

A mixed-method approach to interpret Rasch misfitting behaviour in large-scale Math assessment

Clelia Cascella and Giorgio Bolondi

Clelia Cascella is Marie Curie Fellow at the University of Manchester and researcher at the Italian National Institute for the Evaluation of Educational System

cascella.clelia@gmail.com

Giorgio Bolondi is Full Professor of Didactics of Mathematics at the Freie Universität of Bozen

giorgio.bolondi@unibz.it



Within Rasch framework, misfitting behaviour is source of information

- Research hypothesis

Explaining gap between observed data and model's expectations

- Research aim

FROM A MEASUREMENT POINT OF VIEW → Picking source of misfit to avoid them in the future

- Implication n. 1

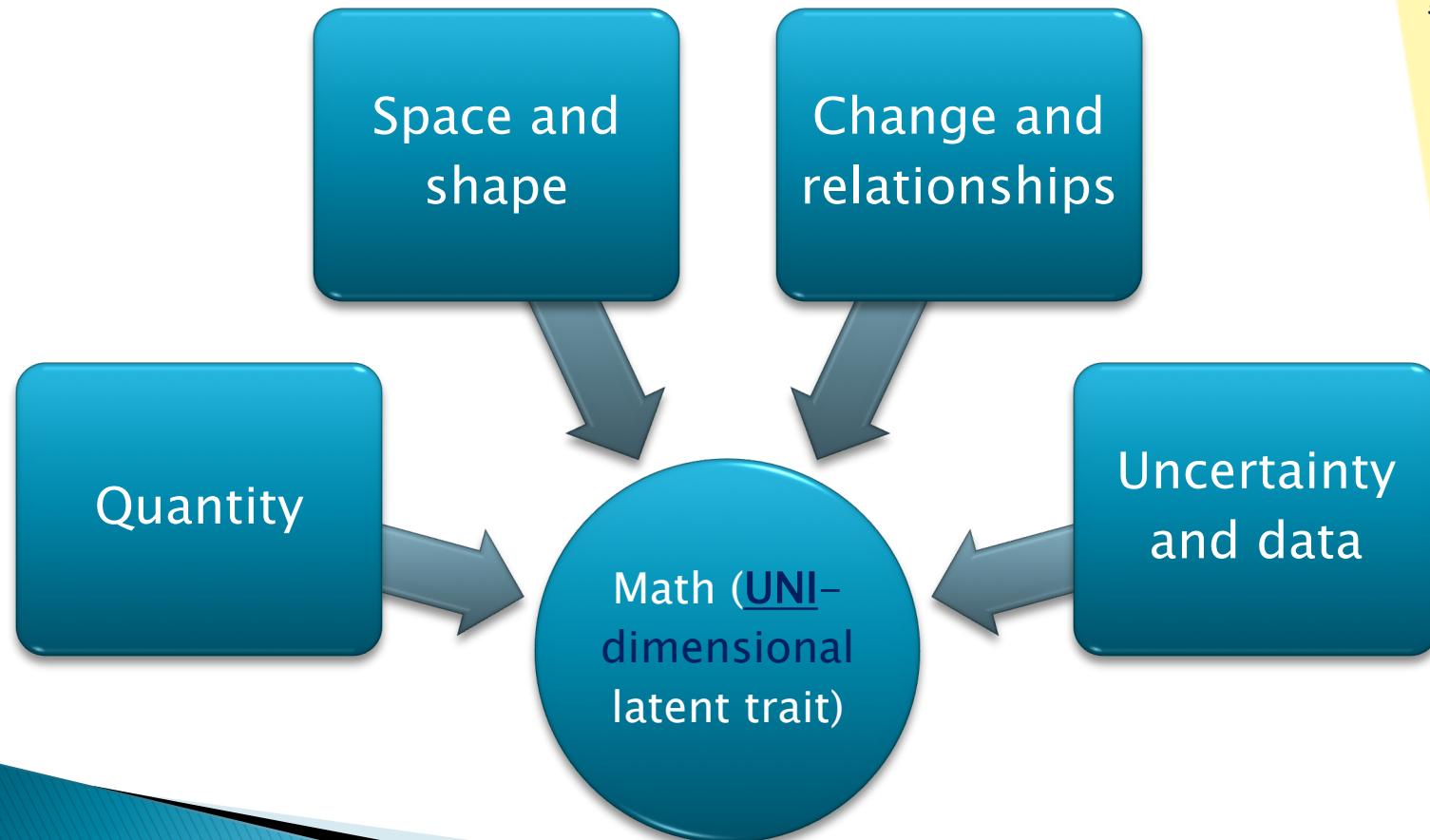
FROM A DIDACTICAL POINT OF VIEW → Collecting information to formulate hypotheses about possible effects of didactical praxes on students' performance

- Implication n. 2

Research pathway

INVALSI

Math achievement tests



The INVALSI system administers a Math achievement test aimed at assessing mathematical competence, i.e. «the ability to develop and apply **mathematical thinking** in order to solve a **range of problems in everyday situations**. Building on a sound mastery of numeracy, the emphasis is on process and activity, as well as knowledge. Mathematical competence involves, to different degrees, the ability and willingness to use mathematical modes of thought (**logical and spatial thinking**) and presentation (**formulas, models, constructs, graphs, charts**)» (European Recommendation 2006/962/EC, p. 6).

