European Journal of Science and Mathematics Education, 2024, **12**(1), 11-20
ISSN: 2301-251X (Online)

OPEN ACCESS

Research Article

Management of semiotic representations in mathematics: Quantifications and new characterizations

Federica Ferretti [1]
0000-0002-4956-4136

Alessandro Gambini [2]
0000-0002-7779-6591

Camilla Spagnolo [3*]
0000-0002-9133-7578

[1] University of Ferrara, Ferrara, ITALY
[2] Sapienza University of Rome, Rome, ITALY
[3] Free University of Bozen-Bolzano, Bolzano, ITALY
[*] Corresponding author: Camilla.Spagnolo@unibz.it

Citation: Ferretti, F., Gambini, A., & Spagnolo, C. (2024). Management of semiotic representations in mathematics: Quantifications and new characterizations. *European Journal of Science and Mathematics Education, 12*(1), 11-20. https://doi.org/10.30935/scimath/13827

ARTICLE INFO

Received: 5 Apr 2023

Accepted: 9 Oct 2023

ABSTRACT

As highlighted in the literature, one of the main difficulties in mathematics is the management of different semiotic representations. This difficulty occurs in verticals throughout schooling and is often an obstacle to the proper learning process of mathematics. The present study aims to investigate the different facets of these difficulties with regard to mathematical tasks in secondary school. In particular, questions from Italian large-scale mathematics assessments are analyzed and interpreted through the theoretical lens of Duval's (1993) theory. Statistical analyses on a robust national sample allow a framing of the main difficulties and provide valuable information in this field.

Keywords: different representations, large-scale assessment, mixed methodology, mathematics education

INTRODUCTION

This work fits within a broader strand of research into the informed and formative use of large-scale mathematics assessments (De Lange, 2007; Looney, 2011). Along with the development of the most important international experiences (IEA-PIRLS, IEA-TIMSS, and OECD-PISA), in recent decades almost all countries have introduced large-scale national surveys of learning in primary and secondary schools (Eurydice, 2016). As highlighted in the literature (see for example, Doig, 2006), the institutional impact of large-scale assessments also has strong consequences at the classroom level, thus on teaching-learning processes. These surveys, constructed with the goal of assessing mathematics learning at the system level, are having increasing implications from educational, instructional, cultural-historical, and policy perspectives both locally and globally (Atkin, 1998; Breakspear, 2012; Kanes et al., 2014; Tasaki, 2017). All this has led to large-scale standardized assessments becoming, in different directions, the subject of research in mathematics education (De Lange, 2007; Kloosterman et al., 2015; Meinck et al., 2017; Suurtamm et al., 2016). In particular, several studies highlight the extent to which *mixed-method research* conducted from the analysis of large-scale mathematics assessment questions can characterize, quantify and measure the extent of several phenomena already studied in the literature framed by internationally shared constructs and contextualize them, within different educational systems (e.g., Facciaroni et al., 2023; Ferretti et al., 2022; Garuti & Martignone, 2019;

Mazza & Gambin, 2023). In this paper we focused on a macro-phenomenon that emerged in large-scale Italian evaluation regarding the handling of different semiotic representations in Duval's (2006) sense.

As highlighted by literature, the management of different semiotic representations represents one of the main obstacles in mathematics learning (Presmeg et al., 2016). The nature of mathematical objects, from both epistemological and ontogenetic points of view, means that they require an interpretation of different semiotic representations if they are to be understood. As some fundamental education studies report, the management of transformations within semiotic registers plays a crucial role (e.g., Duval, 2008). As we will see in the case study featured in this paper, the management of transformations within different semiotic registers causes difficulty for many pupils at different scholastic levels. Understanding the nature, structure and organization of learning in mathematics is a crucial problem of research in mathematics education.

In details, our study focuses on analysis of responses given by grade 8 and grade 10 pupils in Italian national large-scale tests (INVALSI) to two mathematical questions that require the management of different semiotic representations in different registers. Analysis of INVALSI data allows national quantification of phenomena highlighted in the literature in mathematics education (e.g., Casalvieri et al., 2023; Ferretti & Giberti, 2020; Spagnolo et al., 2021a). For each question, the INVALSI commission delivers the results of a national sample group with detailed statistical analysis that also considers students' mathematical competence. As we will explore further hereafter, a *mixed-method research* study (Hart et al., 2009) conducted with combined qualitative and quantitative analysis, allowed the measurement and quantification of the phenomena under investigation as well as a new definition in terms of student competence (Ferretti & Bolondi, 2021). As also shown in Gambini et al. (2017), macro-phenomena that emerged in INVALSI standardized assessment allows us to study and characterize the management of different semiotic representations.

