

FOSTERING THE EVALUATION CAPACITY BUILDING FOR SCHOOL IMPROVEMENT

VIII Seminar "INVALSI data: a tool
for teaching and scientific research"

edited by
Patrizia Falzetti

FrancoAngeli 



INVALSI PER LA RICERCA
STUDI E RICERCHE



INVALSI PER LA RICERCA

La collana Open Access INVALSI PER LA RICERCA si pone come obiettivo la diffusione degli esiti delle attività di ricerca promosse dall'Istituto, favorendo lo scambio di esperienze e conoscenze con il mondo accademico e scolastico.

La collana è articolata in tre sezioni: "Studi e ricerche", i cui contributi sono sottoposti a revisione in doppio cieco, "Percorsi e strumenti", di taglio più divulgativo o di approfondimento, sottoposta a singolo referaggio, e "Rapporti di ricerca e sperimentazioni", le cui pubblicazioni riguardano le attività di ricerca e sperimentazione dell'Istituto e non sono sottoposte a revisione.

Direzione: Roberto Ricci

Comitato scientifico:

- Tommaso Agasisti (Politecnico di Milano);
- Gabriella Agrusti (Università LUMSA, sede di Roma);
- Cinzia Angelini (Università Roma Tre);
- Giorgio Asquini (Sapienza Università di Roma);
- Carlo Barone (Istituto di Studi politici di Parigi);
- Maria Giuseppina Bartolini (Università di Modena e Reggio Emilia);
- Giorgio Bolondi (Libera Università di Bolzano);
- Francesca Borgonovi (OCSE•PISA, Parigi);
- Roberta Cardarello (Università di Modena e Reggio Emilia);
- Lerida Cisotto (Università di Padova);
- Alessandra Decataldo (Università degli Studi Milano Bicocca);
- Patrizia Falzetti (INVALSI);
- Michela Freddano (INVALSI);
- Martina Irsara (Libera Università di Bolzano);
- Paolo Landri (CNR);
- Bruno Losito (Università Roma Tre);
- Annamaria Lusardi (George Washington University School of Business, USA);
- Alessia Mattei (INVALSI);
- Stefania Mignani (Università di Bologna);
- Marcella Milana (Università di Verona);
- Paola Monari (Università di Bologna);
- Maria Gabriella Ottaviani (Sapienza Università di Roma);
- Laura Palmerio (INVALSI);
- Mauro Palumbo (Università di Genova);
- Emmanuele Pavolini (Università di Macerata);
- Donatella Poliandri (INVALSI);
- Arduino Salatin (Istituto Universitario Salesiano di Venezia);
- Jaap Scheerens (Università di Twente, Paesi Bassi);
- Paolo Sestito (Banca d'Italia);
- Nicoletta Stame (Sapienza Università di Roma);
- Gabriele Tomei (Università di Pisa);
- Roberto Trincherò (Università di Torino);
- Matteo Viale (Università di Bologna);
- Assunta Viteritti (Sapienza Università di Roma);
- Alberto Zuliani (Sapienza Università di Roma).

Comitato editoriale:

Andrea Biggera; Nicola Giampietro; Simona Incerto; Francesca Leggi; Rita Marzoli (coordinatrice); Daniela Torti.



OPEN ACCESS FrancoAngeli solution

This volume is published in open access format, i.e. the file of the entire work can be freely downloaded from the FrancoAngeli Open Access platform (<http://bit.ly/francoangeli-oa>).

On the FrancoAngeli Open Access platform, it is possible to publish articles and monographs, according to ethical and quality standards while ensuring open access to the content itself. It guarantees the preservation in the major international OA archives and repositories. Through the integration with its entire catalog of publications and series, FrancoAngeli also maximizes visibility, user accessibility and impact for the author.

Read more: [Publish with us \(francoangeli.it\)](http://francoangeli.it)

Readers who wish to find out about the books and periodicals published by us can visit our website www.francoangeli.it and subscribe to “[Keep me informed](#)” service to receive e-mail notifications.

FOSTERING THE EVALUATION CAPACITY BUILDING FOR SCHOOL IMPROVEMENT

VIII Seminar "INVALSI data: a tool
for teaching and scientific research"

edited by
Patrizia Falzetti



FrancoAngeli 

The opinions expressed are solely of the author(s). In no case should they be considered or construed as representing an official position of INVALSI.

Assistant Editor: Francesca Leggi.

Isbn: 9788835181385

Isbn e-book Open Access: 9788835191087

Copyright © 2026 by FrancoAngeli s.r.l., Milan, Italy.

This work, and each part thereof, is protected by copyright law and is published in this digital version under the license Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0).

Text and Data Mining (TDM), AI training and similar technologies rights are reserved.

By downloading this work, the User accepts all the conditions of the license agreement for the work as stated and set out on the website
<https://creativecommons.org/licenses/by-nc-nd/4.0>

Index

Introduction <i>Patrizia Falzetti</i>	pag. 7
1. Deepening teachers' beliefs, practices, and school contexts during Emergency Remote Teaching. The role of INVALSI data <i>Marco Giganti, Emanuele Marcora</i>	» 9
2. Supporting teachers' self-evaluation and competence development for the use of artificial intelligence in the classroom: the Erasmus AI4T project <i>Sara Mori, Jessica Niewint, Francesca Storai</i>	» 39
3. Exploring school personnel perspectives on a development program for school self-evaluation <i>Giuseppe C. Pillera, Letizia Giampietro, Donatella Poliandri</i>	» 63
4. Small schools for self-evaluation <i>Graziella Arazzi</i>	» 98
5. What makes a good reader? Evidence from PIRLS 2021 <i>Valeria F. Tortora, Paola Giangiacomo, Michele Cardone</i>	» 125
The authors	» 147

*1. Deepening teachers' beliefs, practices,
and school contexts
during Emergency Remote Teaching.
The role of INVALSI data*

by Marco Giganti, Emanuele Marcora

The contribution presents a multiple case study of three omnicomprehensive schools in Lombardy aimed at observing and describing their context, teachers' beliefs, and practice statements about formative assessment, learning, and student engagement in the Emergency Remote Teaching activated during the critical period of the Covid-19 pandemic, and the relationship between them and the context. Specifically, the qualitative-quantitative analyses carried out on the results emerging from the semi-structured interviews conducted with Headmasters, document analysis (the three-year plan of educational offerings, self-evaluation report, circulars, policy acts, etc.), the questionnaire submitted to teachers, and focus groups conducted with the HM, the internal evaluation team and selected teachers by school order will be presented. In the research in general and in the introduction to the last phase in particular, INVALSI data from the 2021 surveys were used in order to propose a comparison and initiate reflections on the performance of the schools. The case study is part of a larger research project of which hints will be provided.

Il contributo presenta un caso studio multiplo su tre Istituti omnicomprendivi in Lombardia finalizzato a osservare e descrivere il loro contesto, le credenze degli insegnanti e le dichiarazioni di pratiche sulla valutazione formativa, l'apprendimento e l'impegno degli studenti nell'Emergency Remote Teaching attivato durante il periodo critico della pandemia di Covid-19 e la relazione tra loro e il contesto. Nello specifico, saranno presentate le analisi quali-quantitative effettuate sui risultati emersi dalle interviste semi-strutturate condotte con i presidi, dalle analisi documentali (PTOF, RAV, circolari, atti di indirizzo ecc.), dal questionario somministrato agli insegnanti e dai focus group condotti con il dirigente, il NIV e due insegnanti selezionati per

ordine scolastico. Nella ricerca in generale e nell'introduzione all'ultima fase in particolare, sono stati utilizzati i dati INVALSI delle indagini del 2021 per proporre un confronto e avviare riflessioni sulle prestazioni delle scuole. Lo studio di caso fa parte di un progetto di ricerca più ampio di cui saranno forniti alcuni riferimenti.

1. Introduction and theoretical background

The centrality of assessment in the learning paradigm is a theme widely developed in the literature and finds particular expression in formative assessment (Scriven, 1967; Bloom, 1971; Black and William, 1998; Weeden, Winter and Broadfoot, 2009; Brown *et al.*, 1992).

The pedagogical potential and social effects of this kind of assessment have also long been considered by UNESCO and the Council of Europe (CoE), which, during the first phase of Covid-19 spread, saw fit to incentivize its use to support the engagement of students then their learning process (Engzell, Frey and Verhagen, 2020; UNESCO, 2020c, 2020d, 2020e; Hughes, 2020).

In spring 2020, national governments decided to close schools and abruptly switch to distance learning as measures to counteract infection and not disrupt educational activities. In most cases, school systems have adopted Distance Learning (DL) through e-learning or videoconferencing platforms (UNESCO, 2020a, 2020b). Some research sees formative assessment as a useful tool to foster student engagement and inclusion, especially in online education. A study by Z. Chen, J. Jiao, and K. Hu (2021), carried out in the pandemic period, finds that temporally and spatially separated students and teachers are subject to long-term criticality typical of distance education.

