



Standardized test in Switzerland: discussion of process and results

**First joint international meeting of the Italian and
Spanish mathematical societies, July 2014**

Alberto Crescentini, Miriam Salvisberg, Silvia Sbaragli, Giovanna Zanolla
alberto.crescentini@supsi.ch

July 4, 2014

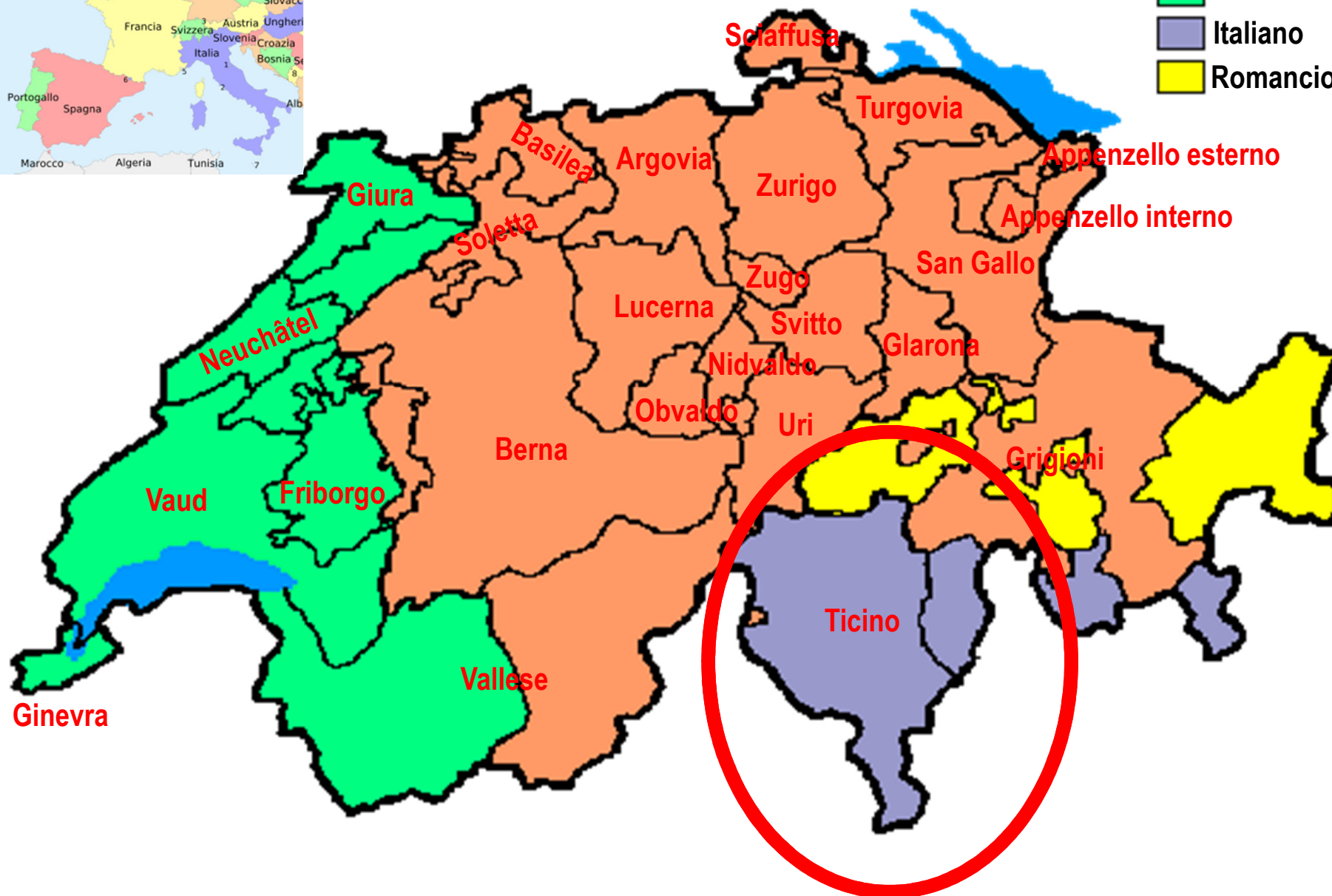
Handwritten signature and date: 1 July 2014

Beginning





- Tedesco
- Francese
- Italiano
- Romancio



Swiss and Ticino starting point



2007: several cantons sign the Harnos agreement.

2012-13: first visible changes (age of compulsory school beginning; 4).

2015-16: full implementation (common learning aims, length of compulsory school, definition of fundamental competencies, number of languages learned at school).

It is compulsory to evaluate such implementation.



No tradition of standardized test.

High level of autonomy of teachers.

Program of primary and first level secondary planned together.

The project

Since 2010 a project with the aim of producing and administering a standardized test (Woolfolk, 2007) to evaluate mathematical competencies in the fourth class of primary school has been running in Ticino.

Is the first time that such a test is produced in Ticino.

Aims:

- Give some information to the teacher to allow a reflection on a «objective» base;
- Produce a monitoring instruments for stakeholders.

Theoretical framework: Standardized test



Some words about standardized tests

- Standardized tests are administered, scored and interpreted in a standardized manner, i.e. with the same directions, time limits and scoring for all (Woolfolk, 2007).
- In the psychometrical jargon, the use of the term “standardized” when referring to tests is tautological, because a test needs to be standardized to give useful information and to be considered a test at all (De Battisti, Salini & Crescentini, 2006).
- Tests can be standard-referenced or criterion-referenced. In the first case the scores are determined by comparing how well individuals performed in the test against other individuals who took the same test. In the second case, scores are compared to a certain predetermined criterion (Popham, 2011).

And about the debate

Standardized testing is a highly controversial and widely debated topic. If, on the one side, standardized tests are a **relatively objective** tool for measuring student achievement that consumes **little class time** and produces useful information on which teachers can work, on the other side, according to some authors, they only reveal students' knowledge during the very short timeframe in which the tests are administered (Boaler, 2003).

More on the debate

- Some students may not do well in standardized tests for reasons connected to anxiety or to the time pressure (Buck, Ritter, Jenson & Rose, 2010);
- Some authors (e.g. Moses and Nanna, 2007) state that, standardized tests reflect the inequities that already exist within schools;
- Students undergo hours of intense test preparation in the classroom (Barrier-Ferreira, 2008), and teachers devote a large amount of time and resources to prepare students for the standardized test (the so-called phenomenon of “teaching to test”) (Popham, 2011).

Research



The Harnos 2010 fundamental competencies mathematic model

	Sapere, riconoscere e descrivere	Eeguire e applicare	Utilizzare strumenti	Presentare e comuni- care	Matema- tizzare e modellizzare	Argo- mentare e giustificare	Interpretare e riflettere sui risultati	Esplorare e tentare
	ASPETTI DI COMPETENZA							
AMBITI DI COMPETENZA	Numeri e calcolo		★			★		
	Geometria	★	★					
	Grandezze e misure		★					
	Funzioni							
	Dati e probabilità	★						

Selected field (“Ambiti”)

Field of competence	Aspect of competence
Data and relations	Knowing, recognizing and describing
Dimensions and measures	Executing and use
Geometry	Executing and use
Geometry	Knowing, recognizing and describing
Numbers and calculating	Arguing and Justifying
Numbers and calculating	Executing and use

Primary Ticinos' school data

- 9 neighborhoods each one with a manager called inspector
- 239 schools
- 44 directors
- 786 sections
- 8 districts
- 15'616 students (around 3'000 per year)
- 1'366 teachers

Main Phase

September-November 2010

Selection of aspects of mathematics to be evaluated

January – June 2011

Construction of items

October 2011

Pilot test → sample of 110 classes, 1591 students of the last class of primary school

→ 2 booklets one hour

March – august 2012

Item analysis; selection of the best performing items

Main phases

October 2012

Test→ all the classes (196) and students

2 test 1 hour

December 2012

Feedback to each teacher on his own class and students

Feedback to all the people involved in school management

October 2013

Final report

Pilot test



Pre pilot phase

- Selection of topics;
- Items construction.

