An exploratory study on the connection between INVALSI assessment and mathematics teaching and learning processes at the Primary School level

Eleonora Faggiano[1], Annarita Monaco[2], Ottavio G. Rizzo[3], Valentina Vaccaro[4]

[1]*University of Bari Aldo Moro*, [2] *IC Francesca Morvillo di Roma.*, [3]*University of Milan*,

[4]*University of Oviedo, INVALSI - Roma.*

Abstract

This paper deals with an interdisciplinary research project aiming at identifying the training needs of teachers at national level and to propose guidelines for the improvement of teaching practices, regarding the use of mathematics INVALSI standardized tests. In order to study the connection between INVALSI assessment and mathematics teaching and learning processes at Primary School level, a survey was designed and administered to a total sample of 526 primary school teachers. Early results of the survey are presented and discussed showing the existence of a meta-didactical conflict concerning discourses about didactical processes like assessment, students' abilities and mistakes, etc.

Introduction

Data from Large Scale Assessment can be considered as tools that teachers can use, not only in a systemic perspective, for the design and implementation of meaningful teaching and learning activities, but also to give back to students, through the practice of formative assessment, detailed information on their learning and thus build reflexive and meta-cognitive paths, functional to a real teaching based on skills (Wiliam, 2010). This is also true, in our view, for the Italian standardized assessment program developed by the INVALSI Institute. The experience of the INVALSI standardized tests has led to the emergence of a series of macro-phenomena in the school system relating to the disciplines, to disciplinary teaching and more generally to educational aspects, also connected to school and to the teachers' evaluation culture. These are phenomena of a very different nature that can be analysed and used in different directions, both with sociological research tools, particularly supporting them with a statistical methodological basis. In particular, the INVALSI standardized assessment program has provoked and provokes a series of research questions and issues concerning: the reading and interpretation of data; the analysis of teachers' training needs; the analysis of how the two variables listed above may affect the teachers' attribution of meaning to the various INVALSI items. An interdisciplinary research project was started in 2017 highlighting the need to interpret the above-mentioned complex phenomena, with the aim of identifying the training

needs of teachers at national level and to propose guidelines for the improvement of teaching practices, regarding the use of mathematics INVALSI items. The project is conducted by the "INVALSI Group - Didactics and Disciplinary Knowledge" of the S.I.R.D. (Italian Society for Educational Research) on general education and disciplinary education, composed of disciplinary experts and pedagogists. Among the various elements to be analysed, there are undoubtedly factors related to the perceptions and opinions of teachers that can facilitate or inhibit the didactic impact of the tests. One crucial interest is therefore the understanding of the attitude and meanings that teachers attribute to the INVALSI assessment. This contribution shows the first results of the project aimed at investigating, through the voice of teachers, the link between the INVALSI math assessment and the mathematics teaching-learning processes at Primary School level.

The research project

Since the beginning, the goal of the mixed group of researchers involved in the project was to start an exploratory study to investigate the meaning that teachers attribute to the INVALSI items, which seems to be rarely used with a view to formative assessment. With this purpose, researchers shared the aim to build a research tool, i.e., a survey, through which investigating the perceptions of mathematics primary teachers with respect to the INVALSI tests.

The tool (that will be described below) is a survey that was firstly administered to 105 teachers (Try Out). This initial phase made it possible to test the questions in the survey. Then, based on the analysis of the data collected with the Try Out, the survey was partially modified. The new version of the survey was administered to 427 teachers (Main Study). In this paper we used only the data relating to the Main Study which has a total of 421 valid cases. Data collected in the two campaigns were encoded and analysed using a statistics software for data analysis (SPSS). Early results will be presented and discussed below.

Terminology

In order to avoid any possible confusion, let's establish the following convention in the use of terms.

An INVALSI standardized *test* is composed of *items*, some items may be subdivided into *parts*, *students answer* to the items, eventually choosing among *options,* and the national-wide results constitute INVALSI *national data*.

