The assessment of statistics learning by large-scale standardized tests: some evidences in the Italian primary and secondary education

Stefania Mignani*
Department of Statistical Sciences, University of Bologna, Italy, stefania.mignani@unibo.it

Roberto Ricci
National Institute for the Evaluation of the Educational System, Italy, roberto.ricci@invalsi.it

Over the past decade there has been an increasing request for statistical literacy, statistical reasoning, and statistical thinking that need to be addressed if future adults have to become more informed citizens. In Italy the teaching of Statistics has been integrated into the mathematics curriculum as a fundamental topic at all grades and types of schools.
In this paper we present the results of student performances coming from national standardized tests. In the test for assessing mathematics competences, there are several items devoted to statistical learning. We consider the data coming from the national tests administered at the fifth, eighth and tenth grade of education by the National Institute for the Evaluation of the Educational System (INVALSI) to Italian students. The results show that students perform well as regard to the statistical literacy but have some lacks on statistical thinking.

Keywords: ability evaluation, statistical literacy, student outcomes, teaching statistics

1. Introduction

Over the past decade there has been an increasing request for statistical literacy, statistical reasoning, and statistical thinking that need to be addressed if adults (or future adults) have to become more informed citizens (delMas, 2002; Gal, 2002).
In Italy since 2001 the teaching of Statistics has been integrated into the mathematics curriculum as a fundamental topic at all levels and types of schools (Ottaviani, 2005).
The emphasis in this paper is on recent issues in statistical learning outcomes. In particular we focus on the results of assessments coming from national large-scale standardized tests. 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 and to offer cost efficiency in terms of development, scoring, and administration. 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.
The key questions of our work are: how does the student performance differ over the different school levels? Are the assessment tools consistent with the objectives of statistical learning?
To answer these research aims we consider the results coming from tests administered by the National Institute for the Evaluation of the Educational System, (INVALSI) to Italian students at the end of primary school, at the end of lower secondary school, and at the end of the second year of higher secondary school. The test assess Mathematics and Italian language skills of students. In the mathematics test there are several items to evaluate statistical competences.
For the first time at national level, in this paper, the statistics items are analyzed in detail considering both the outcomes of student achievements and the characteristics of the items.
The results can explain the quality of student learning and can allow to improve curricula and suggest more effective pedagogical strategies.

2. The assessment of the statistical competence

The assessment of a skill (or a competence) is a process involving several steps.

Firstly, it is fundamental to identify the student learning objectives for the program. Secondly, the design for assessment is arranged to detect how a student might demonstrate that achieved a particular learning objective. These steps involve collecting and analyzing the appropriate assessment data. Taking into account what the data reveal, it is possible to propose improvements to either the curriculum or teaching strategy to implement. Developing national guideline for learning outcomes, designing appropriate assessment items, collecting and evaluating the data, are tasks requiring thought and effort.
As regards the first step, recent researches in statistics education (delMas, 2002; Garfield and Ben-Zvi, 2007, 2009; Garfield and delMas, 2010) 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.
With respect to the second step, considering these cognitive statistical learning outcomes, the assessment should balance procedural proficiency, conceptual understanding, and the use of contexts of statistical investigations and should reflect the values of the disciplines of statistics emphasizing data exploration, rather than mathematical computations.
Consequently, in assessing student learning in statistics it is important to (Garfield and Franklin, 2010):

1) include real data and problem context;
2) include recognizing and understanding variability, that is not more important than the trend or pattern in the data;
3) include the opportunities to select methods of graphing and analyzing data. This is more important than actual computations and calculations that are used do carry on the procedure;
4) maintain a balance between items assessing probability concepts and understanding statistical concepts;
5) when possible, require students to provide interpretations of data analysis as well as justifications for their analysis and conclusions.

Taking into account these suggestions there is a need for accessible and high quality assessment instruments to allow a deeper comparison of students who belong to different curricula or different educational settings. To this aim, in this work we focus on large-scale assessments designed to provide reliable and comparable scores for individuals considering the goals defined in the curriculum at different school levels. Tasks are generally of the type that can be presented in paper-and-pencil format, that student can answer quickly and can be scored reliably and efficiently.

3. Learning of Statistics in Italian schools

In Italy the process of education consists of three cycles: the first one is divided into primary school (five years) and lower secondary school (three years); the second one is the higher secondary school (five years) and the third is tertiary education.
To establish contents and development, according to the specificities of each school grade, the Ministry of Education prepares National Guidelines. Statistics is incorporated into the mathematics curriculum under the domain defined as "Uncertainty and data". We focus on the objectives for the first and the second cycle of education.

Learning Objectives
<u>First cycle of education - primary school</u>
-To develop procedure to collect, organize and interpret data through tables and charts.
-To promote understanding of patterns and trend in data by drawing inference from the frequencies and measure of central tendency (mode and mean).

-In the field of probability to identify characteristics of random events from real situations.

