

Statements, declared practices and beliefs about INVALSI assessment at primary school level: an exploratory study

Enunciados, prácticas declaradas y creencias sobre la evaluación INVALSI en el nivel de Educación Primaria: un estudio exploratorio

VALENTINA VACCARO ^A, GIORGIO BOLONDI ^B AND LUIS JOSÉ RODRÍGUEZ MUÑOZ ^C

^A INVALSI and University of Oviedo, ^B Free University of Bozen-Bolzano, ^C University of Oviedo

^A valentina.vaccaro@invalsi.it, ^B Giorgio.Bolondi@unibz.it, ^C luisj@uniovi.es

^B <https://orcid.org/0000-0003-4125-7745>, ^C <https://orcid.org/0000-0001-8702-8361>

Received/Recibido: November 2024. Accepted/Aceptado: December 2024.

How to cite/Cómo citar: Vaccaro, V., Bolondi, G. & Rodríguez-Muñoz, L. J. (2024). Statements, declared practices and beliefs about INVALSI assessment at primary school level: an exploratory study. *Edma 0-6: Educación Matemática en la Infancia*, 13(2), 47-59. DOI: <https://doi.org/10.24197/edmain.2.2024.47-59>

Open access article under a [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/) / Artículo de acceso abierto distribuido bajo una [Licencia Creative Commons Atribución 4.0 Internacional \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

Abstract: In this paper we retrace the main stages of an interdisciplinary research project with the aim of highlighting the views of mathematics teachers on large-scale standardized tests. The researchers analyse the responses of over 500 Italian primary school teachers to a questionnaire prepared for this purpose using correlational and principal component analysis (PCA). The comparison between the questions relating to the statements and those relating to classroom practices allowed us to let the attitudes and beliefs emerge. Teachers' recognition of the validity of the INVALSI data in building classrooms practices clashes with the belief that large scale standardized assessments are used by Institutions for the evaluation of the single teacher.

Keywords: Perceived difficulty; math education; MTSK model.

Resumen: En este artículo, repasamos las principales etapas de un proyecto de investigación interdisciplinar que tiene el objetivo de destacar las visiones que tienen los profesores de matemáticas sobre pruebas estandarizadas a gran escala. Se han analizado las respuestas de más de 500 profesores de Educación Primaria de Italia a un cuestionario diseñado con este propósito, usando análisis de correlación y análisis de componentes principales (ACP). La comparación entre las preguntas vinculadas con los enunciados y aquellas relacionadas con las prácticas de aula han

permitido al equipo investigador hacer emerger tanto actitudes como creencias del profesorado participante. El reconocimiento por parte de los docentes de la validez de los datos del INVALSI para la construcción de prácticas de aula choca con la creencia de que las instituciones utilizan evaluaciones estandarizadas a gran escala para la evaluación de cada docente en particular.

Palabras clave: Dificultad percibida; Educación Matemática; Modelo MTSK.

INTRODUCTION

This research is part of a wider project, developed nationally and internationally, concerning how large-scale assessments are received, metabolized, and used by the system (Faggiano et al., 2023; Looney, 2011; Scheerens, 2016). Within the specific line of research concerning mathematics education (De Lange, 2007; Meinck et al., 2017), the research project focuses on the perception and use of mathematics tests by teachers (Abrams et al., 2003; Di Martino & Baccaglini-Frank, 2017; Spagnolo et al., 2022). Large-scale assessments have a strong system-wide impact which is highly disputed due to its political and pedagogical implications.

In Italy, the national institute for the evaluation of the education and training system (INVALSI - www.invalsi.it), administers tests every year at the census level to assess skills of students of grades 2, 5, 8, 10 and 13 in Italian, mathematics, and English (reading and listening). The data of a large national sample representative of the national population are returned to all stakeholders by July of the same year. These assessments have a considerable impact on the teaching-learning processes. For this reason, it is very important to investigate these dynamics. With specific training, the data deriving from large scale assessments can become tools, not only for a systemic perspective, but also to develop skills-based teaching (Rizzo & Vaccaro, in press). These data, so solid from a statistical point of view, can, in fact, through teacher's mediation, become a useful resource for the realization of formative evaluation (Bolondi & Ferretti, 2021), going in the direction of solving the problem of how to use the data of a large-scale standardized assessment for the formative assessment of students (Bolondi, 2015).