THEORETICAL FRAMEWORK

As highlighted by Presmeg et al. (2016), semiotics has been at the center of various studies targeted at improving the processes involved in mathematics learning. One of the main research topics in mathematics education has been the management theories of different semiotic representations and their management within a semiotic system (Duval, 2008). The key concept is that the student does not encounter the mathematical object itself but one of its representations during the mathematics learning process. During the mathematics learning process, it is vitally important that all teachers clearly understand this concept, as otherwise, the students could become confused and mistake the semiotic representation for the mathematical object (Ferretti et al., 2022). When and if it is later necessary to modify the semiotic representation of the same "object", the student will not have the means (critical or cultural) at his/her disposal to perform the transformation, thus producing a cognitive gap. Thus, it is essential in mathematics education to guarantee good management of semiotic representations. According to Duval (2008), two distinct processes are involved: treatment and conversion. Treatment is a semiotic transition from one representation into another within the same register, whereas conversion is a semiotic transformation from one representation into another in distinct registers (Duval, 2008). According to D'Amore (2003), precisely the union of the choice of the distinctive features that we want to highlight of the concept, the treatments, and the conversions are the three "actions" on a concept that represents the "construction of knowledge in mathematics". This approach forms part of a constructivist perspective, where mathematics learning is strictly connected to the ability to use semiotic representations, which emerges in representing them in a given register, treating them in the same register, or converting them if they are in two different registers. It therefore turns out to be necessary, for the construction of mathematical learning, to go through the awareness and mastery of the three components of semiotics that are considered today as explicit goals of the teaching-learning process and that are considered "transversal practices", being external to any practice and specific to the overall management of mathematics teaching-learning. it should be considered that the consequences of Duval's (1993) paradox could occur, according to which there is a risk that the student does not arrive at noetic (conceptual learning) but remains at semiotic handling. This could happen when the teacher makes the student handle an excess of semiotic representations of the same concept, or when he or she does not provide a sufficient variety of representations. The paradoxical situation mentioned above, makes it necessary to pay close attention from an educational point of view to this topic, analyzing in depth the students' beliefs, the distinctive features they choose, the motivations behind these choices, the meaning they attach to these representations, the different semiotic registers and the different representations proposed (Duval, 1998). As we will see in this paper, the management of semiotic representations within the same register is a source of much difficulty for students. The questions that guided our research are: among macro-phenomena that have emerged in large-scale assessments that highlight widespread difficulties, can the theoretical lenses of semiotic representation management provide suitable interpretive lenses? If so, can data analysis provide quantification and further characterization in terms of treatments and conversions in Duval's (1993) sense?

METHODOLOGY

Our research focuses on analysis of results from standardized national assessment tests, INVALSI tests. In the Italian context, we have the possibility to track some students' difficulties over time thanks to INVALSI tests (tests with the purpose of measuring students' levels of competence in relation to the Italian curricular Guidelines), which were administered since in 2008 in grade 2, grade 5, grade 8, grade 10, and grade 13 from the National Institute for the Evaluation of the Educational System (from 2009 to 2013, the tests also covered grade 6).

The Italian Ministry of Public Education established that INVALSI (www.invalsi.it) would have been responsible for the standardized assessment of the Italian educational system to conduct yearly surveys of all students in second and fifth years of primary school (grade 2 and grade 5), third year of middle school (grade 8), and second and fifth years of high school (grade 10 and grade 13). The national evaluation system, in particular, is run by the INVALSI Institute, which also periodically and systematically assesses students' knowledge and skills (relating to reading comprehension, grammatical knowledge, and mathematical competency) and the overall quality of educational opportunities provided by schools and vocational training institutions. The major objective of INVALSI standardized tests, which were developed for system evaluation. The tests are administered every year at census level and student results are provided to each school institution. Results and questions of INVALSI tests are considered as a resource also for researchers in the field of mathematics education (Garuti & Martignone, 2019) and are used in national and international research (e.g., Spagnolo et al., 2021b, 2022).

