 (matematyka komputerowa): Style odpowiedzi w dwóch najtrudniejszych pytaniach



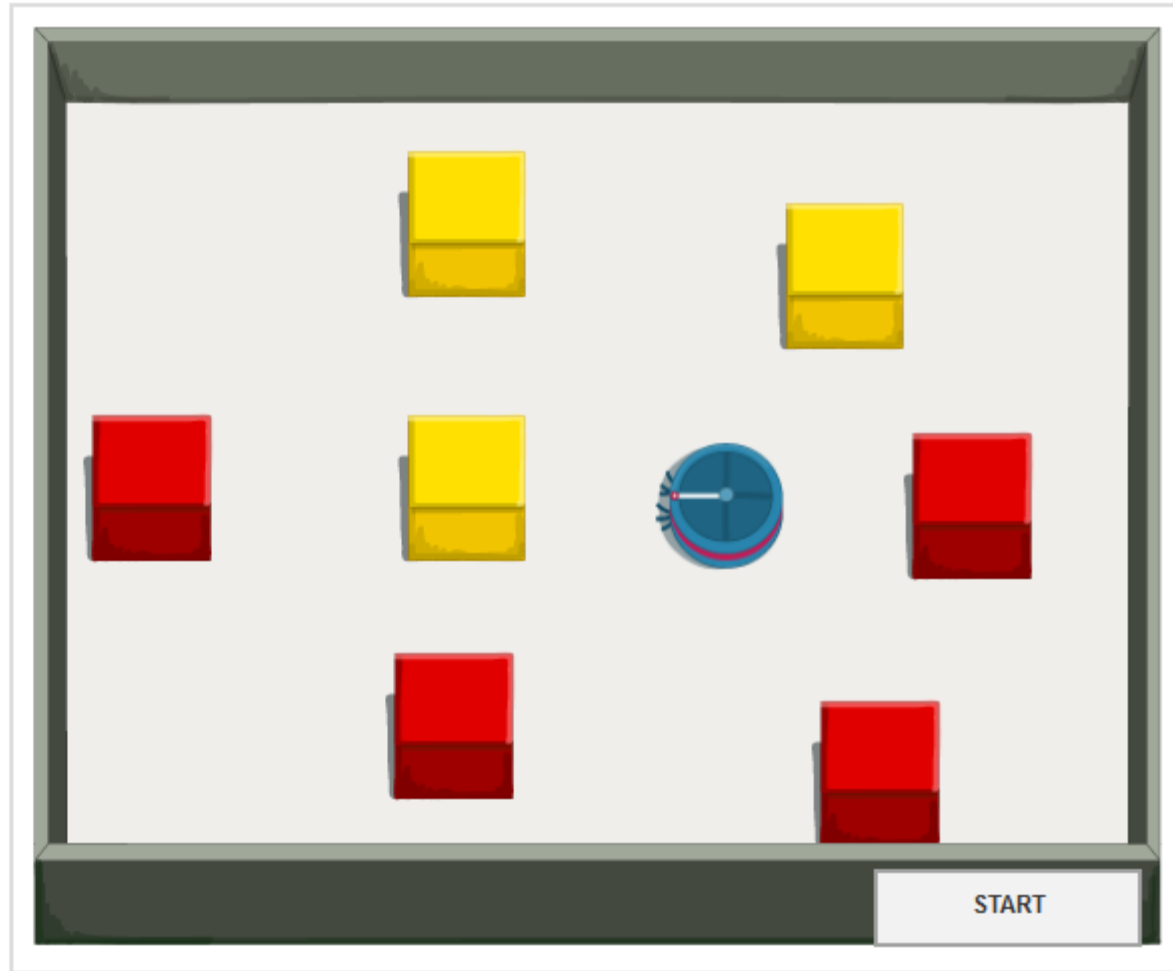
ROBOT CLEANER

<http://www.oecd.org/pisa/test/testquestions/question6/>

The animation shows the movement of a new robotic vacuum cleaner. It is being tested.

Click the START button to see what the vacuum cleaner does when it meets different types of objects.

