INVALSI DATA: METHODOLOGIES AND RESULTS

III Seminar “INVALSI data: a research tool”

edited by
Patrizia Falzetti

FrancoAngeli
OPEN ACCESS

PON 2014-2020

INVALSI

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.

ISBN 9788835113850

Il presente volume è pubblicato in open access, ossia il file dell'intero lavoro è liberamente scaricabile dalla piattaforma **FrancoAngeli Open Access** (http://bit.ly/francoangeli-oa).

FrancoAngeli Open Access è la piattaforma per pubblicare articoli e monografie, rispettando gli standard etici e qualitativi e la messa a disposizione dei contenuti ad accesso aperto. Oltre a garantire il deposito nei maggiori archivi e repository internazionali OA, la sua integrazione con tutto il ricco catalogo di riviste e collane FrancoAngeli massimizza la visibilità, favorisce facilità di ricerca per l'utente e possibilità di impatto per l'autore.

Per saperne di più:
http://www.francoangeli.it/come_pubblicare/pubblicare_19.asp

ISBN 9788835113850

INVALSI DATA: METHODOLOGIES AND RESULTS

III Seminar "INVALSI data: a research tool"

edited by
Patrizia Falzetti

INVALSI

FrancoAngeli
OPEN ACCESS
ISBN 9788835113850

Le opinioni espresse nei lavori sono riconducibili esclusivamente agli autori e non impegnano in alcun modo l'Istituto. Nel citare i contributi contenuti nel volume non è, pertanto, corretto attribuirne le argomentazioni all'INVALSI o ai suoi vertici.

Grafica di copertina: Alessandro Petrini

ISBN 9788835113850

Index

ISBN 9788835113850

ISBN 9788835113850

Introduction

by Patrizia Falzetti

Science is made of data like a house is made of stones.
But a mass of data is no more science than a pile of stones is a real house.
(Henri Poincarè)

Data is part of the process of building scientific knowledge. They are the scales with which to weigh one‘s hypotheses. They are the building blocks with which to build one’s contribution to knowledge on a given topic.

Over the years, interest in data has always grown and, aware of their centrality, many institutions, both public and private, share their data to facilitate the work of all those who wish to use them to interpret phenomena. In the education field, the data produced by INVALSI undoubtedly have a leading role, both at a sample and census level. The availability of data on the learning achievements and socio-economic conditions of students (the so-called "context data"), as well as on the professional and operational conditions of teachers and School Managers, collected through specific questionnaires, is a valuable source of information based on which it is possible not only to plan improvement interventions in the didactic but also to undertake stimulating paths of educational research.

This volume hosts four research papers, presented within the III Seminar "INVALSI data: a research tool", which took place in Bari from 26 to 28 October 2018. Thanks to the INVALSI data, the authors conducted interesting in-depth analysis of various aspects relating to the Italian education system.

In the first chapter, dedicated to kindergartens, the authors give their definition on the quality of kindergarten, in terms of long-term children's learning outcomes, in a quasi-longitudinal perspective. The second chapter focuses on Mathematics learning, examining the relationship between the results of 8th-grade students at the international survey Trends in International Mathematics and Science Study (TIMSS), the results of the same students at the National Mathematics Test, and the school grades in the same discipline, also discussing some possible implications for the Italian school system.

ISBN 9788835113850

The authors of the third chapter use INVALSI data to contribute to research on bullying, unfortunately a very topical issue, describing the characteristics of the students who suffer it and verifying for these students the possible impact on short and medium-term academic performance already in primary school.

In the last chapter, the authors contribute to the studies on the identification of those factors which, more than others, influence students' academic performance: by comparing two methods of variable selection tree-based, they attempt to identify the most relevant predictors of the Italian language INVALSI test results of students in the last year of lower secondary school and, at the same time, the order of importance concerning the predictive power of the selected variables.

ISBN 9788835113850

1. What Do We Know about Preschool Quality in Italy? Preschool Effects on Child Outcomes: A Pseudo-longitudinal Exploration

by Cristina Stringher, Clelia Cascella

According to OECD analyses (2015), monitoring children's learning outcomes to measure quality of Early Childhood Education and Care (ECEC) is increasingly common internationally. By contrast, in Italy, monitoring ECEC quality is carried out only at local, not at national level yet. Thus, it is impossible to measure quality with national monitoring data. Therefore, in this chapter, after an overview of the Italian ECEC system, we expose our definition of preschool quality as long-term child outcomes. We aim to answer the following research questions: a) whether there are and how large are differences in long-term child outcomes in Text Comprehension and Mathematics between primary students that have previously attended preschool or not; b) how child outcomes vary over time for different groups of students (clustered by gender, socio-cultural background and territorial level). In order to answer the first question, we comparatively analysed long-term outcome data (from 2012 to 2015) in a pseudo-longitudinal perspective. For the second question, we disaggregate national data at the provincial level. Students that have previously attended preschool do show positive differences in their outcomes in Text Comprehension and Mathematics both at grade level 2 and 5 compared to those that did not attend it. Such differences are statistically significant and they are clearer only when we examine the provincial level. Implications for further research and policy in Italy are discussed, along with indications for applying our pseudo-longitudinal methodology in countries where no national ECEC evaluations are available.

Secondo analisi dell'OCSE (2015), il monitoraggio dei risultati di apprendimento dei bambini per saggiare la qualità dei servizi per l'infanzia è sempre più diffuso a livello internazionale. Tuttavia, in Italia, il monitoraggio della qualità dei servizi sembra effettuato attraverso pratiche regolate

ISBN 9788835113850

localmente anziché a livello nazionale. Ciò rende impossibile misurare la qualità attraverso dati di monitoraggio nazionale. Di conseguenza, in questo capitolo, dopo una panoramica del sistema infanzia italiano, forniamo la nostra definizione di qualità della scuola dell'infanzia in termini di risultati di apprendimento dei bambini a lungo termine, rispondendo alle seguenti domande di indagine: a) se esistono e quanto grandi sono le differenze nei risultati a lungo termine dei bambini in Comprensione del Testo e Matematica tra studenti di primaria che hanno precedentemente frequentato o meno la scuola dell'infanzia; b) come variano i risultati nel tempo per differenti gruppi di studenti (suddivisi per genere, provenienza socio-culturale e livello territoriale), e se si può scoprire un'eterogeneità latente. Per rispondere alla prima domanda, abbiamo confrontato i risultati a distanza (dal 2012 al 2015), in una prospettiva quasi-longitudinale. Per la seconda domanda, disaggreghiamo i dati nazionali a livello provinciale. Gli studenti che hanno frequentato in precedenza una scuola dell'infanzia mostrano differenze positive nei risultati di Comprensione del Testo e Matematica in seconda e quinta primaria rispetto a chi non l'ha frequentata. Tali differenze sono statisticamente significative e sono più chiare solo quando si esamina il livello provinciale. Discutiamo quindi le implicazioni per ricerche future e per le politiche in Italia, fornendo altresì delle indicazioni per applicare la metodologia quasi-longitudinale in Paesi dove non sono disponibili valutazioni nazionali sulla qualità dei servizi per l'infanzia.

1. Introduction

Neuro-science, Social sciences and Econometric research all support one fundamental point: Early Childhood Education and Care (ECEC) matters greatly (Blair *et al.*, 2002; Burger, 2010; Heckman, 2008, 2013). It does so for children's development, learning and well-being in the short-term and it creates the building blocks for improving later long-term life outcomes, especially against odds (Chambers *et al.*, 2010; Heckman, 2008, 2013; Melhuish *et al.*, 2015). This is why ECEC represents one of the best investments (Pianta *et al.*, 2009) against social inequality worldwide.

According to the Economist Intelligence Unit (EIU, 2012), European ECEC is exemplar for its accessibility, affordability and quality. All but four top 20 positions in the EIU's index are taken by European systems. Among these, Italian ECEC has a long-standing tradition of active pedagogies aiming at the empowerment of children with one-hundred languages as in Reggio (Malaguzzi, 1993, 1998), and with what is currently referred to as

ISBN 9788835113850

self-determination and key competencies (Montessori, 1999, 2000; Deci and Ryan, 2002; Stringher, 2014). In addition, preschool in Italy is practically universal: approximately 95% of in-scope children attend it (12th position in the EIU ranking out of 45 countries). Although not a legal entitlement, in Italy preschool for children aged 3 to 5 years is free of charge for parents and its affordability is probably one of the strategic levers policy makers used to promote universal access (8th position in the EIU ranking).

However, we know little about the penetration of active pedagogies in the Italian ECEC system and the overall quality of provision remains uncertain (Economist Intelligent Unit, 2012; Del Boca and Pasqua, 2010). Quality in ECEC is certainly a quite complex concept to grasp, one that cannot be defined only through structural indicators, such as those that are largely proposed in the EIU index. ECEC system quality in Italy is not evaluated nationally (OECD, 2015) and this can leave open interpretations concerning its level, in spite of the international reputation of Italian world-exported excellences.

Our aim is to start shedding some light on the level of preschool quality in Italy. In order to do this, we need to expose our definition of quality, a concept that is being highly debated in international fora nowadays (Anders, 2015; European Commission, 2014; European Commission/EACEA/Eurydice/Eurostat, 2014; IEA ECES, 2016; Love, 2003; Moser *et al.*, 2014; OECD, 2015, 2017a, 2017b). After our definition of preschool quality, we explore the quality of the Italian preschool as emerging from national studies. In the empirical part of our work, we explore long-term child outcome data in Text Comprehension and Mathematics to examine the quality of Italian preschool. Our primary sources are student performance data as measured by national standardized assessments carried out yearly by INVALSI.

2. Theoretical framework

The aim of this and of the following two sections is to synthesize a review of studies dealing with the concept of quality in ECEC in order for us to sustain our choice of independent variables in the empirical section of our study and to discuss our findings in the light of international and Italian literature.

Quality in ECEC is a quite complex and contested concept (Dahlberg *et al.*, 1999). Following Pascal and Bertram (1999), we assume quality to be indexed by settings' effectiveness in producing a positive impact on children's lives. The existence of ECEC services and preschools would otherwise serve other purposes than fostering children's development, and would not be any different

ISBN 9788835113850

from an unprofessional babysitting service or familial child rearing. If governments are to place value on ECEC, quality of provision should be paramount.

By positive impact for children, we mean an impact in terms of balanced, holistic development: physical development, wellbeing, learning and learning dispositions, as described by Italian national curricular guidelines (MOE, 2012). In line with INVALSI's definition of quality preschool, we consider of good quality those ECEC services having a positive impact on learning, wellbeing and development of children both during and after exposure to such services (Stringher, 2016).

To activate this positive long-term impact on children, three types of ECEC quality are generally considered by international sources: structural quality of regulation, service standards and materials available in a setting; process quality of the ECEC environment, of the enacted curriculum and of relationships built within a classroom; and quality of teachers' professional orientations with roots in their attitudes, beliefs and values shaping their interactions and relationships with children (Anders, 2015; Litjens, 2013; Moser *et al.*, 2014; OECD, 2015). Following Anders, orientation quality includes "teachers' pedagogical beliefs such as their definition of their professional role, their educational values, epistemological beliefs, attitudes with regard to the importance of different educational areas and learning goals" (2015, p. 9). Being closest to children, setting process quality, orientation quality and quality of the enacted curriculum seem paramount in fostering positive outcomes for children, both in the cognitive and socio-affective dimensions (La Paro and Pianta, 2000; Ljtiens, 2013; Pascal *et al.*, 2013; Pontecorvo *et al.*, 1990; Slot *et al.*, 2015).

Different authors define these types of ECEC quality in diverse ways (Pianta *et al.*, 2009). Anders (2015) includes in structural quality those aspects (such as setting and classroom size, staff/child ratio, teacher credentials) which can be regulated by policy and funding, while Pianta and colleagues (2009) also include the adoption of a particular curriculum as part of structural quality, probably for the absence of a national ECEC curriculum in the USA. Structural quality seems to function as an ecological precondition mediating process quality that exerts direct influence on child development and outcomes (Mashburn *et al.*, 2008).

Process quality refers primarily to pedagogical interactions between staff and children, among children and between staff and parents (Anders, 2015), but also interactions between children and materials available to them in their settings and the types of activities available therein (Pianta *et al.*, 2009). Process quality, being proximal to children, is thought to have the most direct impacts on their learning and wellbeing outside the Home Learning Environ-

ISBN 9788835113850

ment (HLE) and this claim is supported by a wealth of research internationally (Bronfenbrenner and Morris, 2006; Mashburn *et al.*, 2008; Pianta *et al.*, 2009). Teachers enact their pedagogy based on their pedagogical option and this in turn is influenced by their orientation. Orientation quality includes teachers' educational values, pedagogical beliefs and setting and individual epistemology (Anders, 2015).

In order to predict later outcomes for children in school and beyond, other quality factors need to be considered, some proximal to the child, others quite distant yet pervasive: in particular, quality of the HLE and quality of ECEC system in terms of normative arrangements, national curricular guidelines and access opportunities. Figure 1 offers a graphic representation of ECEC quality factors affecting child outcomes.

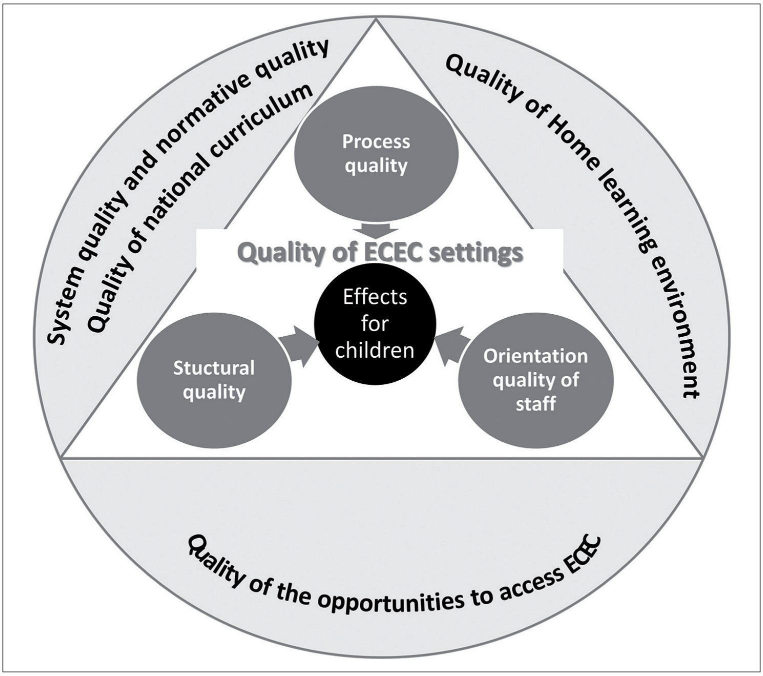

Fig. 1 – ECEC factors impacting children's outcomes

Source: Adapted from Stringher (2016)

How all these factors interact to display their influence on later child outcomes in a particular cultural context is still largely unknown, and it certainly is for the Italian context, where few national studies have addressed the importance of quality of ECEC provision for children's long-term trajectories

ISBN 9788835113850

(see Del Boca and Pasqua, 2010; Montie *et al.*, 2006; Pontecorvo *et al.*, 1990 for notable exceptions).

2.1. Effects of quality ECEC

All children are genetically programmed to learn (Shonkoff and Phillips, 2000; Bingham and Whitebread, 2012), but environmental factors impacting development and learning display possibly an even stronger influence than inherited assets: *The accident of birth is a principal source of inequalities in America*, according to Nobel Prize recipient James Heckman (2008, 2013, p. 3) and his remark is directly referring to the determinant role of socio-economic factors in shaping child development.