In recent decades, large-scale national and international evaluations have developed and are increasingly coming into the school world on different levels (De Lange, 2007). While the comparisons with regional, national, and sometimes other countries are increasingly at the center of political, social, and educational debates in the school world, including at the decision-making level, timely returns at the class and school level can provide valuable information to individual contexts. INVALSI is the research body that annually collects data regarding Italian students' learning at different scholastic levels in Italy. INVALSI census tests are administered throughout Italy and the data are analyzed on statistically significant samples gathered nationwide; various research studies in mathematics learning have developed due to phenomena identified from INVALSI data (Casalvieri et al., 2023; Ferretti & Bolondi, 2021; Spagnolo et al., 2021a). Theoretical framework of INVALSI mathematics test consider both the Italian national guidelines and the key concepts of both national and international mathematics education (INVALSI, 2018). Indexes of classical test theory and the application of the Rasch model (Callingham & Bond, 2006; Rasch, 1960) serve as validation of INVALSI tests. This logistic parameter model enables an approximated assessment of each item's difficulty and student ability. Rasch model expresses probability of giving the right response to an item in terms of both the item's intrinsic difficulty and the student's ability as determined by the complete test. Characteristic curve has been denominated by the logistic curve that the model thusly generated. Specific graphs that allow a comparison between the item characteristic curve and model output with empirical data, particularly the trend of each answer option while considering the student's ability levels as demonstrated by the overall test, provide crucial information on the item behavior. Both percentage terms (probability of choice) and characteristic curve obtained by applying the Rasch model to INVALSI sample group will be used to present the item outcomes in this study. In further detail, the research methodology we used in this study is mixed, qualitative-quantitative (Hart et al., 2009), and the purpose of combining the research components is to integrate the

The square root of **64^{2016}** is

A. ☐ 8^{2014}

B. ☐ 8^{1008}

C. ☐ 64^{2014}

D. ☐ 64^{1008}

Figure 1. Task administered nationwide in 2016 INVALSI test, grade 10 (Source: https://www.gestinv.it/Index.aspx; translation edited by the authors)

results obtained in each (Schoonenboom & Johnson, 2017). According to Johnson and Onwuegbuzie (2004), QUAN→Quan+QUAL scheme is incorporated into the research method. Using Duval's (1993) semiotic theory lens technique, the initial research step involved a search among the large-scale assessment outcomes (for examples different representation conversion problems). This study was conducted using the research tool Gestinv (www.gestinv.it), INVALSI database already used in previous research studies (e.g., Ferretti et al., 2022). Gestinv allows to search, among all INVALSI tasks administered to Italian students, for answers to tasks that showed the lowest percentage of correct answers nationwide (QUAN stage). Results were used to address research question of this study, examining the ability to manipulate representations of mathematical objects in different registers, specifically the ability to perform conversion as laid down by Duval (2006). In last stage, Quan+QUAL stage, national results and their respective statistical analysis were integrated with the qualitative approach through a theoretical lens. Analysis was carried out specifically on the data related to the percentage of chosen options and the percentage of correct answers. Finally, a characteristic curve interpretation allowed for a deeper examination of student performance in relation to their ability determined through entire test.

RESULTS

Tasks chosen in this study are INVALSI tasks (as also specified in the previous section). In our study we focus on INVALSI tasks of grade 8 and grade 10, which in Italy are the transition years from middle school to high school. In addition, we chose tasks from the Gestinv database (www.gestinv.it) that focused on perform an operation on a number.

The task shown in **Figure 1** was set for all Italian students in grade 10, the second year of secondary school, during the scholastic year 2015/16. To answer this question correctly, it is necessary to identify which number present in the four answer options represents the square root of the power stated in the text. The necessary operation (the square root) is stated in verbal form.

The content and skills required to complete this task are in accordance with the national guidelines (MIUR, 2012). The numbers in all four options are transformations of the number in the question text, with manipulations of either the base or exponent. The correct answer is option D. The question was administered to approximately 550,000 students in the second year of secondary school. The results were analyzed from a significant sample of approximately 34,000 students. Only 26.4% of students provided the correct answer, 69.1% were wrong, and 4.5% failed to provide an answer or offered invalid replies. The percentage of wrong answers was very high, resulting as one of the highest percentages of wrong answers from the entire INVALSI tests given to grade 10. Among the wrong answers, B was the most popular. Option B was chosen by 49.6% of students, by more than two-thirds of the students who made a mistake. In the number of option B, the power exponent in question text was correctly divided, but the base was also mistakenly divided. The numbers in option B and the question text are represented in the same semiotic register, but the required operation (the square root) is expressed verbally. To answer this question, it is necessary to perform a transformation of two representations in different semiotic registers (from verbal to numeric register), in what Duval (2006) terms a conversion.