Our survey, on the other hand, is based on a *questionnaire* composed of *questions* (we will use the notation Q*n* to refer to the *n*th question), usually composed of *options* or *ranking scale*, to which survey *participants* gave *responses.*

The research tool

The purpose of the empirical, descriptive and correlational survey was to analyse the knowledge, teaching experiences and beliefs available to primary school teachers to read and interpret INVALSI items and data, in particular in the mathematical field. Specifically, the aim of the survey is: to investigate the beliefs of teachers regarding the knowledge and skills detected by the INVALSI standardized tests; to explore the proximity/distance between the functions and contents of the INVALSI items, on the one hand, and beliefs and statements about the teaching practices of teachers, on the other hand. In order to specify the different research variables we were interested in, and the research hypothesis concerning the relationships among the variables, we built the framework in figure1.
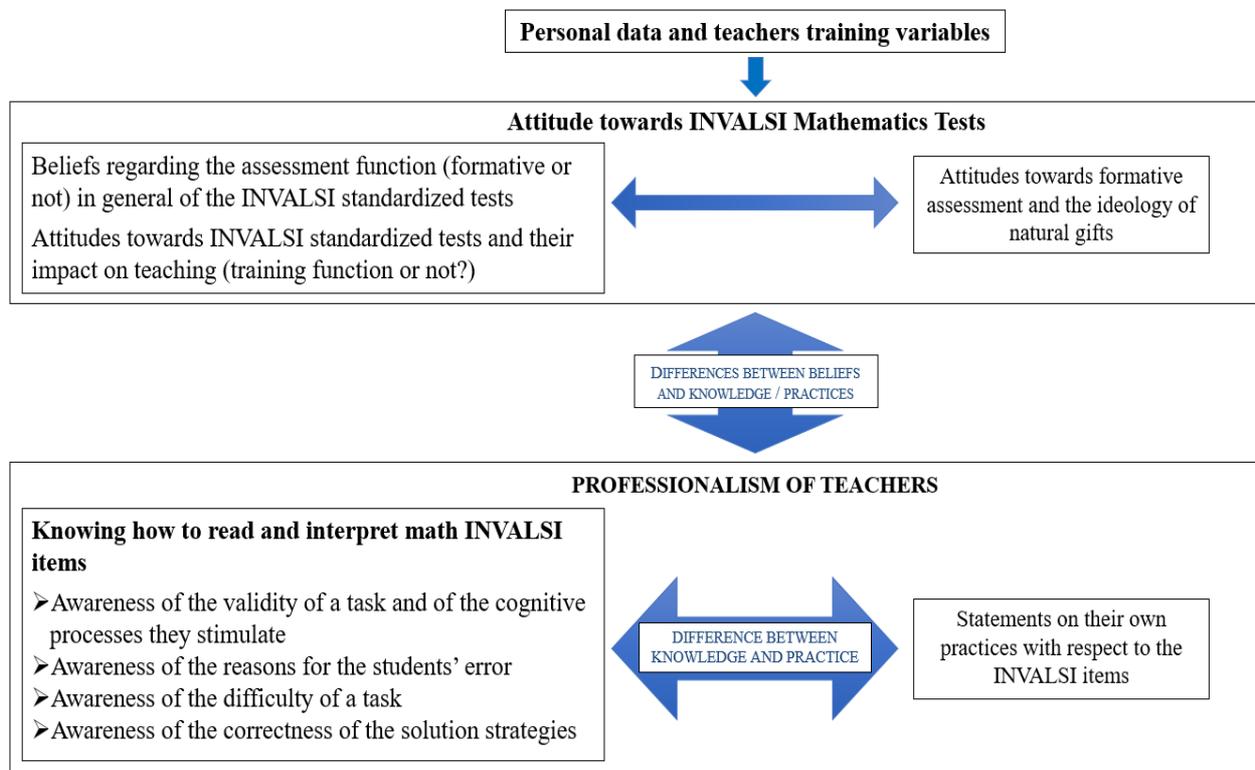

Fig. 1: The research variables framework on which the survey is based

According to the research variables framework, the questionnaire consists of the following three sections:

- one concerning mathematics education (how teachers interpret the Invalsi items and their results);

- one relating to aspects of general education (which beliefs and attitudes teachers have and how they pour them into teaching practices);

- one that collects personal data and context information.