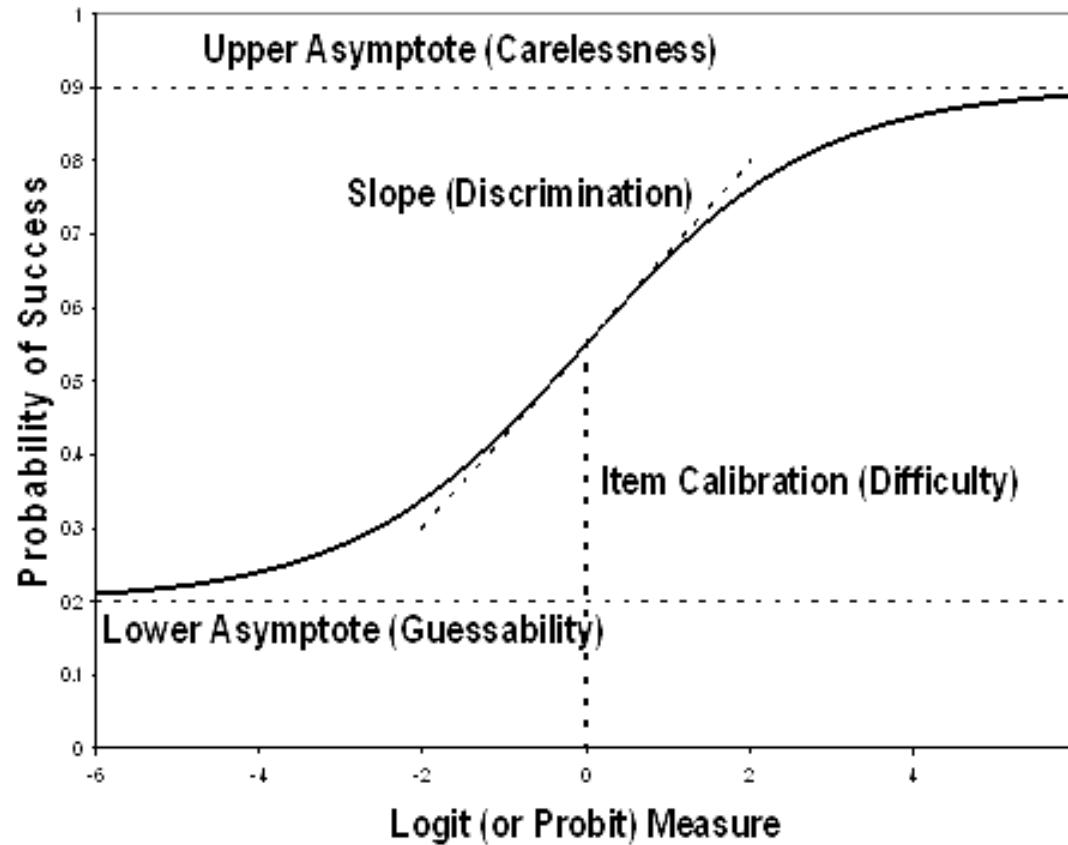


Data

Grade	Scholastic year									
	2008/09	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	
G2	29.204	30.420	31.842	31.773	24.89	26.356	20.922	24.125	23.001	
G5	27.514	29.854	31.875	30.843	24.679	25.331	21.049	23.011	22.014	
G6	-	32.642	40.497	39.668	27.504	-	-	-	-	
G8	-	-	523.111	519.01	520.918	520.917	520.92	519.145	518.998	
G10	-	-	43.458	41.812	38.060	36.932	27.393	28.635	28.362	

From 2011 up to 2017, INVALSI administered **7 Math achievement tests (354 items)**, at both census and sample data.
 These achievement tests are **not linked** and thus each achievement test has been **analysed separately**.