The graph shown in **Figure 2** is the characteristic curve of this question, showing the probability of choice for each option according to the students' skill level throughout the test. The performance of the correct answer (dotted violet line) is in accordance with the model prediction (continued violet line), which suggests that the question is valid from an assessment point of view (Rasch, 1960). Option D (dotted light blue line) was

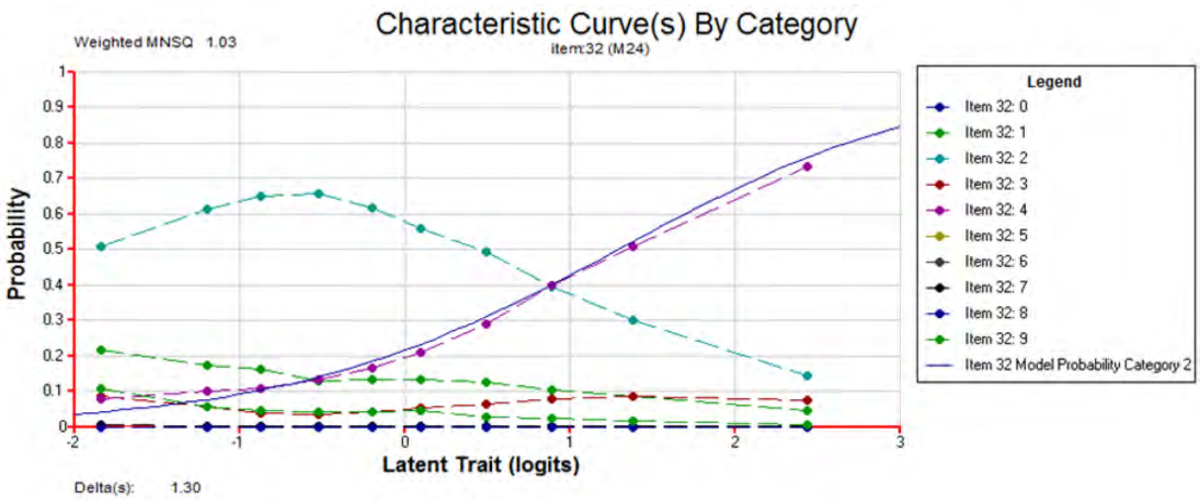

Figure 2. Characteristic curve of answers nationwide referring to grade 10 task investigated (Source: https://www.gestinv.it/Index.aspx)

The tenth part of 10^{20} is

A. ☐ 10^{10}

B. ☐ 1^{20}

C. ☐ 100

D. ☐ 10^{19}

Figure 3. Task administered nationwide in 2012 INVALSI test, grade 8 (Source: https://www.gestinv.it/Index.aspx; translation edited by the authors)

the most popular choice of students with the highest skill levels (as far as the eighth decile) and among medium skill levels it was chosen by more than 60.0% of students. This suggests that the difficulty of managing conversion between two different semiotic representations is limited to low-level mathematics students and (particularly) students who demonstrate medium/good skill in mathematics. The correct answer was chosen mostly by students in the top two deciles of competence.

The task shown in **Figure 3** was administered to all Italian students in the third year of middle school as part of the INVALSI grade 8 test in 2012. The task calls for manipulating semiotic representations in two different registers, which Duval (2006) terms a conversion. The task verbally instructs the student to calculate a power to the tenth (expressed numerically). The correct answer is option D.

In 2012, INVALSI test was part of the state-managed middle school leaving certificate and so student performance was generally always good in comparison with other tests (means that INVALSI test at grade 8 are considered by students as "higher-stake tests", compared with tests in other grades). The task was set for approximately 590,000 students nationwide in the third year of middle school. The statistical analysis of the data was carried out on a nationwide sample of around 10,000 students. Only 26.2% of students provided the correct answer, 2.9% failed to answer or provided an invalid response, while 70.9% selected the wrong answer. In this case, such low numbers of correct answers and high numbers of wrong answers places this task among one of the most difficult sets in INVALSI tests for grade 8. However, the skills and knowledge required are following those established for the end of third-year middle school by the national guidelines (MIUR, 2012). Again in this case, as we can see in the characteristic curve in **Figure 4**, the correct answer (dotted violet line) is following the model prediction (violet curve). For very low skill levels, all three wrong answers had a 20.0% probability of being chosen–specifically, almost half the students in the first decile were likely to choose option A (green line). Option A, where "10 is subtracted" from the exponent of the power present in the text, was the most popular option among the very high levels (as far as seventh decile). Option