In the first section, seven INVALSI items of grade 5 or 6 are presented in their original formulation. For each of them, questions are proposed aimed at detecting the pedagogical knowledge of the mathematical content — the so-called Shulman (1986)'s PCK — by the teachers (misconceptions, recurring errors, level of difficulty). In addition, comparative questions are proposed on the proximity/remoteness of the seven items from teaching practices and National Guidelines (Indicazioni Nazionali, 2012; Italian Ministry of Education, 2018) and on the effectiveness of the considered INVALSI items in assessing certain skills.

The second section proposes three sets of questions regarding: the opinions of teachers on the INVALSI assessment programme; the educational usefulness of the INVALSI items; the didactic practices connected to the INVALSI items; attitude towards the ideology of natural gifts (Ciani & Vannini, 2017).

The data collected in the third section relates to professional training as well as personal data. For example, we asked to indicate: how many years have they been working as teachers; how long have they been teaching in the current school; how long have they been in the role; what duties do they hold in their school; etc.

Early results

First of all, thanks to the analysis of the third section of the questionnaire (personal data and context information) it is possible to outline the characteristics of the sample: 68% of the survey participants received an invitation to fill in the questionnaire by their School Headmaster; 71% of the survey participants teach in Piemonte or Emilia–Romagna (two north regions, which constitute together 15% of Italian population); 90% of the survey participants are tenured teachers; 21% of the survey participants actively participate in the school administration (members of the senior leadership team). Although the sample, even large, cannot be considered to be representative, the data collected can give us a wide range of different information to reflect on.

Herein we present some early results of teachers' responses to the first section of the survey. INVALSI items and survey questions were translated by the paper authors from Italian into English for the ease of the reader.

To give a first insight into the richness of information gathered through the survey, we start dealing with an INVALSI item (figure 2) that in 2009 resulted to be quite a difficult item for 5th grade students: indeed, only 33% of Italian students gave the correct answer.

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

    A. 12.702

    B. 120.702

    C. 12.72

    D. 120.72

Fig. 2: Item 10, Grade 5 Mathematics INVALSI test (2009)

We were interested in investigating teachers' understanding of the difficulty of this item. For this purpose, without informing participants about the percentage of the correct answer given by the students, we asked:

Q6. On a 1 (very easy) to 10 (very difficult) ranking, how difficult do you think the item is for 5th grade students?

As it can be seen in figure 3, 79.5% of teachers estimated the difficulty to be at most 5, hence, although the item required to manage a non-trivial conversation transformation between two different semiotic registers (Duval, 1993), we can say that this item was not considered to be a difficult item.
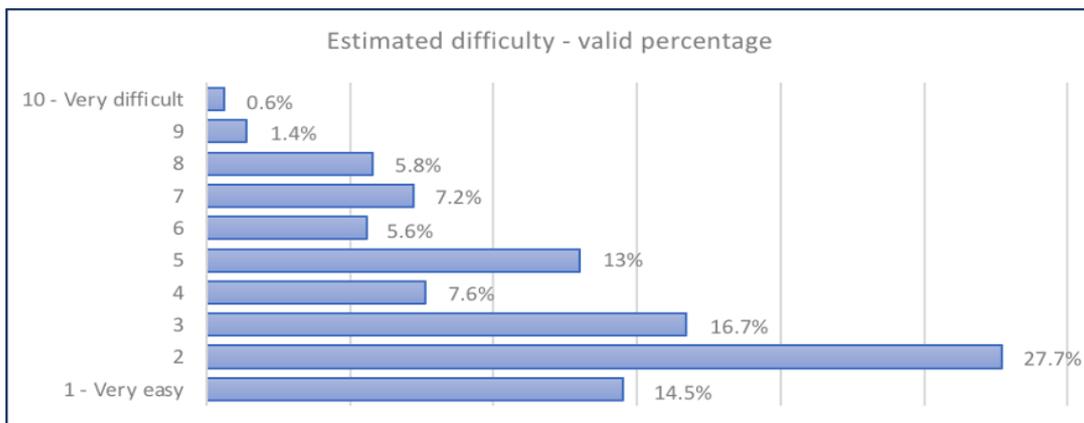

