TEACHING STATISTICS AT SECONDARY EDUCATION IN ITALY: SOME ISSUES ON LARGE SCALE STANDARDIZED TEST RESULTS

Stefania Mignani[1], M. Gabriella Ottaviani[2] and Roberto Ricci[3]
[1]Department of Statistical Sciences, University of Bologna, Italy
[2]Università di Roma, Sapienza, Italy
[3]National Institute for the Evaluation of the Educational System, Italy
stefania.mignani@unibo.it

This paper focuses on recent issues in statistics learning outcomes coming from the national large-scale standardized tests administered to all students in 2013 by the National Institute for the Evaluation of the Educational System *(INVALSI) at the end of 10th grade. Utilizing a representative sample of around 32000 units, we study several items assessing statistics skills that are part of the test concerning mathematics. The INVALSI data set allows replying issues as: which kind of contents do students learn with more difficulty? Do males perform better than females? Does the level of achievement differ among type of schools? Do the results show differences between mathematics and statistics outcomes according school types? The answers to these issues help to explain the quality of student learning with the view to improve taught curricula and suggest more effective pedagogical strategies for statistics.*

INTRODUCTION

In Italy since 2001 there was a proposal carried on by the Italian Mathematics Union and the Italian Statistical Society under the auspices of the Ministry of Education integrating the teaching of Statistics and Probability into the mathematics programs at all school levels under the domain “Data and predictions” (Ottaviani, 2004). More recently in 2010 the Ministry of Education after the reform of primary and secondary school has officially redefined the learning objectives for each curriculum. On that occasion Italian National Guidelines introduced Statistics and Probability contents for each school grade recognizing the importance for young people to acquire skills that are crucial for the professional and cultural challenges of the new century.

At the same time, the community at large demands that schools provide evidence of acceptable standards of students performances. To reach this aim the Policy makers made compulsory to assess learning achievements through national large-scale standardized tests at various stages of schooling, particularly in the areas of language and numeracy. The main goal is to compare the educational curriculum with the achieved one, where the gap should depended on the taught curriculum. Through assessment feedback students are able to measure their own progress and teachers are able to offer guidance and help. The purposes are to identify intervention programs and develop appropriate teaching strategies.

Large-scale summative assessments focus on central aspects of learning in a domain as identified by national guidelines and informed by cognitive research and theory. Large-scale assessments are designed to provide reliable and comparable scores for individuals. To meet these kinds of demands, designers typically create assessments that are given at a specified time, with all students being given the same (or parallel) tests under strictly standardized conditions. Tasks are generally of the type that can be presented in paper-and-pencil format that students can answer quickly.

The emphasis in this paper is on recent issues in statistics and probability learning outcomes. In particular we focus on the results of assessments coming from national large-scale standardized tests administered in 2013 by the National Institute for the Evaluation of the Educational System (INVALSI) at the end of the second year of upper school (10^{th} grade students). The tests assess mathematics and Italian language skills of students. In the mathematics test several items allow to evaluate statistical competences.

Considering both the outcomes of student achievements and the characteristics of the items, this paper tries to answer the following key questions: which kind of contents do students learn with more difficulty? Do males perform better than females? Does the level of achievement differ among type of schools? Do the results show differences between mathematics and statistics outcomes according to kind of school?

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

LEARNING OF STATISTICS IN THE UPPER SECONDARY SCHOOL

In Italy, upper secondary school (high school) lasts 5 years and it is divided into three cycles: the first two last 2 years and the third lasts one year with final examination. The programs are defined according to the type of school: academic, technical and vocational, and for each cycles the learning objective are established by national guidelines (Ministry of Education, 2011a, 2011b). “Data and predictions” belong to all types of school, and have been introduced as a domain within mathematics.

In this paper we consider the learning objectives for the first cycle of high school (10th grade) as defined by INVALSI:

- to develop and implement a plan to collect and organize data; to select and use appropriate statistical methods; to evaluate and reflect on the procedures;
- to choose the appropriate measure of central tendency or variability;
- to understand and apply basic probability concepts; to master with the notion of chance and uncertainty;
- to build tree diagram and using combinatorial analysis to calculate probability and conditional probability.