Pupils for booklet

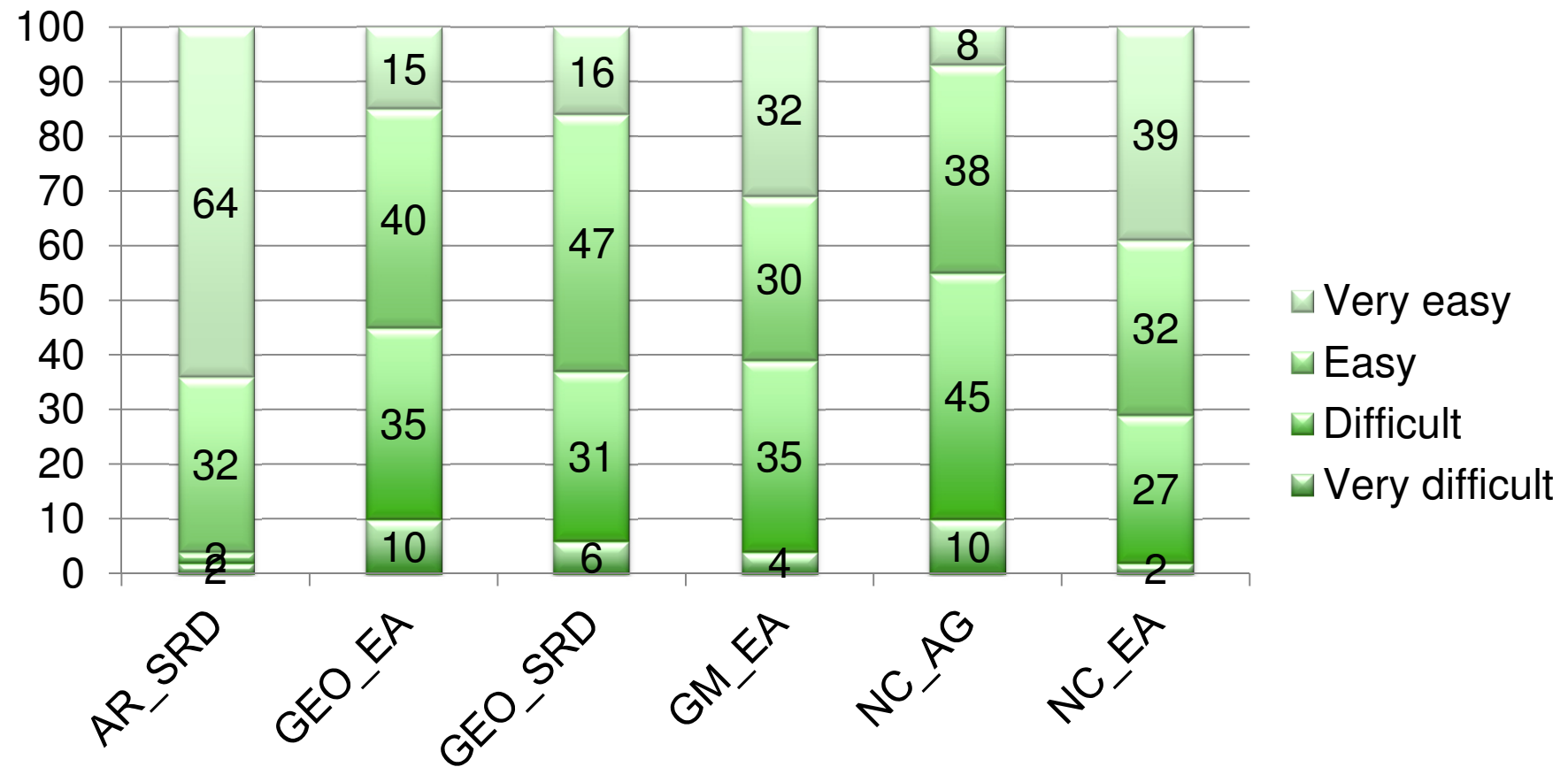
Booklet	Number of pupils			Booklet	Number of pupils		
	T1	T2	Total		T1	T2	Total
A	163	157	320	F	164	159	323
B	163	165	328	G	161	163	324
C	163	163	326	H	157	153	310
D	153	162	315	I	157	160	317
E	159	164	318	K	162	154	316

How we choose the items

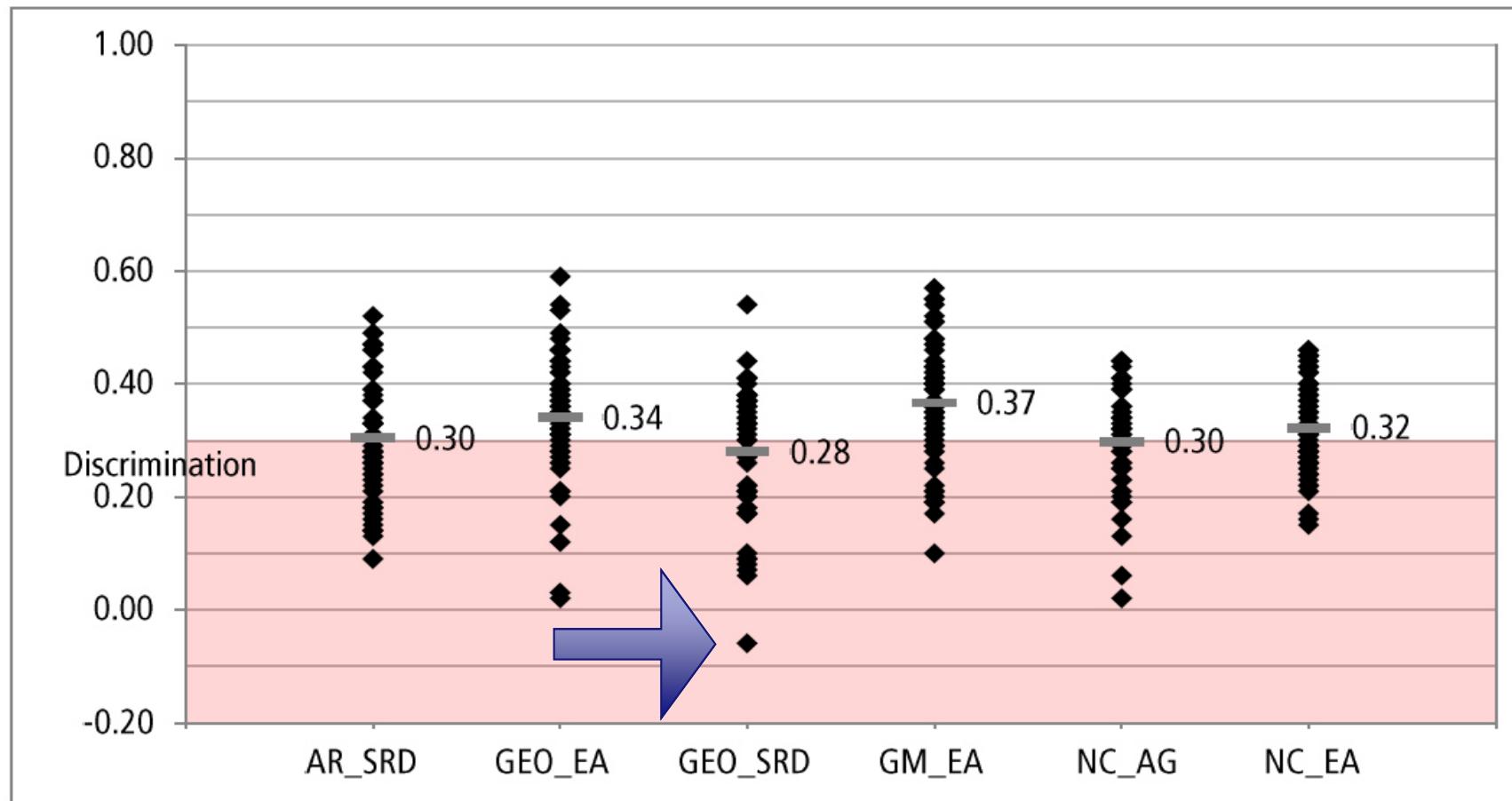
- IRT (Item response theory)
- Rasch model

A main dimension called « mathematic » has been identified and the six subdomain maintains their autonomy