Estimated difficulty - valid percentage

10 - Very difficult	0.6%
9	1.4%
8	5.8%
7	7.2%
6	5.6%
5	13%
4	7.6%
3	16.7%
2	27.7%
1 - Very easy	14.5%

Fig. 3: "How difficult do you think the question [in figure 2] is for 5th grade students?"

This result confirms what was found in the Try Out (Arzarello and Ferretti, in press): teachers' perception of students' difficulties does not correspond to the INVALSI national data. Despite this discrepancy, results also confirm that the item is, among the seven used in the survey, the one which is considered the "most suitable for assessing learning" (with 86.2% of teachers which evaluate its suitability ranking it 3 or 4 over 4) and one of the "most commonly used in assessment tests" (with 87.6% of teachers stating to be using this type of item in their assessment tests — ranking it 3 or 4 out of 4 in the relevant question). At a meta-didactical level, this reveals an apparent inconsistency that is under investigation with quantitative and qualitative methods.

Another example of the questions in the survey is the one concerning the INVALSI item in figure 4.

D3. Look at this sequence of figures
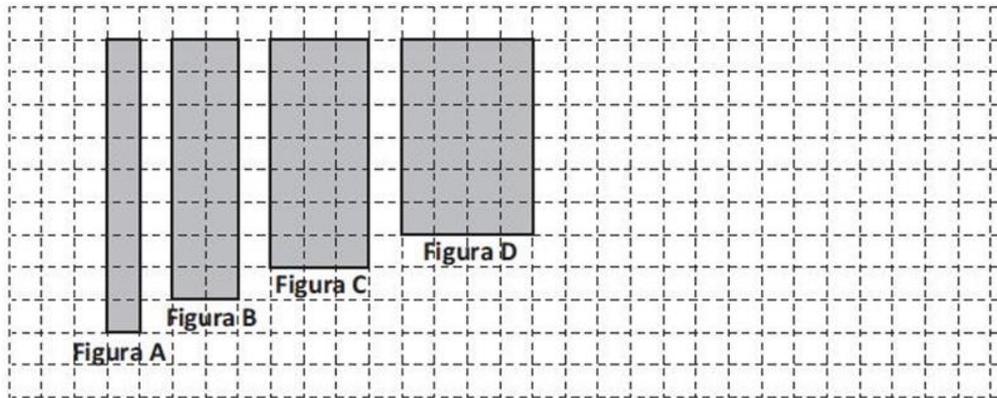
a. **Draw on the squares, next to Figure D, the next figure in the sequence.**

b. **Which of the following sentences is true?**

    A. The area of the figures is always the same

    B. The area of the figures doubles at each step

    C. The perimeter of the figures is always the same

    D. The perimeter of the figures increases at each step

Fig. 4: Item 3, Grade 6 Mathematics INVALSI test (2012)

The correct answer is 'C', and the expected mathematical process is: at each step the altitude is reduced by one while the width increases by one, so the perimeter stays the same. According to the nation-wide results, 85.8% of students correctly drew the next rectangle in the sequence. This shows that it is clear (at least in an intuitive way) what happens to the altitude and the width of the rectangles at each step of the sequence. The fact that more than half of the students that passed part 'a' failed part 'b' indicates that the issue at stake is the common misconception that areas and perimeters should behave in the same way (see, for example, Fandiño-Pinilla & D'Amore, 2006).