Methodology



- ▶ Item Characteristic Curve
- ▶ INFIT
- ▶ DIF
- ▶ Test response dependance

Source: <https://www.rasch.org/rmt/rmt181b.htm>

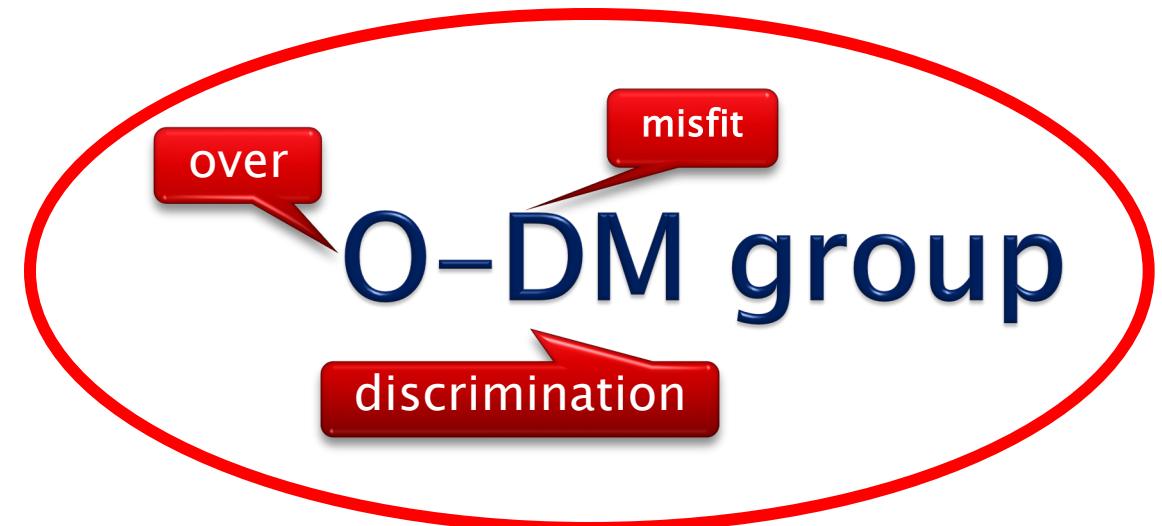
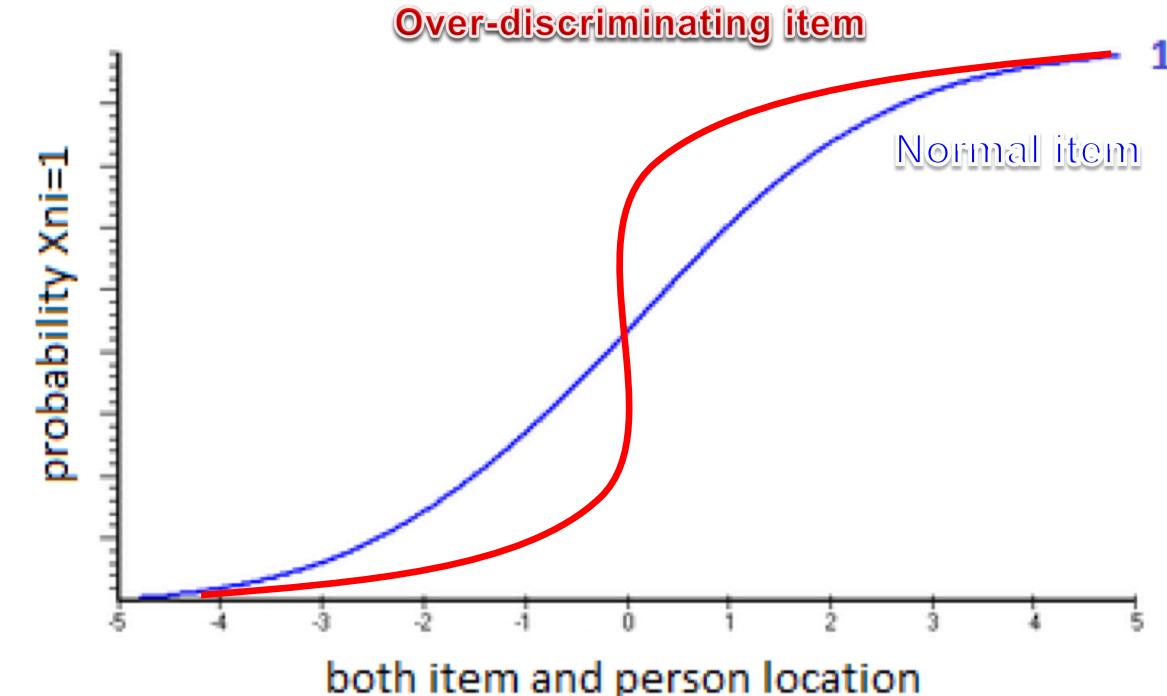
What ICC says...

Over-discrimination

- ▶ 354 items analyzed (administered at grade 10)
- ▶ 59 items over 354 show an overdiscriminating behaviour
- ▶ For each math achievement test, from 2011 up to 2017, when an item is overdiscriminating, it always shows recurring further features!

SELECTION CRITERIUM

- 1) Weighted MNSQ (used to test the violations of the model assumptions of no guessing and homogeneity of item discrimination) < 0.90



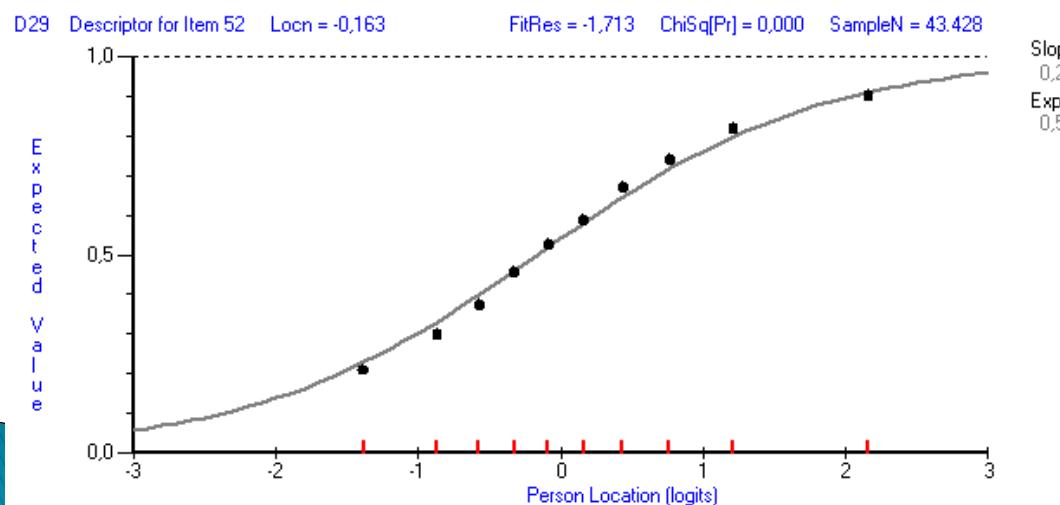
Examples of fitting items

[space and shape]