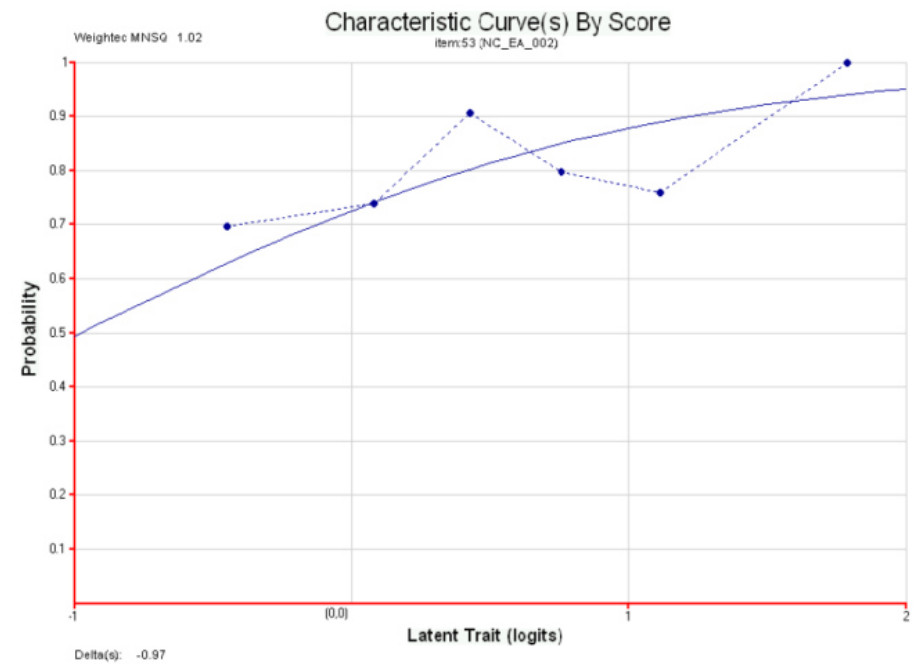
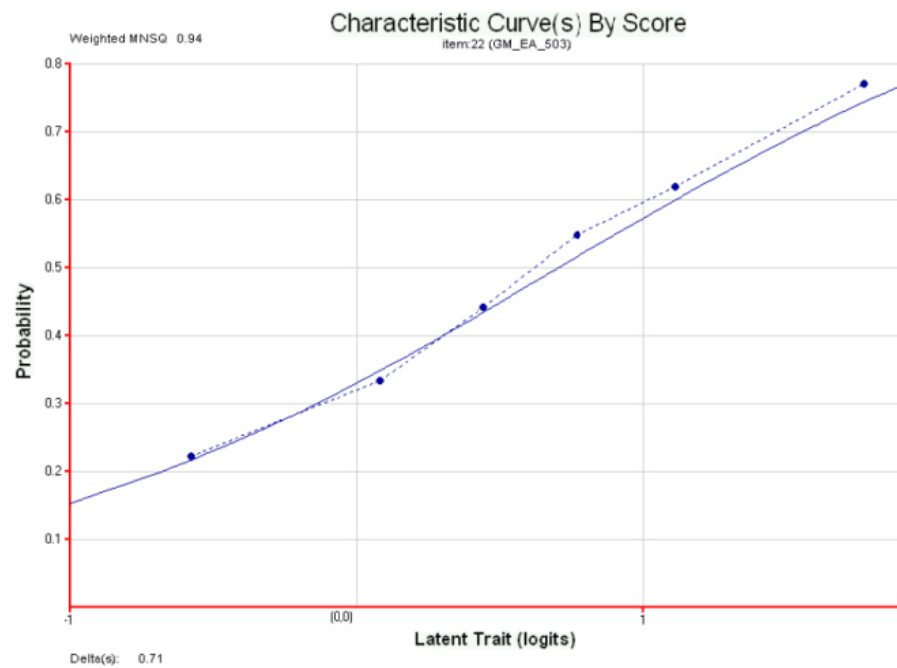
For each dimension has been identified the distribution of items



Discrimination: how to choose items (0.3-1)



Fit value



Tessin Testjahr 2011 alle Items (Februar 2012) Fri Feb 17 14:15 2012
 MAP OF LATENT DISTRIBUTIONS AND RESPONSE MODEL PARAMETER ESTIMATES

Terms in the Model (excl Step terms)		+item
3		38 72 261
	X	31 60
	XXX	119 130 171 275
2		226 84 245 30 177 260
	XXXX	278
	XXXXXXXXXX	35 36 75 81 127 160 178 211 298
1		26 46 96 172 220 27 67 80 175 41 118 120 150 179 276 293
	XXXXXXXXXX	79 148 201 300
	XXXXXXXXXX	48 62 144 161 200 213
0		16 23 93 98 246 49 59 63 70 71 221 234
	XXXXXXXXXX	29 102 106 122 206 222 290 299
	XXXXXXXXXX	15 20 99 167 227 236 243 259 279
-1		22 57 58 145 176 202 203 237 242 32 45 47 56 68 78 139 174 228
	XXXXXXXXXX	44 88 97 151 159 210 212 232 240
	XXXXXXXXXX	40 69 74 95 156 168 193 214 215
-2		11 42 43 64 109 165 223 229 250 110 128 137 197 217 224 251 256
	XXXXXXXXXX	39 61 90 107 108 180 183 194 195
	XXXXXXXXXX	1 85 91 111 116 205 209 252 277
		21 28 55 87 129 158 162 186 199 19 73 77 89 125 157 164 273 288
	XXXXXXXXXX	37 189 190 208 248 253 274 294
	XXXXXXXXXX	123 155 169 216 238 239 255 271
		25 34 113 126 154 207 225 33 65 114 170 198 249
	XXXXXXXXXX	2 13 76 117 153 187 204 218 270
	XXXXXXXXXX	52 121 173 191 192 254 268
	XXX	17 50 82 94 147 219 296
	XXX	24 53 54 92 166 235 267 284 297
	XX	3 112 124 141 149 244 285
	XXX	14 115 188 281
	X	184
	X	103 280
		12 18 134 138 140 233 241 257 5 7 66 146 152 282
	X	86 101 104 131 181 196
		51 182 185
		100 9 136 105 266
	X	8 83
		230 231 262 263 264 265 283
		4 6 10 132 133 135 142 143 163

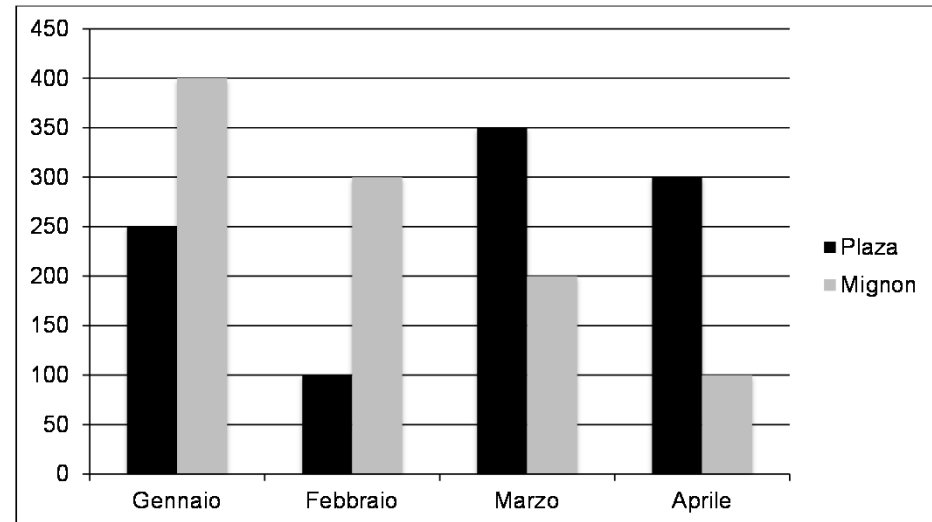
Each 'X' represents 2.7 cases
 Some parameters could not be fitted on the display

Test



Data and relations; Knowing, recognizing and describing

This graph represents the number of viewers of Plaza and Mignon cinemas during the first four months of the year



Which one of the following statements is true?

- a) During February and April Plaza has had the same number of viewers
- b) In March the two cinemas have had the same amount of viewers
- c) During the four months together the two cinemas have had the same number of viewers
- d) Each month Mignon has had more viewers than Plaza

Numbers and calculating; Arguing and Justifying

Carla must execute mentally 123×3

She make $1 \times 3 = 3$; $2 \times 3 = 6$ and $3 \times 3 = 9$

Then she write the results one following the other and she find the number 369 that is the result of the calculus

The Carla's procedure don't work always.