The Italian experience has brought out a series of macro phenomena relating to the culture of evaluation, for whose analysis the “INVALSI Group - Didactic and Disciplinary Knowledge” of the SIRD (Italian Society of Didactic Research) conducted an interdisciplinary research project, in which this paper is framed. One of the aims of the project is to

study statements, practices, and beliefs on the Italian National Standardized Assessment (by INVALSI). Within the research project, a questionnaire, designed for quantitative analysis, was administered. From the comparison between the variables related to the statements and those related to practices it is possible to find out about the beliefs of the Italian mathematics teachers on large scale assessment. We know that standardized tests influence didactic practices (e.g., Rodríguez-Muñiz et al., 2016) but here we want to investigate where the beliefs fit into this context and how they interact with the practices.

In this paper we want to investigate the relationships that exist between the variables relating to statements and those relating to practices using both the point of view, disciplinary and pedagogist.

1. FRAMEWORK

We show in Figure 1 a summary of the framework based on which the main research project was carried out. The framework has its roots in those used by INVALSI and the OECD (Organization for Economic Co-operation and Development) for their large-scale standardized tests. Our results are in the two parts of comparing in Figure 1; both the one in which the declared practices are compared with the knowledge and the one in which they are compared with attitudes and beliefs.

Before explaining in detail the type of analysis carried out for this paper, it is necessary to summarize some results already published to understand why an analysis of the main components on some clusters was necessary and why we wondered whether or not there were correlations between statements on the practices of systematic use of INVALSI items and utility of INVALSI tests to analyse, to reflect on, and to design the teaching experience.

In Faggiano et al. (2023), researchers confirmed the presence of a metadidactic conflict (Arzarello & Ferretti, 2021) between classroom practices and what teachers consider valid for assessment. We show, as an example, one of the items we need both to clarify the origins of the conflict and to introduce another finding already published. In the questionnaire, researchers show the INVALSI item in Figure 2 without the percentage of correct answer and ask to the teachers: “On a 1 (very easy) to 10 (very difficult) ranking, how difficult do you think the item is for 5th grade students?”