However, literature (e.g. Jacquet, 2000; D'Amore & Fandiño-Pinilla, 2005) shows that the building of a satisfactory knowledge of the relationship between perimeter and area has not only an epistemological nature but also a didactical nature, and cognitive nature: the former has been investigated by Jacquet (2000), D'Amore & Fandiño-Pinilla (2005); the latter can be framed within what Stavy & Tirosh (1999, p. 59) call the 'sameA-sameB' intuitive rule, used by students of different ages, who are asked to make comparison tasks. For this reason, in order to understand teachers' awareness of the origin of student errors, we asked the participants to give their interpretation of the

nation-wide results in the item. In particular, we asked them (Q3) to choose one of the reasons why, although 85.8% of students answered correctly to part 'a', only 35.7% of students correctly chose 'C' in part 'b', while almost the same number chose 'D'.

During the design of the questionnaire, we chose the following particular options to recognize different approaches by the teachers:

1. "Pupils do not pay attention while reading the text": we consider this as a boilerplate answer that we expect to be used by a teacher not knowledgeable of the didactical and epistemological issues at play.

2. "Pupils do not know area and perimeter formulae well": we can assume that most 5th grade students have a working knowledge of computing areas and perimeters of such rectangles (it is drawn on square paper, so it suffices to count the squares and make a simple sum or multiplication!), but on the other hand the item does not ask for any explicit numerical result. Hence, we hypothesize that teachers who chose this option reduce the idea of "perimeter" and "area" to the computation of their values using the appropriate formulae, instead of considering the more general geometrical concept involved in the question.

3. "Pupils are led astray by the picture": this option is very similar to the first but might be chosen by teachers who recognize that the item is about geometry.

4. "Pupils believe that the area increases while the perimeter increases": this is the answer we expect from a teacher aware of the didactical and epistemological issues at play.

Table 1 shows how participants answered Q3. It can be seen that only 21.5% of teachers recognize that the reason for students' error is connected with the misconception that areas and perimeters should behave in the same way.

Tab. 1: Distribution of the teachers' choices in answering Q3.

	Options	Percent
1	Pupils do not pay attention while reading the text	34.2%
2	Pupils do not know area and perimeter formulae well	2.1%
3	Pupils are led astray by the picture	32.7%
4	Pupils believe that the area increases while the perimeter increases	21.5%
5	Other	9.5%

Further elements are unveiled analysing teachers' responses to the next two questions of the survey regarding the INVALSI Item in figure 4: the first (Q4) was meant to investigate teachers' awareness of the suitability (on a 1 – not at all – to 4 – completely – ranking) of the item in order to assess students' learning at 5th grade; the second (Q5) aimed to know to what extent (on a 1 – never – to 4 – regularly – ranking) teachers' claim to use this kind of item in their ordinary assessment test. Percentage of responses are shown in Table 2.

Tab. 2: Teachers' responses to Q4 and Q5 with respect to the INVALSI item in figure 4

Q4: How suitable do you find the Item to assess students' learning of your 5th grade students?		Q5: How often do you use this kind of items in your assessment tests?	
Rank	**Percent**	**Rank**	**Percent**
1 (Not at all)	4.8%	1 (Never)	8.6%
2	24.5%	2	37.3%
3	47.3%	3	47.9%
4 (Completely)	23.4%	4 (Regularly)	6.3%

As it can be seen in table2, although 23.4% of teachers considered the item completely suitable to assess students' learning, the percentage of teachers who declared to regularly use this kind of items in their classroom assessment tests is limited to 6.3%.

Using the Spearmans' Rho we also analysed the correlation between Q4 and Q5: the SSPS computation returns a correlation coefficient of 0.485 with 0.01 significance. In other words, teachers state they use the kind of item consistently with how much they deem it suitable to assess students' learning. On the other hand, as it could be expected, there is a very good correlation between perceived suitability and declared use in the classroom.

Finally, we consider worthy of note the participants' answers to one of the questions of the second section of the questionnaire, when the overall results are compared with those obtained restricting the sample to the 21.5% of teachers that recognized the reason of students' error: table 3 shows that the percentage of teachers who completely agreed with the claim that "analysing INVALSI items can help teachers understand what mathematics learning aims are to be achieved" increases from 22.2% to 33.7% if we look at those teachers that answered by choosing option "4" to Q3.