Justify this statement writing a calculus in which the procedure don't work

.....

.....

.....

.....

.....

- 2 booklets of 60 items
- 20 for each “ambito”
- Administered to all the population
- A week between the first and the second booklet
- Each item has a specific weight on the basis of the difficulties evaluated during the previous analysis

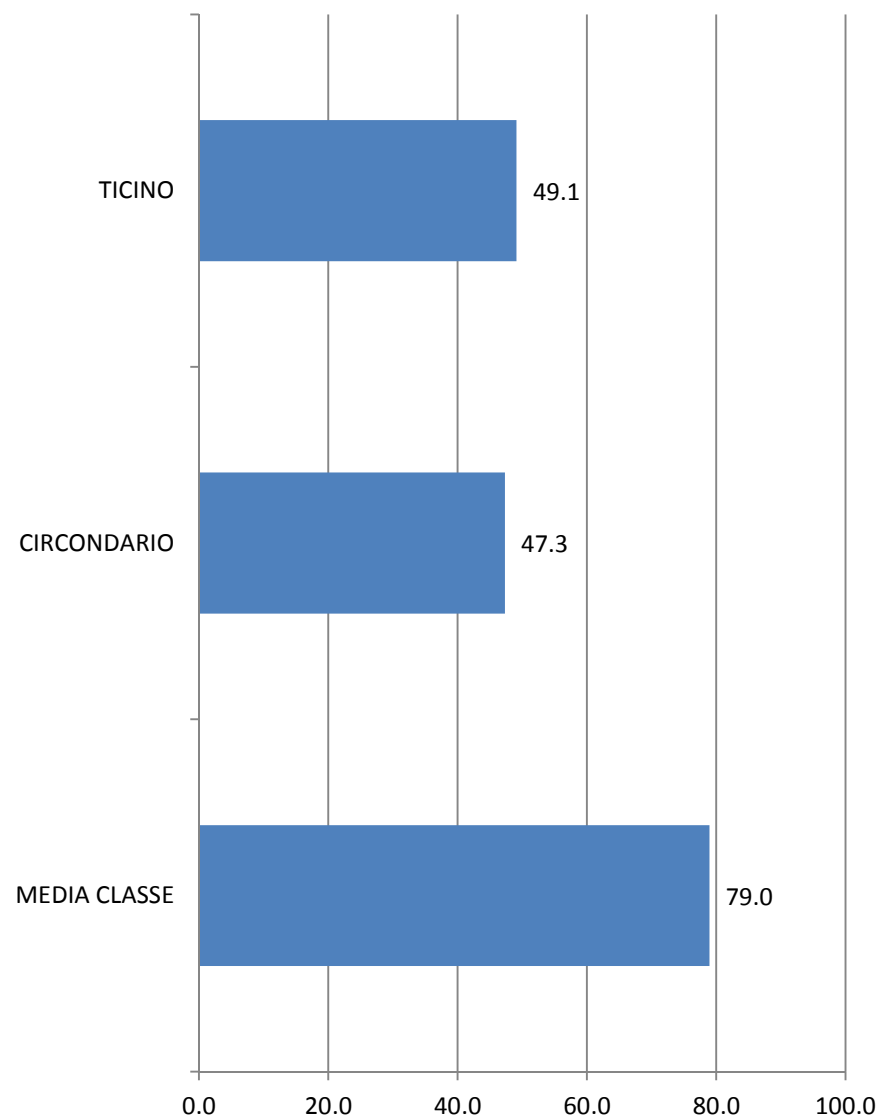
Feedback



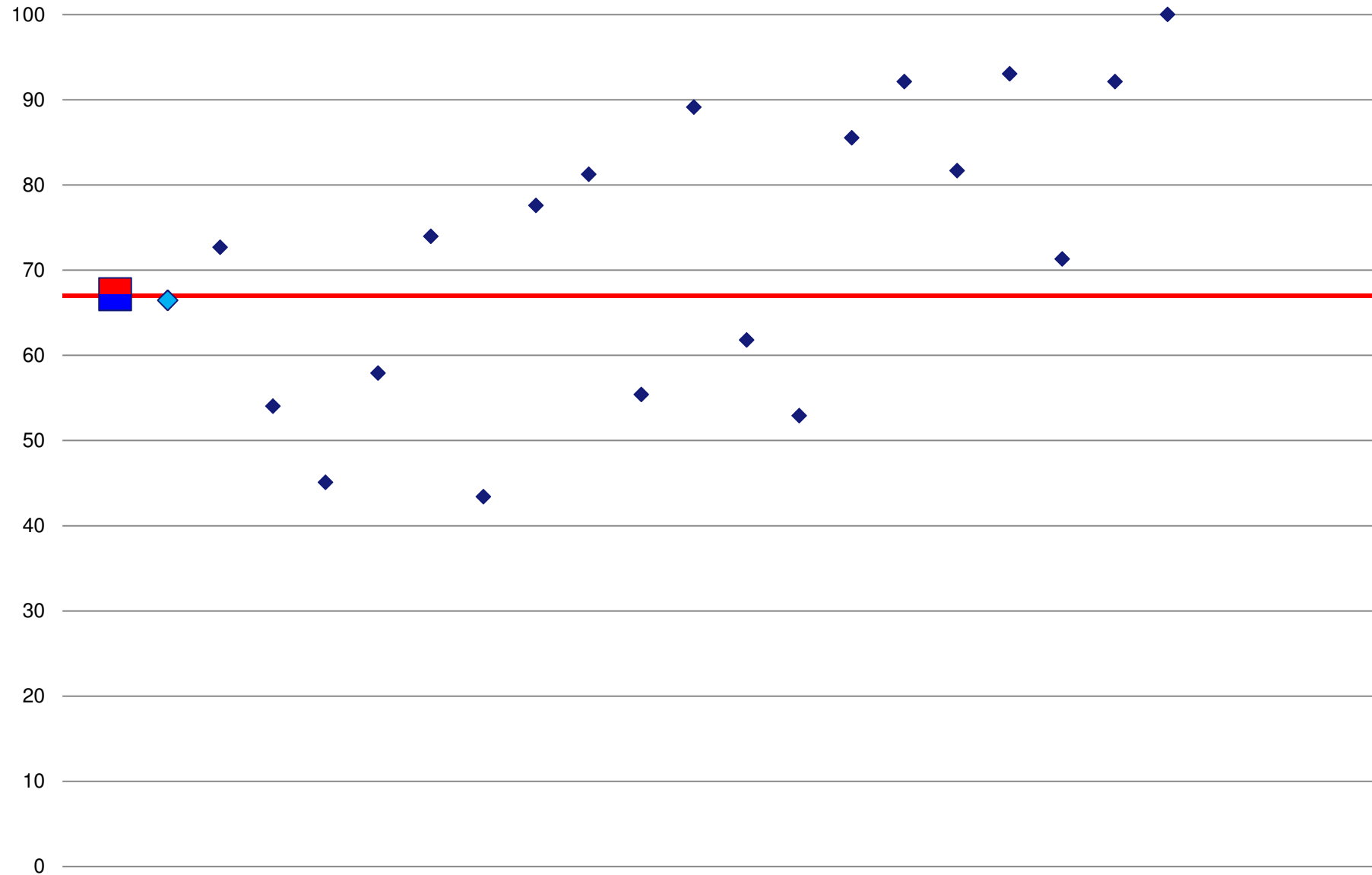
Output

- Teachers feedback → **report**
- Stakeholder report

Mathematics: Class mean, region, Canton



Data and relations - Knowing, recognizing and describing



Some empirical results



1. General Results

1.1 Mean points of the six «Ambiti» and in «Mathematic».

	Data and relations; Knowing, recognizing and describing	Geometry; Knowing, recognizing and describing	Numbers and calculating; Executing and use	Mathematic	Geometry; Executing and use	Dimensions and measures; Executing and use	Numbers and calculating; Arguing and Justifying
Mean	67.23	51.39	50.25	48.90	48.57	44.62	41.16
Standard deviation	22.50	19.48	21.56	17.59	21.04	22.99	19.64
N	2929	2929	2909	2903	2909	2910	2929

