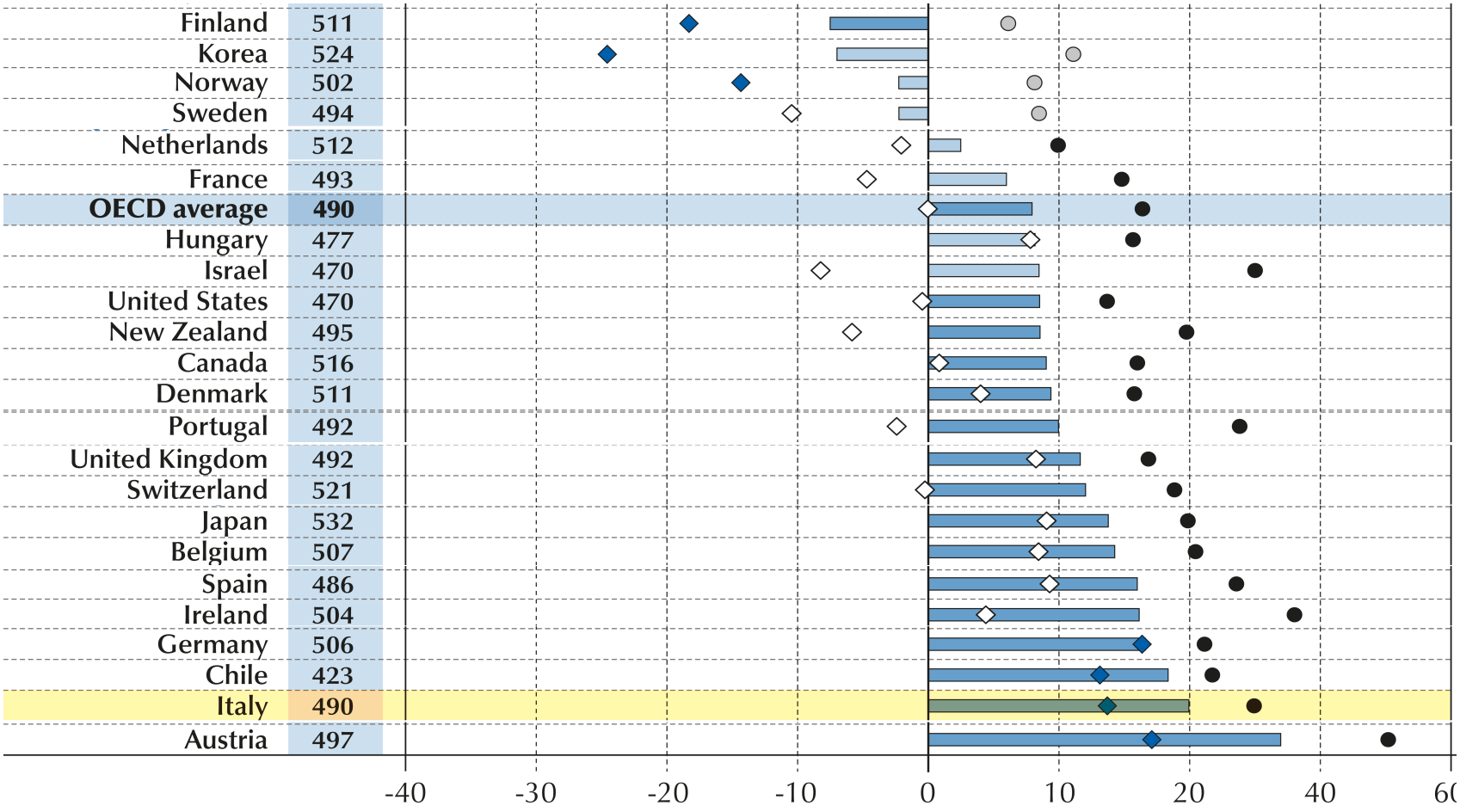
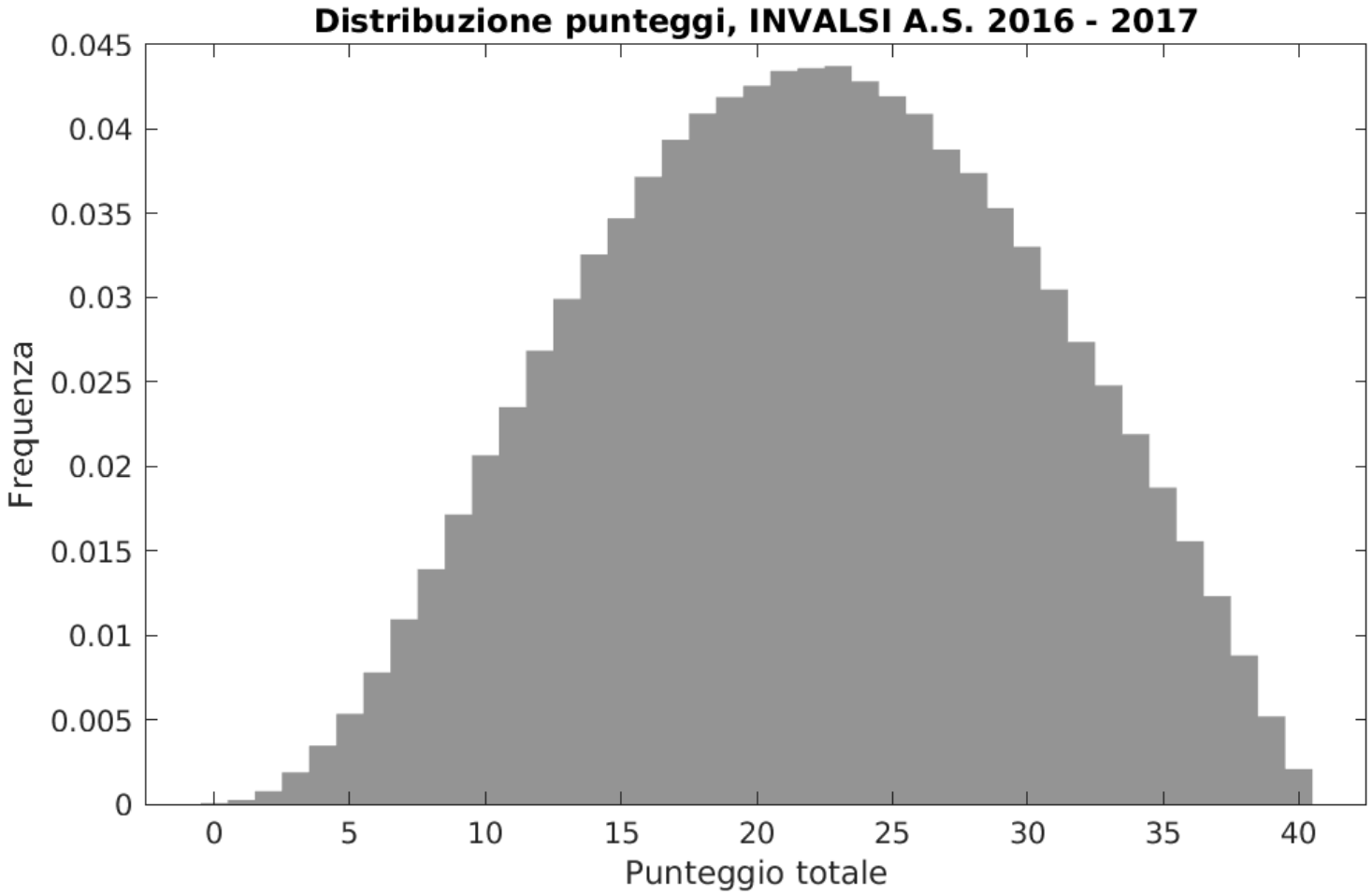