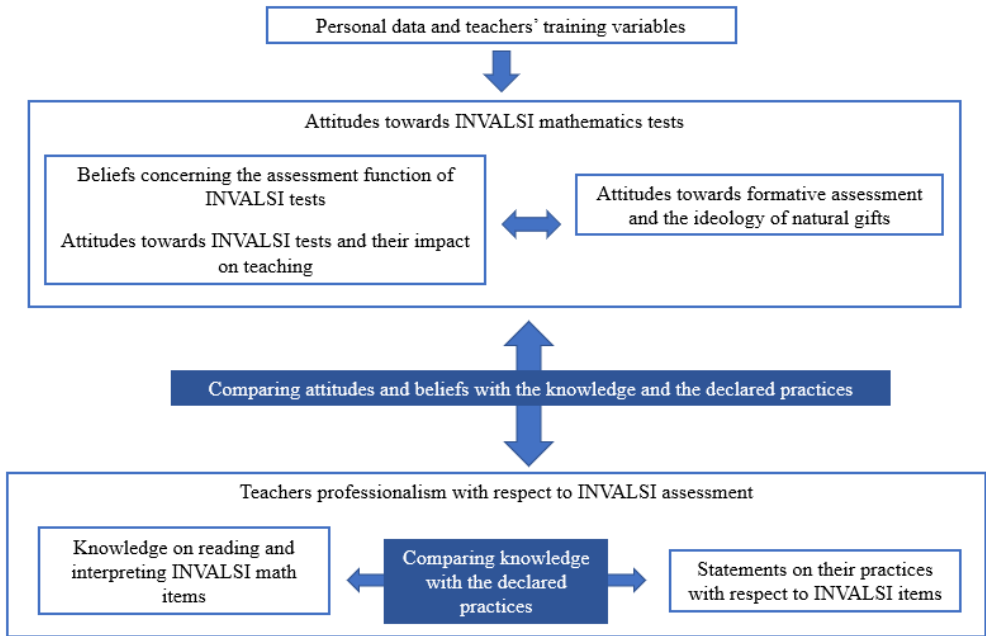


Figure 1. Research variables framework elaborated by the authors

The percentage of correct answers in Figure 2 tells us that the item is difficult for Italian students and this is not surprising given that mastery of the positional number system was required between two different semiotic registers (Duval, 1993). What surprised the researchers was that 79.5 % of teachers estimate the difficulty of the item to be at most 5. The conflict arises from the fact that 87.6 % said they often used similar items in their assessment tests.

D10. To which number does “12 tens, 7 tenths, 2 thousandths” correspond?

- | | |
|------------|-------|
| A. 12.702 | 44.5% |
| B. 120.702 | 33% |
| C. 12.72 | 18.6% |
| D. 120.72 | 3.3% |

Figure 2: Item 10, Grade 5 Mathematics INVALSI test (2009) with percentage of students' answers (in red incorrect, in green right one). Original in Italian.

The goal of the current paper is to compare the variables related to the statements and those related to practices in order to explore if they are

consistent, since in Faggiano et al. (2023) inconsistencies were identified between the questions that teachers found valid and their ability to interpret the reasons for students' errors.

In Rizzo and Vaccaro (in press), the principal component labelled "Utility of INVALSI tests to analyse, to reflect on, and to design the teaching experience" was constructed conducting a PCA analysis on this 9-item cluster. This component summarizes items from 3 to 9.

We are interested in understanding if, and how much, INVALSI questions could be useful for the effective teaching and learning of Mathematics: how much do you agree with the following statements?

3. *Classroom analysis of INVALSI questions helps pupils to understand the importance of reflecting about their learning*
4. *The analysis of INVALSI questions helps teachers to understand which learning goals in Mathematics should be attained*
5. *Teachers must refer to INVALSI outcomes when defining their learning goals*
6. *Didactical design must include activities dealing with INVALSI questions*
7. *The assessment tests used in the classroom should include only questions like the INVALSI ones*
8. *School departments should take in consideration also INVALSI outcomes of the pupils in their work on the vertical development of curricula*
9. *The INVALSI outcomes of pupils should not be overlooked while designing the school curricula*

We need another principal component constructed from a 6-item cluster by Rizzo and Vaccaro (in press): the statements on the use of the INVALSI data to develop the skills of the students ("Teaching with the test"):

For each of the following classroom practices, state how often you use them:

3. *I have the pupils thinking over fast and smart strategies to solve the INVALSI tests*
4. *I draw from INVALSI tests the inspiration for activities about argumentation in Mathematics*

5. *I draw from INVALSI tests the inspiration for activities about problem solving*
6. *I draw from INVALSI tests the inspiration for activities about the justification of answers*

This component summarizes items from 3 to 6 and positively correlates with “Utility of INVALSI tests to analyse, to reflect on, and to design the teaching experience”.

2. INSTRUMENT

The questionnaire is, in fact, divided into three main sections:

- In the first section, seven INVALSI items are presented and for each of them we built a set of three question. The first question is about the teachers’ understanding of the purpose of the question and the hidden reasons behind the students’ errors (Gagatsis et al., 2019); the second question is about the use in the classroom of questions of the same type as the item presented and the third question is about how they believe those items presented may be suitable for assessing the learning of their students.
- The second section, mediated using validated scales, investigates the teachers’ opinions on the formative evaluation and on the ideology of natural gifts (Ciani & Vannini, 2017). In this section there are also clusters on teachers’ opinions towards the INVALSI assessment programme.
- In the third section we have a set of questions on professional training, reading habits, professional context, and personal data.