2. The context

2.1 Mean point of the 9 regions

Region		Data and relations; Knowing, recognizing and describing	Geometry; Executing and use	Geometry; Knowing, recognizing and describing	Dimensions and measures; Executing and use	Numbers and calculating; Arguing and Justifying	Numbers and calculating; Executing and use	Mathematic
1	N	301	299	301	299	301	299	299
	Mean	64.08	40.32	45.76	38.44	36.51	44.79	43.04
2	N	240	240	240	240	240	240	240
	Mean	69.44	53.73	58.34	47.36	44.81	50.75	52.58
3	N	335	329	335	330	335	329	328
	Mean	64.29	48.91	49.10	42.01	39.22	48.76	47.27
4	N	398	396	398	396	398	396	395
	Mean	66.69	48.18	50.43	44.98	41.11	47.96	48.31
5	N	311	308	311	308	311	308	308
	Mean	69.00	47.06	51.99	42.93	40.14	48.92	48.06
6	N	313	314	313	314	313	314	313
	Mean	66.61	46.93	49.49	42.84	42.66	53.20	48.59
7	N	389	385	389	385	389	385	385
	Mean	68.75	50.44	50.94	47.45	41.91	52.58	50.39
8	N	407	403	407	403	407	403	401
	Mean	69.33	52.14	55.79	49.66	43.73	53.37	52.60
7/4/2014 9	N	235	235	235	235	235	235	234
	Mean	66.53	49.04	51.19	44.07	39.91	51.27	48.68

2.2 Urbanization

Only in the case of “Geometry; Executing and use” there is a significant difference with a better result for child from the smaller villages.

3. School

3.1 Dimension

Only in the case of “Geometry; Executing and use” there is a significant difference with a better result for child from the smaller institutes (with less class).

4. Class

4.1 Dimension

No significant relation between number (12 to 24) of students and performance

4.2 multi-class

In «Numbers and calculating; Executing and use» there is a significant difference and students from multi-class have better results.

4.3 teacher gender

Teacher gender seem to be not relevant to differentiate performances if we consider the whole population.

Female child have better performance in «Numbers and calculating; Executing and use» if they have a male teacher.