First cycle of education - lower secondary school
-To collect and representing data also using an electronic spreadsheet; to go beyond the reading of the information for comparing data and making decisions.
-To produce and interpret a number of statistical measure of central tendency (mode, median and mean) and measure of variability (mean deviation, standard deviation, variance).
-To master the language of probability; to associate statistics with observed results and frequencies of corresponding events and make use of frequencies to evaluate the probability of an event; to recognize the concepts of complementary, independent and incompatible events.

Second cycle of education
The curricula adds new topics to the objectives of the previous levels. In particular it emphasizes the encouragement of the statistical thinking and reasoning beyond the mere procedure.
First two years
-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.
Last three years
-To detect patterns in univariate or two-way frequency data.
-To identify the different type of relationships between two variables.
-To calculate and interpret the dependence through a linear model.
-To understand the meaning of random variable.
-To know and apply the most familiar probability distribution (binomial and normal).

The content for the objectives can be classified into categories increasing in intensity and complexity as student progress through the school levels:

1) Organizing, representing and reading data or graphs.
2) Interpreting the analysis.
3) Producing frequencies, measures of tendency and variability.
4) Applying probability reasoning tools.

The types of assessment should be appropriate to measure the desired learning outcomes, and to work backward, thinking about instruction and activities that will lead to these goals.

4. The INVALSI tests

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) and from the school year 2010/2011 also to student at the second year of higher secondary school (10th grade). Currently, INVALSI is arranging the plan to assess the competence at the end of higher school. In this paper we focus on the 5th, 8th and 10th grades of mathematics tests. In detail, the item domains deal with functions and relationships, geometry, number, measurement and data. The last domain includes statistical and probability questions, that are the elements of our analysis. Several types of items are designed: multiple-choice (MC) with four alternatives with only one correct answer, true-false (TF), and open-ended items that ask students to give a univocal (numerical or qualitative) answer (U), or the justification for the answer (J). The items are formulated on the basis of learning objectives in according with the indications of the international large-scale tests (PISA and TIMSS) about statistical literacy with the aim to capture the extent of students' statistical reasoning and thinking.

5. Results

Since the school year 2007-2008, the analyses have been progressively increased to give more information about several aspects. We illustrate some of the results of the last test administered (school year 2011-2012). The analyses regard both the students' performances and the psychometric properties of each item. In particular, using an Item Response Theory approach (Hambleton *et al.*, 1991) to each student is assigned a score measuring the ability based on a Rasch model. The score is transformed in a scale with mean equal to 200 and standard deviation equal to 40. For each item the "difficulty parameter" is calculated with range between -4;+4 where higher values mean more difficult items.

-School year 2011-2012

We focus on the results of each item to highlight the impact of type and content in terms of difficulty. Tables from 1a to 1c show some summary information on the item for the different school levels. The 5th level test consists of 44 items, with 6 (14%) statistics items; the 8th level test consists of 40 items, with 7 (18%) statistics items and finally the 10th level test consists of 45 items with 9 (20%) statistics items.

Table 1a 5th grade

Content	**Type of item**	**Difficulty**
Organizing data	TF	0.89
Interpreting a graph	J	0.53
Reading a graph	U	-0.97
Produce a mean	U	0.54
Produce simple event probability	TF	-1.48
Produce events space	U	-0.81

Table 1b 8th grade

Content	**Type of item**	**Difficulty**
Interpreting data	MC	0.53
Reading data	J	-0.81
Produce a mean	U	-0.11
Produce a median	MC	1.05
Produce events space	U	-1.79
Produce simple event probability	MC	0.05
Produce a mean	U	-0.14

As expected, the number of statistics items in the test increases with the level, coherently with the intensification of content. All the three tests present an appropriate balance between items assessing probability concepts and understanding statistical concepts of the topics. All the tests include items with problem context and for 10th grade also real data; besides, when possible, items require students to provide interpretations of data analysis as well as justifications for their analysis and conclusions. In all the three tests items of different types are present, with a slight prevalence of open-ended answers. As regard the difficulty of the topics we can observe interesting evidences. At the 5th grade the students perform better in reading simple graphs and in applying elementary concepts of probability, as remarked by a low difficulty parameter of the corresponding items. The items requiring to organize data and produce a mean are more critical showing higher difficulty parameters. At the 8th grade the results change slightly. The more difficult item requires to produce median while calculating a mean is quite easy. The elementary probability concepts don't show problems.

