Is the Tower of PISA Really Leaning or is it a Matter of Perspective?

A Case of Comparative Survey Design

Lars Lyberg & Can Tongur

Stockholm University & Statistics Sweden
CSDI Workshop, London UK, March 2015

facebook.com/statisticssweden   @SCB_nyheter

Outline

Statistics Sweden

Statistiska centralbyrån

PISA Surveys: Seeking for Answers

Background

- OECD, Every 3 years from 2000,
  15 year olds, Non-curricular = real-life
- 400 000+ pupils, 30-60+ countries,
  Extensive data collection, elaboration
- Ranking lists & Intense debate

Foreground

- Accepted "proof" of educational status/school status
- Big & costly enough to be the truth (compare Wikiality)
- But…not really accepted by the academic community:
  Issues with equivalence…

	Mean score in PISA 2012
OECD average	494
Shanghai–China	613
Singapore	573
Hong Kong–China	561
Chinese Taipei	560
Korea	554
Macao–China	538
Japan	536
Liechtenstein	535
Switzerland	531
Netherlands	523
Estonia	521
Finland	519
Canada	518
Poland	518
Belgium	515
Germany	514
Viet Nam	511
Austria	506
Australia	504
Ireland	501
Slovenia	501
Denmark	500
New Zealand	500
Czech Republic	499
France	495
United Kingdom	494
Iceland	493
Latvia	491
Luxembourg	490
Norway	489
Portugal	487
Italy	485
Spain	484
Russian Federation	482
Slovak Republic	482
United States	481
Lithuania	479
Sweden	478
Hungary	477
Croatia	471
Israel	466
Greece	453
Serbia	449
Turkey	448
Romania	445

Switzerland

Sweden

| | On the overall reading scale | On the reading subscales | | | | | On the mathematics scale | On the science scale |
		Access and retrieve	Integrate and interpret	Reflect and evaluate	Continuous texts	Non-continuous texts		
OECD average	493	495	493	494	494	493	496	501
Shanghai-China	556	549	558	557	564	539	600	575
Korea	539	542	541	542	538	542	546	538
Finland	536	532	538	536	535	535	541	554
Hong Kong-China	533	530	530	540	538	522	555	549
Singapore	526	526	525	529	522	539	562	542
Canada	524	517	522	535	524	527	527	529
New Zealand	521	521	517	531	518	532	519	532
Japan	520	530	520	521	520	518	529	539
Australia	515	513	513	523	513	524	514	527
Netherlands	508	519	504	510	506	514	526	522
Belgium	506	513	504	505	504	511	515	507
Norway	503	512	502	505	505	498	498	500
Estonia	501	503	500	503	497	512	512	528
Switzerland	501	505	502	497	498	505	534	517
Poland	500	500	503	498	502	496	495	508
Iceland	500	507	503	496	501	499	507	496
United States	500	492	495	512	500	503	487	502
Liechtenstein	499	508	498	498	495	506	536	520
Sweden	497	505	494	502	499	498	494	495
Germany	497	501	501	491	496	497	513	520
Ireland	496	498	494	502	497	496	487	508
France	496	492	497	495	492	498	497	498
Chinese Taipei	495	496	499	493	496	500	543	520
Denmark	495	502	492	493	496	493	503	499
United Kingdom	494	491	491	503	492	506	492	514
Hungary	494	501	496	489	497	487	490	503
Portugal	489	488	487	496	492	488	487	493
Macao-China	487	493	488	481	488	481	525	511
Italy	486	482	490	482	489	476	483	489
Latvia	484	476	484	492	484	487	482	494
Slovenia	483	489	489	470	484	476	501	512
Greece	483	468	484	489	487	472	466	470
Spain	481	480	481	483	484	473	483	488
Czech Republic	478	479	488	462	479	474	493	500
Slovak Republic	477	491	481	466	479	471	497	490

Survey Sampling: Design

- 2-folded design
  - 2 stage survey sampling: schools and pupils Schools sampled according to a stratified PPS
  - Clusters of test packages ("booklets") Randomization pattern of 4 cluster areas in three main fields: mathematics, reading & science

Area	2000	2003	2006	2009	2012	2015
Reading	**Major**	Minor	Minor	**Major**	Minor	Minor
Maths	Minor	**Major**	Minor	Minor	**Major**	Minor
Science	Minor	Minor	**Major**	Minor	Minor	**Major**

Table based on *Baird et al. (2011)*

Survey Sampling: Design

- Target age 15: age vs. schooling years
  - **a**) Curricular years, **b**) reception years, **c**) repetition
    PISA applies time-window of birth
    Large variability in **a**, **b c**
  - Intra-country vs. inter-country comparability

- Demographical distributions
  - Biased school samples (in both directions!)
    UK upwards, U.S downwards
    Non-response
  - Socio-economic factors

Survey Items
- *Item Response Theory*

- *Differential Item Functioning*
  A (possible) consequence of multicultural use of basically <u>one</u> unison concept

  1. Language within country:
     Target population and translating population differ in semantics

  2. General aptitude test

  3. Contextual diversities between target populations

Survey Items
- *Modeling/Scaling*

- The Rasch Model – Lost in Translation?

$$P_i\left(Y_i = y_i | \Theta = \theta\right) = \frac{\exp(\theta - b_i)}{1 + \exp(\theta - b_i)}$$

1) Local indepedence between questions
2) No DIF: $Y_i \perp$ (Country & Auxilliary information)$| \Theta$

Difficulty b_i is considered fix: assumptions valid?
Can items be non-sensitive to translation?
Dodgy items *country-wise*. Language causality
assumption

- Data Augmentation – Plausible Values

  Minor areas through Majors as anchor
  Input Vector based on θ. Bayesian approach

The Swedish Case
- The Most Remarkable Case?