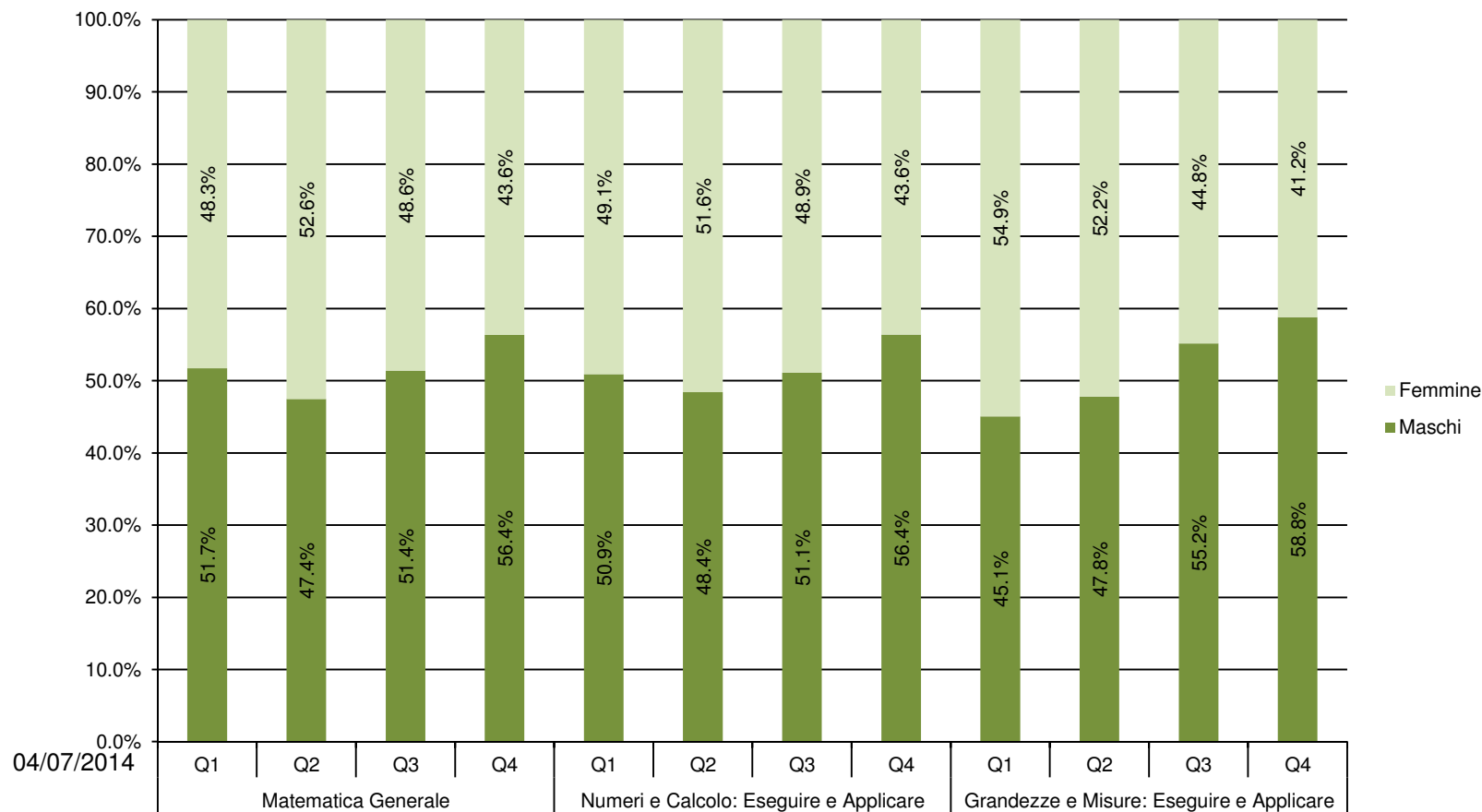
4.4 teachers experience

There are none significant relations

5. Student

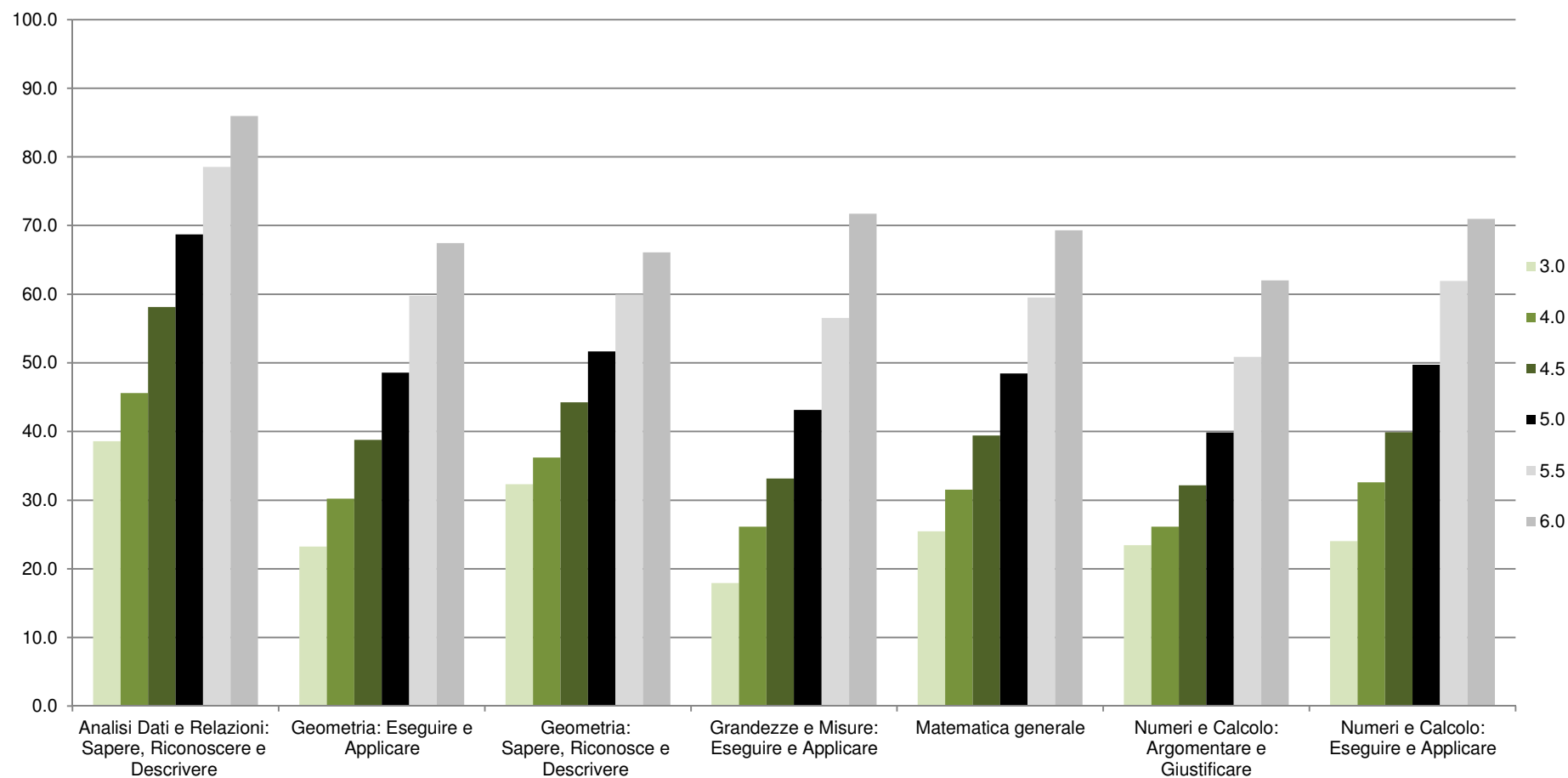
5.1 gender

Better male performances in: «Dimensions and measures; Executing and use» and «Numbers and calculating; Executing and use»



5.2 End year evaluation and test scores

Good correlation



5.3 Students age

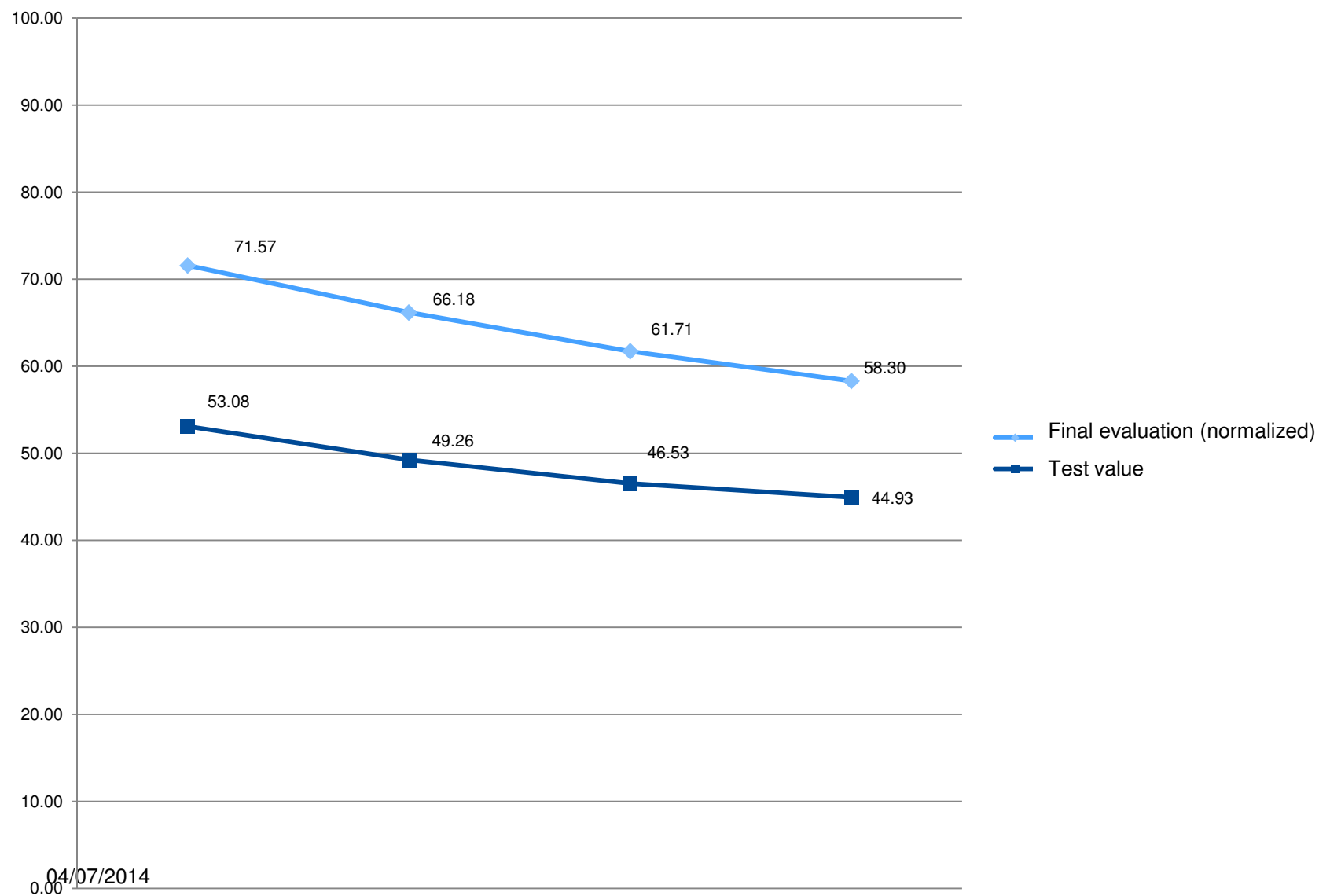
Negative correlation between age and performance.

		Data and relations; Knowing, recognizing and describing	Geometry; Executing and use	Geometry; Knowing, recognizing and describing	Dimensions and measures; Executing and use	Numbers and calculating; Arguing and Justifying	Numbers and calculating; Executing and use	Mathematic
Age	Pearson	-.198**	-.151**	-.153**	-.143**	-.148**	. -.145**	-.182**