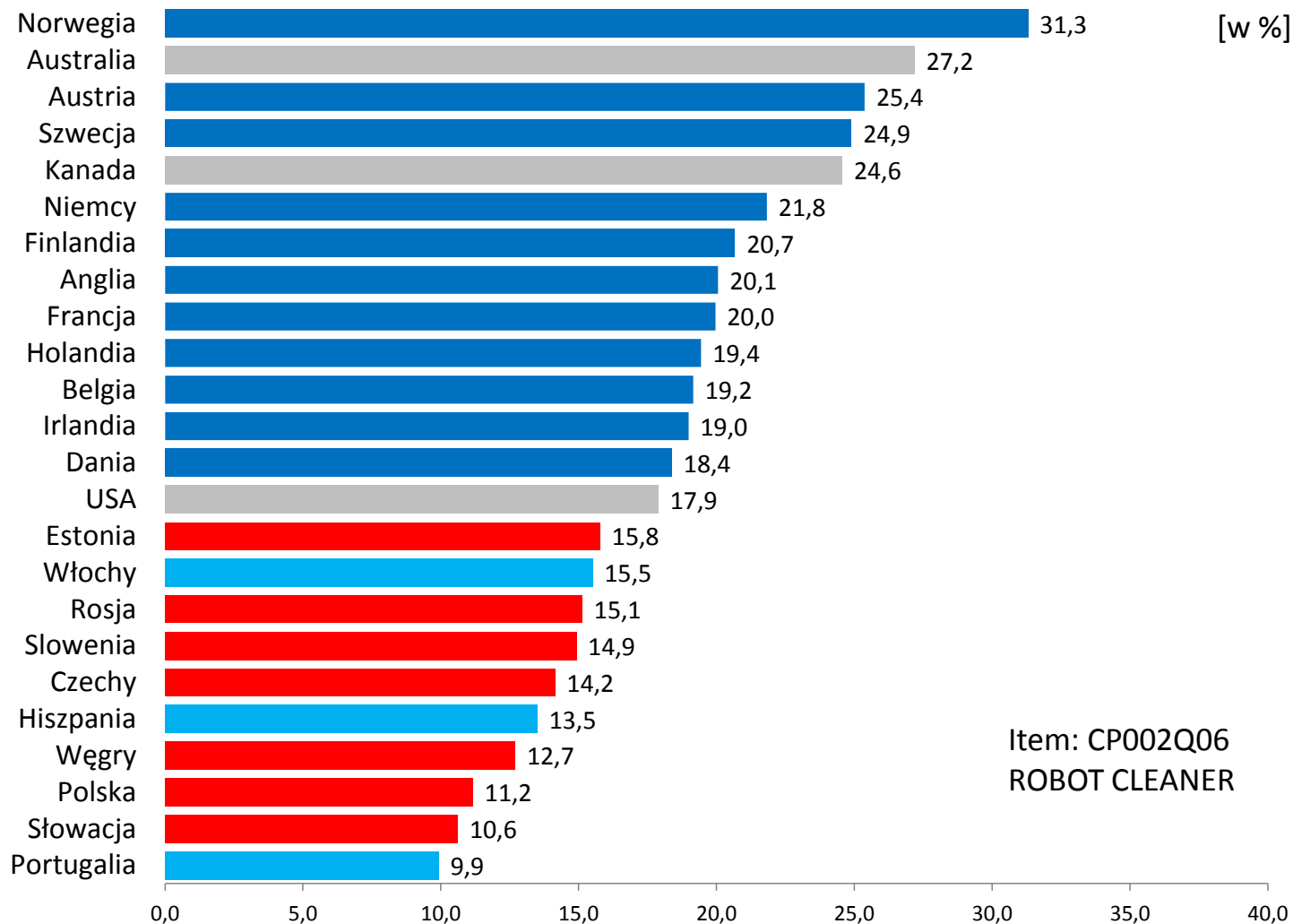
You can use the RESET button to place the vacuum cleaner back in its starting position at any time.



Question : ROBOT CLEANER

The vacuum cleaner's behaviour follows a set of rules. Based on the animation, write a rule that describes what the vacuum cleaner does when it meets a yellow block.

PISA 2012 (test komputerowy, rozwiązywanie problemów): Udział odpowiedzi kompletnych wśród odpowiedzi poprawnych



PSYCHOMETRIKA
2013
DOI: 10.1007/s11336-013-9347-z

ANALYSES OF MODEL FIT AND ROBUSTNESS. A NEW LOOK AT THE PISA SCALING MODEL UNDERLYING RANKING OF COUNTRIES ACCORDING TO READING LITERACY

SVEND KREINER

UNIVERSITY OF COPENHAGEN

KARL BANG CHRISTENSEN

UNIVERSITY OF COPENHAGEN

This paper addresses methodological issues that concern the scaling model used in the international comparison of student attainment in the Programme for International Student Attainment (PISA), specifically with reference to whether PISA's ranking of countries is confounded by model misfit and differential item functioning (DIF). To determine this, we reanalyzed the publicly accessible data on reading skills from the 2006 PISA survey. We also examined whether the ranking of countries is robust in relation to the errors of the scaling model. This was done by studying invariance across subscales, and by comparing ranks based on the scaling model and ranks based on models where some of the flaws of PISA's

scaling model are taken into account. Our analyses provide strong evidence of misfit of the PISA scaling model and very strong evidence of DIF. These findings do not support the claims that the country rankings reported by PISA are robust.

Dyskusja

- Alternatywa: harmonizacja ex post
 - Korzyści
 - pozwala uwzględnić krajową specyfikę
 - Zagrożenia
 - brak metodologii porównań między krajami
- Strategie mieszane?

Dziękuję Państwu za uwagę!

z.sawinski@ibe.edu.pl

zsawins@ifispan.waw.pl