Notwithstanding the debate over nature and nurture (Shonkoff and Phillips, 2000), environment – broadly conceived – is malleable, thus responsibility of parents, schools and society at large for optimal child outcomes should not be underestimated, especially in Italy, where a wide competence gap persists in primary and secondary education between Northern and Southern areas of the country (INVALSI, 2016). This is not to say that socio-economic family background is an inexorable determinant of children's futures: in fact, Sylva and colleagues (2004, p. 1) teach us that "the quality of the home learning environment is more important for intellectual and social development than parental occupation, education or income". In other words, to a certain extent and in non-extreme situations of poverty and deprivation, *what parents do is more important than who parents are* for their children's optimal growth and disadvantaged children may benefit from the additional support of quality ECEC.

Researchers around the world have been actively trying to explain the mechanisms of ECEC's influence in the life of children as they grow up (Pianta *et al.*, 2009; Melhuish, 2011; Melhuish *et al.*, 2015; Sylva *et al.*, 2004; Zellman and Karoly, 2012). Quality Rating and Improvement Systems (QRISes) for ECEC in the USA, for instance, rely on improvement in the input and processes of an ECEC program to produce improved child outcomes (Zellman and Karoly, 2012). However, this direct link between processes and outcomes for children is unclear, as many studies are correlational and the overall theory maintaining that improved program quality yields better child functioning "has not yet been tested" (Zellman and Karoly, 2012, p. 10). One major problem with this equation is the difficulty in improving process quality and quality of teachers' interactions with children (Pianta *et al.*, 2016), especially on a large national scale.

ISBN 9788835113850

In spite of this, internationally, participation in ECEC (broadly defined as encompassing education and care for children 0-6) seems to produce short and long-term effects (European Commission/EACEA/Eurydice/Eurostat, 2014; OECD, 2013; 2015). Better early learning outcomes in basic competencies are perhaps among the most researched effects. Early learning outcomes, in turn, have a positive impact on individuals' educational trajectories and this may lead to favorable long-term life outcomes, such as: reduced involvement in criminal activities, less likelihood of risky behavior and more likelihood to enjoy good mental and physical health, all of which constitute long-term earnings of education for the individual and society (Heckman, 2008, 2013; Cingano and Cipollone, 2009).

Recently, Dumčius and collegues (2014) demonstrated the possible effective use of quality ECEC even in preventing early school leaving in Europe. Specular results by Pianta and colleagues (2009) support such findings. These scholars maintain that preschool increases children's rates of upper secondary school completion. These studies would corroborate the value of quality ECEC as a protective factor, especially for disadvantaged or minority groups.

Internationally, quality ECEC seems to have a positive impact on all children, boys and girls, and even more so on children belonging to lower socio-economic backgrounds, thus contributing to reduce social inequalities (Melhuish, 2011; Melhuish *et al.*, 2015; Pianta *et al.*, 2009; OECD, 2013). Possibly this is because effective ECEC can make up for children's poor socio-cultural conditions, less stimulating Home Learning Environment (HLE), poor parental interactions and insecure patterns of attachment. Pianta and colleagues (2009) claim that also well-off children gain substantially from preschool education, as high as 75% of gains accrued for disadvantaged groups. Interestingly, among positive and significant outcomes for children, several studies indicate also less grade repetition and less likelihood of special education placement (Pianta *et al.*, 2009).

Quality ECEC seems to be both cost-efficient and effective in aiding child development in a number of areas, both cognitive and socio-affective-motivational. In addition, the notion that cognitive and socio-emotional-motivational factors interact in children during development is generally well-accepted (Blair *et al.*, 2002), with some scholars even pointing to the prevalence of socio-emotional-motivational factors over cognitive abilities. Executive functions, as indexed by behavioral self-regulation, appear to be directly and positively related to emergent Literacy and Mathematics skills in the USA and elsewhere (McClelland *et al.*, 2007, 2014; Wanless *et al.*, 2011; Størksen *et al.*, 2014).

ISBN 9788835113850

When further narrowing down ECEC's impact on children's cognitive outcomes, international studies confirm that quality of preschool matters for specific desirable long-term educational outcomes (Aunio *et al.*, 2008; Chambers *et al.*, 2010; Linder *et al.*, 2013; Marcon, 2002), such as Literacy and Numeracy. However, complexity arises when different cultures are taken into account in the analyses of early learning outcomes, as in a study of early Numeracy among Finnish, English and Chinese 5-year-olds (Aunio *et al.*, 2008): culture and informal Mathematics seem to be better predictors of early Numeracy for Chinese children than other factors, such as a six-months exposure to pre-maths curriculum (in England). Confucian values and the informal Mathematics Chinese children absorb from their culture (both at home and in preschool) seem to be key in their superior results compared to European counterparts. In part, such difference is also reflected in the differential curricular emphasis given to Mathematics in the three countries, but unfortunately, the authors did not elaborate much on such curricular aspects.

Cultural differences in the values placed by parents and teachers on early Mathematics could well be encountered in different parts of Italy and they could account for the early differences in performance in the early primary grades. All in all, children who struggle in early Mathematics before they enter formal schooling are expected to see their gap increase over the years and cultural differences are already there at age four or five (Aunio *et al.*, 2008). These and other researchers also touched upon gender differences, with Finnish girls outperforming boys in early maths, indicating that cultural factors again may have stronger effects over genetic factors (Cortázar, 2015). Neuroscientists in fact claim no difference between the genetic endowment of girls and boys at birth (Dehaene, 1997).

To add to the complexity of factors affecting children's early learning outcomes, different pre-Literacy programs yield differential effects, both in a short and long-term perspective (NELP, 2008). Chambers and colleagues (2010) found that comprehensive cognitive developmental approaches, broader than purely academic programs, yield better long-term effects on social outcomes such as reductions in delinquency, welfare dependency, and teenage pregnancy, and increases in educational and employment levels. For these authors, it is also notable that effects of preschool exposure on Literacy can be detected later in children's school career, since gains in vocabulary breadth influence reading ability. Academic programs, with pedagogies specifically aimed at developing Literacy and Numeracy skills, display their results in the short term, yet their effects tend to fade as children progress through primary grades (Marcon, 2002).

ISBN 9788835113850

These results seem shared also by other analysts as reported in a Eurydice study (2009, p. 23): holistic center-based programs coupled with parental support display effect sizes on Intelligence Quotient (IQ) and achievement in the 0.6-0.8 range, conventionally considered medium to strong. Linder and colleagues (2013) not only support again these results, but also maintain that the development of logical and mathematical skills during preschool years is particularly effective for children with disadvantaged backgrounds. In addition, these scholars point to teacher's practices enhancing logical and mathematical skills in children: not only playing with numbers, direct instruction of pre-mathematical concepts and construction blocks, but also free exploration of children in their environment and self-initiated activities are needed; teachers celebrating new acquisitions of children; teacher's protection of the child from disapproval or inappropriate punishment and appropriate teacher guidance and limitation of children's inappropriate behaviour. All elements that corroborate the importance of the socio-emotional-affective components of teacher-child quality interactions within an ECEC setting. The Italian system does not support one approach over another in the national curriculum, yet the pedagogical discourse generally emphasizes the holistic development of children through playful activities and focus on socio-emotional skills, rather than pre-academic programs.

Which groups of children benefit from ECEC exposure is another question that the international literature has addressed over time. In their study on early Numeracy, Aunio and Niemivirta maintain that no gender difference exists at birth in primary numerical ability and that preschool age children are those benefitting the most from early Numeracy activities, since "children's competence seems to transit from biologically primary qualitative skills to more complex and culturally bound, biologically secondary number, counting and arithmetic skills" (Aunio and Niemivirta, 2010, p. 428). In addition, the authors claim that socio-economic gaps in Mathematics performance is well documented in the preschool and early primary grades. For kindergarten children, nonverbal task formats are less sensitive than verbal task formats to socioeconomic variation, and this seems to reflect better HLE and language development support in families with higher versus lower Socio-Economic Status (SES), as also Melhuish and colleagues (2008b) explain. These authors in particular underline the power of HLE, over and above parental education and SES, on educational attainment. Vulnerability in school readiness for learning is stronger in unhealthy children versus children in good health, in boys versus girls and in lower-income families (Janus and Duku, 2007); children reared in broken versus intact families also have higher odds of being vulnerable at school entry.

ISBN 9788835113850

Dosage of preschool programs also seems a quite important element contributing to measurable results: Sylva and colleagues (2004) maintain that an early start in attendance (even under 3 years of age) results in better intellectual outcomes. However, full or part-time attendance does not seem to make a great difference. In Argentina, a study by Berlinsky and colleagues (2006) found that one year of preschool increases children's achievement measured in third grade in mother tongue and Mathematics by 8%, or 23% of a standard deviation, compared to average. In Italy, preschool starts at age 3, with a duration of 3 years, and several models co-exist for parents to choose from: morning-only time tables for 6 week days (which internationally could be considered part-time attendance) can be opted in *versus* longer daily schedules, up to 40 hours per week.

When ECEC provision is of poor quality, what happens to children's development and learning outcomes? Several studies have not only found no effect (Melhuish, 2011), but also negative effects associated with poor quality ECEC provision (Alexander *et al.*, 1997; Melhuish *et al.*, 2015). In particular, for disadvantaged groups, poor quality ECEC provision greatly limits the possibility to close the achievement gap in pre-academic skills and in basic competencies once children are in school (Pianta *et al.*, 2009). Aspects of ECEC quality negatively associated with child outcomes are teachers' low qualifications and overall quality of teaching, with a stronger focus on routines, on large group activities and on parental or staff needs, rather than on children's (Melhuish *et al.*, 2015; Montie *et al.*, 2006; Pianta *et al.*, 2009).

Finally, in the model illustrated in Figure 1, individual factors affecting children's outcomes are only implicit. However, child characteristics (such as geographical origin, family socio-economic status, gender, genetic endowments and temperament), along HLE, are among the strongest predictors of short- and long-term outcomes (La Paro and Pianta, 2000; Melhuish *et al.*, 2008a, 2008b; Moser *et al.*, 2014; Son and Peterson, 2016). Internationally, countries monitoring quality ECEC with process and children's outcome measures are increasing (OECD, 2015). Tools for monitoring child outcomes mainly include local observation and narrative assessments rather than national direct assessments. In Italy no national monitoring exists yet and only locally applied tools are in place.

2.2. Quality of Italian ECEC and research questions

National studies examining the quality of Italian childcare or preschool and their impact on children's outcomes are scarce and to our knowledge

ISBN 9788835113850

have never been conducted on nationally representative samples. Preschool process quality has been locally measured (Bondioli, 2001; D'Ugo, 2013; Harms and Clifford, 1994) generally separate from children's outcomes (Coggi and Ricchiardi, 2014; Commodari, 2013; Corsaro and Molinari, 2008; ERR, 2014; Zanetti and Miazza, 2002; Zanetti and Cavioni, 2014), especially though not exclusively in preschools. A deficiency of Italian research even on renowned pedagogies, such as that of Reggio Emilia, results in lack of empirical verification in terms of positive child outcomes they purport to sustain. Only one longitudinal comparative study examined the relationship of Italian ECEC quality with longer-term child outcomes data (Montie *et al.*, 2006). However, country-level data for Italy for this study are not available. If children are to fully benefit from quality preschool, positive effects of preschool quality are to be valued by primary education. Only if primary teachers are able to link their action to preschool, children will experience a smooth transition and will build upon early years' acquisitions (Pontecorvo *et al.*, 1990), taking advantage of quality preschool to enrich their acquisitions in primary school.

Another design approach to the study of ECEC quality is the use of later child outcome data to infer quality of ECEC retrospectively. To our knowledge, only few studies concentrated on how parental inputs (of time and choice of 0-3 provision) affect long-terms child outcomes (Del Boca and Pasqua, 2010). In their study, Del Boca and Pasqua investigated the effects of maternal care time and childcare attendance on children's behavioral and cognitive development in primary school and beyond. Using three different data sets and a set of econometric techniques, the authors demonstrated a positive effect of childcare on later cognitive and behavioral outcomes, such as: performance in national Text Comprehension and Mathematics tests in grade 2 and 5. A study carried out with more recent INVALSI data show similar results of ECEC services effects on Text Comprehension scores, yet no effects on Mathematics scores (Brilli *et al.*, 2016), thus encouraging additional analyses on childcare effects in Italy.

National monitoring of preschool quality does not exist in Italy yet. To fill in this gap, INVALSI is currently introducing the new national Preschool Self-Evaluation Report Format (PSERF), experimented during 2019. Thus, we do not have information on process quality of ECEC settings. Our research question is thus: what is the level of preschool quality and how is it distributed across Italy before the introduction of PSERF, i.e., in the absence of process and outcome measures? As ECEC quality is key to combat early inequalities, we try to provide an initial contribution to this lack of information, concentrating our attention on long-term preschool effects. Our defi-

ISBN 9788835113850

nition of preschool quality is thus: quality of primary school outcomes for children exposed or not exposed to preschool. To measure such long-term outcomes, we use INVALSI national test results in Text Comprehension and Mathematics from grade level 2.

3. Methodology

3.1. Research questions

In the present study, we tried to answer the following research questions: a) whether there are and how large are differences in long-term child outcomes in Text Comprehension and Mathematics between children that have previously attended or not attended preschool; b) how outcomes of children (clustered by gender, socio-cultural background and territorial level) vary over time. In order to answer our questions, we examined long-term outcome data (from 2012 to 2015), with a pseudo-longitudinal design. For the second research question, we disaggregated national data and analysed them at both macro-geographical and provincial levels.

3.2. Data sources

INVALSI tests are administered at the end of each scholastic year to the entire population of children attending grades 2 and 5 in primary education. A sample of students is drawn in order to limit the bias in test scores due to student and teacher cheating during the testing sessions. Competence in Mathematics and Text Comprehension estimated on this basis is net of cheating effects, because an INVALSI inspector, who guarantees the accuracy and fairness of the testing procedure, conducts the administration (INVALSI, 2012a, 2012b).

3.3. A Pseudo-Longitudinal Design

A pseudo-longitudinal design (also known as pseudo-panel) is substantially a repeated cross-sectional study of the same birth cohort over time, that thus allows for the calculation of valid estimates of changes at the population level from independent samples (Steel, 2011). Such a design is not properly longitudinal as it does not track the same individuals over time, yet it allows longitudinal results to be examined at the systemic level, as samples are rep-

ISBN 9788835113850

resentative of the same student population. The INVALSI national sample is statistically representative of the Italian population over time at both national and regional level, yet not at provincial level. Therefore, results may not be generalized to the universe. In this analysis, for example, we compared answers given by children attending second grade (in 2012) and by children attending fifth grade (in 2015) (Table 1). In fact, data collected at grade 2 and at grade 5 are both statistically representative of the same birth cohort, i.e. it is representative of the population born in 2005.