In this case, it is appropriate to take up the distinction postulated by Hodges *et al.* (2020) between online distance learning (DL) and emergency remote teaching (ERT). Unlike educational experiences designed to be online, ERT is a temporary and alternative mode of instructional delivery due to crisis circumstances. It involves fully distance teaching strategies; education, which would otherwise be delivered face-to-face or as blended or hybrid teaching, will revert to the traditional format when the emergency has receded. In this context, the main goal is not to recreate a robust educational ecosystem but to provide temporary access to education and learning supports, while also preventing possible risks of disengagement typical of emergencies.

Research shows that students feel isolated in online education (Hammond, 2009), resulting in high dropout rates (Hodges and Kim, 2010), high boredom rates, and poor performance (Fredricks, 2015). According to Chen,

Jiao, and Hu, the use of formative assessment in such contexts can be an effective solution. Implemented with forum discussion tasks, quizzes, and tests, it is effective in improving engagement and increasing learning. The studies mentioned explicitly refer to a mixed mode of teaching between online and presence, specifically concerning DL. According to the authors, there is limited evidence that this also works in the typical ERT mode.

Central to the management of such didactic has been the teaching profession, led by beliefs and transmuted into practice. The most problematic aspects emerging from empirical research (Richardson, 2002) concern the link between conceptions and effectiveness of teacher training (Balduzzi and Vannini, 2008). In this regard, there are studies inherent to teacher change (Floden, 2002; Richardson and Placier, 2002) and how beliefs and practices change: do the former follow the latter (and vice versa) or are they mutually interacting and synergistic (Goffman, 1973; Peterman, 1993)?

This contribution aims to present a multiple case study of three omnicomprehensive schools (primary to high school) in the Lombardy Region, to investigate the context in which teachers operated during COVID-19, their beliefs, and practice statements about formative assessment, learning, and engagement during ERT, what are the links between context and beliefs and practices. In the first stage, semi-structured interviews were conducted with principals, document analysis (the three-year plan of educational offerings, self-evaluation report, policy acts, circulars, etc.), and qualitative analysis was conducted on them. In the second stage, a questionnaire was administered to all teachers in the identified schools, aimed at investigating the constructs examined, and the data were analyzed quantitatively and qualitatively. In the third phase and from a constructivist perspective, a focus group was conducted with the school principal, the Internal Evaluation Core, and teachers selected by order of school to jointly interpret the data that emerged in the first two phases, and the data were analyzed qualitatively.

Analyses show that teachers favored formative assessment convinced that it could support student learning and engagement in a context considered emergent. Specifically, statistical analyses show that beliefs and practice statements about formative assessment are related to constructivist learning and student engagement; they are found to be consistent and applicable in the ERT context. It is interesting to note that the age and experience of the teachers are a conditioning factor as well as the order of teaching and the emergency. Also, the school influenced the formative assessment and learning beliefs: the smaller one, identifiably stronger with accompanying practices and teacher training was decisive on the agreement averages. The primary school also proved to be a favorable context for formative assessment (FA),

student engagement, and constructivist learning. Other hints of the analyses conducted will be given during the presentation of the contribution.

2. Research objectives and hypothesis

The main objectives are referred to three different areas: *theoretical-conceptual* to identify whether and what relationship exists between teachers' beliefs and practices on formative assessment in ERT; *methodological* to analyse and describe the school context, investigate teachers' beliefs and practices on formative assessment, student engagement and learning, and ERT; and *transformative* to engage school principals and teachers as active participants in the research and, at the end of the data analysis, design possible training paths in relation to the issues that emerged from the research. Three general questions were identified to guide the study: what is the school context in which teachers might have used formative assessment during ERT? (RQ1); What are teachers' beliefs and practice statements about formative assessment, engagement, and learning in the ERT context? Is there a relationship? What kind? (RQ2); how does the school context relate to teachers' beliefs and practice statements? (RQ3). After defining the problem and research questions, two general research hypotheses were formulated: there is a correlation between teachers' beliefs and practice statements about FA, learning, and student engagement in the ERT context (GHp1); school context is related to teachers' beliefs and practice statements (GHp2). The formulation of the following specific research hypotheses is intended to assume precision and concreteness with respect to the general ones: teachers' beliefs about learning are related to the application of FA in ERT context (SHp1); teachers' beliefs about student engagement are related to the application of FA in ERT context (SHp2).

3. Materials and methods

The multiple case study (fig. 1) took place in the context of the Lombardy Region (Italy) – specifically in the metropolitan city of Milan – and involved, through a non-probabilistic sampling of volunteers, three omnicomprehensive school institutes (from primary to high school, from first to thirteenth degree). The first is a state (girl's) boarding school in the city center of Milan and has 62 teachers, the second is a private Catholic inspired institute and is also located in the center and has 69 teachers, the third is in two cities in the province of Milan and has 147 teachers.

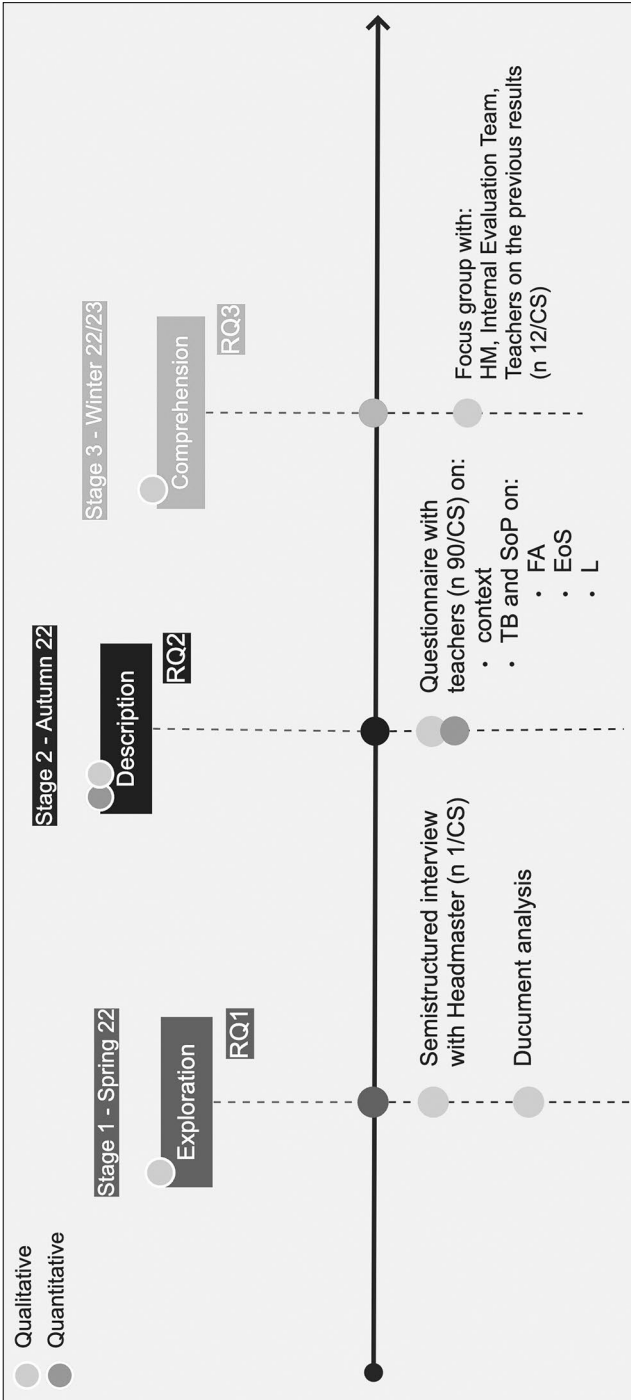


Fig. 1 – Research design

The study aims to investigate the context, beliefs, and statements of practice of teachers during two specific periods, both of which are part of the broader ERT period: remote learning in March-June 2020 (RL) and integrated digital education in September 2020-June 2021 (IDE).

Three semi-structured interviews were conducted on the Institutes' premises, one for each school principal and/or deputy, recorded and transcribed *verbatim*. The first of these, lasted one hour and forty minutes, was held at the School 3 (S3) HM office in May 2022 in the presence of the deputy at the same time; the second, lasted thirty-eight minutes, in the School 1 (S2) faculty room, in June 2022, with the presence of the deputy alone; the third was more articulated, took place in February 2023, and the first part of the interview was conducted with the rector's collaborator in his office in the presence of the principal of the junior high school (forty-four minutes), the second with the primary school principal in a classroom (thirty-five minutes), the third with the secondary school principal in his office (thirty-two minutes), the latter was also rector *pro tempore* at the time of ERT.

The focus group is aimed at sharing the analysis carried out by the researcher and based on this, at carrying out further investigations and reflections with the identified stakeholders. The analyses of the data collected are qualitative and are carried out using the MaxQDA software. The S1 focus group was 9 people, the S2 11, and the S3 7, and were conducted in spring 2023.

Transcribed interviews, focus groups, and documents were analyzed with the MaxQDA software using a code system (tab. 1) structured *ex ante* by reference to the variables identified for the questionnaire and *ex-post* because of the recurrence of certain themes important for the study of the context and the situations that occurred (Duverger, 1961; Rositi, 1971; Bruschi, 1999).