- D29. Which of the following decimal numbers represent the expression

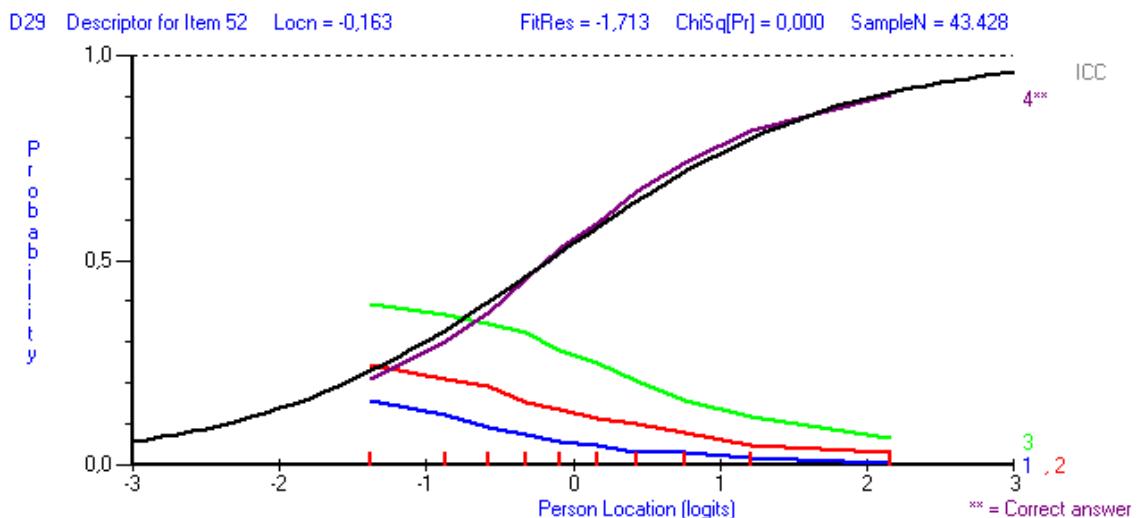
$$\frac{9}{10} + \frac{8}{10^2} + \frac{7}{10^4} + \frac{2}{10^5}$$

- A. 98.72
- B. 9.8072
- C. 0.9872
- D. 0.98072



ITN ConQuest 4.0

Cases	43357	Discrimination	0.45				
Threshold	-0.13	Weighted MNSQ	0.97				
Item delta	-0.13						
Label	Score	Count	% of tot	Pt Bis	T (p)	PV1 Avg:1	PV1 SD:1
1	0.00	2544	5.87	-0.17	-35.29(0.000)	-0.62	0.77
2	0.00	5261	12.13	-0.17	-36.68(0.000)	-0.43	0.81
3	0.00	10146	23.40	-0.21	-44.13(0.000)	-0.35	0.83
4	1.00	22701	52.36	0.45	103.59(0.000)	0.40	0.97
9	0.00	2705	6.24	-0.16	-33.82(0.000)	-0.59	0.87



Example 1

[space and shape]

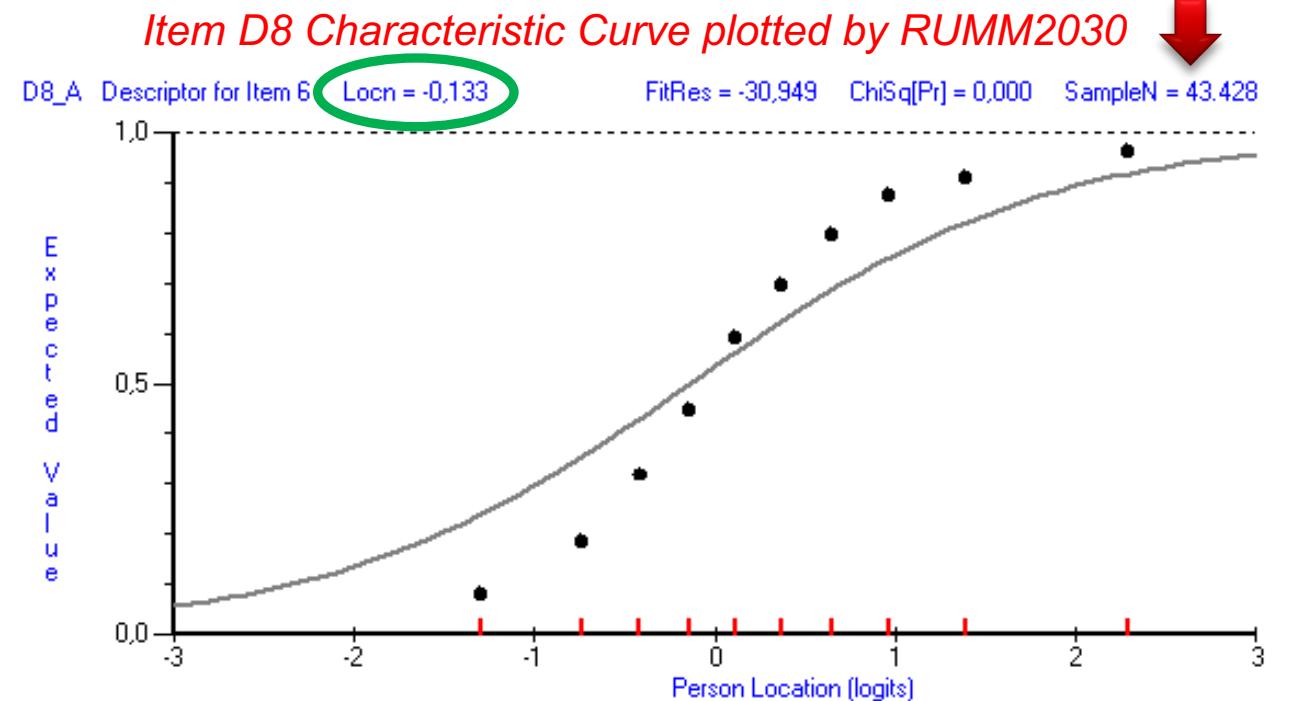
Item administered in 2011 at grade 10

D8. A TV dimension is the measure of its screen's diagonal in inches (1 inch = 2.54 centimeters). For new generation televisions, the relationship between height and width is equal to 16:9.