Tab. 3: Participants level of agreement with the claim: "analysing INVALSI items can help teachers understand what mathematics learning aims are to be achieved"

Level of agreement	Frequency (participants who answered "4" to Q3)	Valid Percent	Valid percent of the whole sample of participants
1 - Completely disagree	5	5.1	8
2	20	20.4	23.8
3	40	40.8	46
4 - Completely agree	33	33.7	22.2
Missing	15		

The last example we present in this paper is the one concerning the INVALSI item in figure 5.

D25. Look at the picture
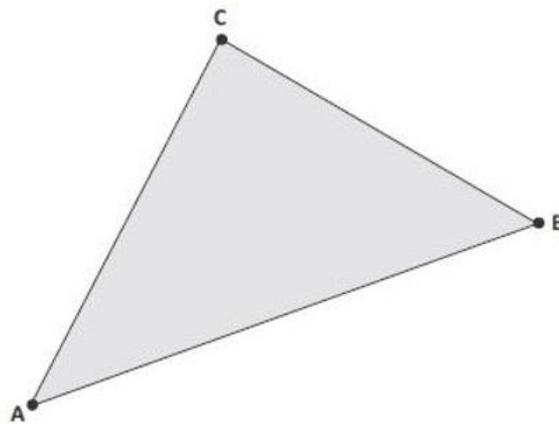
In the triangle ABC, draw the altitude CH relative to the side AB

Fig. 5: Item 25, Grade 6 Mathematics INVALSI test (2012)

According to nation-wide results, only 51.5% of students were able to answer the item correctly, drawing a line perpendicular to the side AB. The difficulty of drawing the altitude of a triangle given in a non-standard position is well known in literature (Gutierrez and Jaime, 1999): 5th grade students are known to believe that altitudes have to be vertical, and that, even if they appear to satisfy the formal definition, "if I want it to become an altitude, I must turn the book and put it straight" (translated by the authors from Martini and Sbaragli, 2005); this phenomenon could be explained by the fact that books and teachers almost constantly show vertical altitudes and this "overexposure to prototypes may impede the growth of fuller concept acquisition" (Tsamir, Tirosh and Levenson, 2008).

Understanding teachers' awareness of the origin of student errors is, in this case, even more important since when a teacher proposes a strong persistent and convincing image, the image turns into an intuitive model and, the stronger the intuitive model, the more difficult it is to break it to accommodate a new image (Martini and Sbaragli, 2005). We should not discount either that such a misconception is rooted in a teacher's deficiencies in MCK, that "modern teacher training is slowly (and partly!) fighting" (Alatorre and Sáiz, 2010).

Similarly, to the previous item, we asked the participants to give their interpretation of the nation-wide results in the item. In particular, we asked them (Q15) to explain why only 51.5% of 6th grade students were able to answer correctly.

During the design of the questionnaire, we chose these particular options to recognize different approaches by the teachers:

1. "Pupils do not pay attention while reading the text": we consider this as a boilerplate answer that we expect to be used by a teacher not knowledgeable of the didactical and epistemological issues at play.

2. "Pupils do not know the definition of altitude of a triangle well": we expect that a teacher that chooses such an option just pieced together the keywords 'altitude' and 'correctly'. The child referred to by Martini and Sbaragli shows that children that do know a correct definition of altitude could nevertheless require it to be vertical.

3. "Pupils are led astray by the picture": we hypothesize that the teacher that gives such an answer has clearly some insight into the epistemological and didactical issues at stake; but, given the following and much more specific option, we also expect that such a teacher has not fully connected these issues with the misconception of "vertical" altitudes.

4. "Pupils think that the altitude should be vertical": this is the answer we expect from a teacher aware of the didactical and epistemological issues at play.

In table 4 we present the distribution of the teachers' choices in answering Q15.