- Years 2000/2003:
  Sweden *significantly* above OECD PISA average
  8/9 OECD countries *significantly* better than SWE

- Years 2006/2009:
  Sweden not *significantly* outside OECD PISA avg.
  10/14 OECD countries *significantly* better than SWE

- Year 2012:
  Sweden *significantly* below OECD PISA avg.
  25 OECD countries *significantly* better than SWE

The Swedish Case (con't)
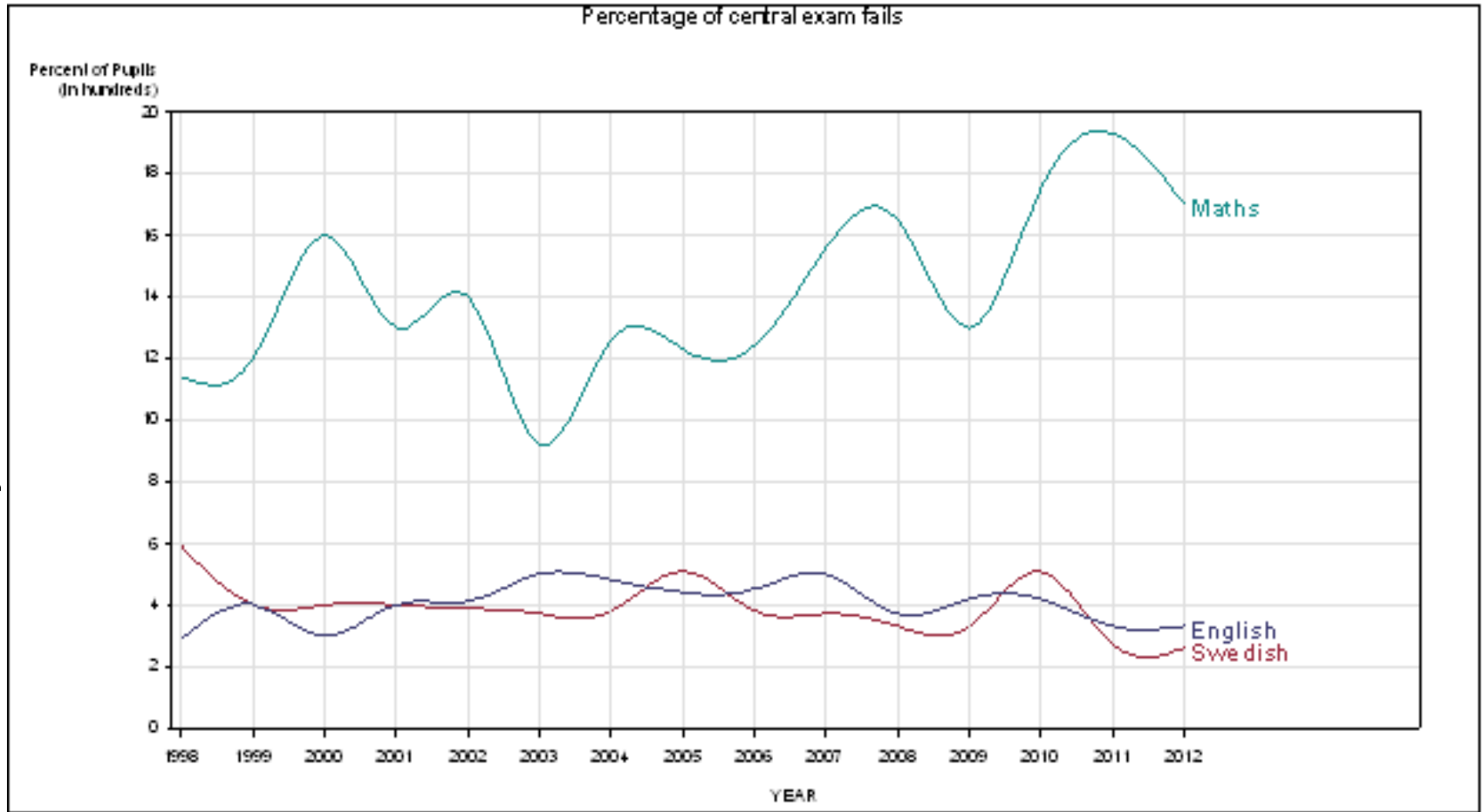
Percentage of central exam fails of Swedish 9th graders between [1998, 2012]

To Conclude

- PISA surveys are big and elaborated, high effort
- PISA offers comparisons with some loose ends: Samples & Items afflicted with issues
- Item Response Theory not well addressed
- The Swedish Case: no reasons identified
- No self-assessment/criticism by the OECD/PISA
- Survey conditions rather diversified
- Could it be done better? Should it be done at all? Perhaps too many dimensions?

References

- Carnoy & Rothstein (2013)
- Goldstein, Harvey: Several publications
- Blomberg, P. & Lundahl, A., Bachelor thesis Stockholm University, superv. L. Lyberg, (2014)
- Kreiner, S., Dept. University of Copenhagen (2011)
- Klein, R. A Reanalysis of PISA results.. (2011)
- Prais, S.J., Oxford Review of Education (2003)
- Serder, M. & Jakobsson, A. (2014).

- OECD: PISA Technical Reports
- Skolverket (www.skolverket.se), Swed. Edu. Board

Statistics Sweden

Statistiska centralbyrån