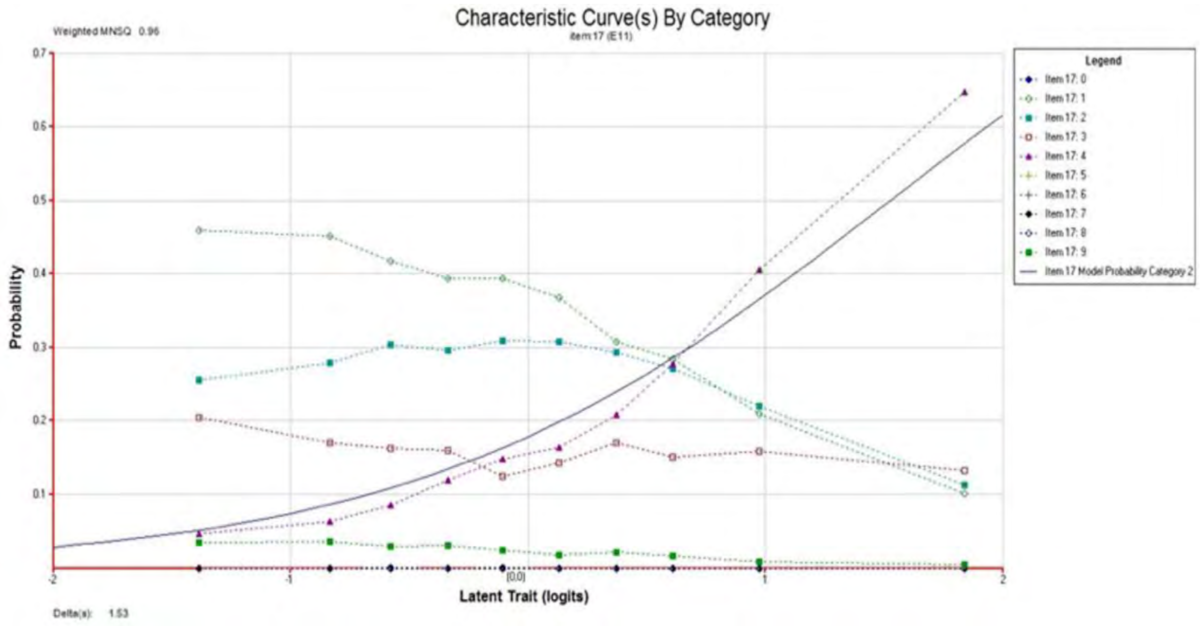

Figure 4. Characteristic curve of answers nationwide referring to grade 8 task investigated (Source: https://www.gestinv.it/Index.aspx)

B was also popular, where the "tenth part" of the base stated in the text is calculated; indeed, as the graph shows, this was chosen by more than 30.0% of medium-ability students. The correct answer was popular only with students in the top two ability deciles.

DISCUSSION

This paper analyzed answers provided by students in a large-scale national mathematics assessment test for grade 8 and grade 10. In the tasks analyzed in this study, selecting the correct answer requires a strong mastery of manipulating two semiotic representations of the same mathematical object in numeric form. The majority of students in both scholastic levels investigated provided an answer, and there were not many missing or incorrect answers; this suggests that the subject was well-known to the students and that they felt competent in answering the task. For this task, student does not have the contextual meaning of the activity, and manipulable objects have various semiotic registers. Most of students who chose incorrectly, according to Duval (2008), did so by selecting options, where it is obviously difficult to execute the conversion between two semiotic registers. This is shown through analysis of students' answers. In both the tasks investigated, performing an operation (described verbally) regarding a power expressed in the question text is necessary. To answer correctly, it is necessary to manipulate the exponents; for both the scholastic levels in question, the majority of students who made a mistake carried out the "verbally-expressed operation" also, or only, to the base number. As can be seen from the statistical data, in the task for grade 10 (**Figure 1**), most students who made a mistake chose option B, where the power has the correct exponent but the wrong base. Indeed, the base number results from the "verbally-expressed operation" of the power base in the task. The same trend can be seen when analyzing the task results for grade 8. Also, in this case, most students who answered incorrectly opted for A and B. The power in option A has the same base as the correct answer but a different exponent: in this case, there has probably been a misunderstanding of the operation expressed in verbal language. The mistake made by students who chose this option echoes the most frequent mistake by grade 10 students: in both cases, verbally-described operation is carried out on base number instead of exponent.