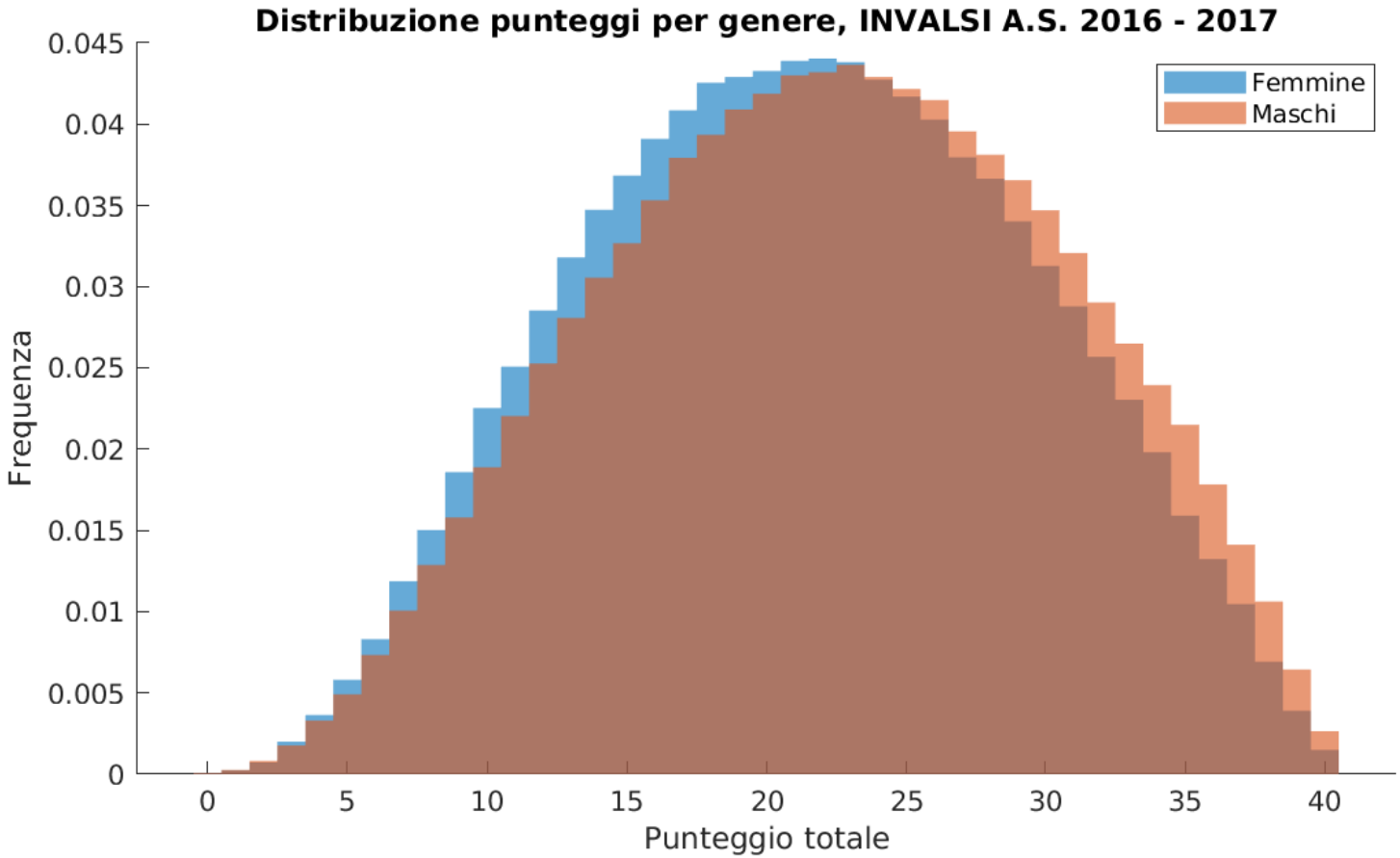
A QUANTITATIVE MODEL FOR GENDER GAP IN G8 STANDARDIZED MATHEMATICS TESTS IN ITALIAN SCHOOLS