Tab. 4: Distribution of teachers' choices in answering Q15.

	Options	Percent
1	pupils do not pay attention while reading the text	8.2%
2	pupils do not know the definition of altitude of a triangle well	21.5%
3	pupils are led astray by the picture	28.8%
4	pupils think that the altitude should be vertical	34.6%
5	Other	6.8%

It can be seen that participants who recognize that the reason for students' error is connected with the misconception that the altitude should be vertical are 34.6%.

Also, for this item, we were interested in investigating teachers' awareness of its suitability in order to assess students' learning at 5th grade (Q16) and to know to what extent teachers claim to use this kind of item in their ordinary assessment test (Q17). Looking at the responses we found that 50% of teachers considered this item completely suitable to assess students' learning and that 39% of the participants declare to regularly use this kind of item. Moreover, it can be seen that 35% of the teachers gave the maximum rank to both Q16 and Q17 (and in particular that they made up 70% of those who considered the item completely suitable to assess students' learning), however, 64.2% of them were not able to identify the reason for students' errors in answering the item. That is, 22.6% of participants, even considering this item to be completely suitable to assess students' learning and declaring they regularly use this kind of item in their assessment tests, did not recognize the reason for students' error.

Finally, we consider worth noticing that, even if participants recognizing the reason for students' errors make up 21.5% for the item in figure 4 and 34.6% for the item in figure 5, those who were able to recognize both the issues were only 9.2%. However, 64.8% of teachers consider the two items suitable (partially or completely) to assess 5th grade students' learning and 49.9% of teachers declare that they make use (often or regularly) of both these kinds of items in their assessment practices.

Discussion

Presenting the first findings from the Try Out of the survey, Arzarello and Ferretti (in press) highlighted how teachers' responses revealed the presence of a three-fold *meta-didactical conflict,* concerning discourses about didactical processes like assessment, students' competencies and mistakes, etc. According to this point of view, the conflict can be analysed by focusing on three different components. The first component concerns teachers' perception of students' difficulties in tackling INVALSI items: our results showed how teachers often have a perception which is not in tune with the INVALSI national data. The second component concerns the teachers' interpretation of students' answers and mistakes: with this respect, teachers' responses to the Try Out also revealed a discrepancy with the national data. The third component refers to the contradictory responses of the teachers to the questions of the survey dealing with the overall rationale of the INVALSI assessment, such as the suitability to assess students' learning or the compliance with the curriculum national guidelines (Indicazioni Nazionali, 2012).

Teachers' responses to Q6 confirm the existence of the first component of the meta-didactical conflict as it already arose analysing the results of the Try Out. Results, indeed, reveal an apparent inconsistency: on the one hand, teachers failed in considering the difficulty of the items; on the other, they claimed this item is suitable to assess students' learning and it is often used. We believe that this inconsistency requires further investigating with quantitative and qualitative methods, in order to clarify its origin and its nature. Moreover, the teachers' responses to the survey questions presented above seem to confirm the existence of the meta-didactical conflict, and particularly of the second and the third components. Evidence of the existence of the second component of the meta-didactical conflict is given by participants' responses to Q3 and Q15: only 21.5% of teachers identify that students' difficulty in answering INVALSI item in figure 4 is due to a wrong construction of the meaning concerning the relationship between area and perimeter while 34.6% of teachers identify that students' difficulty in answering INVALSI item in figure 5 is due to the misconception that the altitude should be vertical. Hence, we can say that participants seemed to have some difficulties in recognizing the reasons for the students' error, especially if we consider also that, as highlighted before, participants who discern the issues at play in both the proposed items make up only the 9,2% of the total.

On the other hand, based on the participants' responses to Q3 and Q4 and on the described correlation between answers to Q4 and Q5, there is a need to further investigate the meaning teachers attributed to the expression "same kind" when they answered Q5: the apparent discrepancy in the teachers' answers indeed seems to bring to the fore the existence of the third component of the meta-didactical conflict. This hypothesis, indeed, seems also to be confirmed by the fact that the awareness of the

students' errors presents a positive correlation with the awareness that analysing INVALSI items can help teachers understand what mathematics learning aims are to be achieved.