If the screen width is around 57.5 centimeters, what is its approximate height?

Answer: _____ centimeters

- The item does not show any DIF.
- Response dependence can be excluded.
- Medium difficulty.
- High percentage of missing values.
- NO high difference in ability between students' providing a wrong answer and students' not providing any answer.
- Pt. bis. correlation is higher for students not providing any answer than for students giving a wrong answer.



Item analysis carried out by ConQuest 4.0

Analysis of item D8

Cases for this item	43458	Discrimination	0.63			
Item Threshold	0.33	Weighted MNSQ	0.82			
Item Delta	0.33					
Label	Score	% of tot	Pt Bis	t (p)	PX1 Avg:1	PV1 SD:1
0	0.00	30.09	-0.28	-61.95(0.000)	-0.40	0.75
1	1.00	42.81	0.63	168.34(0.000)	0.67	0.87
9	0.00	26.64	-0.40	-91.53(0.000)	-0.62	0.74

Note. 0 = wrong answer; 1 = correct answer; 9 = missing values.

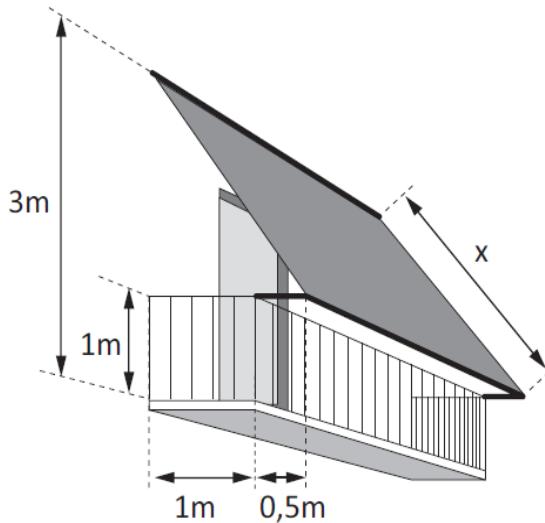
Source: Our elaboration on SNV INVALSI data collected in 2011 at the 2nd grade of upper secondary school

Example 2

[space and shape]

Item administered in 2012 at grade 10

D24. A sunblind has to be tailored. It has to be fixed on the wall three meters away from the floor on the balcony. The balcony is one meter wide. Sunblind has to stick out 0.5 meters from the handrail which is one meter high.

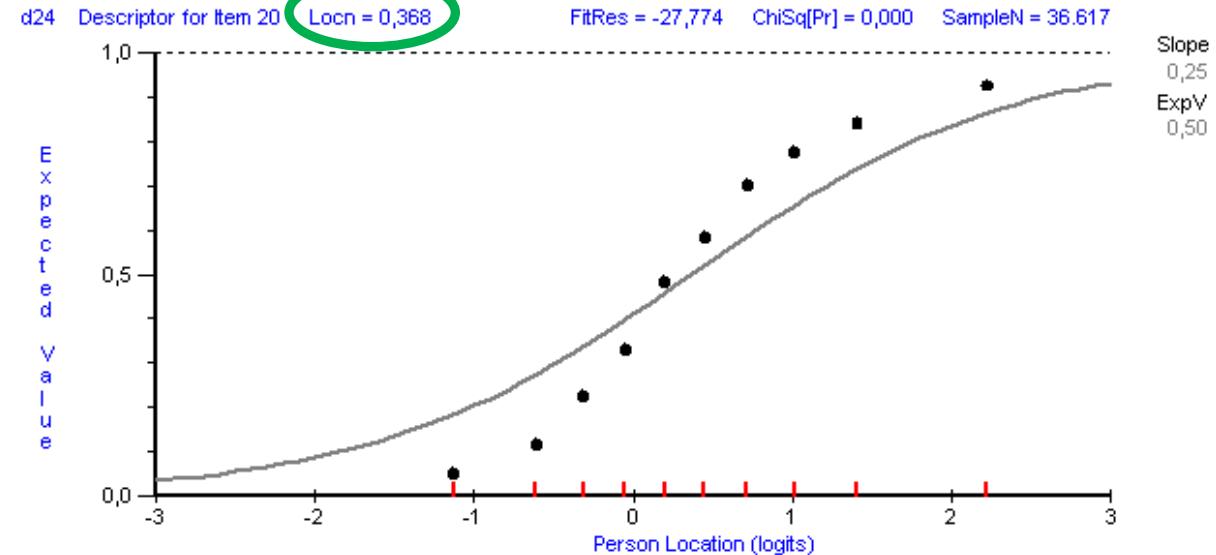


Write the algebraic operation that you perform to calculate the sunblind's length (x) and write the result of this calculation

.....
.....

Result: $x = \dots$

Item D8 Characteristic Curve plotted by RUMM2030



Item analysis carried out by ConQuest 4.0

Analysis of item D24

Cases for this item	41812	Discrimination	0.59			
Item Threshold	1.15	Weighted MNSQ	0.84			
Label	Score	% of tot	Pt Bis	t (p)	PX1 Avg:1	PV1 SD:1
0	0.00	27.64	-0.13	-26.80(0.000)	-0.19	0.83
1	1.00	27.65	0.59	148.86(0.000)	0.89	0.85
9	0.00	43.79	-0.41	-90.62(0.000)	-0.44	0.85

Note. 0 = wrong answer; 1 = correct answer; 9 = missing values.