Orlando Riccardo
ISIS Einaudi, Dalmine (BG)
orlando.riccardo.95@gmail.com
Rizzo Ottavio Giulio
Università degli Studi di Milano

VI SEMINAR
"INVALSI DATA: A TOOL FOR TEACHING AND SCIENTIFIC RESEARCH"
ROME, NOVEMBER 25TH - 28TH, 2021


Finland 511
Korea 524
Norway 502
Sweden 494
Netherlands 512
France 493
OECD average 490
Hungary 477
Israel 470
United States 470
New Zealand 495
Canada 516
Denmark 511
Portugal 492
United Kingdom 492
Switzerland 521
Japan 532
Belgium 507
Spain 486
Ireland 504
Germany 506
Chile 423
Italy 490
Austria 497
-40
-30
-20
-10
0
10
20
40
60


Distribuzione punteggi, INVALSI A.S. 2016 - 2017
Frequenza
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0
5
10
15
20
25
30
35
40
Punteggio totale


Distribuzione punteggi per genere, INVALSI A.S. 2016 - 2017
Femmine
Maschi
Frequenza
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0
5
10
15
20
25
30
35
40
Punteggio totale

What is done in the italian literature?

- Item level analysis

What is done in the italian literature?

- Item level analysis
  Interesting, but only a few items at a time

What is done in the italian literature?

- Item level analysis
  Interesting, but only a few items at a time
- Large-scale statistical analysis

What is done in the italian literature?

- Item level analysis
  Interesting, but only a few items at a time
- Large-scale statistical analysis
  Comprehensive, but hard to gain insight

The data

- Grade 8 INVALSI Math tests

The data

- Grade 8 INVALSI Math tests
- 8 years of tests (2009 - 2017)

The data

- Grade 8 INVALSI Math tests
- 8 years of tests (2009 - 2017)
- 342 items

The data

- Grade 8 INVALSI Math tests
- 8 years of tests (2009 - 2017)
- 342 items
- Split equally into a Construction Set and a Validation Set

Why grade 8?

- Need items from a single grade

Why grade 8?

- Need items from a single grade
- Gender gap grows with grade

Why grade 8?

- Need items from a single grade
- Gender gap grows with grade
- Highest grade in italian common curriculum

Why grade 8?

- Need items from a single grade
- Gender gap grows with grade
- Highest grade in italian common curriculum
- Highschool choice heavily correlated with gender

Quantifying discrimination: Differential Item Functioning

The student population is split in slices
according to their total test score.