The request in both items can be described, as follows: we have a task, which, in both cases, is "perform an operation on a number" (as we have already pointed out). The operation is expressed in the verbal register, the number in the symbolic one. The answer must be chosen among numbers expressed in symbols. Conversion must be performed on operation: pupils must covert the operation from one register to the other.

In both cases, the most "popular" option is the one, where pupils perform a conversion of the operation expressed in words, which results to be confused and wrong since *there are meaningless interference of repeated school practices*. In the first case, "square root" is converted, in option B, in "divide exponents" (which is related to a correct conversion) + "square root of the basis" (which is a "literal" conversion of the verbally expressed task). In the second case, "divide by ten" is converted, in the most popular option, in "subtract exponents". So, we may say that these tasks show how repeated treatment practices within one register may affect the ability of converting between different registers.

The phenomenon investigated is revealed at two different scholastic levels, with the same characteristics: the difficulties investigated lie in the management of conversion (as outlined by Duval, 2008) between verbal and semiotic registers. The problem is highlighted in a significant sample–around 272,800 students chose option B in grade 10 INVALSI task (16,800 from the sample analysis group). Furthermore, characteristic curve's analysis also provides accurate information on how students behave, and their mathematical ability measured across whole test. Students seem significantly more inclined to select incorrect response up to medium-high ability levels, especially when selecting options examined above, as shown by both characteristic curves (**Figure 2** and **Figure 4**). This highlighted that a large number of students who reflect higher ability can be affected by this phenomenon, in addition to "weaker" students. The fact that for both scholastic grades the curves are similar (in fact, the graphs show that the issue under investigation appears more commonly with reference to the grade 10 task), means that the phenomenon does not "disappear" as students move up the school, i.e., it persists despite the acquisition of greater mathematic skills and knowledge.

CONCLUDING REMARKS

The management of semiotic representations represents a widespread difficulty that often affects mathematics learning (Duval, 2008; Presmeg et al., 2016). The difficulties highlighted and widely studied in the literature are studied and quantified nationwide in our study's large-scale assessment test. Identification of phenomena and examination of specific features were made possible by analyzing test answers from a significant sample of Italian students (Ferretti & Giberti, 2020; Ferretti & Santi, 2023; Spagnolo et al., 2022). Thus, these tools do not simply let to assess results of tasks, but also the processes that students engage in. While examination of the characteristic curve (Rasch, 1960) emphasized particular behavior according to the ability level of students as specified in the test, the percentage numbers of the individual alternatives provided accurate information about mistakes made by students. The fact that the research focused on two different scholastic grades, which correspond to two different scholastic levels in Italy, also allows the compilation of vertical data. In fact, analysis of the two task's results shows similar trends and peculiarities regarding the low percentage of correct answers and the wrong options chosen. Specifically, the results highlight the same difficulties in conversion (Duval, 2008) between two semiotic registers. The characteristic curves reveal that this issue also regards students with medium-high levels of mathematical ability.

Data collection and analysis of the outcomes of large-scale national and international assessment tests can offer an opportunity to examine results through a quantitative lens and suggest (as in our case) specific issues about the macro-phenomena that emerge at the systematic level. This has already been mentioned in previous studies (e.g., Giberti et al., 2023; Spagnolo et al., 2021b, 2022). These tests frequently reveal "phenomena" of students' behavior in response to various math tasks, as well as information about their learning habits and processes, especially in terms of the dynamics and potential sources of their difficulties. Data analysis and statistical measurement offer details on typical student difficulties as well as global achievements at school system level. In order to study and develop key constructs in mathematics education in the future, this study suggests starting points for additional research, where qualitative and quantitative analysis carried out on extensive assessment tests may be crucial.

Author contributions: FF, AG, & CS: contributed to design & implementation of research, analysis of results, writing–original draft, writing–review & editing, & conceptualization; **FF & CS:** data curation; **AG & CS:** methodology; & **AG:** supervision. All authors approved the final version of the article.