Questionnaire was administered in two campaigns (Rizzo & Vaccaro, in press), to a total sample of 526 Italian primary school teachers, a large national-scale sample but geographically imbalanced. All questionnaire responses were coded and analysed with a statistical software for data analysis (IBM® SPSS® Statistics 27).

3. METHODOLOGY

First and foremost, two indices were constructed for synthetising the information from the questionnaire on the INVALSI items to perform a correlational analysis.

Then, we performed a Pearson correlational analysis using a statistical software for data analysis. Given two variables, the Pearson correlation coefficient measures the strength and direction of the linear relationship between the two variables. In the analysis performed we decided to consider only Pearson correlational coefficient with a significance level of 0.01 or less. The correlational analysis was performed both on the indices constructed in this paper and on the principal components found in Rizzo and Vaccaro (in press).

Finally, we performed a principal components analysis (PCA) to reduce the number of original variables in a questionnaire cluster and to be able to perform correlations between this principal component and the other variables found. PCA is, in fact, a dimensionality-reduction method, in which a set of variables can be replaced by a minimum set of variables, the principal components (Jolliffe & Cadima, 2016).

We use a statistical software for data analysis (IBM® SPSS® Statistics v. 28) to conduct the analysis of the principal components and the correlation analysis between the components thus obtained that we needed for this article.

4. RESULTS

In the first section of the questionnaire, respondents were asked to declare on a scale from 1 to 4, how much they usually used in their assessment tests questions like the ones in six of the seven INVALSI items used in the questionnaire. Using these questions, we built an index labelled: “Statements on the practice of systematic use”. This index is in the “Statements on their practices with respect to INVALSI items” section of the framework. For each of the seven INVALSI items in the first section of the questionnaire, there is another interesting question: “On a scale from 1 to 4, how suitable do you find the Item to assess students’ learning of your 5th grade students?”. Using these questions, we built a second index labelled: “Validity awareness”. This index is in the “Knowledge on reading and interpreting INVALSI math items” section of the framework.

In both indices, the variables have been added and the same weight has been attributed to each of the variables. The two indices correlate positively with a Pearson's coefficient of .593.

The principal component labelled "Utility of INVALSI tests to analyse, to reflect on, and to design the teaching experience", located in the "Attitudes towards INVALSI tests and their impact on teaching" section of the framework, correlates with the "Validity awareness" index with a .442 Pearson's coefficient.

We want to consider now the principal component labelled "Teaching with the test" that positively correlates both with "Utility of INVALSI tests to analyse, to reflect on, and to design the teaching experience" (Rizzo & Vaccaro, in press) and with the index "Statements on the practice of systematic use". Pearson's correlation coefficients are, respectively, .515 and .315. Finally, the principal component "Teaching with the test" also correlates with the "Validity awareness" index with a .378 Pearson's coefficient.

We use the PCA on the following 4-item cluster:

*Please indicate your degree of agreement with the following statements.
The INVALSI tests are intended to evaluate*

1. *the Italian school system*
2. *the learning in mathematics of each student*
3. *the professional preparation of the individual teacher*
4. *the didactic efficacy of schools*

From the PCA analysis, two components are obtained which have a total variance explained equal to 71.8 %.