These contents seem to be consistent with recent researches in statistics education (see for example Garfield and BenZvi’s operative proposals, 2009, or the didactic units on "Dati e previsioni" in the M@t.abel project) which suggest the following categorization of cognitive statistical learning outcomes: a) statistical literacy, understanding and using the basic language and tools of statistics; b) statistical reasoning, reasoning with statistical ideas and making sense of statistical information; c) statistical thinking, recognizing the importance of examining and explaining variability and connecting data analysis to the larger context of statistical investigation. However, point c) is not an actual goal for Italian schools.

THE INVALSI TEST

The use of standardized tests to assess students’ learning and compare pupils’ performances nationally and across countries is widespread, as large-scale assessments such as PISA, TIMSS, and PIRLS demonstrate. In the Italian context, the use of standardized assessment has assumed an increasing importance only recently, thanks to the annual surveys conducted by INVALSI at different school levels (http://www.invalsi.it/). Since the scholastic year 2007/2008, the INVALSI has developed standardized national tests to assess pupils’ reading comprehension, grammar knowledge and mathematics competency, and has administered them to the primary school students (2nd and 5th grade), and lower secondary school students (6th and 8th grade). Since the school year 2010/2011 the test was administered also to the students at the second year of upper secondary school (10th grade). As already said, we focus on the mathematics tests for 10th grade. In detail, the item domains deal with functions and relationships, geometry, numbers, data and predictions. The last domain includes statistical and probability questions, that are the objects of our analysis. Several types of items are designed: multiple-choice with four alternatives with only one correct answer, true-false, and open-ended items that ask students to give a univocal (numerical or qualitative) answer, or the justification for the answer. The items are formulated on the basis of the learning objectives described before also according to the guidelines of the international surveys (PISA and TIMSS) about statistical literacy with the aim to capture the extent of students’ statistical reasoning.

DATA AND ANALYSIS METHODS

Since the school year 2007-2008, the analyses carried on by INVALSI have been progressively increased to give more information about several aspects. Our paper regards the results of the last test that was administered in school year 2012-2013. The analyses concern both the students’ performances and the psychometric properties of each item. In particular, we use some standard tools of the Classical Test Theory. Also an Item Response Theory (IRT) approach (Hambleton *et al.*, 1991) is applied to assign a score measuring student’s ability and to classify each item difficulty. The focus of IRT is on the specification of the relationship between item psychometric properties (such as difficulty and discrimination) and the latent, non-observable, trait (ability) measured by the candidate's responses. IRT models express the probability of an item response as a function of the latent student's variable and the item properties.

The most common IRT models are based on the one-dimensionality assumption, according to which responses to a set of items can be explained only by a single latent ability on the local independence assumption, which implies that item responses are uncorrelated and statistically independent, conditional to the latent trait.

IRT models can be distinguished on the basis of the number of item parameters affecting response probabilities. In this paper we consider the well-known Rasch model (one-parameter logistic model) for binary data (i.e. correct/incorrect responses) and use the expected a posterior (EAP) procedure to estimate ability for each student. For each item the estimates of the "difficulty parameter" ranges between -4 and +4 where higher values mean more difficult items. The student's achievement scale obtained typically varies between -3 and 3, due to the normality assumption of ability.

THE RESULTS

The test was administered to all the students attending the second year of high school (560487 students in 26200 classes). We consider a sample built by INVALSI through a complex sampling. Within the whole population INVALSI draws a sample (38274 students in 2.575 classes) where the test administration is entirely ensured by external persons, so that test conditions are very similar for all students.

Table 1 presents some characteristics of the students (gender and type of school) that we have successively taken into account for interpreting the learning outcomes.

Table 1. Percentage of students by gender and by type of school

Female	49,1%
Male	51,9%
Academic	42,9%
Technical	23,6%
Vocational	33,5%

The mathematics test is composed of 54 items classified by the four domains: geometry 12, functions and relationships 12, data and predictions 13, numbers 17. The time allotted for answering the full test is 90 minutes.

We carried on analyses for the 13 statistics items and for 41 mathematics items separately. With reference to data and predictions, table 2 reports some descriptive indicators for each item. Under item, number and letter allow connection with the full test that can be read in Italian at the site http://www.invalsi.it/areadati/SNV/12-13/strumenti/SNV2013_MAT_10_Fascicolo1. The second column contains the percentages of the correct answers and the last one the difficulty parameter estimates based on the Rasch model. As we said, the answers have been dichotomized correct/not correct. The items are classified in groups according to the contents and the required task, and they are ordered by the difficulty parameter in ascending order within each group.

The test has an appropriate balance between items assessing probability and statistical concepts and includes items with problem context and real data. Besides some items require students to provide interpretations of data analysis as well as justifications for their analysis and conclusions.