Tab. 1 – Sample characteristics

	Ita02		*Mat02*		*Ita05*		*Mat05*	
	Preschool attendance							
	Yes	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*
Father's occupation								
1. Unemployed	1,144	154	1,161	197	808	110	840	116
2. Homemaker	84	1,045	86	251	76	22	80	24
3. (Executive) manager, lecturer/ professor	728	174	733	102	564	54	583	54
4. Entrepreneur	0	1,209	0	0	879	120	909	125
5. Professionals (including self-employed), e.g., lawyer, doctor, researcher	2,950	42	2,937	485	2,140	357	2,194	378
6. Own-account worker, e.g., shop keeper, artisan, mechanic	4,843	519	4,851	588	3,175	386	3,290	394
7. Teacher, employee	4,556	0	4,561	726	3,082	465	3,190	482
8. Worker, member of a cooperative	6,839	52	6,858	957	4,337	595	4,469	620
9. Retaired	135	3,288	132	25	129	21	138	19
Missing	1,207	0	3,925	3,308	0	4	0	5
Total	22,486	6,483	25,244	6,639	15,190	2,134	15,693	2,217
Father's education								
Primary school	694	162	701	156	382	61	389	62
Lower intermediate school	8,228	1,245	8,280	251	5,089	712	5,271	755
3-years diploma (Professional qualification)	2,293	178	2,298	102	1,579	114	1,654	121
5-years diploma	7,905	1,069	7,894	0	5,694	855	5,900	902
Other qualification higher than Diploma	437	31	433	485	267	23	282	22
Degree, Master, Ph.D.	3,107	387	3,101	588	2,331	380	2,358	400
Not available	2,502	3,359	2,537	982	1,532	2,214	0	0
Total	25,166	6,431	25,244	2,564	16,874	4,359	15,854	2,262

ISBN 9788835113850

Tab. 1 – Sample characteristics (to be continued)

	Ita02		*Mat02*		*Ita05*		*Mat05*	
	Preschool attendance							
	Yes	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*
Mother's occupation								
Unemployed	1,200	177	1,209	197	868	98	908	300
Homemaker	7,984	1,549	8,019	251	5,009	791	5,175	4,177
(Executive) manager, lecturer/professor	262	74	261	102	198	29	201	1,603
Professionals (including self-employed), e.g., lawyer, doctor, researcher	0	0	0	0	279	38	288	6,589
Own-account worker, e.g., shop keeper, artisan, mechanic	1,964	278	1,968	485	1,394	222	1,436	459
Teacher, employee	6,028	811	1,772	588	1,250	170	1,294	2,999
Worker, member of a cooperative	3,301	352	6,029	726	4,309	612	4,446	1,340
Retaired	34	16	3,299	957	2,171	276	2,238	0
Missing	2,628	2,929	37	25	4,309	9	26	0
Total	23,401	6,186	22,594	3,331	19,787	2,245	16,012	17,467
Mother's education								
Primary school	618	154	701	156	298	65	4177	65
Lower intermediate school	6,517	1,045	8,280	251	4,011	620	1,603	655
3-years diploma (Professional qualification)	2,143	174	2,298	102	1,535	98	6,588	99
5-years diploma	9,064	1,209	7,894	0	6,394	979	459	1,034
Other qualification higher than Diploma	636	42	433	485	437	28	2,996	29
Degree, Master, Ph.D.	4,036	519	3,101	588	2,925	446	2,995	458
Missing	2,152	3,288	2,537	982	1,274	2,123	1,340	2,218
Total	25,166	6,431	25,244	2,564	16,874	4,359	20,158	4,558
Student's sex								
Male	12,754	3,322	12,785	3,403	8,560	2,257	8,891	2,361
Female	12,412	3,109	12,459	3,187	8,314	2,102	8,576	2,197
Total	25,166	6,431	25,244	6,590	16,874	4,359	17,467	4,558

Source: our elaboration on INVALSI SNV data 2012 and 2015

3.4. Measures

Children's ability in Mathematics and Text Comprehension was estimated by using the Rasch model (Rasch, 1960, 1961, 1977, 1980). This model is particularly adequate for the purposes of the present study because of its

ISBN 9788835113850

property of measurement invariance: each item difficulty in a test is sample-free and vice versa. In other words, each child's ability is *test free* because measurement loses randomness due to possible variations in children's or questions' samples, thus it becomes an "invariant" feature of the model (Wilson and Engelhard, 1995; Masters, 2001). This property allows robust statistical comparisons between sub-groups of children clustered as a function of relevant variables. We compared test scores of children attending and not attending preschool, by Socio-Cultural Index (SC-Index). We also included a descriptive comparison of long-term outcomes at provincial level[1].

As clustering variables, we used gender and socio-cultural background of children's families. We used information about parents' education and occupation to construct a measure of socio-cultural status (named SC-index). Highest parental education and occupation have been combined into three and five categories respectively and then combined as shown in Table 2.

Tab. 2 – The construction of SC-index based on highest parental education and occupation

Employment/ Education	*Unemployed*	*Housewife*	*Worker*	*White collar worker*	*Entrepreneur or self employed*
Low	Low	Low	Low	Medium	Medium
Medium	Low	Low	Low	Medium	High
High	Medium	Medium	Medium	High	High

Source: our elaboration on INVALSI SNV data 2012 and 2015

In case of missing data on either education or occupation, the highest information available is considered.

3.5. Analytical strategy

We initially computed mean Rasch scores in Mathematics and Text Comprehension on all of INVALSI samples available in 2012 and 2015. We run a t-test to compare students' attainment depending on preschool attendance. Then, we explored differences in attainment between attending and not attending students across regions and provinces.

[1] INVALSI samples are statistically representative at national and regional level but not at provincial level. Therefore, comparison based on data aggregated at provincial level cannot be inferred to the population.

ISBN 9788835113850

4. Results

4.1. Pseudo-longitudinal analysis

In the following table, we show differences in Rasch test scores in Mathematics and Text Comprehension between attending and not attending students, after having controlled for gender and socioeconomic status. In addition to statistically significance, we reported effect size (values around 0.20 or below indicate small effect size, values around 0.50 indicate medium effect size, and values around or larger than 0.80 indicate large effect size) (Cohen, 1988).

In grade 2, for Text Comprehension we observe that attendance to preschool increases test scores for children with low SC-index, whose performance is similar to that of children with higher background. Performance is higher in children attending preschool. Three years later, at grade 5, it seems that preschool effect lessens (although children who have attended tend to have higher scores still), while primary school seems unable to counter socio-cultural gaps, on the contrary, it seems to reproduce inequalities and to introduce a gender gap: females' test scores are higher than those of male students. Nevertheless, differences in test scores between children attending and not attending preschool, though always statistically significant, are generally so small to be negligible (always less than a quarter of a standard deviation).

In grade 2, for Mathematics, both females and males attending preschool have better Mathematics test scores than those not attending. Notably, there does not seem to be a gender gap at this stage. However, Mathematics test performance is increasing in the higher SC-index levels, with a similar pattern observed between children attending and not attending. This is a notable difference compared to Text Comprehension. Three years later, attending students' advantage in Mathematics test scores is confirmed. In addition, generally test scores increase over time, yet a gender gap starts to be visible, with boys outperforming girls in all socio-cultural levels, except in the higher SC-index group of children not attending preschool. Nevertheless, differences in test scores between attending and not attending preschool are generally so small to be considered negligible (always less than a quarter of a standard deviation).

ISBN 9788835113850

Tab. 3 – Differences in Rasch test scores in Mathematics and Text Comprehension, at grade 2 and 5, by preschool attendance, gender, and SC-index

		Rasch test scores							
		Grade 2							
		Mathematics				*Text Comprehension*			
		Preschool attendance				*Preschool attendance*			
		Yes	*No*	*(2-tails) Sig.*	*Cohen's d*	*Yes*	*No*	*(2-tails) Sig.*	*Cohen's d*
Male	SC-index 1	188.69	176.30	.000	.30	205.37	184.42	.000	.56
	SC-index 2	191.21	188.50	.000	.07	201.40	189.21	.000	.31
	SC-index 3	205.91	204.14	.000	.04	204.80	185.35	.000	.49
Female	SC-index 1	186.85	185.35	.172	.03	210.67	190.50	.000	.59
	SC-index 2	188.26	183.88	.000	.11	208.24	177.60	.000	.78
	SC-index 3	202.01	197.75	.000	.11	205.83	196.15	.003	.23
		Grade 5							
		Mathematics				*Text Comprehension*			
		Preschool attendance				*Preschool attendance*			
		Yes	*No*	*(2-tails) Sig.*	*Cohen's d*	*Yes*	*No*	*(2-tails) Sig.*	*Cohen's d*
Male	SC-index 1	181.33	166.83	.132	.36	189.28	179.68	.000	.24
	SC-index 2	200.58	196.50	.089	.10	199.82	196.26	.000	.09
	SC-index 3	213.88	202.51	.000	.28	214.03	191.33	.000	.57
Female	SC-index 1	178.35	180.83	.781	-.06	193.51	175.44	.000	.45
	SC-index 2	192.40	183.09	.000	.23	203.28	193.16	.000	.25
	SC-index 3	206.17	200.90	.005	.13	216.37	205.82	.000	.26

Source: our elaboration on INVALSI SNV data 2012 and 2015

ISBN 9788835113850

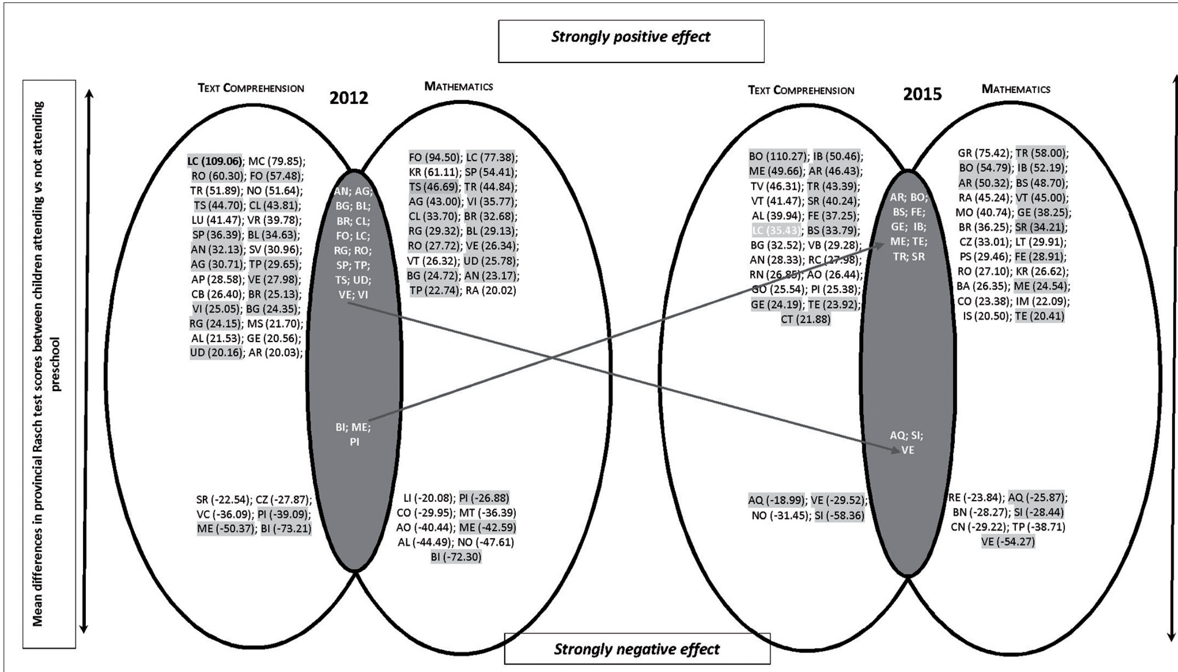

Fig. 2 – Longitudinal effect of Italian preschool on later learning outcomes in Text Comprehension and Mathematics by year and by province

Highlighted in grey, provinces with differences > 1/2 S.D.; in the intersection set, provinces with differences > 1/2 S.D. in both competences; and highlighted in grey, one province with differences > 1/2 S.D. in both competences at grade 2 and longitudinally in Text Comprehension. Provinces ranked in order of magnitude in Rasch scores differences.

Source: our elaboration on INVALSI SNV data 2012 and 2015. For province acronyms, see https://www.istat.it/it/archivio/6789

ISBN 9788835113850

On average, preschool attendance had a positive long-term effect on children's performance in both Mathematics and Text Comprehension (but larger on students' performance in Text Comprehension than in Mathematics), at both grade 2 and 5. So far, we only reported the national level, while at the provincial level we want to know whether there are territorial differences. Figure 2 reports Italian provinces where difference in test score is equal or greater than 20 points on the Rasch scale, i.e. more than half of a standard deviation. For Text Comprehension and Mathematics, again at grade 2 and 5, we sort provinces as a function of observed differences in Rasch test scores between children attending and not attending preschool.

At grade 2, we observe strong positive effects on children's test scores, in favour of those attending preschool, in 33 Italian out 103 provinces of the INVALSI sample. At this level, 16 provinces show positive effects over ½ of a standard deviation in both competences. Similarly, at grade 5, 23 provinces show a strong positive effect in Rasch test scores and 10 provinces show this effect in both competences. In both years, we found three provinces showing a negative preschool effect on children's outcomes. Some specific provinces are worth mentioning. For example, the outstanding Lecco (LC), where results show excellent long-term outcomes for children attending preschool in both competencies and longitudinally in Text Comprehension. In contrast, Reggio Emilia preschools do not display positive effects as it could be expected, and in grade 5 a negative effect is observed for Mathematics (of approximately half of a standard deviation). In addition, interesting preschool provinces seem to be those of Messina (ME) and Venice (VE): they obtain reverse results in second and fifth grade.

Taken together, these results highlight strong differences at provincial level and thus confirm that a latent geographical heterogeneity actually exists. Moreover, we have identified many other provinces where differences in Rasch test scores are just a notch below our cut-off criterion.

5. Discussion and conclusions

Our aim with this study has been to start understanding preschool effects on children's long-term outcomes in basic competencies in order to infer the quality of Italian preschool. Empirical analyses aimed at answering the following research questions: a) whether there are and how large are differences in long-term child outcomes in Text Comprehension and Mathematics between students that have previously attended or not attended preschool; b) how child outcomes vary over time for different groups of students (clustered by gender, socio-cultural background and territorial level).

ISBN 9788835113850

In order to measure the preschool effect on children's achievement in the long-term, obviously we need longitudinal data. In order to do this, usually, cohort designs are employed. In the absence of panel data, cross-sectional designs are generally carried out, but they have several relevant limitations: first and foremost, they do not allow to explicitly study the evolution of a phenomenon over time. In the absence of longitudinal data, in our study we used data of the Italian National Student Assessment samples, selected by INVALSI, in 2012, for grade 2 and in 2015 for grade 5. The comparison between data collected in 2012 and in 2015 provides more accurate information compared to a cross-sectional design, because these samples are statistically representative of the same population over time. This methodology actually allows longitudinal results at the systemic level. This is a quite good level of analysis in order to pursue our research aims on long-term systemic preschool effects in Italy.

We recognize that our methodological solution is somewhat tautological in that it assumes, according to international literature, that only quality preschools yield positive long-term outcomes for children. In principle, this tautology could be considered a limitation of studies with our research design because we do not measure preschool quality directly and we can only infer it indirectly. However, we can detect null or negative preschool effects also with our methodology, especially when we explore geographical differences. Thus, we believe our proxy methodology to be a promising avenue for those countries that lack national analyses on the quality of their ECEC systems.

Primary school children that have previously attended preschool do show differences in their outcomes in Text Comprehension and Mathematics at both grade 2 and 5 compared to their not-attending peers. Such differences are statistically significant taking into account our clustering variables. In particular, socio-cultural and gender differences seem revealed by our national elaborations, while territorial heterogeneity seems evident in view of our geographical analyses. However, compensation effects may occur in big cities, such as Rome, where a zip code analysis could reveal a similar heterogeneity observed at provincial level. We do not have this possibility yet, but it could be worth exploring this aspect in the future, especially in Rome, where the difference between children attending and not attending preschool is close to zero, as shown by our results.