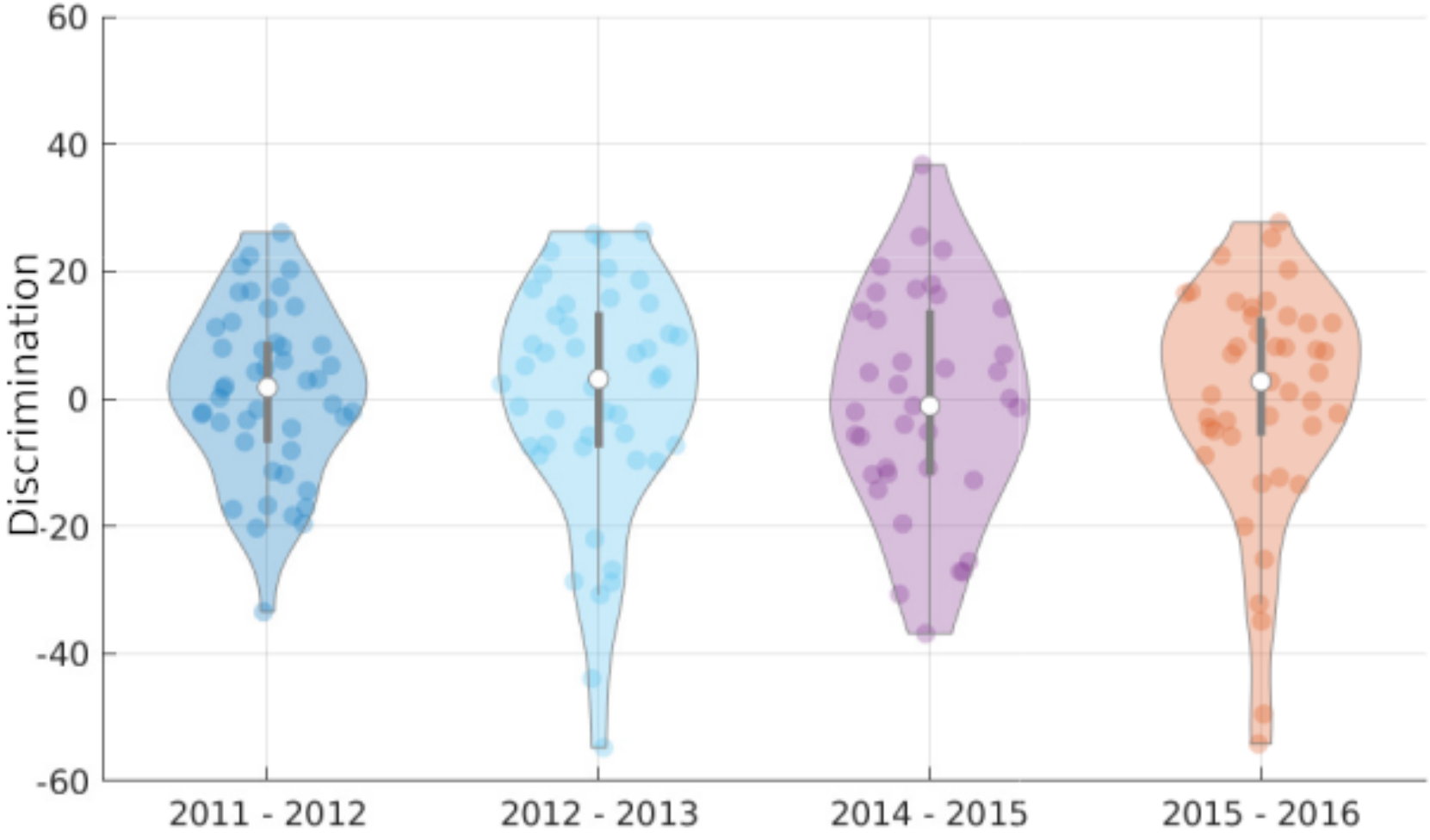
$$E = \frac{\sum_i \frac{m_{c,i} f_{e,i}}{N_i}}{\sum_i \frac{m_{e,i} f_{c,i}}{N_i}}$$

Quantifying discrimination: Differential Item Functioning

The student population is split in slices
according to their total test score.

$$E = \frac{\sum_i \frac{m_{c,i} f_{e,i}}{N_i}}{\sum_i \frac{m_{e,i} f_{c,i}}{N_i}}$$

$$D = -100 \log E$$


60
40
20
0
-20
-40
-60
Discrimination
2011 - 2012
2012 - 2013
2014 - 2015
2015 - 2016

Model assumptions

- Certain properties of items cause their discrimination

Model assumptions

- Certain properties of items cause their discrimination
- Items belong to *Categories*: $i \in C$

Model assumptions

- Certain properties of items cause their discrimination
- Items belong to *Categories*: $i \in C$
- Each Category has a discrimination value: $discr(C)$

Model assumptions

- Certain properties of items cause their discrimination
- Items belong to *Categories*: $i \in C$
- Each Category has a discrimination value: $discr(C)$

$$discr(i) = \sum_{C \ni i} d(C)$$

- d item discrimination

- $\boldsymbol{d}$ item discrimination
- M Classification matrix

- $\boldsymbol{d}$ item discrimination
- M Classification matrix
- $\boldsymbol{c}$ Categories' discrimination