As regards the difficulty of the topics we can observe interesting evidences. The easiest contents are reading and reasoning on graphs and applying elementary concepts of probability, as remarked by their low difficulty parameters. The items requiring to do a comparison using relative change formula (20b) and to recognize the necessity to calculate an harmonic mean (14), are the most difficult.

Table 2. Descriptive indicators for each item

Item	% correct answers	Contents/tasks	Difficulty
4b	84%	Reading a graph	-1.86
4c	69%	Reading a graph	-0.89
4a	62%	Reading a graph	-0.55
31	59 %	Reading a graph	-0.42
4d	57%	Reasoning on a graph	-0.33
20a	63%	Computing: comparisons of quantities (actual change)	-0.62
12a	45%	Computing a frequency from a graph	0.21
20b	15%	Computing: comparisons of quantities (relative change	1.96
28	36%	Reasoning on the estimation of population size	0.64
14	17%	Computing and reasoning about a weighted mean	1,96
11a	69%	Reading a simple probability	-0,91
12b	40%	Computing a probability from a graph	0,46
11b	26%	Computing and reasoning on a complex probability	1.16

As regard probability topics, computing and reasoning on a complex probability (11b) are again the main critical aspects while applying elementary concepts of probability is an easier task. Students seem to have acquired the ability of reading and producing indicators in standard situations, but they also seem to show some weaknesses to solve problems in unusual contexts both in statistics and in probability topics. For what concerns probability it seems to appear the need to find ways to link ideas of chance and data, rather than studying probability as a formal mathematical theme (11b versus 11a).

To give evidence to different cognitive performances in relation to several aspects we study more deeply the achievement outcomes. Firstly, we analyze the performances by gender. It is well known that males outperform females in mathematics skills. This issue is confirmed in our analysis, where the mean of the estimated ability is 0,140 (sd. 0.841) for males and -0.149 (sd. 0,786) for females.

Considering the percentages of correct answers we find other interesting evidences.

Table 3. % correct answers of each item by gender

	M (% correct answers)	F (% correct answers)	M-F (difference)
4b	85.7	82.5	3.2
4c	72.1	65	7.1
4a	6.6	57.8	7.8
31	58.9	59.1	-0.2
4d	62.6	51.7	10.9
20a	67.4	58.8	8.6
12a	48.9	41.,5	7.4
20b	18.4	10.8	7.,6
28	42.2	30.3	11.9
14	16.7	15.6	1.1
11a	72.5	65.4	7.1
12b	46.2	33.5	12.7
11b	30	22,6	7.4

As it is usual for mathematics males show the higher percentage of correct answers for all the items. The differences in the answers tend to increase as the difficulty of items increase. In particular items that require reasoning skills show the greatest differences.

Other interesting suggestions on learning behaviors can be derived from the comparison between the estimated ability on the statistics and mathematics items. As we already said we estimated two models, one for the 13 statistics items and one for the 41 mathematics items. Table 4 shows the correlations of the Rasch score in mathematics and in statistics with the marks given by teachers during the scholastic year in oral and written mathematics as well as in Italian language exams. As we can see, the correlations between Rasch scores and marks are higher for mathematics than for statistics. Furthermore the correlations between scores and oral marks are slightly higher than the ones with written exams, both for mathematics and Italian language. INVALSI tests assess the skills based on cognitive process in a way that seems to be more similar to the process using during an oral examination by teachers. This issue may be due to the structure of the tasks that not require merely an application of a formula or of standard procedures, but involve a reasoning procedure, as occurs more frequently during an oral exam.

Table 4. Correlations among Rasch scores and marks

	Rasch for statistics items	Rasch for maths items
Rasch for statistics items	---	0.687
Maths written marks	0.307	0.392
Maths oral marks	0.318	0.399
Italian language written marks	0.268	0.324
Italian language oral marks	0.280	0.345

These suggestions are confirmed by lower correlations between statistics scores and marks, because statistics items are less standard then the mathematics ones and consequently more dissimilar from the assessing approach generally used by teachers, especially in written exams.