Tab. 1 – System of codes

<i>System of codes</i>	<i>Annotation</i>	<i>Frequency</i>
ENG	Beliefs and practices about engagement	0
PRAXENG	Practices to support student engagement	19
CONENG	Student engagement beliefs	8
APP	Learning beliefs and practices	4
CONAPP	Beliefs about student learning	42
INVALSI tests	Discussions about the results for INVALSI tests	15
PRAXAPP	Practices to support student learning	11
APP post-ERT	Learning after ERT	13
VAL	Beliefs and practices about assessment	1
Training VAL	Assessment training	15
CONVAL	Assessment beliefs	82
PRAXVAL	Assessment practice statements	34
Assessment in ERT	How to apply assessment during ERT	48
VAL post-ERT	Post-ERT assessment modalities	8
Emotions	Pandemic emotions	13
Emotions +	Emotions positives	16
Emotions -	Emotions negatives	45
Context	Background information	83
REL TER	School relations with the territory	20
REL FAM	Relations between schools and families	44
DS and collaborators' role	Management's role in the application of the FA	42
Projects	Projects developed during Covid	9
ERT	Emergency Remote Teaching	6
ERT training	ERT training	9
DIGICOMP	Digital competences	14
CONERT	Teachers' beliefs during the ERT	33
INDICMIN	Relationship to ministerial or policy direction	10
PRAX ERT	Teacher practice during the ERT	45
STUD in ERT	Attitudes of students during the ERT	15

The questionnaire is subdivided into three blocks and aims to collect data on teachers' socio-professional status (16 questions) and, according to their point of view, the school context during the ERT period (9 questions). In addition, teachers' convictions and declarations of practice concerning the constructions and variables in question are recorded using eight scientifically validated and specially constructed Likert scales with four requests

for examples. A first version of the questionnaire was drafted, which was submitted to a tryout phase with teachers from different types of institutes and school orders who sent observations and suggestions. The questionnaire was administered online by sending a cover letter and using Qualtrics software, initially for two weeks, extended then for another; schools 2 and 3 were sent to winter 2022 and school 2 in spring 2023. The analysis of the data collected is quantitative, using SPSS-IBM and R software, and qualitative.

To further strengthen the methodological rigor of the study, it is important to clarify how the questionnaire data were treated.

Aware of the ordinal nature of Likert-type scales, we initially conducted descriptive analyses and parametric tests (ANOVA, etc.) as an exploratory step, in line with a strand of educational research that treats Likert data as approximately interval when sample sizes are sufficiently large (Norman, 2010; Carifio and Perla, 2008; de Winter and Dodou, 2010; Harpe, 2015). In a second step, in order to ensure robustness, we complemented these analyses with non-parametric tests (Kruskal-Wallis for multiple groups, Mann-Whitney for two-group comparisons) and ordinal logistic regression models. The results obtained through these procedures were largely consistent with the preliminary analyses, thus reinforcing the reliability of the findings. This multimodal approach allowed us to triangulate quantitative and qualitative evidence, thereby reducing the risk of distorted interpretations that might result from treating ordinal data as continuous.

4. Results¹

4.1. *Some findings from stage 1: INVALSI data 2021*

During the document analysis process, it was crucial to access the INVALSI data of 2021 in order to observe the differences between schools. The results of grades 2 and 5 (average percentage score net of cheating), 8 and 13 (student outcomes at the same scale as the national report) for the subjects of Italian, Mathematics, English (listening and reading) were asked to the school principals (figures 2-15).

¹ In this paper not all results from the analyses (including interviews and focus groups) are considered. This section presents the main results of the statistical analysis of the questionnaire and INVALSI data. References can be found in the discussion section; for more information, please contact the author.

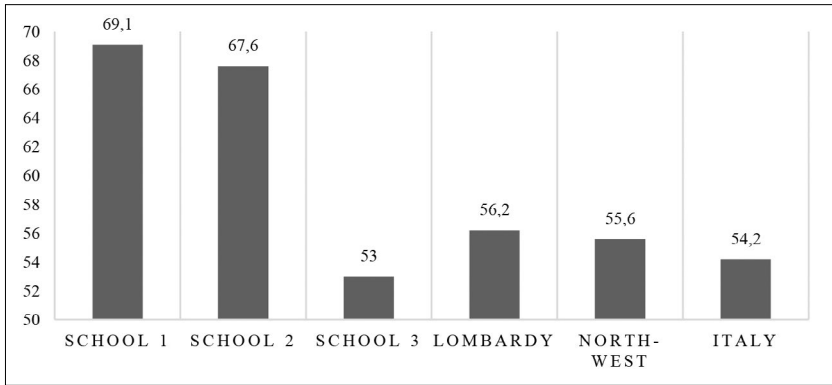


Fig. 2 – Italian (grade 2)

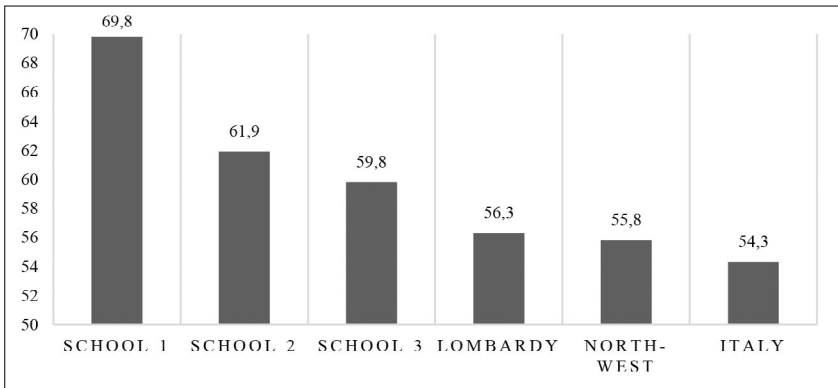


Fig. 3 – Italian (grade 5)

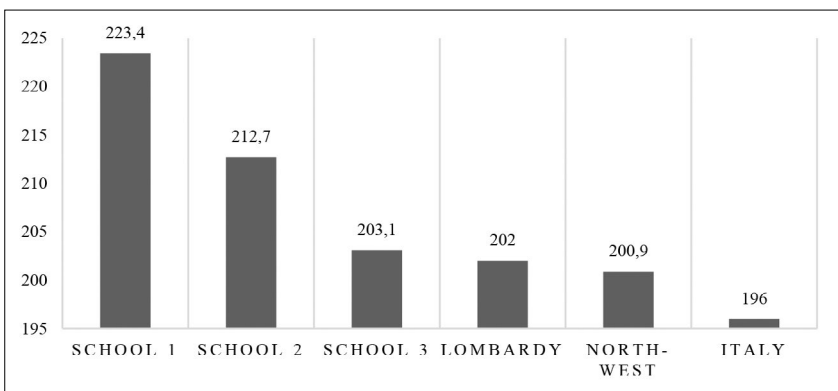


Fig. 4 – Italian (grade 8)

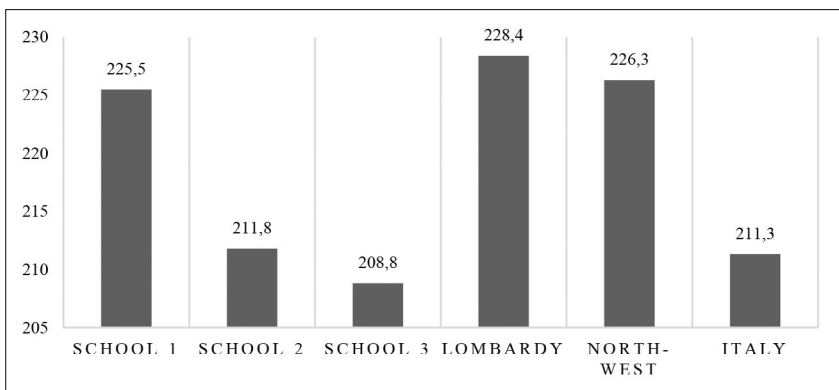


Fig. 5 – Italian (grade 13)

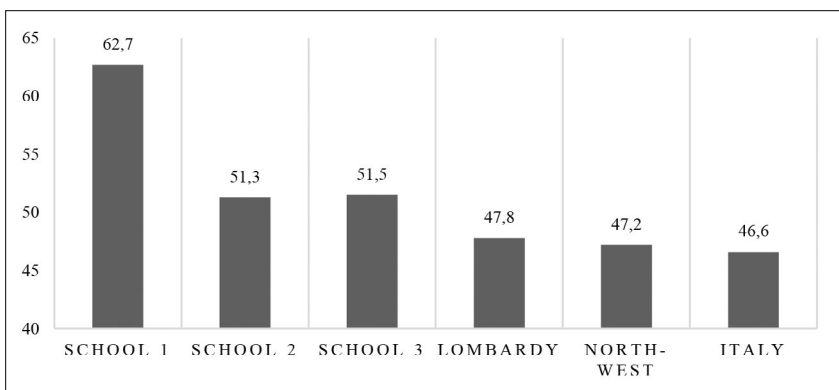


Fig. 6 – Mathematics (grade 2)