Source: Our elaboration on SNV INVALSI data collected in 2012 at the 2nd grade of upper secondary school

Example 3

[Quantity / numeracy]

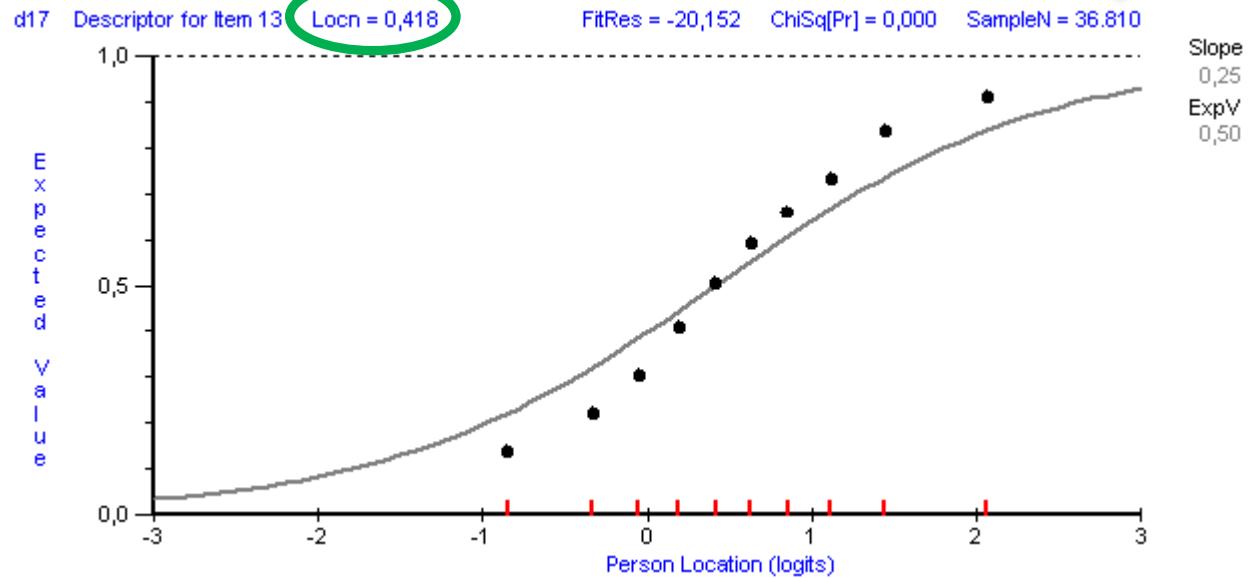
Item administered in 2012

D17. In the equation $(2k-3)x-1-k=0$, x is the unknown quantity and k is a real number.

The solution is equal to 1 for k equal to



Item D17 Characteristic Curve plotted by RUMM2030



Item analysis carried out by ConQuest 4.0

Analysis of item D17

Cases for this item	41812	Discrimination	0.59			
Item Threshold	1.15	Weighted MNSQ	0.84			
Label	Score	% of tot	Pt Bis	t (p)	PX1 Avg:1	PV1 SD:1
0	0.00	27.64	-0.13	-26.80(0.000)	-0.19	0.83
1	1.00	27.65	0.59	148.86(0.000)	0.89	0.85
9	0.00	43.79	-0.41	-90.62(0.000)	-0.44	0.85

Note. 0 = wrong answer; 1 = correct answer; 9 = missing values.

Source: Our elaboration on SNV INVALSI data collected in 2012 at the 2nd grade of upper secondary school

Example 3bis

[Quantity / numeracy]

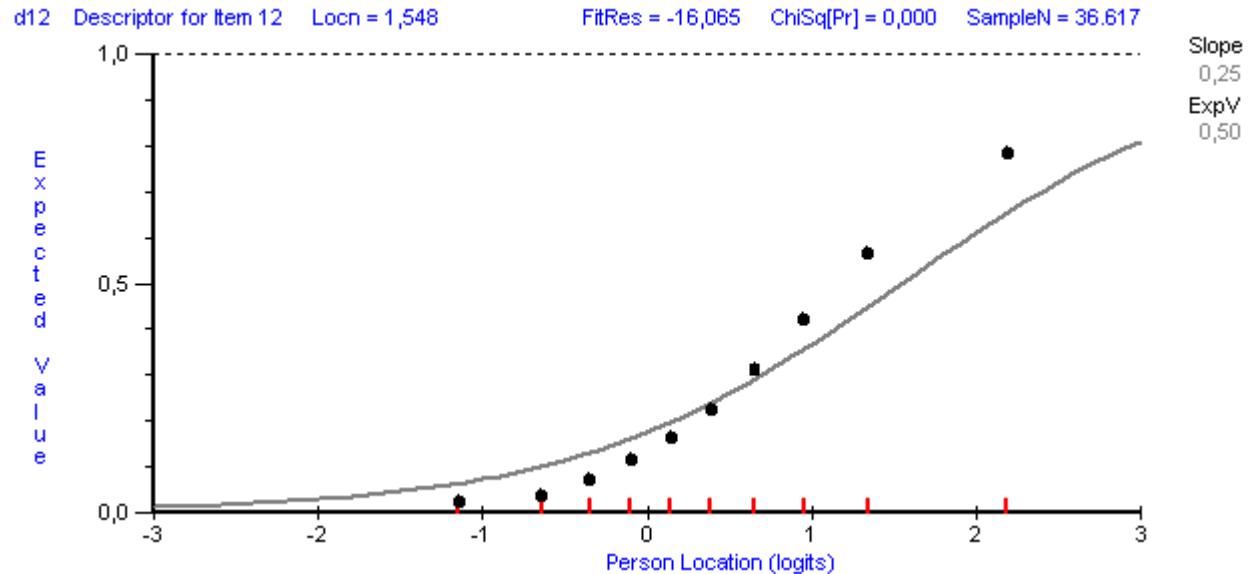
Item administered in 2014

D12. In the equation $(3k-6)x-5x+2=0$, x is the unknown quantity and k is a real number.