Overall, our analyses reveal positive long-term effects of preschool on children's competencies in Text Comprehension and (to a lesser extent) Mathematics in all areas of the Country. This is also not surprising, in light of the strong pedagogical tradition of preschools in Italy and of the international literature. Interestingly, for Text Comprehension preschool seems to

ISBN 9788835113850

be able to even out children's scores, irrespective of gender and socio-cultural background. Mathematics effects are less evident, possibly due to a lack in pre- and in-service preschool teacher training in early Numeracy. In addition, the provincial territories with highly positive preschool effects are spread across the Country (there seems to be no clear quality divide between Northern and Southern preschools), and the numerosity of "positive provinces" seems to give credit to the positive image that Italian preschools have abroad. Such widespread positive cases seem independent from the coordinators' qualification level (only Emilia Romagna region seems to have almost all coordinators with tertiary degree according to R-ER – Regione Emilia Romagna, 2003), and this represents a counter-intuitive finding that would be worth exploring further, when data on coordinators are nationally available.

Less obvious are other geographically relevant findings, to be taken with care due to the non-representative sample at provincial level: according to our descriptive statistics, none of the renowned Italian local pedagogies seems to stand out (Reggio Emilia, Roma, Pistoia, and Modena). Possibly, an explanation could be that primary schools in these areas do not capitalize on the wealth of positive experiences that children receive in preschools. Primary school, in order to build on preschool quality, should work in continuity and should minimize transition effects for children in the passage from preschool to primary education. An additional explanation within the Reggio Emilia province is a compensation effect between preschools actually applying the Reggio Children approach compared to those that do not: it might be that good results in preschools with the Reggio approach are nullified by worse results in preschools without this pedagogical approach.

The reverse effects from grade 2 to 5 in two provinces could be interpreted as the ability of primary school to make up for a low-quality preschool (in Messina), while the opposite works for Venice, where primary schools do not seem to capitalize on the benefits of good quality preschools. Such local cases need further exploration, especially considering our non-representative sample at provincial level, in order to better understand these results in light of preschool process data when they become available.

These territorial examples and generally our national results seem to corroborate the idea that Italian preschools seem to protect children against social inequalities at least up to grade level 2, even though the trend seems to differ between Text Comprehension and Mathematics. Overall, our results seem to confute international studies on school readiness, that tend to find significant socio-culturally determined gaps in children's competencies at primary school entry level with baseline assessments (Hair *et al.*, 2006;

ISBN 9788835113850

Janus and Duku, 2007; Pianta *et al.*, 2007). However, it seems that Italian preschool teachers are more successful in fostering language and pre-Literacy skills than logic and early Numeracy: children's social background seems to be a stronger predictor of success for Mathematics than it is for the acquisition of Text Comprehension, and this may pose questions of preschool teacher initial and in-service training. Also worth noting is that Italian primary school seems to widen socio-cultural disparities and gender differences in these competencies. Questions worth exploring with further research are to be differentiated by school level: are preschool teachers equipped to teach logical skills or early Mathematics in order to prevent the development of early social inequalities? Moreover, is primary school able to capitalize on the positive preschool effects during the transition of children from preschool to primary? Territorial differences at provincial level could also be further investigated in the future.

Our study has several limitations. First, we do not have information on the duration of children's exposure (dosage of preschool, at least expressed in years of attendance, as suggested by Sylva *et al.*, 2004 and Berlinsky *et al.*, 2006). In addition, we do not have information on the type of preschool attended (state, private and municipal), nor on their observed level of process quality. These three factors constitute a limitation to the possible interpretation of our results, either positive or negative, and we can only hypothesize the reasons for the differences observed at provincial level. We suggest that such information be made available in the future. Specifically, when data from INVALSI's Preschool Self-Evaluation Report Format (PSERF) are accessible, it would be very useful to replicate our analyses correlating results with observed structural and process quality indicators, especially in the provinces where strongly positive or strongly negative long-term child outcomes are found. The literature suggests in fact that negative preschool effects could even result in later student disengagement and dropout. Thus, with high dropout rates in Italy, this point merits attention.

Another relevant limitation of our study design, relying on already collected data not representative at provincial level, is the lack of information on children's individual characteristics, such as family socio-economic background, child temperament or learning outcomes at preschool completion. We propose that such information be collected in the future, in order to better appraise the quality of preschool and its impact also in the short-term.

ISBN 9788835113850

References

Alexander K.L., Entwisle D.R., Horsey C.S. (1997), "From first grade forward: Early foundations of high school dropout", *Sociology of Education*, 70, pp. 87-107.

Anders Y. (2015), *Literature review on pedagogy. For a review of pedagogy in Early Childhood Education and Care (ECEC) in England (United Kingdom)*, OECD, Paris.

Aunio P., Aubrey C., Godfrey R., Yuejuan P., Liu Y. (2008), "Children's early numeracy in England, Finland and People's Republic of China", *International Journal of Early Years Education*, 16, 3, pp. 203-221.

Aunio P., Niemivirta M. (2010), "Predicting children's mathematical performance in grade one by early numeracy", *Learning and Individual Differences*, 20, pp. 427-435.

Berlinsky S., Galiano S., Gertler P. (2006), "The effect of pre-primary education on primary school performance", *The institute for fiscal studies WP06/04*.

Bingham S., Whitebread D. (2012), *School Readiness. A critical review of perspectives and evidence*, Tactyc, no place, retrieved on September, 15, 2020, from http://scmeyegroup.weebly.com/uploads/1/8/4/9/1849450/readiness_review_12312.pdf.

Blair C. (2002), "School Readiness: Integrating Cognition and Emotion in a Neurobiological Conceptualization of Children's Functioning at School Entry", *American Psychologist*, 57 (2), pp. 111-127.

Bondioli A. (2001), *AVSI – Autovalutazione della scuola dell'Infanzia*, FrancoAngeli, Milano.

Brilli Y., Del Boca D., Pronzato C.D. (2016), "Does child care availability play a role in maternal employment and children's development? Evidence from Italy", *Review of Economics of the Household*, 14 (1), pp. 27-51.

Bronfenbrenner U., Morris P.A. (2007), "The Bioecological Model of Human Development", *Handbook of Child Psychology*, I, 14, pp. 793-828.

Burger K. (2010), "How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds", *Early Childhood Research Quarterly*, 25 (2), pp. 140-165.

Chambers B., Cheung A.K., Slavin R.E., Smith D., Laurenzano M. (2010), *Effective early childhood education programs: A systematic review*, Johns Hopkins University, Center for Data-Driven Reform in Education, Baltimore (MD).

Cingano F., Cipollone, P. (2009), "I rendimenti dell'istruzione", *Banca d'Italia Occasional Papers*, 53.

Cohen J. (1988), *Statistical Power Analysis for the Behavioural Series*, 2nd edition.

Coggi C., Ricchiardi P. (2014), "La 'school readiness' e la sua misura: uno strumento di rilevazione per la scuola dell'infanzia", *ECPS Journal*, 9, pp. 283-309.

Commodari E. (2013), "Preschool teacher attachment, school readiness and risk of learning difficulties", *Early Childhood Research Quarterly*, 28 (1), pp. 123-133.

Corsaro W.A., Molinari L. (2008), "Policy and Practice in Italian Children's Transition from Preschool to Elementary School", *Research in Comparative and International Education*, 3 (3), pp. 250-265.

ISBN 9788835113850

Cortázar A. (2015), “Long-term effects of public early childhood education on academic achievement in Chile”, *Early Childhood Research Quarterly*, 32 (3), pp. 13-22.
Dahlberg G., Moss P., Pence A. (1999), *Beyond quality in early childhood education and care: Postmodern Perspectives*, Routledge, London.
Dehaene S. (1997), *The number sense. How the minds creates Mathematics*, Oxford University Press, New York.
Deci E., Ryan R. (2002), *Handbook of Self-Determination Research*, The University of Rochester Press, Rochester (NY).
Del Boca D., Pasqua S. (2010), “Esiti scolastici e comportamentali, famiglia e servizi per l’infanzia”, *FGA Working Paper*, 36.
D’Ugo R. (2013), *La qualità della scuola dell’infanzia: la scala di valutazione PraDISI (Vol. 7)*, FrancoAngeli, Milano.
Dumčius R., Peeters J., Hayes N., Van Landeghem G., Siarova H., Ceneric I., Hulpia H., Masiulienė L. (2014), *Study on the effective use of early childhood education and care in preventing early school leaving. Final report*, Publications Office of the European Union, Luxembourg.
Economist Intelligence Unit, Lien Foundation (2012), *Starting well. Benchmarking early education across the world. A report from the Economist Intelligence Unit*, Author, Singapore.
ERR – Emilia Romagna Region (2014), *Kids in places – Validazione EDI*, Bologna, Courtesy of the authors.
Eurydice (2009), *Early Childhood Education and Care in Europe: Tackling Social and Cultural Inequalities*, EACEA P9 Eurydice, Brussels.
European Commission/EACEA/Eurydice/Eurostat (2014), *Key Data on Early Childhood Education and Care in Europe. 2014 Edition. Eurydice and Eurostat Report*, Publications Office of the European Union, Luxembourg.
European Commission – DG Education and Culture (2014), *Proposal for key principles of a Quality Framework for Early Childhood Education and Care*, European Union, Brussels.
Hair E., Halle T., Terry-Humen E., Lavelle B., Calkins J. (2006), “Children’s school readiness in the ECLS-K: Predictions to academic, health, and social outcomes in first grade”, *Early Childhood Research Quarterly*, 21, pp. 431-454.
Harms T., Clifford R.M. (1994), *SOVASI – Scala per l’osservazione e la valutazione della scuola dell’infanzia*, Edizioni Junior, Azzano S. Paolo (BG).
Heckman J., National Bureau of Economic Research (2008), “Schools, Skills, and Synapses”, *NBER Working Paper*, No. 14064.
Heckman J. (*ed.*) (2013), *Giving Kids a Fair Chance: A Strategy That Works*, MIT Press, Cambridge.
IEA ECES (2016), *Early Childhood Policies and Systems in Eight Countries – Findings from IEA’s Early Childhood Education Study*, Author, Hamburg.
INVALSI (2012a), *Rilevazione nazionali degli apprendimenti. Il quadro di sistema*, INVALSI, Frascati.
INVALSI (2012b), *Rilevazioni nazionali sugli apprendimenti 2011-2012. Rapporto tecnico*, INVALSI, Frascati.

ISBN 9788835113850

INVALSI (2016), *Rilevazioni nazionali sugli apprendimenti 2015-2016*, INVALSI, Frascati.

Janus M., Duku E. (2007), "The School Entry Gap: Socioeconomic, Family, and Health Factors Associated With Children's School Readiness to Learn", *Early Education and Development*, 18 (3), pp. 375-403.

La Paro K., Pianta R. (2000), "Predicting Children's Competence in the Early School Years: A Meta-Analytic Review", *Review of Educational Research*, 70, 4, pp. 443-484.

Linder S.M., Ramey M.D., Zambak S. (2013), "Predictors of School Readiness in Literacy and Mathematics: A Selective Review of the Literature", *Early Childhood Research & Practice*, 15 (1), retrieved on September, 15, 2020, from: https://ecrp.illinois.edu/v15n1/linder.html.

Litjens I. (2013), *Literature Review on Monitoring Quality in Early Childhood Education and Care (ECEC)*, OECD, Paris.

Love J., Harrison L., Sagi-Schwartz A., van IJzendoorn M.H., Ross C. (2003), "Child Care Quality Matters: How Conclusions May Vary With Context", *Faculty Publications, Department of Child, Youth, and Family Studies*, paper 41, retrieved on September, 15, 2020, from http://digitalcommons.unl.edu/famconfacpub/4.

Malaguzzi L. (1993), "For an education based on relationships", *Young Children*, 11, 93, pp. 9-13.

Malaguzzi L. (1998), "History, Ideas and Basic Philosophy: An Interview with Lella Gandini", in C. Edwards, L. Gandini, G. Forman (*eds.*), *The Hundred Languages of Children – The Reggio Emilia Approach – Advanced Reflections*, Ablex Publishing Co., Greenwich (CT).

Marcon R. (2002), "Moving up the Grades: Relationship between Preschool Model and Later School Success", *ECRP Early Childhood Research and Practice*, 4, 1, no pages available.

Mashburn A.J., Pianta R., Hamre B., Downer J., Barbarin O., Bryant D., Howes C. (2008), "Measures of Classroom Quality in Prekindergarten and Children's Development of Academic, Language, and Social Skills", *Child Development*, 79 (3), pp. 732-749.

Masters G.N. (2001), *The key to objective measurement*, Australian Council on Educational Research, retrieved on September, 15, 2020, from: https://www.rasch.org/memo70.pdf.

McClelland M.M., Cameron C.E., Connor C.M., Farris C.L., Jewkes A.M., Morrison F.J. (2007), "Links between behavioral regulation and preschoolers' literacy, vocabulary, and Mathematics skills", *Developmental Psychology*, 43, pp. 947-959.

McClelland M., Cameron C., Duncan R., Bowles R., Acock A., Miao A. (2014), "Predictors of early growth in academic achievement: the head-toes-knees-shoulder task", *Frontiers in Psychology*, 5, 599, pp. 1-14.

Melhuish E.C., Sylva K., Sammons P., Siraj-Blatchford I., Taggart B., Phan M., Malin A. (2008a), "Preschool influences on Mathematics achievement", *Science*, 321 (5893), pp. 1161-1162.

ISBN 9788835113850

Melhuish E.C., Phan M.B., Sylva K., Sammons P., Siraj-Blatchford I., Taggart B. (2008b), "Effects of the home learning environment and preschool center experience upon literacy and numeracy development in early primary school", *Journal of Social Issues*, 64 (1), pp. 95-114.

Melhuish E.C. (2011), "Preschool matters", *Science*, 333 (6040), pp. 299-300.

Melhuish E., Ereky-Stevens K., Petrogiannis K., Ariescu A., Penderi E., Rentzou K., Leseman P. (2015), "A review of research on the effects of early childhood education and care (ECEC) upon child development. CARE project", *Curriculum quality analysis and impact review of European early childhood education and care (ECEC)*.

Ministry of Education, MOE (2012), *Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo di istruzione*, Rome.

Montessori M. (1999), *La scoperta del bambino*, Garzanti, Milano.

Montessori M. (2000), *L'autoeducazione*, Garzanti, Milano.

Montie J.E., Xiang Z., Schweinhart L.J. (2006), "Preschool Experience in 10 Countries: Cognitive and Language Performance at Age 7", *Early Childhood Research Quarterly*, 21, pp. 313-331.

Moser T., Melhuish E., Petrogiannis K., Pastori G., Slot P., Leseman P. (2014), "Initial framework for evaluating and monitoring ECEC quality and wellbeing", *CARE: Curriculum & Quality Analysis and Impact Review of European Early Childhood Education and Care*, retrieved on September, 15, 2020, from http://ecec-care.org/resources/publications/.

NELP (2008), *Developing Early Literacy Report of the National Early Literacy Panel*, National Center for Family Literacy, Jessup (MA).

OECD (2013), *Literature Review on Monitoring Quality in Early Childhood Education and Care (ECEC)*, OECD, Paris.

OECD (2015), *Starting Strong IV: Monitoring Quality in Early Childhood Education and Care*, OECD, Paris.

OECD (2017a), *Starting Strong 2017: Key OECD Indicators on Early Childhood Education and Care*, OECD, Paris.

OECD (2017b), *Starting Strong V: Transitions from Early Childhood Education and Care to Primary Education*, OECD, Paris.