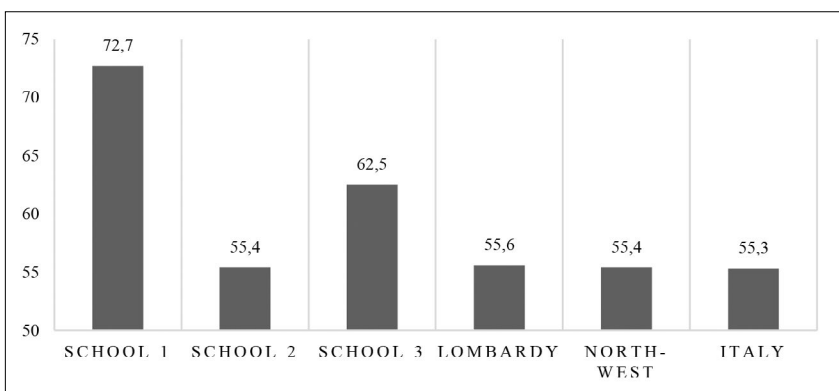


Fig. 7 – Mathematics (grade 5)

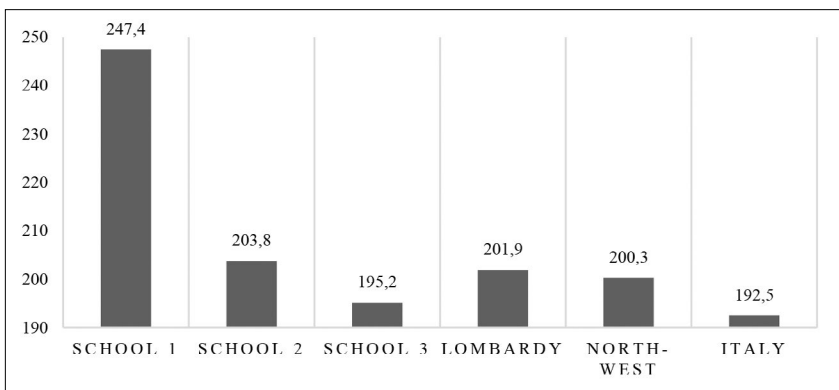


Fig. 8 – Mathematics (grade 8)

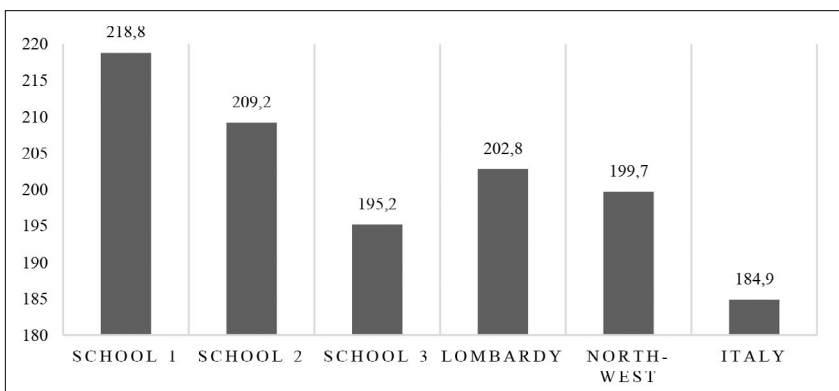


Fig. 9 – Mathematics (grade 13)

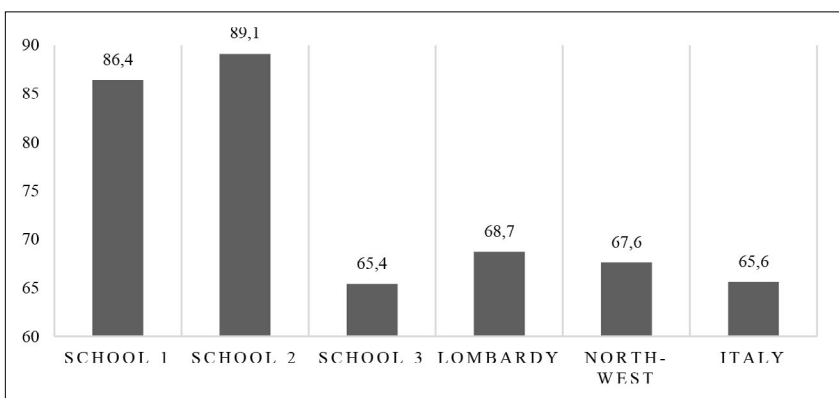


Fig. 10 – English (listening – grade 5)

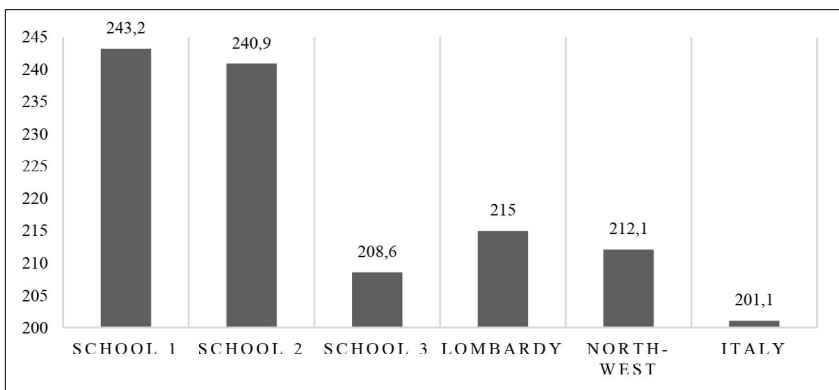


Fig. 11 – English (listening – grade 8)

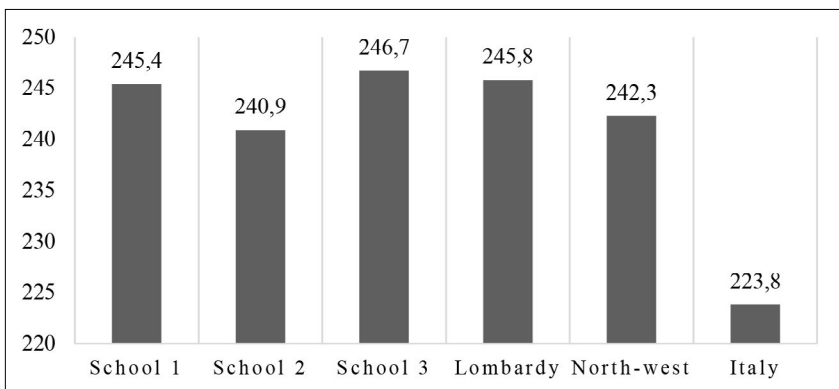


Fig. 12 – English (listening – grade 13)

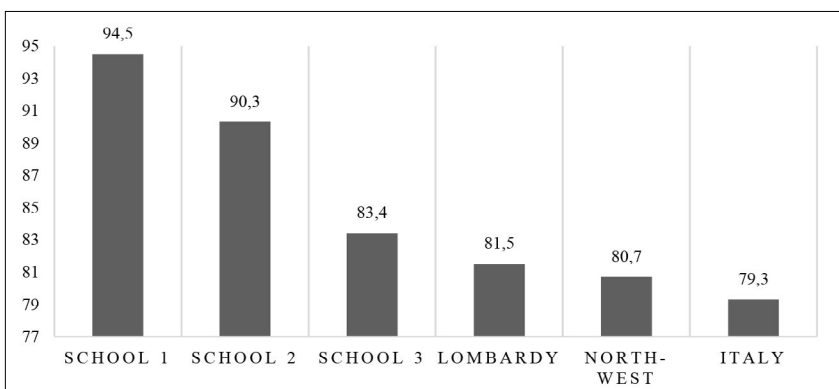


Fig. 13 – English (reading – grade 5)

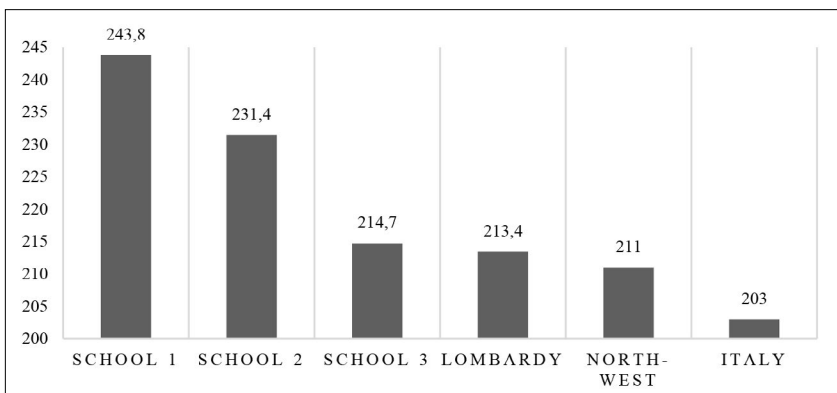


Fig. 14 – English (reading – grade 8)

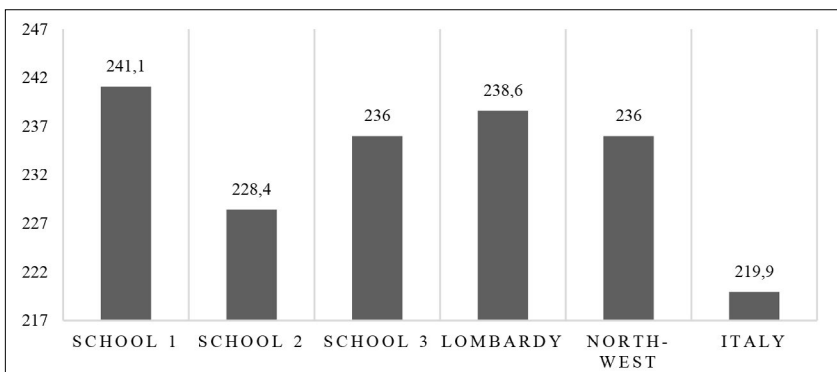


Fig. 15 – English (reading – grade 13)

In order to contextualize the interpretation of the graphs, it is necessary to briefly explain the rationale for using INVALSI data.