The solution is equal to 1 for k equal to



Item D12 Characteristic Curve plotted by RUMM2030



Item analysis carried out by ConQuest 4.0

Analysis of item D24

Cases for this item	41812	Discrimination	0.59			
Item Threshold	1.15	Weighted MNSQ	0.84			
Label	Score	% of tot	Pt Bis	t (p)	PX1 Avg:1	PV1 SD:1
0	0.00	27.64	-0.13	-26.80(0.000)	-0.19	0.83
1	1.00	27.65	0.59	148.86(0.000)	0.89	0.85
9	0.00	43.79	-0.41	-90.62(0.000)	-0.44	0.85

Note. 0 = wrong answer; 1 = correct answer; 9 = missing values.

Source: Our elaboration on SNV INVALSI data collected in 2012 at the 2nd grade of upper secondary school

Example 4

[Uncertainty and data]

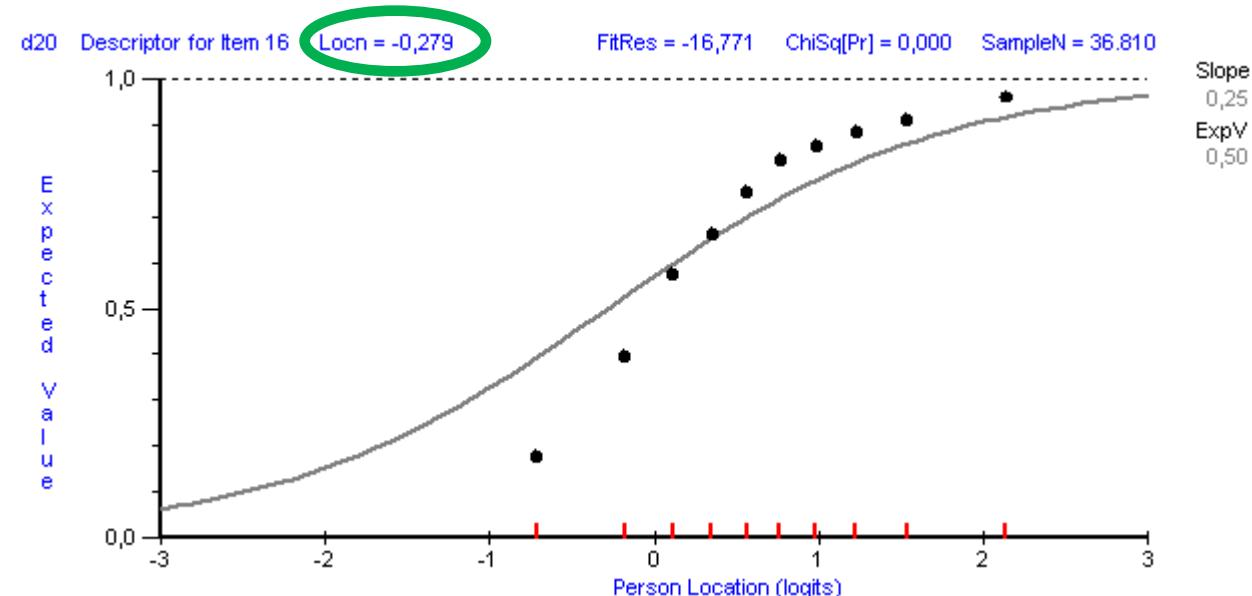
D20. From a control of quality, you noticed that a machinery has produced 14 defective pieces over a production of 1200 pieces. What estimate is reasonable to make about the number of defective pieces over a production of 2150?

Write the calculations you do to find the solution and then write the result (approximate it to the unit)

.....
.....
.....
.....

Result (approximate it to the unit):

Item D8 Characteristic Curve plotted by RUMM2030



Item analysis carried out by ConQuest 4.0



Analysis of item D17

Cases for this item	36932	Discrimination	0.57				
Item Threshold	0.87	Weighted MNSQ	0.88				
Label	Score	% of tot	Pt Bis	t (p)	PX1 Avg:1	PV1 SD:1	
0	0.00	28.1	-0.07	12.65(0.000)	-0.11	0.72	
1	1.00	31.81	0.50	110.74(0.000)	0.63	0.79	
9	0.00	39.38	-0.41	86.35(0.000)	-0.43	0.69	

Note. 0 = wrong answer; 1 = correct answer; 9 = missing values.