Table 1. Rotated component matrix (own elaborated)

	First component	Second component
The INVALSI tests are intended to evaluate the Italian school system	.852	
The INVALSI tests are intended to evaluate the learning in mathematics of each student		.896
The INVALSI tests are intended to evaluate the professional preparation of the individual teacher	.452	.635
The INVALSI tests are intended to evaluate the didactic efficacy of schools	.822	

The distribution of the item “The INVALSI tests are intended to evaluate the professional preparation of the individual teacher” on the two components is not surprising as the public debate has often brought out the teachers’ belief that the INVALSI tests evaluate teachers at individual level. We labelled the second component as: “Utility to evaluate the individual” and performed a correlational analysis with “Utility of INVALSI tests to analyse, to reflect on, and to design the teaching experience”. The Pearson coefficient of this correlation is .262. On the other hand, the first component labelled “Utility to evaluate the system” also correlates with “Utility of INVALSI tests to analyse, to reflect on, and to design the teaching experience”. The Pearson coefficient of this correlation is .460.

5. DISCUSSION AND CONCLUSIONS

The first correlation found between the indices “Statements on the practice of systematic use” and “Validity awareness” indicates that there is agreement between the statements on the use that teachers make of INVALSI items and how they consider them suitable for evaluating the learning of their pupils. Although with a lower coefficient, the “Validity awareness” index also correlates with the main component “Utility of INVALSI tests to analyse, to reflect on, and to design the teaching experience”. These correlations tell us that the teachers not only declare

that they are aware of the validity of the items for evaluating the knowledge, skills, and competences of each of their students but also that they consider them useful for the organization of the contents and for the design of their didactic action.

The correlation between the design of teaching and the use of the INVALSI data for the development of a teaching based on skills (“Teaching with the test”) also tells us that those who use INVALSI items for the design of their teaching do so in the way best possible and, that is, to develop a teaching based on the development of skills. Some teachers therefore declare that they use the information deriving from large-scale standardized tests in their teaching both for its design and for the development of skills.

The conflict found in Faggiano et al. (2023) between the belief about that standardized assessment items are suitable for evaluates students (at individual level) and the difficulty on understanding the reasons for the students’ errors seems to extend to the aim of large-scale standardized assessment. In fact, the “Validity awareness” index correlates with the component labelled “Utility of INVALSI tests to analyse, to reflect on, and to design the teaching experience”, extending the usefulness of the INVALSI data from individual to a school level. On the other hand, the “Utility of INVALSI tests to analyse, to reflect on, and to design the teaching experience” also positively correlates with the “Utility to evaluate the individual” and this belief conflicts with statements on usefulness of the INVALSI items and data in the construction of didactic activities and evaluation. The profile of a teacher who, although he/she recognises the validity of the INVALSI data in building classrooms practices, also believes that large scale standardized assessments are used by Institutions for the evaluation of the single teacher, emerges.

Fortunately, the correlational analysis between the components “Utility to evaluate the system” and “Utility of INVALSI tests to analyse, to reflect on, and to design the teaching experience” suggests the presence in the sample of a good number of teachers who, despite having of the uncertainties about identifying the reasons for students’ errors, he/she declares himself/herself aware of the usefulness of the items of the standardized tests and their data for both evaluating the school system and evaluating the single student in class.

ACKNOWLEDGMENTS

The authors thank all the members of the research group led for the disciplinary part by Professor Ferdinando Arzarello and for the pedagogical part by Professor Ira Vannini. Luis J. Rodríguez-Muñiz is funded by Grant PID2021-122180OB-100 by the State Agency of Research (Spain) and he belongs to MTSK Network (<https://redmtsk.net/>).