Table 1c 10th grade

Content	**Type of item**	**difficulty**
Reading a graph	TF	-0.01
Reading data	U	-1.83
Produce simple event probability	MC	-0.42
Produce conditional probability	U	2.06
Interpreting analysis (comparison between graphs)	TF	0.81
Produce a variability measure	MC	1.62
Interpreting data	U	-1.20
Produce a mean (non standard)	U	1.69
Interpreting analysis (comparison mean and standard deviation)	J	1.18

The results of 10th grade clearly show an increase of the importance of assessing statistical skills and the items that are more pertinent with the learning objectives. As regards the performances, for statistical content the test is not so easy, in fact 5 items have a rather high difficulty parameter (above 0.0). Students seem to have acquired the ability of reading and interpreting data and of producing indicators in standard situations, but they also seem to show some weaknesses to solve problems in unusual contexts, especially for probability topics. There is a need to find ways to link ideas of chance and data, rather than studying probability as a formal mathematical theme**.**

In general we can underline that for each level, coherently with learning objectives, students achieve statistical literacy and reasoning while they show some lacks for the statistical thinking.

Finally, Table 2 illustrates for each level the mean score (determined by Rasch model as we said before) indicating the ability necessary to answer correctly at least to 50% of items, where higher values reveal worst performances.

Table 2

5th	8th	10th
192	182	220

These results confirm the evidences of some weaknesses on the assessment for 10th level.

-*Some results of the previous administered tests*

To conclude, we show in the following tables some information on the tests administered in the previous school year 2010-2011. We make a simple comparison between years and grades. Table 3a shows the structure of statistics items, Table 3b presents for each content the number of items and the min/max of difficulty parameters. The results reveal changes from 2011 to 2012 mainly for the test structure. In the last edition there is more heterogeneity in the types of item, in particular there are open-ended items requiring a justification. As regards the performances, the ability of interpreting data is not improved and it remains the more critical aspect with the increase of the level that requires a major extent of skills.

Table 3a

	5th grade				**8th grade**				**10th grade**			
Percentage of item	*23%*				*21%*				*23%*			
Type	*MC*	*TF*	*U*	*J*	*MC*	*TF*	*U*	*J*	*MC*	*TF*	*U*	*J*
		8	*3*		*2*		*4*	*1*	*7*	*3*	*1*	*-*

Table 3b

	5th grade			8th grade			10th grade		
Content (difficulty)	*N°*	*min*	*max*	*N°*	*min*	*max*	*N°*	*min*	*max*
Probability	*4*	*-2.2*	*-1.6*	*4*	*-1.1*	***2.1***	-	-	-
Reading data and graph	*5*	***-3.3***	*-1.0*	*2*	*-2.2*	***-2.2***	*8*	***-2.2***	*-1.4*
Interpreting data	*1*	*-0,7*	-	*1*	*-2.0*	-	*2*	*-2.2*	***0.2***
Tendency/variability	*1*	***0,0***	-	-	-	-	*1*	*-0.5*	-

6. Concluding comments

The paper focuses on the use of assessments to support learning. We consider the results of national standardized large-scale tests assessing mathematical literacy and we focus on statistics items. Our analysis has a limitation due the items on statistics are not many so it is difficult to draw conclusive remarks. Anyway some general considerations should be pointed out as the analysis involved standardized assessment tools administered to many students with different curricula or in different educational settings.

This study confirms the need to teach students not only how to make a graph but also to focus on what graphs tell us. The students should be given repeated opportunities to compare and reason about multiple representations of the same data set.

Moreover. this study reveals the difficulties with concepts fairly elementary such as mean. The students only state how to find it computationally but demonstrate a lack of understanding of the mean. The results also highlight the problems in connecting measure of central tendency in relation to other core concept such as distribution and comparing groups. About variability topics students present weakness in computing formal measures and also in understanding what these summary statistics represent.

The findings underline the need to implement teaching strategies that focus on enforcing and developing statistical thinking. At this aim further researches should be focused on longitudinal data analysis. The future directions in assessment should take into account progress in learning and tests should be designed to provide information that maps back to the progression.

References

delMas R.C. (2002) “Statistical literacy, reasoning and learning: A commentary”, *Journal of Statistical Education,* 10(3), Online: www.amstat.org/publication/jse.

Gal, I. (2002) “Adults’ statistical literacy: meanings, components, responsibilities”, *International Statistical Review*, *70*(1), 1-51.

Garfield J. and 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. and Ben-Zvi, D. (2009) “Helping student develop statistical reasoning: Implementing a statistical reasoning learning environments”. *Teaching Statistics*, 31(3), 72-77.

Garfield J. and Franklin C. (2010) “Assessment of learning, for learning, and as learning in Statistics education”, in C. Batanero, G. Burril and Reading C. (Eds.) *Teaching Statistics, in school Mathematics – Challenges for teaching and teacher education: A Joint ICMI/IASE Study*, Springer, 133-145

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

Pellegrino, J. W., and Hickey, D. (2006) “Educational assessment: Towards better alignment between theory and practice”. In L. Verschaffel, F. Dochy, M. Boekaerts, and S. Vosniadou (Eds.). *Instructional psychology: Past, present and future trends. Sixteen essays in honour of Erik DeCorte*, Oxford, England: Elsevier, 169-189.

Ottaviani M.G. (2005), “Research into statistics Education as a discipline”, *International Statistical Review*, 73 (2), 207-209