- $\boldsymbol{d}$ item discrimination
- M Classification matrix
- $\boldsymbol{c}$ Categories' discrimination

$$\boldsymbol{d} = M\boldsymbol{c}$$

- $\boldsymbol{d}$ item discrimination
- M Classification matrix
- $\boldsymbol{c}$ Categories' discrimination

$$\boldsymbol{d} = M\boldsymbol{c}$$

$$\boldsymbol{d} \xrightarrow[\text{squares}]{\text{least}} \boldsymbol{c}$$

Choosing the categories

- Based on the items from the Construction Set

Choosing the categories

- Based on the items from the Construction Set
- Considering:

Choosing the categories

- Based on the items from the Construction Set
- Considering:
  - Math skills

Choosing the categories

- Based on the items from the Construction Set
- Considering:
  - Math skills
  - Language skills

Choosing the categories

- Based on the items from the Construction Set
- Considering:
  - Math skills
  - Language skills
  - Didactic contract

Choosing the categories

- Based on the items from the Construction Set
- Considering:
  - Math skills
  - Language skills
  - Didactic contract
  - Item presentation

Choosing the categories

- Based on the items from the Construction Set
- Considering:
  - Math skills
  - Language skills
  - Didactic contract
  - Item presentation

This process yielded 16 Categories.

Example categories

Standard Arithmetic

The item is an arithmetic task that can be completed using only standard procedures.

Example categories

Standard Arithmetic

The item is an arithmetic task that can be completed using only standard procedures.

Redundant information

The item contains redundant or superfluous information.

Example categories

Standard Arithmetic

The item is an arithmetic task that can be completed using only standard procedures.

Redundant information

The item contains redundant or superfluous information.

Implicit assumption

The item contains an implicit assumption, or requires a datum not explicitly mentioned in the text.

Example categories

Standard Arithmetic

The item is an arithmetic task that can be completed using only standard procedures.

Redundant information

The item contains redundant or superfluous information.

Implicit assumption

The item contains an implicit assumption, or requires a datum not explicitly mentioned in the text.

Numeric answer

The item's answer is a single number, eventually with a unit of measure, or it is clearly divided in sub-items that match this definition.

Results

On the Validation Set:

Results

On the Validation Set:

- Relative residue: 0.88

Results

On the Validation Set:

- Relative residue: 0.88
- 3 Categories with insufficient items

Results

On the Validation Set:

- Relative residue: 0.88
- 3 Categories with insufficient items
- 8 Categories with low discrimination

Results

On the Validation Set:

- Relative residue: 0.88
- 3 Categories with insufficient items
- 8 Categories with low discrimination
- 5 Categories with interesting results

Results

On the Validation Set:

- Relative residue: 0.88
- 3 Categories with insufficient items
- 8 Categories with low discrimination
- 5 Categories with interesting results

Coherent with the Didactic Contract

Coherent with the Didactic Contract

- Explain your reasoning ($D = 6.9$)

Coherent with the Didactic Contract

- Explain your reasoning ($D = 6.9$)
  - Need for formal justification

Coherent with the Didactic Contract

- Explain your reasoning ($D = 6.9$)
  - Need for formal justification
- Asymmetric distractors ($D = -9.3$)

Coherent with the Didactic Contract

- Explain your reasoning ($D = 6.9$)
  - Need for formal justification
- Asymmetric distractors ($D = -9.3$)
  - Discard extreme options

Coherent with the Didactic Contract

- Explain your reasoning ($D = 6.9$)
  - Need for formal justification
- Asymmetric distractors ($D = -9.3$)
  - Discard extreme options
- Extrapolate ($D = -6.3$)

Coherent with the Didactic Contract

- Explain your reasoning ($D = 6.9$)
  - Need for formal justification
- Asymmetric distractors ($D = -9.3$)
  - Discard extreme options
- Extrapolate ($D = -6.3$)
  - Only use values explicitly mentioned in the item's text

Example item: "Explain your reasoning" ($D = 6.9$)

D13. **Una medicina viene venduta in scatole da 28 compresse divisibili come quella in figura. Ogni compressa è da 20 mg. La nonna di Piero deve prendere tutti i giorni, per un mese, 30 mg di questa medicina.**

Per quanti giorni la nonna di Piero può prendere la sua dose giornaliera del farmaco utilizzando una sola scatola?