The choice of focusing on INVALSI 2021 data is justified by their temporal proximity to the most critical phase of the pandemic, making them particularly relevant to examine the potential effects of Emergency Remote Teaching on students' achievement. The selected domains (Italian, Mathematics, English) represent the core subjects assessed across all schools and are considered essential for monitoring basic competencies. These results were not treated as absolute measures of instructional effectiveness but rather as a reflective tool to be discussed with principals and teachers during the focus groups. Cross-school comparisons may be influenced by socio-economic and territorial variables. For this reason, INVALSI data were used within a triangulated design, integrated with qualitative findings, to avoid

simplistic or decontextualized conclusions. Moreover, considering the data shared by the principals, it was possible to carry out only descriptive analyses and not to conduct inferential analyses (e.g., regression models) to assess the impact of the school context on performance.

As can be seen in the graphs, there is almost a constant repetition of the placement of school results relative to each other. This suggests deepening the influence that context could have on outcomes. In particular, it must be verified that there is an effect of the school variable. This verification was carried out by analysis of the variance with the variables under investigation.

4.2. Structural description of the sample

A total of 193 teachers responded to the questionnaire. Of these, 41 taught in School 1 (S1), representing 21.2% of the sample and corresponding to a response rate of 66.1% (41 out of 62 teachers). Forty-seven teachers (24.4%) were from School 2 (S2), with a response rate of 68.0% (47 out of 69). The largest group, 105 teachers (54.4%), came from School 3 (S3), with a response rate of 71.4% (105 out of 147). The overall sample was predominantly female (77.7%). In addition, seven teachers (3.6%) preferred not to indicate their gender. In general, the response rate is higher than average (50%) with Computer Assisted Web Interviews (CAWI). Overall, the percentages remain unchanged for each school.

Most respondents are between the ages of 36 and 50 (39.4%), with the smallest being 25 years or younger. It should be noted that the bands identified are not homogeneous. Most teachers are in the 14-25 age group (28.5%). Except for those who have been teaching for about a year – and who would not be on duty during Covid anyway – the other bands have almost the same number of teachers. 69.9% of the sample is held, 22.3% were on an annual mission and 2.6% were on temporary duty during the Covid period. Teachers have been teaching for the longest time since 1980 (S1) and 8 since 2022; the oldest holder of an annual office since 2007 (S3) and 15 since 2022. Most teach in High School (40.4%), and minor in secondary school level I (24.4). It should also be noted that two teachers in secondary I also teach in primary and five in HS.

At the time of Covid, 22 worked in a different order (3 S1, 9 S2, 10 S3) and more specifically 7 in primary (4 S2, 3 S3), 6 sec. of 1 (1 S1, 2 S2, 3 S3) and 9 HS (S2). A little more than half (53.4%) have no missions other than teaching. The most common are class coordination, area or department coordination, instrumental function (pupils with disabilities and special ed-

educational needs, programming, and evaluation, etc.), and project reference. 86.2% report having changed their teaching style during the ERT period.

It is interesting to note that the remaining 13.8% admit that they have not made any changes, essentially reproducing the ordinary modalities in an extraordinary context. More specifically, the highest percentage (17%) is reported by the S2 and the lowest by the S1 (7.5%), and the S3 (14.9%). Those who have varied have acted: on the learning environment that has necessarily been transferred to digital platforms, exploiting their potential such as, for example, the division into small groups to promote interaction; on the duration of the lessons, mainly by reducing their length following the ministerial guidelines; on programming, by limiting the content provided and by choosing those which are most inherent to the daily life of pupils; on evaluation, by giving priority to training; on didactic methods and tools, focusing on the classical room to give more time to interaction and some digital applications to support student engagement; on didactic materials, digitizing all necessary documents and giving preference to audio-video material already present on the web. 72.4% (N 134) of teachers were present in the same institute even during the Covid period (34 S1 – 72%; 69 S2 – 69.7%; 31 S3 – 79.5%); the remaining 27.6% were not asked the context description questions.

4.3. Background description

Relationships with families have been maintained through a variety of tools and modalities: most teachers say they prefer video conferencing platforms that allow eye contact (36.98%), followed by email (28.30%) and electronic communication (21%). Note that 22 teachers (7.07%) also used the private phone. Specifically, those in S1 did not use two tools (face-to-face interviews and the school phone); those in S2 did not identify other instruments and used the electronic register less (10%), and S3 tended to match the average of the total sample.

The electronic register was used primarily to record grades (31%) and, to a slightly lesser extent, for bureaucratic procedures (28%). About one quarter of teachers (23%) reported using it to communicate with families, a result that only partially aligns with the responses to the previous question (21.9%). By contrast, only a small proportion of teachers (15%) indicated that they used the register to provide feedback directly to students. This limited use is particularly relevant when examining formative assessment practices. Specifically, S1 teachers report targets that are consistent with the general sample;

S2 teachers, although they used it less than others, used it primarily to record scores (43%) and less to complete bureaucratic procedures (22%) and communicate with households (13%); S3 less to record votes (27%) and more to communicate with households (27%).

At the end of this section, teachers were asked to indicate some of the projects proposed by the school and considered significant for teaching management in the emergency period. We can distinguish some thematic cores: management of digital didactics and applications with support, STEAM and e-twinning methodology, review of assessment criteria, emotional and psychological support, book club production or architectural appetizers (aimed at student engagement), online dialogues with professionals from different parts of the world, lessons in common between professors from different disciplines.

4.4. Scales for detecting beliefs and practices

This part of the questionnaire is intended to collect teachers' statements on the beliefs and practices on survey themes. The scales² (tab. 2) were defined over four levels by reference to the theoretical constructions studied or to certain ministerial documents: convictions and declarations of practice on assessment, student engagement, learning, and ERT. In addition, exploratory factor analyses were conducted to identify latent factors. The hypotheses that guided this exploratory part of the research are oriented toward the analysis of the principal components and the verification of the validity of the scales for the next steps and analyses of the research. It is expected that the item batteries examined will be internally coherent so that they can be considered as valid scales concerning the theoretical background; they highlight some early correlations with other important research variables to support the assumptions made.

² The scales are shown in the English translation, the administration took place with the Italian one. In this paragraph are presented only the scales to detect the beliefs and practices of teachers about formative assessment and Emergency Remote Teaching.

Tab. 2 – Summary of variables and latent factors

Scale	N. LF	Name of LF	p value	Sat.	Cronbach a
VALFOR (Ciani and Vannini, 2017) <i>Teachers' beliefs about assessment</i>	2	Summative assessment (item 1-3-7-8): μ 1,74 Formative assessment (2-4-5-6): μ 3,25			
PRAXVAL <i>Teachers' practices about assessment</i>	2	Assessment actions of the teacher (item 1-7): μ 3,41 Peer/self-assessment (8-12): μ 2,26	< .001	53.39%	.842
CONERT <i>Teachers' beliefs about ERT</i>	2	Emergency of didactics (item 1-2): μ 3,24 Applicability of DDI (3-11): μ 2,20 → 8 e 10 (< 2)	< .001	58.55%	.791
PRAXERT <i>Teachers' practices about ERT</i>	2	Use of digital tools and materials (item 1-2): μ 3,32 Agreement with the teaching staff (3-5): μ 3,65	< .001	76.86%	.748

Tab. 3 – VALFOR scale

Variable code	Variables related to the hypothesized theoretical construct (VALFOR battery)
VALFOR 1	Sometimes it is necessary to attribute lower marks to urge the student to commit more
VALFOR 2	The main role of assessment during the quarter is to identify the learning difficulties of the student to help him overcome them
VALFOR 3	There is attention and interest in a class only if the student knows that it will be assessed on the concepts expressed by the teacher during the lesson
VALFOR 4	The assessments during the quarter are used to identify not only what the student has learned but also what the teacher must deepen
VALFOR 5	The assessment of students during the quarter allows the teacher to verify the validity of his work
VALFOR 6	Assessment during the quarter should help the student to better understand his learning process
VALFOR 7	The attribution of low marks (or judgments) during the quarter helps the teacher to be better respected
VALFOR 8	The teacher should not let the students know his assessment criteria

Fonte: Ciani and Vannini (2017)

The VALFOR battery items (tab. 3) come from a scale built and validated by A. Ciani and I. Vannini (2017) and are intended to measure formative and summative assessment beliefs. Descriptive battery statistics show medium-low agreement levels (not or not at all agree) with items 1, 3, 4, and 8, and medium-high (somewhat or very agree) with items 2, 5, 6, and 7. There is thus a polarization in two different typologies of assessment designs that the researchers defined by factor analysis that identified two latent factors: formative function (VALFOR 2, 4, 5, 6) and summative (VALFOR 1, 3, 7, 8) of the assessment. Descriptive statistics of the latent factors of the VALFOR battery indicate that this polarization is indeed related to the two different assessment designs during the ERT in favour of the formative function. Furthermore, it appears that there is no “very consensual” level for the summative one.