Pascal C., Bertram T. (1999), *Accounting early for lifelong learning*, in L Abbott, H. Moylett (*eds.*), Early Education Transformed, Falmer Press, London, pp. 93-104.

Pascal C., Bertram T., Delaney S., Nelson C. (2013), *A comparison of international childcare systems*, Department for Education, Centre for Research in Early Childhood Research Report.

Pianta R., Cox M., Snow K. (2007), *School readiness and the transition to kindergarten in the era of accountability*, Brookes Publishing, Baltimore.

Pianta R.C., Barnett W.S., Burchinal M., Thornburg K.R. (2009), "The Effects of Preschool Education: What We Know, How Public Policy Is or Is Not Aligned With the Evidence Base, and What We Need to Know", *Psychological Science in the Public Interest*, 10 (2), pp. 49-88.

Pianta R., Downer J., Hamre B. (2016), "Quality in early education classrooms: Definitions, gaps, and systems", *Future of Children*, 26, pp. 119-138.

ISBN 9788835113850

Pontecorvo C., Tassinari G., Camaioni L. (1990), *Continuità educativa dai quattro agli otto anni*, La Nuova Italia, Scandicci.

Rasch G. (1960), *Probabilistic models for some intelligence and attainment tests*, The University of Chicago Press, Chicago.

Rasch G. (1961), "On general laws and meaning of measurement in psychology", *Proceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Theory of Probability*, University of California Press, Berkeley, pp. 321-333.

Rasch G. (1977), "On Specific Objectivity: An Attempt at formalizing the request for generality and validity of scientific statements", *Danish Yearbook of Philosophy*, 14, pp. 58-94.

Rasch G. (1980), *Probabilistic models for some intelligence and attainment tests (Reprint)*, The University of Chicago Press, Chicago.

R-ER Regione Emilia Romagna (2003), *Il coordinatore pedagogico per l'infanzia nei servizi pubblici e privati dell'Emilia-Romagna*, Bologna.

Shonkoff and Philips (*eds.*) (2000), *From neurons to neighbors. The Science of Early Childhood Development*, National Research Council and Institute of Medicine. Committee on Integrating the Science of Early Childhood Development. Board on Children, Youth, and Families, Commission on Behavioral and Social Sciences and Education, National Academy Press, Washington DC.

Slot P., Lerkkanen M-C., Leseman, P. (2015), *The relations between structural quality and process quality in European early childhood education and care provisions: secondary analyses of large scale studies in five countries*, CARE Contractor Utrecht University, Utrecht.

Son S., Peterson M. (2016), "Marital Status, Home Environments, and Family Strain: Complex Effects on Preschool Children's School Readiness Skills", *Infant and Child Development*, 26, 2, pp. 19-67.

Størksen I., Ellingsen I., Wanless S., McClelland M. (2014), "The Influence of Parental Socioeconomic Background and Gender on Self-Regulation Among 5-Year-Old Children in Norway", *Early Education and Development*, 0, pp. 1-22.

Stringher C. (2014), "What is learning to learn? A learning to learn process and output model", in R. Deakin Crick *et al.*, *Learning to Learn. International Perspectives from Theory and Practice*, Routledge, London.

Stringher C. (2016), "Cosa vuol dire qualità nella scuola dell'infanzia? Qualità per i bambini al primo posto", *Scuola dell'Infanzia*, 5, pp. 14-16.

Sylva K., Melhuish E., Sammons P., Siraj-Blatchford I., Taggart B. (2004), *The effective provision of pre-school education (EPPE) project: Final Report: A longitudinal study funded by the DfES 1997-2004*, Institute of Education, University of London/Department for Education and Skills/Sure Start.

Wanless S.B., McClelland M.M., Acock A.C., Ponitz C.C., Son S.-H., Lan X., Morrison F.J., Chen J.-L., Chen F.-M., Lee K., Sung M., Li S. (2011), "Measuring Behavioral Regulation in Four Societies", *Psychological Assessment*, 23, 2, pp. 364-378.

Wilson M. (1994), *Objective measurement: Theory into practice (Vol. 3)*, Ablex, Norwood (NJ).

ISBN 9788835113850

Zanetti M.A., Cavioni V. (a cura di) (2014), *SR 4-5. School readiness. Prove per l'individuazione delle abilità di base nel passaggio dalla scuola dell'infanzia alla scuola primaria*, Erickson, Trento.

Zanetti M.A., Miazza D. (2002), "Commenti sull'articolo di Clancy Blair – Un approccio multifattoriale all'idoneità alla frequenza scolastica", *Psicologia e Psicologi*, 2, 3, pp. 335-341.

Zellman G.L., Karoly L.A. (2012), *Moving to Outcomes. Approaches to Incorporating Child Assessments into State Early Childhood Quality Rating and Improvement Systems*, RAND Corporation, Santa Monica.

ISBN 9788835113850

2. TIMSS Mathematics achievement, school grades and national test scores: different or similar measures of student learning?

by Laura Palmerio, Elisa Caponera

The aim of this study is to examine the relationship between Mathematics achievement in *Trends in International Mathematics and Science Study* (TIMSS) and school achievement measures, such as grades and national test results in Mathematics. More than 4,000 eighth-grade students who participated in both TIMSS and national surveys in 2015 were considered. We examined the relationship between TIMSS scores, national test scores and grades in Mathematics of the entire sample that took both tests, and we investigated the differences in results between different subgroups of students based on socio-economic and cultural background. The results show that there is a positive relationship between TIMSS Mathematics achievement and national test results. TIMSS Mathematics achievement is also strongly and positively associated with grades but only after considering the geographic area where the students reside and even more after taking into account the school class of the students. The relationships between school grades and TIMSS scores for native students and immigrant students were similar.

Students from advantaged socio-economic and cultural backgrounds performed better overall in TIMSS than those from disadvantaged backgrounds; moreover, the relationship between TIMSS and the other achievement measures varied as a function of socio-economic background.

The results have implications on how one should view the results from TIMSS as a measure of student Mathematics achievement and thus how the results can be used. The possible implications for the Italian school system are discussed.

Obiettivo del presente studio è quello di esaminare la relazione tra i risultati degli studenti di terza secondaria di I grado all'indagine internazionale Trends in International Mathematics and Science Study *(TIMSS) con*

ISBN 9788835113850

i risultati degli stessi studenti alla Prova nazionale di Matematica e con i voti scolastici.

Sono stati considerati esclusivamente gli studenti che hanno preso parte a entrambe le prove (ca. 4,000 studenti). In particolare è stata indagata la relazione tra i risultati in TIMSS 2015, i voti degli studenti e il punteggio alla Prova nazionale, sia considerando gli studenti nel complesso sia verificando la presenza di differenze nei risultati tra diversi sottogruppi di studenti, in funzione del background socio-economico e culturale. I risultati evidenziano una relazione positiva tra il rendimento in Matematica TIMSS e la prova nazionale. Per quanto riguarda invece la relazione con il voto in Matematica, i risultati sono più complessi e l'associazione tra voto e rendimento in Matematica è significativa e positiva, solo dopo aver considerato l'area geografica degli studenti e ancor più dopo aver considerato la classe degli studenti. La relazione tra il rendimento nella prova nazionale e TIMSS è simile per gli studenti con background migratorio e gli studenti autoctoni.

Gli studenti provenienti da un contesto socio-economico e culturale privilegiato hanno conseguito complessivamente risultati migliori in TIMSS rispetto a quelli provenienti da un ambiente socio-economico e culturale svantaggiato. La relazione tra i risultati in TIMSS, prova nazionale e voti varia in funzione del background socio-economico.

I risultati emersi danno informazioni sulla possibilità di utilizzo dei dati TIMSS come una misura del rendimento in Matematica degli studenti; sono discusse alcune possibili implicazioni per il sistema scolastico italiano.

1. Introduction

In the last decade, in Italy, the relevance of standardized tests for the school system has increased; nonetheless, their use has always been controversial (see, e.g., Wang, Beckett and Brown, 2006).

At the international level, different studies were conducted to examine the relationships between standardized tests and teacher grades. In general, several studies evidenced a strong correlation between socio-economic and cultural status (SES) and student achievement, and strong effects of SES were found both in achievement tests and teacher grades.

Standardized test scores are often used as a criterion for admission in the next step of a student's school career. Alternatively, teacher grades are sometimes used as a criterion. Thus, different studies were conducted to identify which criterion is less influenced by the student's SES.

ISBN 9788835113850

The literature shows that SES is related to scores on standardized admission tests, such as the SAT (Scholastic Assessment Test) in the US, to performance on large-scale assessments, such as the National Assessment of Educational Progress, and to other academic measures, including school grades. Consequently, some critiques of testing (e.g., Geiser and Studley, 2002; Rothstein, 2004) have expressed that the correlations between these measures and subsequent grades are basically a secondary outcome of the influence of SES on all of these measures. It should be noted that, although sometimes, in common language, cognitive tests and standardized achievement tests are used as interchangeable terms, the vast amount of literature in this regard has shown that the standardized scores of cognitive abilities tests are often weighted by gender, age and SES.

However, the debate on this subject is still ongoing. Sackett, Kuncel, Arneson, Cooper, and Waters (2009), for example, based on a meta-analysis, found that the association between SAT scores and college grades was virtually undiminished when SES was controlled for. They evidenced that a large part of the test-academic performance relationship was independent of SES. However, Atkinson and Geiser, in 2009, underlined that in that study, Sackett and co-authors did not consider the effects of high school grade point average.

Bridgeman, Pollack and Burton (2004) verified that, after controlling for high school grades and other factors, students with higher scores on standardized tests tend to earn higher college grades, on average, than those with lower scores. Several studies also suggested that high school grades are better predictors of success than standardized test scores and that high school grades seem more accurate in predicting academic achievement than any other factor (Fleming, 2002; Hoffman and Lowitzki, 2005).

Camara, Kimmel, Scheuneman and Sawtell (2003) carried out a predictive validity study in a broad range of colleges and universities and showed that high school grade point average is the best predictor of freshmen grades. However, standardized test scores significantly improved the prediction; thus, the combination of high school grades and test scores is a better predictor of academic achievement than high school grades alone.

With this in mind, let us now consider the literature concerning gender differences. Throughout elementary, middle, and high school, girls obtain higher grades than boys in all major subjects, including Math and Science (Cole, 1997; Corbett, Hill and St. Rose, 2008; Pomerantz, Altermatt and Saxon, 2002), and girls graduate from high school with higher overall grade point averages than those of their male counterparts (US Department of Education, National Center for Education Statistics, 2004). Girls continue to outperform boys at the college and university level (e.g., Mau and Lynn,

ISBN 9788835113850

2001). However, girls do not have higher Intelligence Quotients (IQs), and they often perform lower in Mathematics on standardized tests (Corbett, Hill and St. Rose, 2008). What could explain the different performances of girls and boys in Mathematics depending on the type of measure considered?

Much research has found that there are differences between males and females, for example, in the level of self-discipline (e.g., Duckworth and Seligman, 2006) or consciousness (e.g., Schmitt *et al.*, 2008).

In Italy, several international and national studies documented systematic differences in the Mathematics results of Italian boys and girls in favour of the former (e.g., INVALSI, 2015; Mullis *et al.*, 2016; OECD, 2016). In Italy, differences between boys and girls are always in favour of boys, and this trend is more consistent than in many other participating countries (OECD, 2016; Mullis *et al.*, 2012). It should also be noted that such differences tend to increase with the level of student education and, thus, the complexity of the tests.

Research on "stereotype threat" (Steele and Aronson, 1995) suggests that these gaps may be partly due to stereotypes that dispute the abilities of females in Mathematics. Good, Aronson and Inzlicht (2003) showed that the gap in favour of boys in a standardized Mathematics test has scarcely changed in the past ten years, despite the many programmes designed to increase females' Math and Science outcomes, such as Expanding Your Horizons[1]. Many psychological and educational studies analysing the various factors assumed to underlie gender gaps have concluded that sociological factors, such as teachers' expectations, are often at stake (e.g., Jencks and Phillips, 1998; Klein et al., 1994; Romo and Falbo, 1995; Valencia, 1997).

Furthermore, different studies evidenced that teachers tend to have lower expectations for low-SES students than for middle- or high-SES students (e.g., Auwarter and Aruguete, 2008; De Boer *et al.*, 2010; Ready and Wright, 2011; Timmermans, Kuyper and van der Werf; 2015; Tobisch and Dresel, 2017).

For example, in their study, Tobisch and Dresel (2017) found that a sample of primary school teachers in Germany overestimated students without an immigration background and with high socioeconomic status.

De Boer *et al.* (2010) investigated the effect of teachers stereotypes over five years on students who entered secondary school and they found that teacher stereotypes are reduced over the first two years, and then remain substantially constant.

In the present study the data of the Italian students of the third year of lower secondary school who participated both in the TIMSS Mathematics

[1] http://www.expandingyourhorizons.org/.

ISBN 9788835113850

test and in the national Mathematics test will be used, together with the marks self-referred by the students, to investigate the following:

- there is a difference in the results of the students at three different measures of Mathematics achievement, and
- the relationships among these three different measures vary as a function of students' gender, socio-economic and cultural background and migration background.

2. Methods

2.1. Participants

The analyses presented in this paper were conducted on the TIMSS 2015 data and on the INVALSI 2015 national test scores for eighth-grade students. Only students who participated in both surveys were included in the analyses. Moreover, cases with missing values in one or more explanatory variables were excluded from the analyses. From the original sample of 4,481, the overall sample used in this study consisted of 4,026 students, grouped into 163 schools, representative of approximately 500,000 eighth-grade Italian students.

2.2. Measures

TIMSS Mathematics achievement scale. The scale was developed for the TIMSS project. The overall Mathematics performance scale consists of multiple-choice questions and open-ended questions. The eighth-grade Mathematics content domains included Number, Algebra, Geometry, and Data and chance. The cognitive domains measured were Knowing, Applying and Reasoning. Various combinations of the assessment items were compiled into 14 booklets while maintaining the distribution of items across content and cognitive domains. Using Item Response Theory (IRT) estimates, a score of Mathematics achievement was calculated for each student, drawn from five plausible values: this overall proficiency score was used in the analyses (for a detailed description, see Martin, Mullis and Hooper, 2016).

INVALSI Mathematics achievement scale. The scale was developed by the INVALSI research group for national surveys on learning. The Mathematics scale consists of 42 closed or open-ended questions (for a detailed description, see INVALSI, 2015). In the framework, Mathematics has been defined as con-

ISBN 9788835113850

ceptual knowledge that is derived from the internalization of experience and critical reflection. A central aspect in the definition of the construct was mathematical formalization, defined as the ability to express and use mathematical thinking. The INVALSI research group identified three cognitive domains: Solving Problems, Arguing, Knowing. The four content domains are Numbers, Space and figures, Data and predictions, and Relations and functions.

Teachers' grades. Students answered two questions regarding the last grade they obtained in Mathematics, both written and oral. The variable used in the following analyses is the average score of these two pieces of information.

Socio-economic and cultural status (SES). Based on the answers in the TIMSS student questionnaire, a general index of each student's socio-economic and cultural status was created by IEA: (1) student home environments, including the parents' educational level; (2) the number of resources for study available at home; and (3) the number of books in the home. To compare the index within Italian students, tertile groups were created to group students: 1) students with low socio-economic and cultural backgrounds; 2) students with medium socio-economic and cultural backgrounds; and 3) students with high socio-economic and cultural backgrounds.

Immigration status (Immig). Based on TIMSS student questionnaire answers, a new variable was created to identify native students (students born in the country of the test or with at least one parent born in the country of the test, code 0) and non-native students (students not born in the country of the test or born in the country of the test but with both parents born in another country, code 1).