REFERENCES

- Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the Classroom: Teachers' Opinions of Statewide Testing Programs. *Theory into Practice*, 42, 19-29. https://doi.org/10.1207/s15430421tip4201_4
- Arzarello, F., & Ferretti, F. (2021). The connection between the mathematics INVALSI test and the teaching practices: an explorative study. In P. Falzetti (Ed.), *INVALSI data to investigate the characteristics of student, schools and society: IV Seminar "INVALSI data: a tool for teaching and scientific research"* (pp. 96-109). FrancoAngeli.
- Bolondi, G. (2015) Transformar la evaluación estandarizada en evaluación formativa. In B. D'Amore & M. I. Fandiño Pinilla (Eds.), *Didáctica de la Matemática. Una mirada internacional empírica y teórica* (pp. 109-120). Universidad de La Sabana.
- Bolondi, G., & Ferretti, F. (2021) Quantifying Solid Findings in Mathematics Education: Loss of Meaning for Algebraic Symbols. *International Journal of Innovation in Science and Mathematics Education*, 29(1), 1-15. <https://doi.org/10.30722/IJISME.29.01.001>
- Ciani, A., & Vannini, I. (2017) Equità e didattica. Validazione di scale sulle convinzioni di insegnamento democratico. *Cadmo*, 25(2), 5-32. <https://doi.org/10.3280/CAD2017-002003>
- De Lange, J. (2007). Large-Scale Assessment and Mathematics Education. In F. K. Lester Jr. (Ed.), *Second Handbook of Research on*

- Mathematics Teaching and Learning* (pp. 1111-1142). National Council of Teachers of Mathematics (NCTM).
- Di Martino, P., & Baccaglioni-Frank, A. (2017). Beyond performance results: Analyzing the informational and developmental potentials of standardized mathematics tests. *For the Learning of Mathematics*, 37, 39-44. <https://www.jstor.org/stable/26548470>
- Duval R. (1993). Registres de représentations sémiotiques et fonctionnement cognitif de la pensée. *Annales de Didactique et de Sciences Cognitives*, 5, 37-65.
- Faggiano, E., Monaco, A., Rizzo, O. G., & Vaccaro, V. (2023). An exploratory study on the connection between INVALSI assessment and mathematics teaching and learning processes at the Primary School level. In P. Falzetti (Ed.), *The School and Its Protagonists: The Teachers. V Seminar "INVALSI data: a tool for teaching and scientific research"* (pp. 9-23). FrancoAngeli.
- Gagatsis, A., Michael-Chrysanthou, P., Christodoulou, T., Iliada, E., Bolondi, G., Vannini, I., Ferretti, F., & Sbaragli, S. (2019). Formative assessment in the teaching and learning of mathematics: Teachers' and students' beliefs about mathematical error. *Scientia Paedagogica Experimentalis*, 56(2), 145-180.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Looney, J. W. (2011). Integrating Formative and Summative Assessment: Progress Toward a Seamless System? *OECD Education Working Papers*, No. 58. OECD Publishing. <https://doi.org/10.1787/5kghx3kbl734-en>.
- Meinck, S., Neuschmidt, O., & Taneva, M. (2017). Workshop Theme: "Use of Educational Large-Scale Assessment Data for Research on Mathematics Didactics". In G. Kaiser (Ed.), *Proceedings of the 13th International Congress on Mathematical Education, ICME-13*

Monographs (pp. 741-742). Springer. https://doi.org/10.1007/978-3-319-62597-3_132

Rizzo, O. G., & Vaccaro, V. (in press). Attitudes and beliefs of primary school mathematics teachers: an exploratory study on the use of INVALSI assessment data in a competency-based approach. In P. Falzetti (Ed.), *I dati INVALSI: uno strumento per la ricerca*. FrancoAngeli.

Rodríguez-Muñiz, L. J., Díaz, P., Mier, V., & Alonso, P. (2016) Washback Effect of University Entrance exams in Applied Mathematics to Social Sciences. *PLoS ONE* 11(12): e0167544. <https://doi.org/10.1371/journal.pone.0167544>

Scheerens, J. (2016). *Educational Effectiveness and Ineffectiveness. A Critical Review of the Knowledge Base*. Springer. <https://doi.org/10.1007/978-94-017-7459-8>

Spagnolo, C., Giglio, R., Tiralongo, S., & Bolondi, G. (2022). Formative Assessment in LDL Workshop Activities: Engaging Teachers in a Training Program. In B. Csapó & J. Uhomobhi (Eds.), *Computer Supported Education, 13th International Conference, CSEDU 2021* (pp. 560–576). Springer. https://doi.org/10.1007/978-3-031-14756-2_27