Tab. 4 – PRAXVAL scale

<i>Variable code</i>	<i>Variables related to the hypothesized theoretical construct (PRAXVAL battery)</i>
PRAXVAL 1	I shared with students learning goals before they started working
PRAXVAL 2	Before assigning a test, I clarified to the students what I would assess
PRAXVAL 3	In class, I used different methods of assessment (written, oral, graphic, practical, structured, and unstructured)
PRAXVAL 4	I made sure that the assignments could verify the progress of students' learning
PRAXVAL 5	In addition to the grade (rating), I gave feedback to students
PRAXVAL 6	If I gave feedback on it, I highlighted the strengths or weaknesses of the task
PRAXVAL 7	I provided tips to students to improve their learning
PRAXVAL 8	I asked the students to assess the work of the comrades
PRAXVAL 9	I asked students to provide feedback to fellow students to help them improve
PRAXVAL 10	I asked students to assess their work
PRAXVAL 11	I asked students to identify the strengths and/or weaknesses of their work
PRAXVAL 12	I asked students to propose strategies to improve their work

The PRAXVAL battery items (tab. 4) were defined based on research by D. Wiliam and M. Thompson (2007) and are intended to measure statements of the practical application of the five didactic strategies for the use of formative assessment. Descriptive battery statistics show medium-high agreement levels for almost all items except 8 and 9 that relate to peer review application practices. This could be due to the isolation of students and the didactic typology that limited peer-to-peer work in favour of predominant contact with the teacher. Practices aimed at promoting self-assessment have a medium level, unlike those where the teacher is a major player in which it is high.

Initially, a confirmatory factor analysis of the hypothesis made during the construction of the scale was carried out but did not have the expected results. Subsequently, an exploratory factorial analysis (Varimax method) was performed showing that the Kaiser-Meyer-Olkin sampling adequacy measure is good (.844), the Barlett sphericity test is significant ($< .001$) and indicates two latent factors (PRAXVAL 1 to 7 and 8 to 12) which saturate 53.39% of the total variance. The first covers the actions of exclusive competence of the teacher and combines the first three didactic strategies of Wiliam and Thompson, the second of the practices of the teacher that make the participating student active (peer and self-assessment) and combines the last two strategies. Finally, a reliability analysis of the scale was carried out which has a good Cronbach Alpha (.842), and it is not necessary to delete any element since the result would always be greater than .8. Descriptive statistics of latent factors indicate that teachers have more frequently carried out actions within their exclusive competence and less those involving student intervention.

Finally, the items of the CONERT and PRAXERT batteries are constructed on the studies of C. Hodges *et al.* (2020) and on the ministerial documents (Note 318 of 11.3.2020 and DM 26 June 2020, n. 39). They study teachers' beliefs and practices in emergency remote teaching.

Tab. 5 – CONERT scale

<i>Variable code</i>	<i>Variables related to the hypothesized theoretical construct (CONERT battery)</i>
CONERT 1	RL/IDE or emergency remote teaching is a temporary transition to an alternative mode of providing education due to extraordinary circumstances
CONERT 2	RL involves the use of fully distance learning solutions and education will return to the traditional format once the crisis or emergency has subsided
CONERT 3	The IDE has effectively allowed the integration between teaching in presence and at distance
CONERT 4	The IDE has favored disciplinary and interdisciplinary insights
CONERT 5	The IDE has favored the customization of routes
CONERT 6	IDE has enabled the recovery from learning loss
CONERT 7	The IDE has encouraged the development of disciplinary skills
CONERT 8	The IDE has effectively fostered the development of students' soft skills (personal, social, citizenship, learning to learn, etc.)
CONERT 9	The IDE has effectively improved the correspondence between the teaching action of the teacher and the different learning styles of the students
CONERT 10	The IDE has indeed made it possible to meet the requirements of SEN, linguistic disadvantages, etc.
CONERT 11	IDE has favored the balanced combination of synchronous and asynchronous activity

The descriptive statistics of the CONERT battery (tab. 5) show medium-high levels of agreement (3) for items related to the urgency of the RL/IDE, the first where reference is made to the brevity and the extraordinary character of this didactic. The level for the remaining ten is medium-low (2), especially for the eighth concerning the development of transversal skills where there is no level 4, and the eleventh concerning the satisfaction of the needs of SEN students, etc. where the average is less than 2.

Exploratory factor analysis (by Varimax method) shows that the Kaiser-Meyer-Olkin sampling adequacy measure has a good value (.861), the Barlett sphericity test is significant ($< .001$), and extracted two components that could be defined as beliefs about the emergency of didactics (factor 1, CONERT 1 and 2) and the applicability of IDE as thought by the Department (factor 2, CONERT 3 to 11) that saturate 58.55% of the total variance. The reliability analysis of the scale shows an average Cronbach Alpha (.791); the correlation matrix between the elements has a negative correlation between the first two items and the other nine, plus Cronbach Alpha if removed at a good level (.838). For these reasons, the scale is divided into two distinct factors: CONERT and APPDDI. The descriptive statistics of the two scales identified confirm a medium-high level (3) of agreement with the emergency status of RL/IDE and a medium-low level on the effective applicability of IDE; the difference is about one point. This situation was also noted in the interviews and documents analyzed.

The descriptive statistics of the PRAXERT battery (tab. 6) show high agreement levels (4) for all items except the second where reference is made to the demand and monitoring of individual or grouped digital processing.

Tab. 6 – PRAXERT scale

<i>Variable code</i>	<i>Variables related to the hypothesized theoretical construct (PRAXERT battery)</i>
PRAXERT 1	I used e-learning and/or videoconferencing platforms
PRAXERT 2	I requested and monitored the processing of individual or group digital material
PRAXERT 3	Following the ministerial and college instructions I gave lessons of the appropriate duration (ca 45 min.), and I planned appropriate breaks
PRAXERT 4	Following what has been defined by the faculty, I have adopted the shared criteria for the design/implementation of RL/IDE
PRAXERT 5	Following what was defined by the teachers' college I adopted the organizational elements shared for the RL/IDE (duration of lessons, tools to use, etc.)

Exploratory factor analysis (using the Varimax method) shows that the Kaiser-Meyer-Olkin sampling adequacy measure has sufficient value (.707),

the Barlett sphericity test is significant ($< .001$) and extracted two components that could be defined as the use of tools and the production of digital material (factor 1, PRAXERT 1 and 2) and the agreement with the college of teachers (factor 2, CONERT 3 to 5) which saturate 76.86% of the total variance. The reliability analysis of the scale indicates an average value of Cronbach's Alpha (.748) and if items are deleted, the value will decrease anyway except for item 5 but the difference is .003. The descriptive statistics of latent factors confirm a medium-high level (3-4) of practices regarding the use of digital tools/materials and compliance with ministerial/college guidelines; more specifically, there is a higher average for seconds than for the first.

4.5. Study of correlations and effects of sociodemographic variables

As seen in the literature (Goffman, 1973; Peterman, 1993), in general terms, it can be said that in the three cases considered, there is a correlation between teachers' beliefs and statements of practice concerning assessment and the other constructions involved in the ERT and socio-demographic and contextual variables (tab. 7).

Tab. 7 – Correlations between LF and other variables

	<i>Latent factors (ERT)</i>	<i>Other variables</i>
Beliefs	IDE applicability	Formative assessment teacher actions (.190*, p .022) – Beliefs on constructivist learning (.204*, p .014)
Practices	Use of digital tools and materials	Beliefs on formative assessment (.184*, p .03) – Formative assessment teacher actions (.304**, p < .001) – beliefs on engagement (.279**, p < .001) – teacher practices for engagement (.307**, p < .001) – practices on constructivist learning (.316**, p < .001)
	Compliance with the ministerial indications	Beliefs on formative assessment (.180*, p .033) Formative assessment teacher actions (.291**, p < .001) – beliefs on engagement (.205*, p .014) – teacher practices for engagement (.272**, p < .001) – practices on constructivist learning (.258**, p .002)

Summative assessment beliefs correlates with age (-.171*, p .027), seniority of service (-.184*, p.017), scholastic order (-.184*, p .017); formative assessment beliefs with summative assessment beliefs (-.199*, p .011), assessment actions of the teacher (.320**, p < .001); emergency of RL/IDE with the applicability of the IDE rules (-.204*, p .014); compliance with the ministerial indications with the use of digital tools (.330**, p < .001).