2.3. Data analysis

The descriptive and correlation analyses were conducted using the software IEA IDB Analyzer, a software developed by the IEA Data Processing and Research Center for analysing data from many international surveys. The IDB Analyzer allows for the handling of complex sample designs, using plausible value methodology and calculating correct standard errors when conducting analyses with large-scale surveys. The IDB Analyzer creates an SPSS code that can be used to conduct statistical analysis considering the complex sample and assessment structures of these databases (IEA, 2012).

Descriptive analyses were used to verify whether there were differences in performance in relation to the geographical area, gender, socio-economic and cultural background, and migration status. Data analyses were conducted with respect to means, and deviations from the means, within each test.

ISBN 9788835113850

Pearson's coefficients were calculated across Italian geographic areas, gender, socio-economic and cultural background, and by immigrant background to verify the relationship among all three different measures of Mathematics achievement, and the unique contribute of each measure in the student Mathematics assessment. We also investigated the strength of the relation between the three different measures depending on the student level of socio-economic and cultural background.

3. Results

3.1. Descriptive statistics

Table 1 shows the descriptive statistics divided by geographic area[2], gender and socio-economic and cultural background.

Because of the scarce number of students with immigrant backgrounds, the analyses divided per immigrant/non-immigrant background were conducted considering the entire sample, instead of geographic areas.

Concerning national Mathematics achievement, the differences between students from the north and the south are significant, with a 13 point of difference[3].

The difference between north and south is also significant in TIMSS achievement, with a difference of 42 points[4].

Furthermore, in both the INVALSI and TIMSS tests, the difference between centre and south is significant, even though it is more moderate compared with the difference between north- south difference.

In both achievement tests, the differences between north and centre are not significant.

There is no difference in teachers' grades across geographic areas.

[2] North: Emilia-Romagna, Friuli-Venezia Giulia, Liguria, Lombardia, Piemonte, Trentino-Alto Adige, Valle D'Aosta, Veneto; Centre: Lazio, Marche, Toscana, Umbria; South: Abruzzo, Basilicata, Campania, Calabria, Molise, Puglia, Sicilia, Sardegna.

[3] The standard deviation value in INVALSI test is 40.

[4] The standard deviation value in TIMSS is 100.

ISBN 9788835113850

Tab. 1 – Sample characteristics

Number of Students	*All students*	*Female*		*Male*		*Low SES*		*Medium SES*		*High SES*		*Immigrant background*		*No immigrant background*	
	N	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*
North	2,018	1,024	51	994	49	400	20	617	31	1,001	50	307	15	1,711	85
Centre	530	238	45	292	55	127	24	162	31	241	45	60	11	470	89
South	1,478	740	50	738	50	587	40	461	31	430	29	54	4	1,424	96
Italy	4,026	2,002	50	2,024	50	1,114	28	1,240	31	1,672	42	421	10	3,605	90

Because of rounding, some total percentages may appear inconsistent.

Data source: TIMSS, 2015

Tab. 2 – Mean, s.e. and standard deviation for INVALSI Mathematics achievement, TIMSS and teachers' grades

	North			*Centre*			*South*			*Italy*		
	Mean	*(e.s.)*	*Std. dev.*	*Mean*	*(e.s.)*	*Std. dev.*	*Mean*	*(e.s.)*	*Std. dev.*	*Mean*	*(e.s.)*	*Std. dev.*
INVALSI	203	(1.2)	37.8	201	(3.2)	41.7	**190**	**(1.7)**	**35.3**	198	(1.0)	38.1
TIMSS	515	(3.2)	67.0	503	(6.6)	69.9	**472**	**(4.5)**	**76.2**	496	(2.5)	73.8
Grades	6.9	(0.1)	1.4	6.6	(0.1)	1.4	6.8	(0.1)	1.4	6.8	(0.1)	1.4

In parentheses are the standard errors.
In bold are groups with significant differences with respect to the other groups within type of assessment.

Data sources: TIMSS, 2015; INVALSI, 2015

ISBN 9788835113850

Tab. 3 – Mean, s.e. and standard deviation for INVALSI Mathematics achievement, TIMSS and teachers' grades per gender

		North			*Centre*			*South*			*Italy*		
		Mean	*(e.s)*	*Std. dev.*	*Mean*	*(e.s)*	*Std. dev.*	*Mean*	*(e.s)*	*Std. dev.*	*Mean*	*(e.s)*	*Std. dev.*
INVALSI	Male	206	(1.7)	38.4	205	(2.8)	40.3	**190**	**(1.8)**	**36.0**	199	(1.0)	38.5
	Female	201	(1.7)	37.1	197	(5.1)	42.8	**189**	**(2.2)**	**34.5**	195	(1.0)	37.5
TIMSS	Male	**518**	**(3.6)**	**68.6**	508	(6.3)	69.8	**477**	**(5.1)**	**76.7**	**500**	**(2.8)**	**74.6**
	Female	**512**	**(3.9)**	**65.2**	496	(8.9)	69.4	**468**	**(5.0)**	**75.4**	**492**	**(3.1)**	**72.8**
Teacher grades	Male	6.7	(.07)	1.4	**6.5**	**(.11)**	**1.4**	6.6	(.08)	1.3	**6.7**	**(.05)**	**1.4**
	Female	7.0	(.08)	1.4	**6.8**	**(.16)**	**1.4**	7.0	(.08)	1.4	**7.0**	**(.06)**	**1.4**

In parentheses are the standard errors.
In bold are the gender group with significant differences with respect to the other group within type of assessment.

Data sources: TIMSS 2015; INVALSI 2015

ISBN 9788835113850

Tab. 4 – Mean, s.e. and standard deviation for INVALSI Mathematics achievement, TIMSS and teachers' grades per socio-economic and cultural background

		North			*Centre*			*South*			*Italy*		
		Mean	*(e.s)*	*Std. dev.*	*Mean*	*(e.s)*	*Std. dev.*	*Mean*	*(e.s)*	*Std. dev.*	*Mean*	*(e.s)*	*Std. dev.*
INVALSI	Low SES	**189**	**(2.9)**	**34.1**	191	(2.7)	33.6	**184**	**(2.6)**	**31.4**	**185**	**(2.4)**	**33.0**
	Medium SES	**201**	**(2.0)**	**36.4**	**196**	**(3.1)**	**39.0**	**201**	**(3.2)**	**34.0**	205	(3.0)	36.2
	High SES	**217**	**(2.1)**	**38.4**	**223**	**(5.8)**	**46.6**	**213**	**(5.7)**	**36.6**	217	(7.8)	39.9
TIMSS	Low SES	**480**	**(4.7)**	**67.4**	**472**	**(9.7)**	**69.7**	**442**	**(5.7)**	**70.2**	**458**	**(3.8)**	**71.6**
	Medium SES	**509**	**(3.9)**	**63.1**	**494**	**(6.0)**	**65.5**	**483**	**(4.1)**	**71.6**	**496**	**(2.7)**	**67.9**
	High SES	**535**	**(3.8)**	**61.9**	**526**	**(7.8)**	**65.2**	**508**	**(6.1)**	**72.2**	**526**	**(3.1)**	**66.4**
Teacher grades	Low SES	**6.36**	**(0.1)**	**1.3**	6.16	(0.1)	1.3	**6.34**	**(0.1)**	**1.2**	**6.32**	**(0.1)**	**1.2**
	Medium SES	**6.73**	**(0.1)**	**1.4**	6.36	(0.1)	1.4	**6.94**	**(0.1)**	**1.3**	**6.75**	**(0.1)**	**1.4**
	High SES	**7.17**	**(0.1)**	**1.4**	**7.15**	**(0.2)**	**1.3**	**7.32**	**(0.1)**	**1.4**	**7.21**	**(0.1)**	**1.4**

In parentheses are the standard errors.
In bold are the group with significant differences with respect to the group below within geographic area and type of assessment.

Data sources: TIMSS 2015; INVALSI 2015

ISBN 9788835113850

In the north, there are significant differences in the national test between geographic areas in favour of males.

The difference between male and female students is significant in TIMSS achievement and the INVALSI national test, with male students outperforming female students. The difference is significant only if we consider the entire sample.

In contrast, regarding teachers' grades, female students received better grades than male students in all geographic areas.

In most geographic areas, there are differences in the function of the socio-economic and cultural index: students with higher socio-economic and cultural backgrounds outperform students from more disadvantaged backgrounds in national and TIMSS tests and teacher grades.

Tab. 5 – Mean, s.e. and standard deviation for INVALSI Mathematics achievement, TIMSS and teachers' grades per immigrant background (Italy)

	Immigrant background			*No immigrant background*		
	Mean	*(s.e.)*	*Std. dev.*	*Mean*	*(s.e.)*	*Std. dev.*
INVALSI	**186**	**(3.7)**	**35.0**	**209**	**(5.0)**	**38.2**
TIMSS	**478**	**(5.9)**	**73.3**	**498**	**(2.6)**	**76.5**
Teacher grades	**6.3**	**(0.1)**	**1.4**	**6.8**	**(0.1)**	**1.4**

In parentheses are the standard errors.
In bold are the group with significant differences with respect to the group below within geographic area and type of assessment.

Data sources: TIMSS 2015; INVALSI 2015

In Italy, students with no immigrant background outperformed students with immigrant background both in the national test and in the TIMSS test; moreover, they received higher grades.

3.2. Correlation analyses

Tables 6, 7, 8 and 9 show the association between the three different measures of Mathematics achievement to better understand their relationships.

The association between the three measures of student achievement in Mathematics is high. At the national level, the strongest association is between the INVALSI test and TIMSS. Teachers' grades are strongly associated with the INVALSI test and less associated with TIMSS. Regarding the geographic area, the associations among the three variables are more consistent within the centre area and less consistent in the north area.

ISBN 9788835113850

Tab. 6 – Correlation between the INVALSI test, TIMSS and teachers' grades in Mathematics by geographic area

		TIMSS Test		*INVALSI Test*	
Teachers' grades	North	0.56	(0.03)	0.62	(0.02)
	Centre	0.64	(0.04)	0.69	(0.02)
	South	0.61	(0.02)	0.62	(0.02)
	Italy	0.57	(0.01)	0.62	(0.01)
INVALSI Test	North	0.64	(0.02)		
	Centre	0.73	(0.03)		
	South	0.69	(0.03)		
	Italy	0.68	(0.02)		

All correlations are statistically significant at $p < 0.01$.

Data sources: TIMSS 2015; INVALSI 2015

Tab. 7 – Correlation between the INVALSI test, TIMSS and teachers' grades in Mathematics by geographic area and gender

		TIMSS Test				*INVALSI Test*			
		Male		*Female*		*Male*		*Female*	
		r	*s.e.*	*r*	*s.e.*	*r*	*s.e.*	*r*	*s.e.*
Teachers' grades	North	0.54	(0.04)	0.60	(0.02)	0.62	(0.03)	0.64	(0.02)
	Centre	0.66	(0.04)	0.65	(0.03)	0.69	(0.03)	0.71	(0.05)
	South	0.64	(0.03)	0.60	(0.03)	0.62	(0.03)	0.63	(0.03)
	Italy	0.58	(0.02)	0.58	(0.02)	0.62	(0.01)	0.64	(0.02)
INVALSI Test	North	0.65	(0.03)	0.69	(0.03)				
	Centre	0.74	(0.03)	0.71	(0.05)				
	South	0.69	(0.03)	0.63	(0.04)				
	Italy	0.69	(0.02)	0.67	(0.02)				

All correlations are statistically significant at $p < 0.01$.

Data sources: TIMSS 2015; INVALSI 2015

After dividing the data by gender, the results do not show a substantial difference in the relationships among the three different measures of Mathematics achievement as a function of gender.

ISBN 9788835113850

Tab. 8 – Correlation between the INVALSI test, TIMSS and teachers' grades in Mathematics by geographic area and socio-economic and cultural background

		TIMSS Test						*INVALSI Test*					
		Low SES		*Medium SES*		*High SES*		*Low SES*		*Medium SES*		*High SES*	
		r	*s.e.*	*r*	*s.e.*	*r*	*s.e.*	*r*	*s.e.*	*r*	*s.e.*	*r*	*s.e.*
Teachers' grades	North	0.55	(0.05)	0.49	(0.04)	0.55	(0.04)	0.59	(0.05)	0.54	(0.03)	0.63	(0.03)
	Centre	0.56	(0.08)	0.57	(0.07)	0.64	(0.04)	0.57	(0.06)	0.60	(0.05)	0.72	(0.03)
	South	0.50	(0.04)	0.56	(0.04)	0.65	(0.03)	0.54	(0.05)	0.56	(0.03)	0.65	(0.03)
	Italy	0.55	(0.03)	0.55	(0.02)	0.64	(0.02)	0.55	(0.03)	0.55	(0.02)	0.64	(0.02)
INVALSI Test	North	0.67	(0.03)	0.66	(0.03)	0.67	(0.04)						
	Centre	0.62	(0.08)	0.66	(0.04)	0.76	(0.03)						
	South	0.52	(0.05)	0.60	(0.04)	0.69	(0.03)						
	Italy	0.59	(0.03)	0.63	(0.02)	0.69	(0.02)						

All correlations are statistically significant at $p < 0.01$.

Data sources: TIMSS 2015; INVALSI 2015

ISBN 9788835113850

The results show a different relationship among the three measures of Mathematics achievement as a function of socio-economic and cultural background. The association is strongest within the group of students with high socio-economic and cultural background in Italy and in geographic areas, except for the north, where the associations among the three variables do not change as a function of socio-economic and cultural status.

Tab. 9 – Correlation between the INVALSI test, TIMSS and teachers' grades in Mathematics by immigrant background

	TIMSS Test				*INVALSI Test*			
	Immigrant background				*Immigrant background*			
	No imm. background		*Imm. background*		*No imm. background*		*Imm. background*	
	r	*s.e.*	*r*	*s.e.*	*r*	*s.e.*	*r*	*s.e.*
Teachers' grades	0.57	(0.02)	0.54	(0.04)	0.62	(0.01)	0.61	(0.05)
INVALSI Test	0.69	(0.02)	0.62	(0.04)				

All correlations are statistically significant at $p < 0.01$.

Data sources: TIMSS 2015; INVALSI 2015

The difference in the association among the three measures is small; the INVALSI test and TIMSS have the strongest correlation in both immigrant background students and non-immigrant background students.

To verify whether there were any the differences between grades on one hand and standardized measures of achievement on the other hand, we compared the results in the TIMSS and INVALSI tests with the students grades on maths, based on student socio-economic and cultural background.

At high level of achievement, in TIMSS there are no strong differences between students with high or low SES in all geographic areas, while at low level in TIMSS, students with low SES systematically obtain lower grades at school than students with high SES.

Regarding to the INVALSI tests, low SES students have lower marks in Mathematics compared with high SES students who achieved the same level in the INVALSI test.