Funding: The authors received no financial support for the research and/or authorship of this article.

Acknowledgements: The authors would like to thank all students & teachers involved in experimentation & INVALSI for support in implementation of this study. Nonetheless, views expressed in this work are attributable exclusively to the authors, & thus, in citing work, it is not appropriate to attribute arguments expressed herein to INVALSI.

Ethics declaration: The authors declared that the study was conducted according to guidelines of the Declaration of Helsinki & the Ethical code of the Free University of Bozen-Bolzano. Written informed consents were obtained from the participants.

Declaration of interest: The authors declared no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

Atkin, J. M. (1998). The OECD study of innovations in science, mathematics and technology education. *Journal of Curriculum Studies, 30*(6), 647-660. https://doi.org/10.1080/002202798183369

Breakspear, S. (2012). *The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance*. OECD Publishing. https://doi.org/10.1787/5k9fdfqffr28-en

Callingham, R., & Bond, T. (2006). Research in mathematics education and Rasch measurement. *Mathematics Education Research Journal, 18*(2), 1-10. https://doi.org/10.1007/BF03217432

Campbell, C., & Levin, B. (2009). Using data to support educational improvement. *Educational Assessment, Evaluation and Accountability, 21*(1), 47-65. https://doi.org/10.1007/s11092-008-9063-x

Casalvieri, C., Gambini, A., Spagnolo, C., & Viola, G. (2023). The concept of derivatives through eye-tracker analysis. In *Proceedings of the 15th International Conference on Computer Supported Education* (pp. 378-385). https://doi.org/10.5220/0011941100003470

D'Amore, B. (2003). La complexité de la noétique en mathématiques ou les raisons de la dévolution manquée [The complexity of noetic in mathematics or the reasons for failed devolution]. *For the Learning of Mathematics, 23*(1), 47-51.

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). Information Age Publishing.

Doig, B. (2006). Large-scale mathematics assessment: Looking globally to act locally. *Assessment in Education: Principles, Policy & Practice, 13*(3), 265-288. https://doi.org/10.1080/09695940601035403

Duval, R. (1993). Registres de représentations sémiotique et fonctionnement cognitif de la pensée [Registers of semiotic representations and cognitive functioning of thought]. *Annales de Didactique et de Sciences Cognitives* [*Annals of Didactics and Cognitive Sciences*], *5*, 37-65.

Duval, R. (1998). Signe et object (I). Trois grandes étapes dans la problématique des rapports entre répresentation et objet [Sign and object (I). Three main stages in the problem of the relationship between representation and object]. *Annales de Didactique et de Sciences Cognitives* [*Annals of Didactics and Cognitive Sciences*], *6*, 139-163.

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics, 61*(1-2), 103-131. https://doi.org/10.1007/s10649-006-0400-z

Duval, R. (2008). Eight problems for a semiotic approach in mathematics education. In L. Radford, G. Schubring, & F. Seeger (Eds.), *Semiotics in mathematics education: Epistemology, history, classroom, and culture* (pp. 39-61). Sense Publishers. https://doi.org/10.1163/9789087905972_004

Eurydice, I. U. I. (2016). *La valutazione delle scuole in Europa: Politiche e approcci in alcuni Paesi europei* [*School evaluation in Europe: Policies and approaches in some European countries*]. Eurydice Italia.

Facciaroni, L., Gambini, A., & Mazza, L. (2023). The difficulties in geometry: A quantitative analysis based on results of mathematics competitions in Italy. *European Journal of Science and Mathematics Education, 11*(2), 259-270. https://doi.org/10.30935/scimath/12590

Ferretti, F., & Bolondi, G. (2021). This cannot be the result! The didactic phenomenon 'the age of the earth'. *International Journal of Mathematical Education in Science and Technology, 52*(2), 194-207. https://doi.org/10.1080/0020739X.2019.1670366

Ferretti, F., & Giberti, C. (2020). The properties of powers: Didactic contract and gender gap. *International Journal of Science and Mathematical Education, 19*(8), 1717-1735. https://doi.org/10.1007/s10763-020-10130-5

Ferretti, F., & Santi, G. (2023). Teachers' subjectification as a dialogue with standardized assessment. *For the Learning of Mathematics, 43*(2), 19-21.