Tab. 8 – Analysis of effects

Outcome	Grouping variable	χ^2	df	p	Significant pairwise (Dunn-Holm)
CONVALF (Formative beliefs)	SCHOOL	8.23	2	.016	School 3 < School 1 (p = .014)
CONVALS (Summative beliefs)	ORDER	5.76	2	.050	High school > Primary (p = .050)
PRACTAPP (Learning practices)	ORDER	7.63	2	.022	Primary > High school (p = .025)
PRACTENG_TEACH (Engagement practices)	ORDER	7.64	2	.022	Primary > High school (p = .021)

Note. Kruskal-Wallis χ^2 values are reported with rounded test statistics. Pairwise comparisons are Dunn's tests with Holm correction. Only significant contrasts are shown.

Descriptive analyses indicated that teachers' beliefs and practices were generally stable across age and seniority groups, with median values consistently around 3 on the Likert scale and narrow interquartile ranges, suggesting homogeneous distributions. However, Kruskal-Wallis tests identified significant effects of school context and scholastic order for several constructs. Formative assessment beliefs (CONVALF) varied significantly across schools, $\chi^2(2) = 8.23$, $p = .016$, with Dunn *post-hoc* tests showing that teachers in School 3 reported significantly lower formative beliefs than those in School 1 ($p = .014$). Summative assessment beliefs (CONVALS) also differed by scholastic order, $\chi^2(2) = 5.76$, $p = .050$, with *post-hoc* analysis indicating stronger endorsement among high school teachers compared with primary school colleagues ($p = .050$). In addition, scholastic order significantly affected both constructivist learning practices (PRAXAPP), $\chi^2(2) = 7.63$, $p = .022$, and engagement practices (PRAXENG_TEACH), $\chi^2(2) = 7.64$, $p = .022$. Dunn's *post-hoc* tests revealed that primary teachers reported significantly higher levels of both learning practices ($p = .025$) and engagement practices ($p = .021$) compared with high school teachers. Altogether, these findings suggest that scholastic order – and, to a lesser extent, school context – represents the most relevant contextual factor in shaping teachers' beliefs and practices during Emergency Remote Teaching (tab. 8).

4.6. Some results from ordinal logistic regression

We estimated proportional-odds models for each construct with age (AGE), seniority of service (SENSERV), school (ID_SCUOLA), and scholastic order (ORDER) as predictors (tab. 9). For formative assessment beliefs (CONVALF) the model indicated meaningful between-school and scholastic-order effects: compared with the reference school, teachers in School 3 showed lower odds of endorsing higher categories on CONVALF ($\beta = -1.731$, OR = 0.18, $p < .001$), and teachers in upper-secondary (ORDER3) also had lower odds than the reference order ($\beta = -1.194$, OR = 0.30, $p = .0057$); AGE and SENSERV were not significant. The model's Nagelkerke $R^2 = 0.154$ suggests small-to-moderate explanatory power. For summative assessment beliefs (CONVALS), both School 3 ($\beta = 1.207$, OR = 3.34, $p = .0083$) and upper-secondary ($\beta = 1.315$, OR = 3.72, $p = .0013$) had higher odds of endorsing higher categories, with $R^2 = 0.200$. For engagement beliefs (CONENG), only seniority group 4 (14-25 years) was significant ($\beta = 2.192$, OR = 8.96, $p = .0259$), while other predictors were not; $R^2 = 0.135$. For constructivist learning beliefs (CONAPP), teachers in AGE3 (36-50 y.o.)

($\beta = -3.688$, OR = 0.03, $p = .0085$) and AGE4 (51-60 y.o.) ($\beta = -3.947$, OR = 0.02, $p = .0071$) had lower odds than the reference age group; in addition, School 3 ($\beta = -1.311$, OR = 0.27, $p = .0069$) and upper-secondary ($\beta = -1.350$, OR = 0.26, $p = .0015$) were associated with lower odds; $R^2 = 0.207$. For engagement practices (PRAXENG_TEACH), School 3 ($\beta = -0.911$, OR = 0.40, $p = .046$) and upper-secondary ($\beta = -1.260$, OR = 0.28, $p = .0021$) showed lower odds than their reference categories; $R^2 = 0.117$. Finally, for learning practices (PRAXAPP), both lower-secondary (ORDER2) ($\beta = -1.091$, OR = 0.34, $p = .043$) and upper-secondary (ORDER3) ($\beta = -1.364$, OR = 0.26, $p = .0045$) were associated with lower odds relative to the reference order; $R^2 = 0.098$. Taken together, the results indicate that scholastic order (especially upper-secondary vs. the reference order) and, for some outcomes, school context are consistent correlates of teachers' beliefs and practices, whereas age and seniority exert limited or outcome-specific effects.

Tab. 9 – Significant predictors from ordinal logistic regressions

<i>Outcome (DV)</i>	<i>Predictor</i>	β	OR	<i>p</i>	<i>Nagelkerke R²</i>
CONVALF (Formative beliefs)	ID_SCUOLA3	-1.731	0.18	< .001	.154
	ORDER3	-1.194	0.30	.006	
CONVALS (Summative beliefs)	ID_SCUOLA3	1.207	3.34	.008	.200
	ORDER3	1.315	3.72	.001	
CONENG (Engagement beliefs)	SENSERV4	2.192	8.96	.026	.135
CONAPP (Learning beliefs)	AGE3	-3.688	0.03	.009	.207
	AGE4	-3.947	0.02	.007	
	ID_SCUOLA3	-1.311	0.27	.007	
	ORDER3	-1.350	0.26	.002	
PRAXENG_TEACH (Engagement practices)	ID_SCUOLA3	-0.911	0.40	.046	.117
	ORDER3	-1.260	0.28	.002	
PRAXAPP (Learning practices)	ORDER2	-1.091	0.34	.043	.098
	ORDER3	-1.364	0.26	.005	

Note. β = logit coefficient; OR = odds ratio ($\exp(\beta)$); p = Wald test probability. Nagelkerke R^2 reported per model. Reference categories correspond to the baseline factor levels set in the dataset.

All these results seem to corroborate what was observed by reading the INVALSI 2021 data.

5. Discussion

On the basis of all the data collected, analysed and triangulated during the research (Giganti, 2025), several considerations can be made. The socio-economic level of the territory to which the schools belong, and the characteristics of the institute have been decisive, because where there are more human, instrumental, and economic resources, the smaller size, and stronger relationships have enabled more processes of change and innovation to be activated than other more fragmented, less cohesive, and resource-poor areas, despite the efforts of teachers and school leaders.

It is interesting to note that the age and experience of teachers are a conditioning factor as well as the order of teaching and the emergency. More experienced teachers do not need to use summative assessment to control the class and gain student respect. Similarly, primary school was the framework in which it was possible to offer formative assessment, a constructivist type of learning and to involve pupils more than in other orders, because of the different training of teachers, the reform of assessment introduced in a pandemic period and, as the data from the focus groups show, the tacit belief that older students did not need special support or attention. The school of belonging also influenced the beliefs of formative assessment and learning: the institutes numerically smaller, stronger identity with coaching and teacher training practices have been decisive on agreed averages. In all three schools, however, the management of integrated digital education has been complex but seen as an opportunity to innovate or improve certain teaching practices and to give new impetus to teaching professionalism.

From several points of view (open responses in the questionnaire, reflections emerging in the focus groups) it emerges that convictions and practices are difficult to change, especially if they are conditioned by the experience of the students and because of short and urgent delays. Training is seen as useful for this purpose but often too far removed from the daily practice of teachers, weak in its applicability in didactics, and ill-suited to profound transformation. In this regard, teachers report that they have not received adequate initial training on the subjects covered by the research and on assessment, apart from those in primary education with a university degree, and on digital education; continuing training is also judged in the same way as it was during the pandemic period.

INVALSI data has been fundamental to open questions about the school context and to start a comparison between institutes. At the same time, there were great difficulty for the researcher to access the data of the schools (difficulty for the schools to share, to find or to understand the needed data).

At this stage, it is also useful to comment on the statistical approach adopted.

In line with the INVALSI findings, the non-parametric analyses confirmed significant differences across schools and school levels, suggesting that contextual factors play a substantial role in shaping teachers' beliefs and practices regarding formative assessment. Nevertheless, the ordinal nature of the data requires interpretative caution: results should be understood as coherent tendencies rather than as precise measurements. The integration of quantitative and qualitative sources (interviews, focus groups, and document analysis) enhances the internal validity of the study and provides a more nuanced understanding of the dynamics at play, which numerical evidence alone cannot fully capture.

6. Researcher and Principal considerations

From the researcher point of view, it is possible to consider that an evolutionary process of assessment happened: from the confusion and replication of traditional methods teachers and schools have come to a functional approach. This had consequences on educational redesign in terms of major quality and minor quantity. Teaching practices developed according to the learning paradigm and students' engagement has been supported. The dissonances between implicit and explicit raise questions about the effectiveness of change models. Changes occasioned from contingent situations (such as pandemic) are not sufficient for lasting change. It is necessary to foster the development of paths consistent with the needs and beliefs of implicit/explicit training, at an appropriate time and within the communities of practice.

From the principal point of view. On 1st September 2021 the official report of the INVALSI test results (taken in spring 2021) was released. The accompanying letter from the Institute underlined: «Especially in this school year, the data is returned to the schools at the beginning of September to encourage a wider use of it in supporting the organization of the teaching activities in the school year».