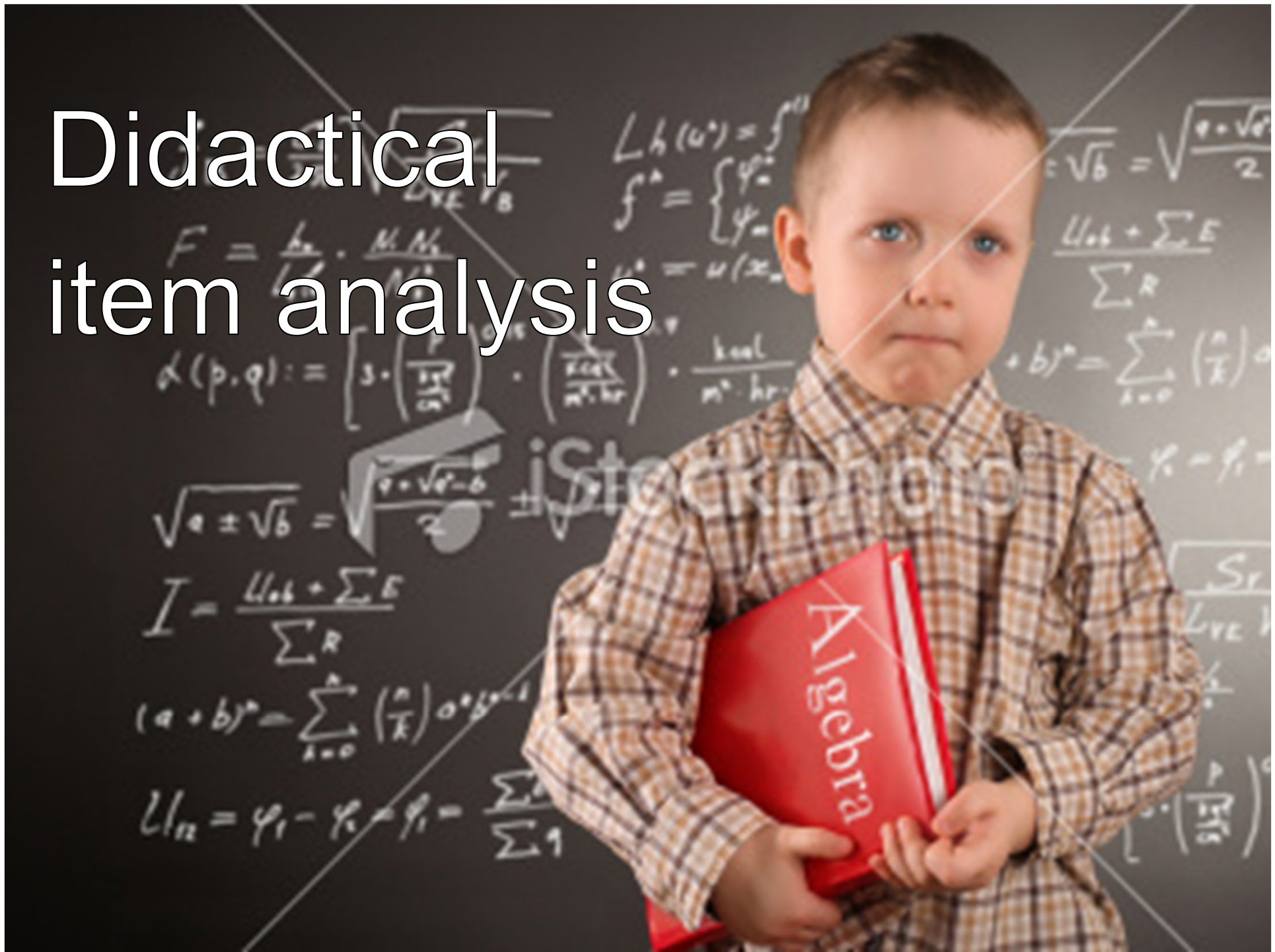
5.4 Socio economical status

Background		Data and relations; Knowing, recognizing and describing	Geometry; Executing and use	Geometry; Knowing, recognizing and describing	Dimensions and measures; Executing and use	Numbers and calculating ; Arguing and Justifying	Numbers and calculating; Executing and use	Mathematic
High status	Mean	71.31	54.56	53.2	45.3	54.05	49.58	53.08
Middle high status	Mean	68.97	52.62	48.38	42.26	49.62	44.09	49.26
Middle low status	Mean	65.46	49.43	46.49	37.83	47.95	42.4	46.53
Low status	Mean	61.91	49.59	45.97	36.05	45.95	39.86	44.93

End year evaluation and test results



Didactical item analysis

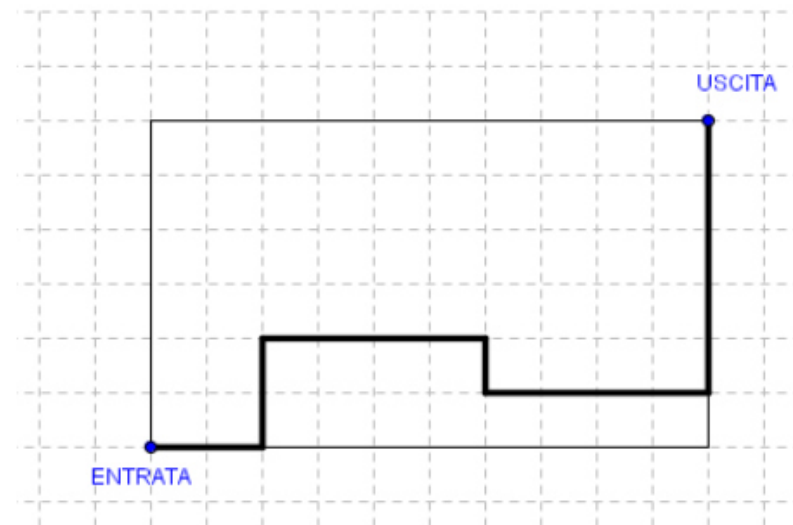


In the two draws below the rectangle represent the map of a park. Roberto and Giada walk through the park following two different paths indicated by the dark line. Both entered from «ENTRATA» and go out from «USCITA»

ROBERTO



GIADA



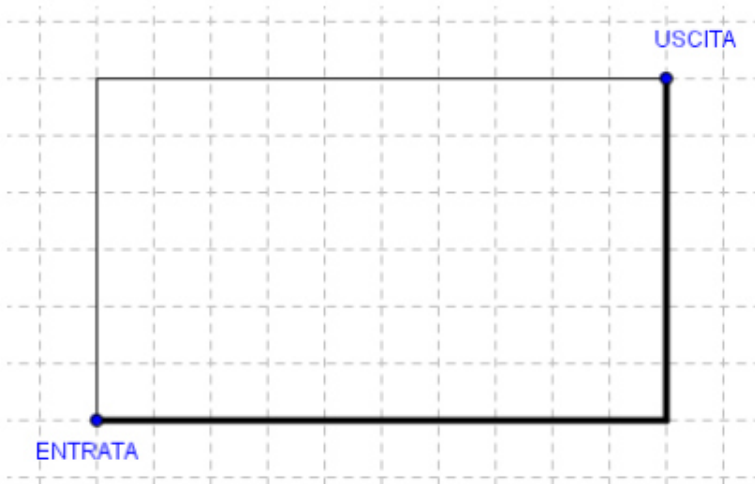
Who make the shortest path?

- a) Roberto
- b) Giada
- c) The two paths are equal

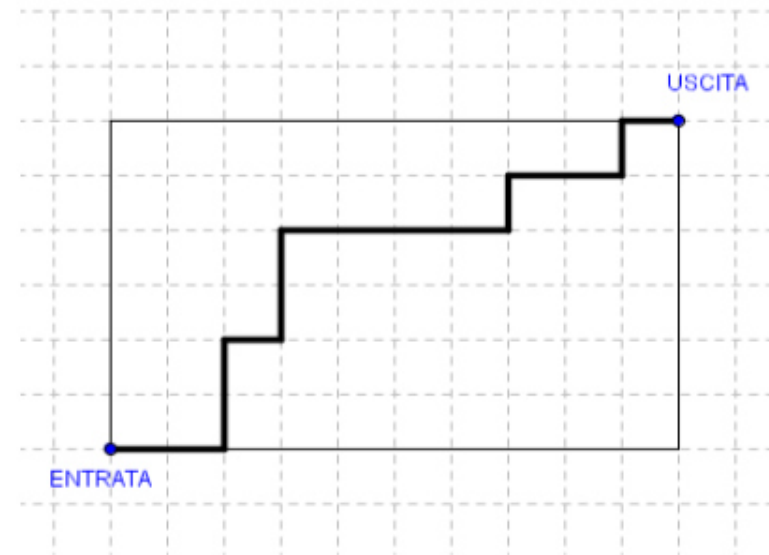
a	b	c	Missing/ not valid
63,9	16,5	17,5	2,1

In the two draws below the rectangle represent the map of a park. Fabio and Matteo walk through the park following two different paths indicated by the dark line. Both entered from «ENTRATA» and go out from «USCITA»

FABIO



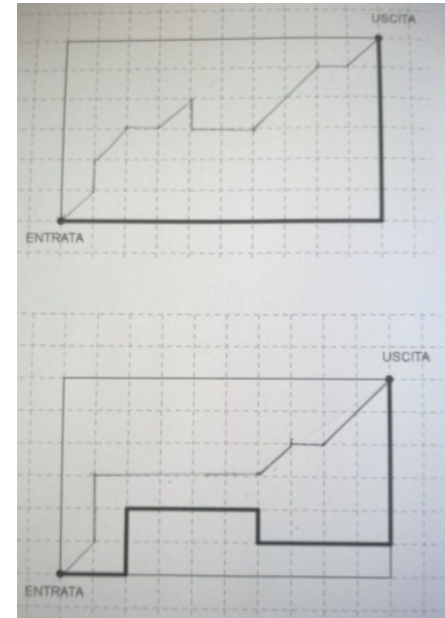
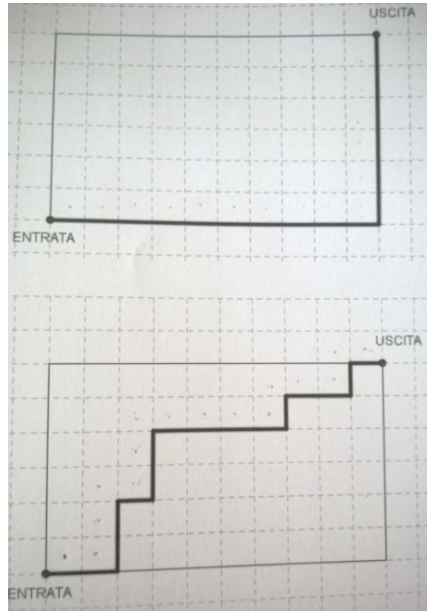
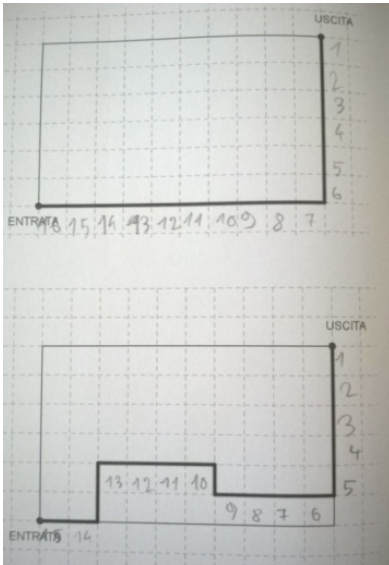
MATTEO