ISBN 9788835113850

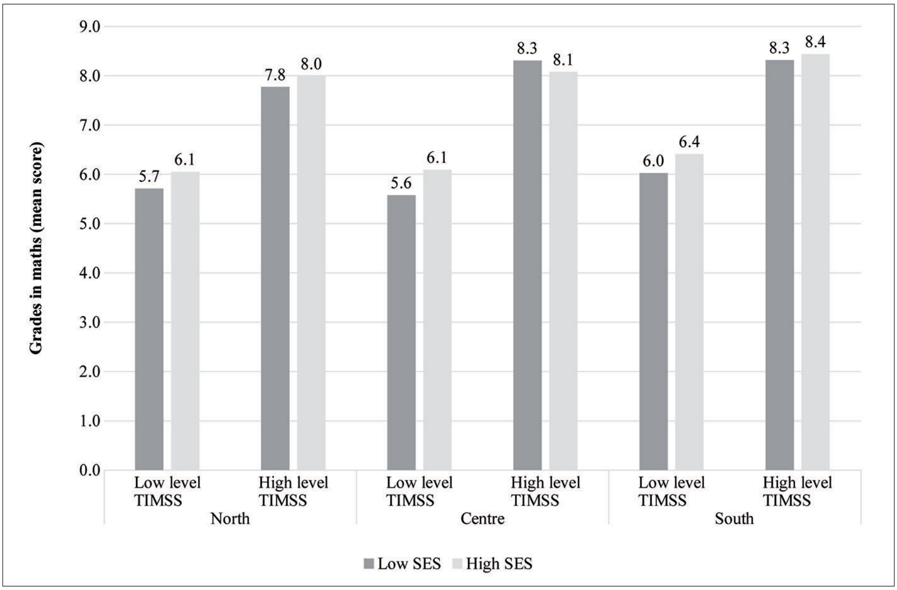

Fig. 1 – Relationships between students' grades and TIMSS achievement

Data sources: TIMSS 2015; INVALSI 2015

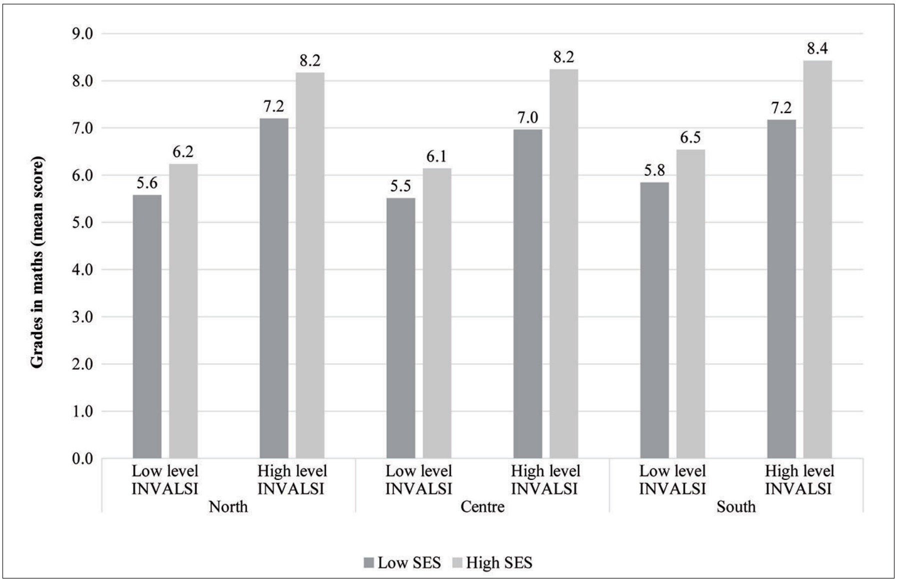

Fig. 2 – Relationships between students' grades and INVALSI achievement

Data sources: INVALSI 2015

ISBN 9788835113850

4. Discussion

The main aim of the present study was to evaluate the relationships among TIMSS scores in eighth grade, INVALSI national test scores and teachers' grades in Mathematics of students participating in both national and international surveys.

According to the literature, female students received better grades in school than male students; on the contrary, boys outperformed girls in the TIMSS test and INVALSI test (Corbett, Hill and St. Rose, 2008; US Department of Education, National Center for Education Statistics, 2004).

Context factors reflecting the availability of economic and cultural resources in the household play a relevant role in determining student performance. As expected and according to previous studies (see, e.g., Chiu and Xihua, 2008; Ismail and Awang, 2008; Levpušček, Zupančič and Sočan, 2013; Sirin, 2005), the analyses showed that a high socio-economic status has a significantly positive effect on student achievement. Compared with their counterparts from a socio-economically disadvantaged background, students from an advantaged background performed better in Mathematics, both on the national test and international test, and obtained better grades in school.

Some limitations to this study should be noted. First, it is necessary to bear in mind that the data used in this study are related to only one school year. Analyses on more than one dataset are needed for a clearer picture of which school factors are associated with Mathematics achievement.

Furthermore, this study did not take into account other factors not strictly related to cognitive performance but relevant in explaining students' achievement (e.g., Poropat, 2009). For instance, all the factors at stake in a self-regulated learning system (metacognitive, affective, motivational, etc.), as well as the individual characteristics, could help clarify the different relations between academic measures as a function of the student's gender.

Bearing in mind these limitations, this study investigated the relationship between three different measures of school achievement in Mathematics at the end of the first cycle of instruction in a large and representative sample of Italian students.

ISBN 9788835113850

5. Conclusion

In many countries, standardized tests are the most important foundation for educational and political decision makers to improve the quality of educational systems and the teaching and learning processes. Although several studies have shown that teacher grades are often better predictors of future academic success (Fleming *et al.*, 2005; Hoffman and Lowitzki, 2005), it should be noted that standardized tests add some relevant information. Students from disadvantaged socio-economic and cultural backgrounds obtained better results in standardized tests than in teachers' grades, perhaps due to an evaluative bias that could influence teachers' perception of the cognitive ability of a student. On the other hand, standardized test scores and teachers' grades were quite consistent for advantaged students.

As far as gender differences are concerned, female students outperform male students in school but not in standardized tests. The literature has shown that female students are usually more self-disciplined and conscientiousness than male students (Duckworth and Seligman, 2006; Schmitt *et al.*, 2008). Further research is necessary to understand in which way students' characteristics play a role in learning at different stages and which factors are involved in the process of classroom evaluation; a deeper understanding of this aspect is relevant to improving the equity of opportunities and students' wellbeing at school.

Another relevant aspect that seems to emerge from this study, and that needs further investigation, is related to the difference in school grades in students with equal performance in the standardized tests: students from socio-economically disadvantaged backgrounds get worse grades at school than those from socio-economically advantaged backgrounds, even where the performance in standardized tests is similar. This may suggest problems of evaluative biases in teachers that the standardized tests allow to keep under control, and thus the use of tests might support teachers in reducing biases.

The results of the present study thus suggest the use of standardized assessment along with school grades to improve educational evaluation in the classroom. It is evident that assessments at school take into account a variety of student factors, not only the cognitive ability, that are relevant to be considered a "good student". Teachers' grades can represent an evaluation of the student as a whole but may suffer from several biases; the results of standardized tests could offer a different point of view to improve the achievement and development of students from disadvantaged backgrounds.

ISBN 9788835113850

References

Atkinson R.C., Geiser, S. (2009), "Reflections on a century of college admissions tests", *Educational Researcher*, 38 (9), pp. 665-676.

Auwarter A.E., Aruguete M. S. (2008), "Effects of student gender and socioeconomic status on teacher perceptions", *The Journal of Educational Research*, 101 (4), pp. 242-246.

Bridgeman B., Pollack J., Burton N. (2004), "Understanding what SAT Reasoning Test™ scores add to high school grades: A straightforward approach", *ETS Research Report Series*, 2004 (2), pp. i-20.

Camara W., Kimmel E., Scheuneman J., Sawtell E.A. (2003), Whose grades are inflated, The College Board, New York.

Chiu M.M., Xihua Z. (2008), "Family and motivation effects on mathematics achievement: Analyses of students in 41 countries", *Learning and Instruction*, 18 (4), pp. 321-336.

Cole N.S. (1997), *The ETS Gender Study: How Females and Males Perform in Educational Settings*, Educational Testing Service, Princeton (NJ).

Corbett C., Hill C., St. Rose A. (2008), *Where the Girls Are: The Facts about Gender Equity in Education*, American Association of University Women Educational Foundation, Washington DC.

De Boer H., Bosker R.J., van der Werf M.P. (2010), "Sustainability of teacher expectation bias effects on long-term student performance", *Journal of Educational Psychology*, 102 (1), pp. 168-179.

Duckworth A.L., Seligman M.E. (2006), "Gives girls the edge: Gender in self-discipline, grades, and achievement test scores", *Journal of Educational Psychology*, 98 (1), pp. 198-208.

Fleming J. (2002), "Who will succeed in college? When the SAT predicts Black students' performance", *The Review of Higher Education*, 25 (3), pp. 281-296.

Fleming C.B., Haggerty K.P., Catalano R.F., Harachi T.W., Mazza J.J., Gruman D.H. (2005), "Do social and behavioral characteristics targeted by preventive interventions predict standardized test scores and grades?", *Journal of School Health*, 75 (9), pp. 342-349.

Geiser S., Studley W.R. (2002), "UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California", *Educational Assessment*, 8 (1), pp. 1-26.

Good C., Aronson J., Inzlicht M. (2003), "Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat", *Journal of Applied Developmental Psychology*, 24 (6), pp. 645-662.

Hoffman J.L., Lowitzki K.E. (2005), "Predicting college success with high school grades and test scores: Limitations for minority students", *The Review of Higher Education*, 28 (4), pp. 455-474.

IEA – International Association for the Evaluation of Educational Achievement (2012), *International database analyzer (version 3.0)*, IEA Data Processing and Research Center, Hamburg.

ISBN 9788835113850

INVALSI (2015), *Rilevazioni nazionali degli apprendimenti 2014-15. Rilevazione degli apprendimenti nelle classi II e V primaria, nella classe III (Prova nazionale) della scuola secondaria di primo grado e nella II classe della scuola secondaria di secondo grado*, retrieved on September, 15, 2020, from https://invalsi-areaprove.cineca.it/docs/attach/035_Rapporto_Prove_INVALSI_2015.pdf.

Ismail N.A., Awang H. (2008), "Differentials in mathematics achievement among eighth-grade students in Malaysia", *International Journal of Science and Mathematics Education*, 6 (3), pp. 559-571.

Jencks C., Phillips M. (1998), "America's next achievement test: Closing the Black-White test score gap", *American Prospect*, 40, pp. 44-53.

Klein S.S., Ortman P.E., Campbell P., Greenberg S., Hollingsworth S., Jacobs J., Sadker M. (1994), "Continuing the journey toward gender equity", *Educational Researcher*, 23 (8), pp. 13-21.

Levpušček M.P., Zupančič M., Sočan G. (2013), "Predicting achievement in mathematics in adolescent students: The role of individual and social factors", *The Journal of Early Adolescence*, 33 (4), pp. 523-551.

Martin M.O., Mullis I.V.S., Hooper M. (*eds.*) (2016), *Methods and Procedures in TIMSS 2015*, TIMSS & PIRLS International Study Center, Boston College, retrieved on September, 15, 2020, from https://timssandpirls.bc.edu/publications/timss/2015-methods.html.

Mau W.C., Lynn R. (2001), "Gender differences on the Scholastic Aptitude Test, the American College Test and college grades", *Educational Psycholog*y, 21 (2), pp. 133-136.

Mullis I.V., Martin M.O., Foy P., Arora A. (2012), *TIMSS 2011 international results in mathematics*, International Association for the Evaluation of Educational Achievement, Herengracht 487, Amsterdam.

Mullis I.V.S., Martinù M.O., Foy P., Hooper M. (2016), *TIMSS 2015 International Results in Mathematics*, TIMSS & PIRLS International Study Center, Boston College, retrieved on September, 15, 2020, from http://timssandpirls.bc.edu/timss2015/international-results/.

OECD (2016), *PISA 2015 Results (Volume I): Excellence and Equity in Education*, OECD, Paris.

US Department of Education, National Center for Education Statistics (2004), *The High School Transcript Study: A Decade of Change in Curricula and Achievement, 1990-2000*, NCES 2004-455, by Robert Perkins, Brian Kleiner, Stephen Roey, and Janis Brown, Project Officer: Janis Brown, Washington DC.

Pomerantz E.M., Altermatt E.R., Saxon J.L. (2002), "Making the grade but feeling distressed: Gender differences in academic performance and internal distress", *Journal of Educational Psychology*, 94 (2), pp. 396-404.

Poropat A.E. (2009), "A meta-analysis of the five-factor model of personality and academic performance", *Psychological Bulletin*, 135 (2), pp. 322-338.

Ready D.D., Wright D.L. (2011), "Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context", *American Educational Research Journal*, 48 (2), pp. 335-360.

ISBN 9788835113850

Romo H., Falbo T. (1995), *Latino high school graduation: Defying the odds*, University of Texas Press, Austin (TX).

Rothstein J.M. (2004), "College performance predictions and the SAT", *Journal of Econometrics*, 121 (1-2), pp. 297-317.

Sackett P.R., Kuncel N.R., Arneson J.J., Cooper S.R., Waters S.D. (2009), "Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance?", *Psychological Bulletin*, 135 (1), pp. 1-22.

Schmitt D.P., Realo A., Voracek M., Allik J. (2008), "Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures", *Journal of Personality and Social Psychology*, 94 (1), pp. 168-182.

Sirin S.R. (2005), "Socioeconomic status and academic achievement: A meta-analytic review of research", *Review of Educational Research*, 75, 3, pp. 417-453.

Steele C.M., Aronson J. (1995), "Stereotype threat and the intellectual test performance of African Americans", *Journal of Personality and Social Psychology*, 69 (5), pp. 797-811.

Timmermans A.C., Kuyper H., van der Werf G. (2015), "Accurate, inaccurate, or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education?", *British Journal of Educational Psychology*, 85 (4), pp. 459-478.

Tobisch A., Dresel M. (2017), "Negatively or positively biased? Dependencies of teachers' judgments and expectations based on students' ethnic and social backgrounds", *Social Psychology of Education*, 20 (4), pp. 731-752.

Valencia S.W., Au K.H. (1997), "Portfolios across educational contexts: Issues of evaluation, teacher development, and system validity", *Educational Assessment*, 4 (1), pp. 1-35.

Wang L., Beckett G. H., Brown L. (2006), "Controversies of standardized assessment in school accountability reform: A critical synthesis of multidisciplinary research evidence", *Applied Measurement in Education*, 19 (4), pp. 305-328.

ISBN 9788835113850

3. Short- and medium-term effects of bullying on academic achievement in Italy

by Elena Demuru, Patrizia Giannantoni, Jana Kopecna

Recently, the phenomenon of bullying has received increasing attention, especially in the field of education. In the literature, numerous studies have focused on this phenomenon and its effect on students' well-being, whereas rare are studies that have analyzed its impact on academic performance. The latter shows that bullying experience leads to poorer school performance (Lacey and Cornell, 2013; Beran *et al.*, 2008). Nevertheless, the majority of the studies are focused on age group 14-19, and in Italy, regardless of age groups, studies on bullying are even much rare.

The aim of this study is to quantify the phenomenon of bullying in primary schools in Italy, by describing the characteristics of bullied students, and seeking the potential short- and medium-term effects on their academic achievement.

Data come from the National INVALSI Assessment of Student academic skills of the 5th grade of primary school for the 2013/2014 and 2014/2015 academic year, linked to the corresponding data for the 8th grade. While these data include the results of the standardized tests, the 5th grade's questionnaire provides the information about the socio-demographic and educational characteristics, and four questions related to bullying: being teased, insulted, beaten up, or isolated.

Firstly, through univariate analysis, the characteristics of the students who suffered bullying will be presented. Results will be enriched with multiple correspondence analysis to trace complete profiles of these students along with context data. Finally, using linear regression models, the effect of bullying on academic achievement will be evaluated, measured as score in the INVALSI tests (5th and 8th grade) and controlling for the principal confounding variables: base ability, gender, citizenship, socio-economic status, and variables related to isolation. The preliminary results indicate that students

ISBN 9788835113850

who suffer from violent forms of bullying are more likely to be males, foreigners, with the medium-low socio-economic status. Moreover, the victims seem to have lower scores in the standardized tests, not only in the observed school year, but also after three years, with a gradient proportional to the frequency of bullying.