There's no need to remember how hard the first two years into the Pandemic were for all the schools, but it's necessary to draw the attention on the different approaches to teaching and learning: in 2020, DAD (that is Emergency Remote Teaching) was the response to lockdown, in 2021, DDI (that is Integrated Digital Teaching) was the only way to face the continued use of quarantines. In the first case, the main obstacle was physical distancing, in

the second one the continuous alternation of setting between remote learning and presence at school of part of the students.

The INVALSI data arrived in a moment full of different emotions, but School staff experienced this moment with feelings of anxiety and fear, rather than confidence and hope.

General results in our school showed elements of fragility and uncertainty in different grades (for instance, in 2nd grade with regard to the Italian test 4 classes out of 6 were below the national average).

Frustration among the teachers, in particular among the most involved in the innovation process, was made evident by the words: «It was all in vain! [...] Our resilience has been unproductive» and so on.

For sure, we focused all our efforts on reaching the aim of improving our students' skills and competencies. But, at this point, I can't fail to point out, here, that there's among teachers some difficulty in reading and interpreting the results: lack of skills in reading complex data make teachers continue to be limited to general results and it stresses a general tendency to be self-centered.

The lack of skills in reading more complex data makes the analysis of the results limited to the mere acknowledgement of the so-called overall scores, easily recognizable through the use of the colors of the traffic lights.

The attempt of going deeper into the reading of the analytical data often leads to finding justifications for individual cases known to the teacher.

Let me conclude with a modest proposal: in addition to training and professional development for teachers, could a task force of experts be created to support schools in a critical reading of their test results?

7. Conclusion

The deepening of constructs has allowed the integration of theoretical literature with empirical studies. In addition, the examination of beliefs and statements of practices provided knowledge to reflect on pre-service and in-service training practices and models.

There are three lines of reflection and development. Teachers need implicit/explicit based training and refresher courses, the feasibility of which must be considered on a large scale, but this is an effective direction that should be pursued. Researchers should deepen teachers' beliefs and practices to design effective training paths; furthermore, they should adopt a rigorous, conscious, and critical use of data also considering their social implications; greater interaction between public research institutions (e.g. INVALSI) and university research centers should be encouraged. Policymakers should val-

idate this type of educational research, help scientific research exit from the retreat on itself, aggregate economic, human, and material resources, stimulate work by geographical areas, and involve philanthropic organizations, enterprises, and the third sector.

In conclusion, the real way to prepare for the extraordinary is to substantiate the ordinary; well-trained and competent teachers are a real resource for the school to fulfil its social mandate.

References

- Balduzzi L., Vannini I. (a cura di) (2008), *Nuovi insegnanti per una scuola nuova? Un'indagine tra i docenti formati alla Scuola di Specializzazione all'Insegnamento Secondario (SSIS) dell'Università di Bologna*, Clueb, Bologna.
- Black P., William D. (1998), *Inside the Black Box: Raising Standards Through Classroom Assessment*, Phi Delta Kappan, Arlington.
- Bloom B.S. (1971), *Handbook of formative and summative evaluation of student learning*, McGraw-Hill, New York.
- Brown A.L., Campione J.C., Webb L.S., McGilly K. (1992), "A new look at assessment and instruction", in B.R. Gifford, M.C. O'Connor (eds.), *Changing Assessment Alternative views of aptitude, Achievement, and instruction*, Kluwer Academic, Boston.
- Bruschi A. (1999), *Metodologia delle scienze sociali*, Bruno Mondadori, Milano.
- Carifio J., Perla R. (2008), "Resolving the 50-year debate around using and misusing Likert scales", *Medical Education*, 42 (12), pp. 1150-1152.
- Chen Z., Jiao J., Hu K. (2021), "Formative assessment as an online instruction intervention: Student engagement, outcomes, and perceptions", *International Journal of Distance Education Technologies*, 19 (1), pp. 1-16.
- Ciani A., Vannini I. (2017), "Equità e didattica. Validazione di scale sulle convinzioni di insegnamento democratico", *Cadmo*, 2, pp. 5-32.
- Dewey J. (1949), *Scuola e società*, La Nuova Italia, Firenze.
- de Winter J.C.F., Dodou D. (2010), "Five-point Likert items: t test versus Mann-Whitney-Wilcoxon (addendum added October 2012)", *Practical Assessment, Research, and Evaluation*, 15 (1), 11.
- Duverger M. (1961), *Méthodes de sciences sociales*, Presses Universitaires de France, Paris.
- Engzell P., Frey A., Verhagen M.D (2020), *Learning Inequality During the Covid-19 Pandemic*, retrieved on December 5, 2024, from <https://ideas.repec.org/p/osf/socarx/ve4z7.html>.
- Floden R.E. (2002), "Research on effects of teaching: a continuing model for research on teaching", in V. Richardson (ed.), *Handbook on research on teaching*, AERA, Washington DC, 4th ed., pp. 3-16.

- Fredricks J.A. (2015), "Academic engagement", in J. Wright (ed.), *The International Encyclopaedia of Social and Behavioral Sciences*, Elsevier, Amsterdam, 2nd ed., vol. 2, pp. 31-36.
- Giganti M. (2025), *La valutazione formative: risorsa per l'apprendimento. Cosa ci ha insegnato l'emergenza?*, FrancoAngeli, Milano.
- Goffman E. (1963), *Behavior in public places: notes on the social organization of gatherings*, Free Press of Glencoe, New York.
- Harpe S.E. (2015), "How to analyze Likert and other rating scale data", *Currents in Pharmacy Teaching and Learning*, 7 (6), pp. 836-850.
- Hodges C., Moore S., Lockee B., Trust T., Bond M. (2020), *The Difference Between Emergency Remote Teaching and Online Learning*, retrieved on December 5, 2024, from <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning>.
- Hodges C.B., Barbour M.K. (2021), "Assessing learning during Emergency Remote Education", *Italian Journal of Educational Technology*, 29 (2), pp. 85-98.
- Hodges C.B., Kim C. (2010), "Email, self-regulation, self-efficacy, and achievement in a college online mathematics course", *Journal of Educational Computing Research*, 43 (2), pp. 207-223.
- Kuh G.D. (2009), "What Student Affairs Professionals Need to Know about Student Engagement", *Journal of College Student Development*, 50 (6), pp. 683-706.
- Norman G. (2010), "Likert scales, levels of measurement and the 'laws' of statistics", *Advances in Health Sciences Education*, 15 (5), pp. 625-632.
- Peterman F.P. (1993), "Staff development and the process of changing: A teacher's emerging constructivist beliefs about learning and teaching", in K. Tobin (ed.), *The practice of constructivism in science education*, Routledge, New York-London, pp. 227-245.
- Piaget J. (1976), *Il linguaggio e il pensiero del fanciullo*, Giunti Barbera, Firenze.
- Richardson V., Placier P. (2002), "Teacher Change", in V. Richardson (ed.), *Handbook of research on teaching*, AERA, Washington DC, 4th ed., pp. 905-947.
- Rositi F. (1971), *Contraddizioni di cultura*, Guaraldi, Rimini.
- Scriven M. (1967), "The methodology of Evaluation", in R. Staker (ed.), *Perspectives of Curriculum Evaluation*, Rand McNally, Chicago.
- UNESCO (2020a), *Covid-19 Education Response. Distance learning strategies in response to Covid-19 school closures*, Issue note, 2.1.
- UNESCO (2020b), *Ensuring effective distance learning during Covid-19 disruption Guidance for teachers*, retrieved on December 5, 2024, from <https://unesdoc.unesco.org/ark:/48223/pf0000375116>.
- UNESCO (2020c), *Covid-19 Education Response Webinar. Ensuring quality of learning and well-being for young children in the context of the Covid-19*, retrieved on December 5, 2024, from <https://unesdoc.unesco.org/ark:/48223/pf0000374005>.
- UNESCO (2020d), *Covid-19 Education Response Webinar. Managing high-stakes exams and assessments during the Covid-19 pandemic*, retrieved on December 5, 2024, from <https://unesdoc.unesco.org/ark:/48223/pf0000373247>.

- UNESCO (2020e), *What have we learnt? Overview of findings from a survey of ministries of education on national responses to Covid-19*, retrieved on December 5, 2024, from <https://unesdoc.unesco.org/ark:/48223/pf0000374702>.
- Vygotskij L.S. (1978), *Mind in Society: The development of higher psychological processes*, Harvard University Press, Cambridge.
- Vygotskij L.S. (1987), *Studi sulla storia del comportamento*, Giunti, Firenze.
- Vygotskij L.S. (2001), *Pensiero e linguaggio*, Laterza, Roma-Bari.
- Weeden P., Winter J., Bradfoot P. (2002), *Valutazione per l'apprendimento nella scuola*, Erikson, Trento.
- William D., Thompson M. (2007), "Integrating assessment with instruction: what will it take to make it work?", in C.A. Dwyer (ed.), *The future of assessment: shaping teaching and learning*, Lawrence Erlbaum Associates, Mahwah, pp. 58-82.