Who make the shortest path?

- a) Fabio
- b) Matteo
- c) The two paths are equal

a	b	c	Missing/ not valid
24,9	33,0	39,1	3,0



B5) What is the perimeter measure of an equilateral triangle with a side of 13 cm?

.....

.....

.....

Possible correct answer

$$13 \times 3 = 39 \text{ (cm)}$$

Results

39 cm o equivalenti	39	13 cm	26 cm	52 cm	Altro	
40,2	14,0	2,4	1,2	1,0	15,0	26,2

Quanto misura il perimetro di un triangolo equilatero con il lato di 13 cm?

$$P: 13 \times 3 = (10 \times 3) + (3 \times 3) = 39 \text{ cm}$$

R: Il perimetro di un triangolo equilatero con i lati di 13 cm è di 39 cm.

Quanto misura il perimetro di un triangolo equilatero con il lato di 13 cm?

$$P = 1 \times 3$$

$$P = 13 + 13 + 13$$

$$P = 39 \text{ cm}$$

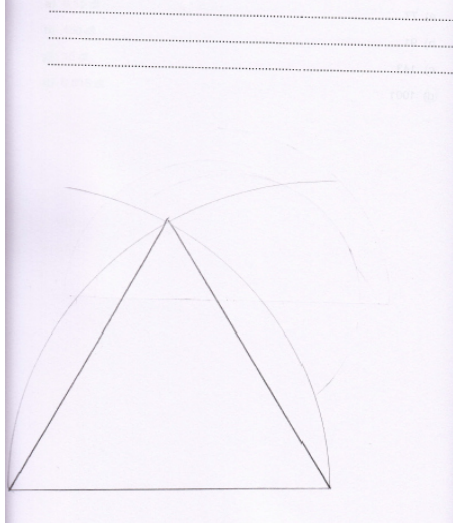
Quanto misura il perimetro di un triangolo equilatero con il lato di 13 cm?

60°

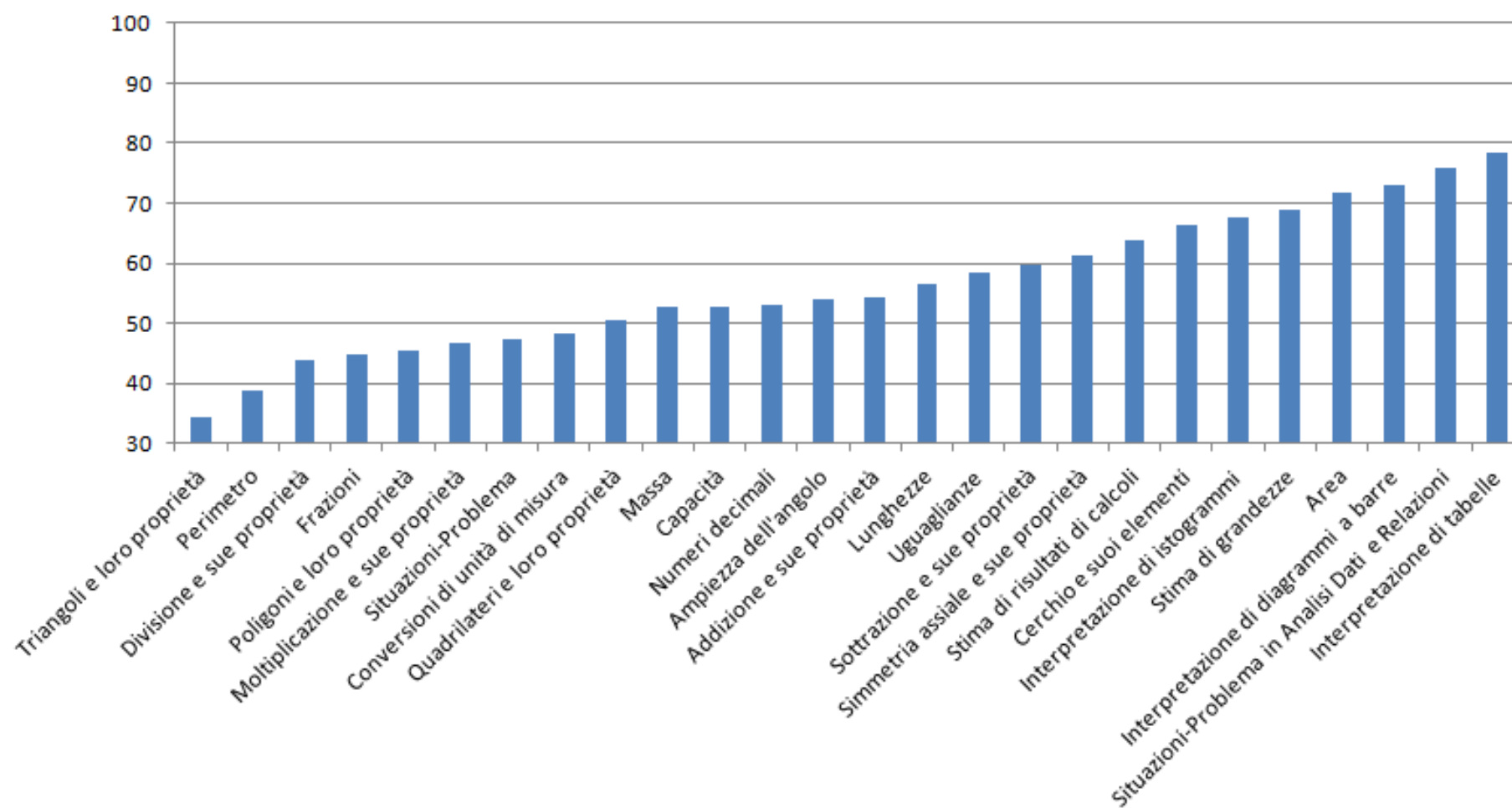
Quanto misura il perimetro di un triangolo equilatero con il lato di 13 cm?

Non si può dire perché il triangolo equilatero è disordinato.

Quanto misura il perimetro di un triangolo equilatero con il lato di 13 cm?



Correct answers on specific arguments (percent)



- Problems in geometrical topics;
- More right answer on triangle than on quadrilateral;
- Lack on language mastering;
- Several problems connected with cognitive more than disciplinal weakness.

Conclusions and perspectives



Conclusions and perspectives (1)

- The information collected stimulate the reflections about system.
- Good collaboration with schools and stakeholders.
- It is possible to identify a difficulty in the argumentation process. It is important to stimulate the work on this topic in the didactical praxis.

Conclusions and perspectives (2)

- There are several difficulties in administer the geometrical items;
- Other projects:
 - develop a regression model to go in deep in the explanation of the differences.
 - Test for the last class of primary school to have a fixed data to compare with secondary school information, fundamental in the Harmos perspective

Conclusions and perspectives (3)

- Reflections:
 - It is important to work on the feedback and in the developing of a system that can help teachers in using the results.
 - It is possible to implement the system of item production.
 - Several meeting of dissemination are in progress with all the stakeholders

Thanks for your time