Ferretti, F., Santi, G. R. P., & Bolondi, G. (2022). Interpreting difficulties in the learning of algebraic inequalities, as an emerging macro-phenomenon in large scale assessment. *Research in Mathematics Education, 24*(3), 367-389. https://doi.org/10.1080/14794802.2021.2010236

Gambini, A., Desimoni, M., & Ferretti, F. (2022). Predictive tools for university performance: An explorative study. *International Journal of Mathematical Education in Science and Technology*. https://doi.org/10.1080/0020739X.2021.2022794

Garuti, R., & Martignone, F. (2019). Assessment and argumentation: An analysis of mathematics standardized items. In U. T. Jankvist, M. van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the 11th Congress of the European Society for Research in Mathematics Education* (pp. 4075-4082). Freudenthal Group & Freudenthal Institute, Utrecht University, and ERME.

Giberti, C., Santi, G., & Spagnolo, C. (2023). The role of metaphors in interpreting students' difficulties in operating with percentages: A mixed method study based on large scale assessment. *European Journal of Science and Mathematics Education, 11*(2), 297-321. https://doi.org/10.30935/scimath/12642

Hart, L. C., Smith, S. Z., Swars, S. L., & Smith, M. E. (2009). An examination of research methods in mathematics education (1995-2005). *Journal of Mixed Methods Research, 3*(1), 26-41. https://doi.org/10.1177/1558689808325771

INVALSI (2018). *Quadro di riferimento delle prove di INVALSI di matematica* [Reference framework for the INVALSI mathematics tests]. https://invalsiareaprove.cineca.it/docs/file/QdR_MATEMATICA.pdf

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26. https://doi.org/10.3102/0013189X033007014

Kanes, C., Morgan, C., & Tsatsaroni, A. (2014). The PISA mathematics regime: Knowledge structures and practices of the self. *Educational Studies in Mathematics, 87*, 145-165. https://doi.org/10.1007/s10649-014-9542-6

Kloosterman, P., Walcott, C., Brown, N. J. S., Mohr, D., Pérez, A., Dai, S., Roach, M., Hall, L. D., & Huang, H.-C. (2015). Using NAEP to analyze eighth-grade students' ability to reason algebraically. In J. A. Middleton, J. Cai, & S. Hwang (Eds.), *Large-scale studies in mathematics education* (pp. 179-207). Springer. https://doi.org/10.1007/978-3-319-07716-1_9

Looney, J. W. (2011). *Integrating formative and summative assessment: Progress toward a seamless system?* OECD Publishing.

Mazza, L., & Gambini, A. (2023). The phenomenon of the gender gap among gifted students: the situation in Italy based on analysis of results in mathematics competitions. *Teaching Mathematics and its Applications: An International Journal of the IMA, 42*(2), 170-186. https://doi.org/10.1093/teamat/hrac009

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

MIUR. (2012). *Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo di istruzione* [*National indications for the curriculum of nursery school and the first cycle of education*]. Le Monnier.

Presmeg, N., Radford, L., Roth, W. M., & Kadunz, G. (2016). *Semiotics in mathematics education.* Springer. https://doi.org/10.1007/978-3-319-31370-2

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research.

Schoonenboom, J., & Johnson, R. B. (2017). How to construct a mixed methods research design. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* [*Cologne Journal of Sociology and Social Psychology*], *69*(Supplement 2), 107-131. https://doi.org/10.1007/s11577-017-0454-1

Spagnolo, C., Capone, R., & Gambini, A. (2021a). Where do students focus their attention on solving mathematical tasks? An eye tracker explorative study. In M. Prasitha, N. Changsri, & N. Boonsena (Eds.), *Proceedings of the 44th Conference of the International Group for the Psychology of Mathematics Education* (pp. 89-96). PME.

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. Uhomoibhi (Eds.), *Proceedings of the International Conference on Computer Supported Education* (pp. 560-576). Springer. https://doi.org/10.1007/978-3-031-14756-2_27

Spagnolo, C., Giglio, R., Tiralongo, S., Bolondi, G. (2021b). Formative assessment in LDL: A teacher-training experiment. In *Proceedings of the 13th International Conference on Computer Supported Education* (pp. 657-664). https://doi.org/10.5220/0010496006570664

Tasaki, N. (2017). The impact of OECD-PISA results on Japanese educational policy. *European Journal of Education, 52*(2), 145-153. https://doi.org/10.1111/ejed.12217

◆❖◆