Source: Our elaboration on SNV INVALSI data collected in 2014 at the 2nd grade of upper secondary school

Common features, in summary

- ▶ None of these items present problems regarding the formulation (task posing), nor do they violate the local independence assumption
- ▶ All of them have a weighted MNSQ between 0.80 and 0.90
- ▶ All of them have very high number of missing answers, especially for students with low ability
- ▶ All of them have a very low guessing level – far lower than that predicted by the model
- ▶ All of items tackle different mathematical contents and cover a wide range of difficulty. Most of them are easy along the Rasch scale.
- ▶ Can require either a single short answer (such as the result of an arithmetic operation) or a longer text passage to discuss and propose a solution.
- ▶ Open-ended (→ missing values)
- ▶ Items in a context (real or mathematical)
- ▶ Always, the correct answer does not follow from a direct application of a knowledge or a procedure → skill required is recognizing in an unusual situation (*for Italian praxis*) the underlying well-known mathematical concepts

in their output

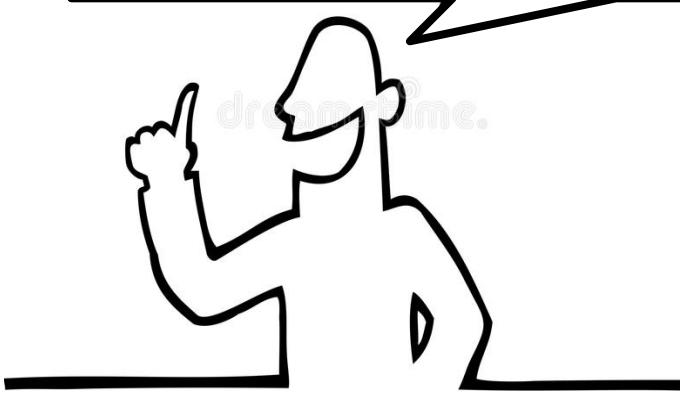
in their formulation and type

Discussion

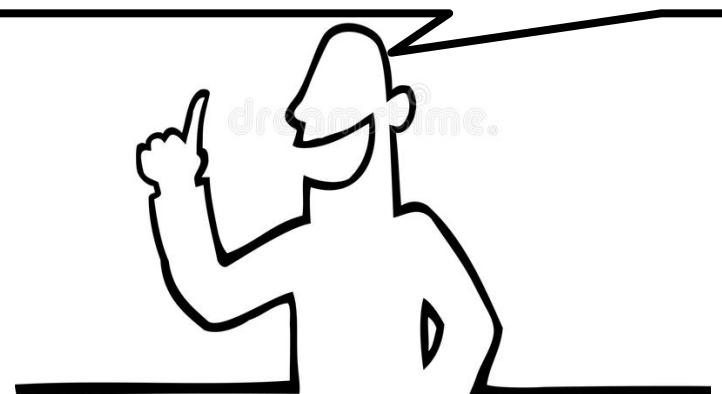
“Missing” correct answers contribute to an overestimation of the difficulty parameter.



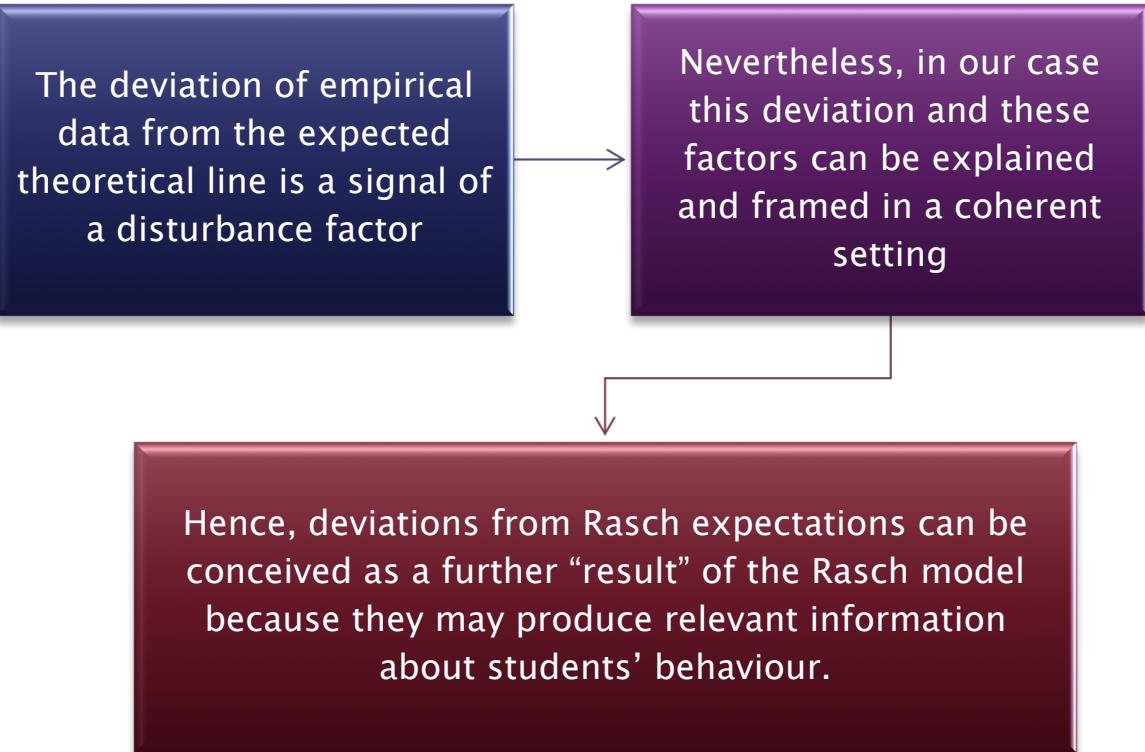
A **displacement** occur, especially for low-ability students when they encounter an “unusual” item → they do not reply or encounter (**mathematization situation**)



Limitations of the study:
Tests are not linked
Comparability based on the use of big data
Qualitative study (interviews with students) misses



Conclusions



- ▶ Just an internal report or a study of interest for a wider audience?