Scrivi come hai fatto per trovare la risposta e poi riporta il risultato.

.....

.....

.....

Risultato: ..... giorni

Example item: "Asymmetric distractors" ($D = -9.3$)

D8. **La seguente tabella riporta i dati sulla cittadinanza degli alunni iscritti a un Istituto Comprensivo:**

TIPO DI SCUOLA	TOTALE ALUNNI	ALUNNI ITALIANI	ALUNNI STRANIERI	PERCENTUALE ALUNNI STRANIERI
INFANZIA	200	120	80	40,0%
PRIMARIA	400	308	92	23,0%
SECONDARIA DI 1° GRADO	400	280	120	30,0%
TOTALE	1 000	708	292	.....

Qual è la percentuale di alunni stranieri dell'intero istituto?

A. ☐ 93,0%

B. ☐ 41,2%

C. ☐ 31,0%

D. ☐ 29,2%

Example item: "Extrapolate" ($D = -6.3$)

D12. **Giovanni versa 100 ml di acqua alla volta nel recipiente che vedi in figura fino a 800 ml. Ogni volta segna con una tacchetta sul recipiente il livello raggiunto dall'acqua.**

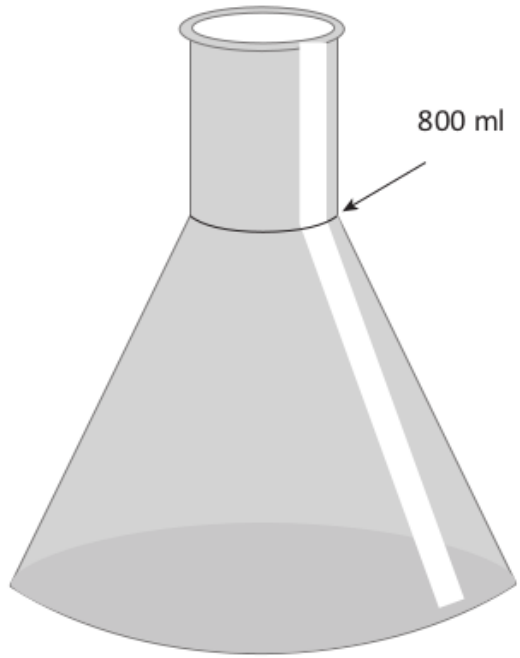
Com'è la distanza tra le tacchette segnate da Giovanni?

A. ☐ La distanza tra le tacchette diminuisce verso l'alto

B. ☐ La distanza tra le tacchette aumenta verso l'alto

C. ☐ La distanza tra le tacchette si mantiene costante

D. ☐ La distanza tra le tacchette prima aumenta e poi diminuisce



Further development

- New Categories

Further development

- New Categories
  - Based on those that worked here?

Further development

- New Categories
  - Based on those that worked here?
  - Based on the didactic contract?

Further development

- New Categories
  - Based on those that worked here?
  - Based on the didactic contract?
- Different grades

Further development

- New Categories
  - Based on those that worked here?
  - Based on the didactic contract?
- Different grades
  - Grades 2 and 5: different Categories?

Further development

- New Categories
  - Based on those that worked here?
  - Based on the didactic contract?
- Different grades
  - Grades 2 and 5: different Categories?
  - Grades 10 and 13: how to handle the differences in curriculum?

Thank you for your attention.