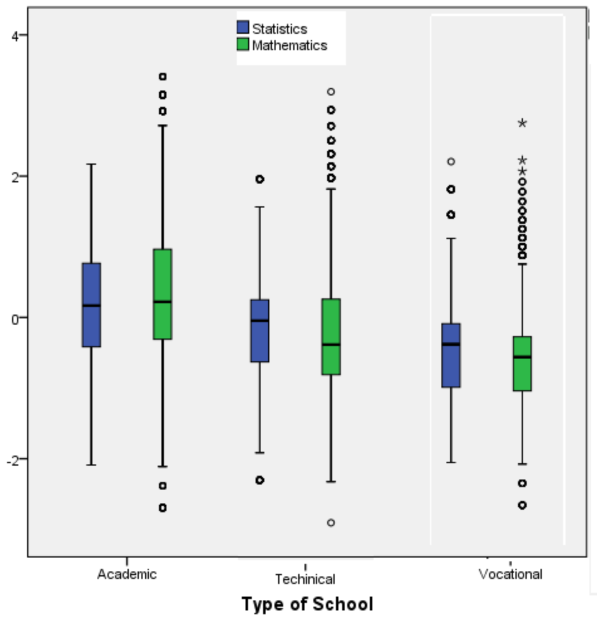

Figure 1. Distributions of abilities by type of school

With reference to the type of attended school, we have also investigated the influence of the curriculum on the students' performances. The learning objectives at the end of second year are the same for each school both for mathematics and statistics competences, but the teaching strategies should be different and affect the learning procedures. Figure 1 shows the distributions of the estimated ability for each type of school by a set of box-plots (based on median).

It is well known that the students' performances vary across the different type of school. The box plots in figure 1 confirm that both for mathematics and statistics students attending academic schools have the best abilities while students of the vocational schools show the worst performances. However, comparing the ability distributions, in technical and vocational schools the results for statistics are better then the ones for mathematics. This suggests that in not-academic schools the teaching strategies of statistics may be more consistent with the learning objectives of curriculum then the mathematics teaching strategies. The cognitive structure of the INVALSI items seems to be more appropriate to evaluate statistical contents that are more empirical and usually based on real context emphasizing the problem solving approach. However, this could also be a signal that the students of technical and vocational schools are less formal, but more "intuitive" and "feeling based" that is to say that they are more keen to explore and discuss data (Snee, 1993).

In figure 1 it is also evident that the distribution of statistics Rasch scores ability is less variable in terms of range than the corresponding mathematics distribution across different type of school. It is as if students are more equal with respect to statistics than with respect to mathematics, and this could be the consequence of the fact that statistics and probability are the new entry of the school curricula.

CONCLUDING REMARKS

The paper focuses on the assessments of statistical learning based on the results of national standardized large-scale tests. Our analysis has a main limitation due the small number of statistics items. Anyway some general considerations can be pointed out as the analysis involved standardized assessment tools administered to many students with different curricula or in different educational settings.

The results suggest a gap between the teaching procedures and the learning objectives. In fact students achieve good results in applying the procedures while they show some lacks for the statistical reasoning especially for not standard tasks. The assessment tool is coherent with contents and produces reliable results.

The findings underline the need to implement teaching strategies that focus on enforcing and developing statistical reasoning (see for example Garfield and BenZvi's, 2009 operative proposals). But to attain this aim it is obviously necessary that teachers too are prepared to meet the challenges of statistics education (Batanero, Burrill & Reading, 2011).

REFERENCES

Batanero C., Burrill G., & Reading C. (Eds.) (2011). *Teaching Statistics, in school Mathematics – Challenges for teaching and teacher education: A Joint ICMI/IASE Study*. Springer.

Garfield J., & Ben-Zvi, D. (2007). How students learn Statistics revisited: a current review of research on teaching and learning statistics. *International Statistical Review*, *75*, 372-396.

Garfield J., & Ben-Zvi, D. (2009). Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment. *Teaching Statistics*, *31*(3), 72-77.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.

M@t.abel project http://risorsedocentipon.indire.it/offerta_formativa/f/

Ministry of Education (2011). *Guidelines for secondary academic school*, http://nuovilicei.indire.it/content/index.php?action=lettura&id_m=7782&id_cnt=10497

Ministry of Education (2011). *Guidelines for secondary technical and vocational school*, http://hubmiur.pubblica.istruzione.it/web/ministero/focus270912

Ottaviani, M. G (2004). Statistics and Probability in the *Mathematics for Citizens* Italian School Curriculum, in SIS, *Atti della XLIII Riunione scientifica*, Università di Bari, Sessioni spontanee, 47-550.

Snee, R. D. (1993).What's missing in statistical education? *The American Statistician*, *47*, 2.