Some more comments can be made analysing participants' responses to questions Q16 and Q17, concerning the INVALSI Item in figure 5. Despite the numbers of teachers that consider the item suitable to assess students' learning and declare they use this kind of item in their classroom, it seems to be contradictory that such a low number of them recognize that this INVALSI item was aimed at detecting the common misconception of the verticality of the altitude of the triangle. This discrepancy again calls to the need for further investigation aiming to shed some light on the nature of the third component of the meta-didactical conflict.

Conclusions

In order to analyse knowledge, teaching experiences and beliefs available to primary school teachers to read and interpret mathematics INVALSI data, we designed and administered a survey. In particular, through the analysis of participants' responses to the survey we were interested in: investigating teachers' beliefs regarding the knowledge and skills detected by the INVALSI standardized tests; exploring the proximity/distance between the functions and objects of the INVALSI standardized tests, on the one hand, and beliefs and statements about the teaching practices of teachers, on the other hand. In this paper we have presented some early results of the survey section specifically concerning mathematics education (the way teachers interpret and use INVALSI standardized tests and data). They have been interpreted using the lens of the meta-didactical conflict by Ararello and Ferretti (in press): results seem to confirm their hypothesis and as next step of the project we are now going to clarify the deep structure and nature of this conflict (e.g. with respect the knowledge and beliefs of teachers) in order to design suitable guidelines for getting rid of it and obtaining a real improvement of practices regarding the use of INVALSI standardized tests in the school.

Acknowledgment

References

Alatorre, S., Sáiz, M. (2010). "Teachers and triangles." CERME 6–WORKING GROUP 10

Arzarello, F., & Ferretti, F. (in press) The connection between the mathematics INVALSI test and the teaching practices: an explorative study. In Falzetti P. (Ed.). I dati INVALSI: uno strumento per la ricerca. Milano: FrancoAngeli.

Ciani, A., Vannini, I. (2017) 'Equità e didattica. Validazione di scale sulle convinzioni di insegnamento democratico', *Cadmo*, 25(2), pp. 5–32. doi: 10.3280/CAD2017-002003.

D'Amore, B., Fandiño, Pinilla, M.I. (2005). Area e perimetro. Relazioni tra area e perimetro: convinzioni di insegnanti e studenti. *La matematica e la sua didattica*, 2, 165-190.

Duval R. (1993). Registres de représentations sémiotiques et fonctionnement cognitif de la pensée. *Annales de Didactique et de Sciences Cognitives*. 5, 37-65.

Fandiño-Pinilla, M.I, D'Amore, B. (2006). *Area e perimetro. Aspetti concettuali e didattici.* Trento: Erickson.

Gutiérrez, A., Jaime, A. (1999) Preservice Primary Teachers' Understanding of the Concept of Altitude of a Triangle. *Journal of Mathematics Teacher Education,* 2, 253–275 https://doi.org/10.1023/A:1009900719800

Italian Ministry of Education (2018). *Indicazioni nazionali e nuovi scenari*

'Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione' (2012) *Annali della Pubblica Istruzione*, special issue.

Jacquet, F., (2000). Il conflitto area-perimetro. *L'educazione matematica*, I parte: 2(2), 66-77; II parte: 2(3), 126-143.

Martini, B., Sbaragli, S. (2005). *Insegnare e apprendere la matematica*. Napoli: Tecnodid.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, *15*(2), 4-14.

Stavy, R., & Tirosh, D. (1999). Intuitive rules: a way to explain and predict students' reasoning. Educational Studies in Mathematics, 38, 51–66.

Tsamir, P., Tirosh, D., & Levenson, E. (2008). Intuitive nonexamples: The case of triangles. *Educational Studies in Mathematics*, *69*(2), 81-95.

Wiliam, D. (2010). Standardized testing and school accountability, *Educational Psychologist*, 45, 2: 107-122.