These results seem to confirm the original hypothesis that being a victim of bullying has a negative impact on academic performance in the short- and medium-term.

Il fenomeno del bullismo sta ricevendo crescente attenzione, soprattutto in ambiente scolastico. In letteratura esistono diversi studi che analizzano tale fenomeno e il suo impatto sul benessere degli studenti, mentre sono meno frequenti gli studi che ne analizzano l'impatto sul rendimento scolastico. Questi ultimi dimostrano che chi subisce episodi di bullismo presenta un rendimento scolastico mediamente inferiore. Tuttavia, la maggior parte degli studi si concentra solo sulla fascia d'età 14-19 anni e in Italia, in particolare, gli studi sul bullismo in qualunque fascia di età sono piuttosto rari.

L'obiettivo di questo contributo è quantificare il fenomeno del bullismo nella scuola primaria in Italia, descrivendo le caratteristiche degli studenti che lo subiscono e verificando per questi studenti l'eventuale impatto sul rendimento scolastico a breve e medio termine.

Sono stati utilizzati i dati delle rilevazioni nazionali INVALSI di quinta primaria (grado 5) degli anni 2013/2014 e 2014/2015, agganciati ai dati di terza secondaria di primo grado (grado 8) a tre anni di distanza. Tali dati includono i risultati nelle prove standardizzate, mentre dal Questionario Studente del grado 5 derivano le informazioni socio-demografiche e scolastiche, e quattro domande relative al bullismo: essere preso in giro, insultato, picchiato, isolato.

In una prima sezione, attraverso analisi univariate, verranno illustrate le caratteristiche degli studenti che subiscono bullismo. I risultati saranno arricchiti dall'analisi delle corrispondenze multiple per tracciare i profili completi di questi studenti insieme a dati di contesto. Infine, attraverso modelli di regressione lineare, verrà valutato l'impatto del bullismo sul rendimento, misurato come punteggio ai test INVALSI (grado 5 e 8), controllando per le principali variabili di confondimento: abilità di base, genere, cittadinanza, status socioeconomico, e per alcune variabili di isolamento.

I risultati preliminari indicano che i bambini che subiscono atti violenti sono più frequentemente maschi, stranieri, di condizione socioeconomica medio-bassa. Inoltre, sembra che le vittime di bullismo riportino punteggi generalmente più bassi nelle prove standardizzate, non solo alla fine dell'an-

ISBN 9788835113850

no scolastico in cui si subisce il fenomeno, ma anche a tre anni di distanza, con un gradiente proporzionale alla frequenza degli episodi.

Questi risultati sembrerebbero confermare che subire episodi di bullismo ha un impatto negativo sul rendimento scolastico a breve e medio termine.

1. Introduction

Even if the attention towards the phenomenon of bullying has a long history (Gini, 2004), in recent years it has grown considerably and not only among scholars and psychology experts but also in the mass media and society as a whole. In fact, in Italy as well as in other parts of the world, bullying episodes seem to increase among children and adolescents. The type of bullying is variable, going from simple – and seemingly innocent – teasing to real insults, up to acts of physical violence. These actions obviously have consequences that can also be very serious, especially in such a fragile stage of life and for particularly sensitive children.

To contrast and prevent bullying, a better knowledge of this phenomenon is necessary in terms of diffusion, main determinants and effects that it may have – both in a short and long term – on the students suffering it. The data gathered by INVALSI (Italian National Institute for the Evaluation of the Educational System of Instruction and Training) allow us to analyse in depth some fundamental aspects. Indeed, personal questionnaires administered to students of 5^{th} and 10^{th} grade (the last class of primary school and the second class of upper secondary school respectively) in the academic year 2014/2015 contain a set of eight questions on the frequency of the acts of bullying, either active or passive (i.e., perpetrated or suffered). This information can be combined with performance of these same students at INVALSI tests, both on the same year and 3 years later.

The objective of this contribution is to exploit the informative power of these data, focusing on primary school, in order to: 1) quantify the phenomenon of bullying in primary school in Italy, both at the national level and at a greater geographical detail; 2) describe the socio-demographic characteristics of students who most frequently suffer bullying; 3) verify the existence and strength of an association between bullying and academic performance in the short and medium term.

ISBN 9788835113850

2. Literature

International literature on bullying mainly focus on the impact of this phenomenon on the well-being of victims, in terms of social integration in the school context and motivation to study (Smith *et al.*, 2004; Barker *et al.*, 2008).

Studies that analyze the impact of bullying on academic performance are less frequent and show in most cases that students who experience bullying have an average lower academic performance. This is what, for example, Beran *et al.* (2008) concluded from their study on data for a sample of 2,084 Canadian students aged 10 and 11, which clearly shows that being bullied is negatively associated with reading and writing as well as with mathematical skills. Such association seems to be stronger among children who receive less support from their parents and are less motivated to study. Brown and Taylor (2007) obtained similar results in their analysis of data from the "National Child and Development Study", an English longitudinal survey that followed a cohort of children born between 3 and 9 March 1958 up until adult age by interviewing them at 7, 11, 16, 23, 33 and 42 years of age. From this study, indeed, it emerged that suffering bullying by schoolmates at 9 and 11 years of age has a strong impact on academic performance at the age of 16. Furthermore, the authors found that the effect of bullying remains even beyond the end of schooling, resulting in lower wage levels among those who had been bullied during childhood. A negative association between bullying and academic performance in terms of mathematical skills was also found in a very recent study by Oliveira *et al.* (2018), based on data relating to 28,983 students enrolled in classes of 6th grade in the city of Recife in Brazil. Contrary to what emerged from the studies discussed so far, Woods and Wolke (2004) found no significant association between being bullied and academic achievement based on their analysis of data for a sample of 1,016 English children enrolled in grades 2 and 4.

Finally, studies analyzing bullying in Italy are still rare and mainly focus on characteristics associated with being bullied or on factors that favour its spread. The very first study was published by Genta *et al.* (1996), who, however, focused their attention on the specific situation of two cities in central and southern Italy (Florence and Cosenza). More recently, Alivernini *et al.* (2017) conducted an analysis at the national level using data released by INVALSI and showed that first and second generation foreign students are more subject to bullying than their schoolmates of Italian citizenship. Finally, an analysis of TIMSS data by Ponzo (2013) showed that being a victim of bullying is significantly associated with persistently lower academic achievement both in 4th and 8th grade.

ISBN 9788835113850

3. Data and methods

We prepared an *ad hoc* database by linking different data sets from national and supplementary INVALSI surveys, using the SIDI (Sistema Informativo dell'Istruzione – Educational Information System) code as linkage key. In particular, we linked the questionnaire given to all students enrolled in the 5th grade in the school year 2014/2015 to scores obtained by the same students in the standardized tests both in that same year and three years later (i.e., to scores obtained in the test in grade 8 in the school year 2017/2018). The resulting data set makes available – for each student – the information collected through the questionnaire together with the socio-demographic data transmitted to INVALSI by the school secretary and the scores obtained in the standardized Italian and Mathematics tests in 5th and 8th grades. Furthermore, in order to be able to control for basic competences in regression models, we also linked to the data set the scores obtained by the same students in 2nd grade tests in the 2011/2012 school year. The richness of this database allowed on the one hand to quantify the spread of bullying in the Italian primary school and contextualize the phenomenon by describing the main characteristics of the students who suffer it, and on the other hand to analyze the consequences of bullying on academic performance in the short and medium term.

The first part of our analysis is aimed at quantifying the frequency of different types of bullying episodes in Italian primary school. Our goal is to describe the general picture of the situation, paying attention to possible differences between geographical areas of the country. For this reason, in addition to the prevalence of bullying at the national level and in 4 macro-areas (North-West North-East, Center, South and Islands)[1], we report maps in which each Italian province is coloured with a grey scale based on the value of the gap between the provincial and national prevalence of each type of bullying considered in the analysis[2]. These maps allow to paint a more complete picture of territorial differences, since they immediately highlight the provinces in which the prevalence of bullying is higher than the national average (dark grey shades) from those in which it is lower (light grey shades).

[1] Each macro-area includes the regions listed below. North-west: Valle D'Aosta, Piemonte, Liguria, Lombardia. North-east: Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna. Center Toscana, Umbria, Marche, Lazio. South: Abruzzo, Molise, Campania, Puglia. South and Islands: Basilicata, Calabria, Sicilia, Sardegna.

[2] Census data allowed us to perform analysis at provincial level.

ISBN 9788835113850

In a second part of this work we illustrate the main socio-demographic characteristics of students who suffer bullying by their peers, first of all with simple descriptive statistics and then by means of a multiple correspondences analysis that allows to trace a complete profile of the victim by summarizing all information given by different variables.

Finally, the last part of our study consists of a multiple regression analysis through which it was possible to quantify the impact of bullying (as independent variable) on academic performance (as dependent variable) while controlling for all major confounding variables: basic academic skills, gender, migration background and socioeconomic status. In particular, we estimated separate linear regression models in which the outcome variable is represented by the score obtained in the grade 5 and 8 tests. For grade 8 only we also estimate logistic regression models in which the outcome is the variable "levels of competencies" attributed to students depending on their results of the grade 8 tests. More specifically, in linear regression models we included as outcome variable the WLE (Weighted Likelihood Estimation) score estimated with the Rasch method, standardized to 200[3]. In logistic regression models, we have included as outcome a dichotomous variable that distinguishes the two lowest levels from all the others[4].

To identify victims of bullying we used the answers given by students to the following four questions: 1) Have you been teased by other students? 2) Have you been insulted by other students? 3) Have you been isolated or excluded by other students? 4) Have you been beaten up by other students? Each of these questions has four possible answers: never, every once in a while, every week, every day. Thus, it is also possible to distinguish victims according to how frequently they have been bullied.

We included in regression models the following variables as covariates: gender, migration background (Italian, first generation foreigner, second generation foreigner), geographical area of residence, socioeconomic background (indicator of socioeconomic status, called ESCS[5] from here on) and

[3] In this regard, it is important to specify that given the structure of CBT – which include questions with the same difficulty level but different for each student and therefore make classroom collaboration between students difficult – it is not necessary to apply to the grade score 8 2017/2018 any correction for cheating. This correction is instead applied to the 2014/2015 grade 5 WLE score and to the 2011/2012 grade 2 WLE score.

[4] As competence levels have been calculated for the first time this year, and they are not available for grade 5 of 2014/2015 school year (INVALSI, 2018).

[5] ESCS index is calculated based on three components: parents' occupational status (HISEI), parents' educational status (PARED) and home possessions of specific resources (HOMEPOS), combined by Principal Components Analysis.

ISBN 9788835113850

regularity of the course of study (anticipating, regular and repeating students). Finally, we chose to describe basic skills in terms of the score obtained by students in INVALSI grade 2 test.

4. Results

4.1. Prevalence and geographical distribution of bullying in Italy

The dataset of the questionnaire administered to students of 5th grade in the 2014/2015 academic year contains the answers of 408,301 children, a figure that does not include students with a SIDI code marked as "not available" or duplicate. The latter, in fact, were previously eliminated from the dataset as they lacked the information necessary to perform the record-linkage with the matrixes containing the scores obtained in 8th and 2nd grade tests by the same student in different school years.

From the original population of 408,301 children of 5th grade 2014/2015 we selected a subsample of 239,587 and 253,551 students for whom it was possible the complete record-linkage of 2nd grade (2011/2012), 5th grade (2014/2015) and 8th grade (2017/2018), for Italian and Mathematics respectively.

For each question that identifies the victims of the four types of bullying investigated in the student questionnaire, the percentage of invalid or missing answers is very low: its value fluctuates between 1.0% and 1.4%. Excluding these answers, it emerges that the most frequent act of bullying is being teased by classmates (Table 1): 8.3% and 7.6% of children declare to experience this behaviour every week and every day during the school year, respectively. Then, verbal insults represent the second most frequent act of bullying (respectively with 6.2% and 5.1%), followed by social isolation (with 4.6% and 3.9% respectively) and physical violence (with 2.0% and 1.6% respectively). These percentages give an idea of how many children in Italy constantly suffer bullying, and therefore are more exposed to its heavy emotional (and not only) consequences. However, it is important to specify that the percentage of children who are even occasionally bullied is fairly high. Considering the most serious case, as many as 17.0% of children who answered the 2014/2015 student questionnaire were beaten at least once by other students. These children are also considered to be at risk.

ISBN 9788835113850

Tab. 1 – Percentage of children enrolled in 5th grade classes by frequency with which they claim to have been teased, insulted, isolated and beaten during the 2014/2015 academic year. Data referable to the entire population*

	During this school year, how often have you been			
	Teased	*Insulted*	*Isolated*	*Beaten*
Never	29.04	46.38	52.36	79.19
Every once in a while	55.09	42.26	39.08	17.17
Every week	8.25	6.23	4.64	2.01
Every day	7.62	5.13	3.91	1.63

* Missing and invalid answers are excluded from the calculation of the percentages.

Source: our processing of INVALSI data

For what concerns the spatial distribution of the phenomenon in different areas of the Italian territory, already from a preliminary descriptive analysis differences emerge between the geographical macro-areas (data not shown). Particularly, a territorial gradient appears according to the severity of bullying episodes: the percentage of children who are most frequently teased is higher in the northern areas in respect to the south and to the islands, but the situation gradually changes to overturn when taking into account the percentages of children being beaten. The latter are in fact higher in the south rather than in the north.

These results are confirmed by the four geographical maps shown in Figure 1, showing the differences between the prevalence of bullying in each Italian province and the average prevalence of bullying at national level. In this case, prevalence includes the forms of bullying suffered regularly (i.e., weekly or daily). From this geographically detailed analysis, it is possible to observe that the distribution of the phenomenon on the Italian territory varies according to the form of bullying declared to be suffered by the victims. Frequencies of direct verbal behaviour by schoolmates, such as being teased and insulted, seem to have a rather similar pattern of territorial distribution, and the same is notable also for indirect behaviours that can somehow lead to exclusion or isolation. The abovementioned forms of bullying occur more frequently among students in the north of the country and increase slightly in the provinces of Emilia Romagna region, in the provinces of south-western Italy, in Northern Sardinia, and in some provinces of eastern Sicily. The areas with the lowest values are instead located between the provinces of central Italy. As for the most severe forms of bullying, such as physical violence, these seem to be more frequent in the north-eastern areas of the country. However, there are also numerous provinces in the centre, south and islands

ISBN 9788835113850

having values above the national average. Physical violence, on the contrary, seems rather rare among students in North-Western Italy. In conclusion, while the North-East has the highest proportions of victims of all the four forms of bullying, in comparison to the rest of the country, in the South clearly prevail more violent, physical actions.

4.2. The profiles of victims of bullying in the Italian primary school

The results of the descriptive analysis shed light on what are the main characteristics of the bullied victims. In addition to the differences by geographical area described in the previous paragraph, statistically significant differences – tested with the Chi-squared test – also emerge by gender, citizenship, regularity of studies and socio-economic background among those who claim to being bullied every week or every day during the academic year (Figure 1). First of all, regardless of the type of bullying considered, the percentage of students who frequently suffer bullying is significantly higher among males rather than among females. Moreover, our results confirm what has already been found for Italy by Alivernini *et al.* (2017), i.e. that first generation foreign students are teased, insulted, isolated and beaten more often than Italians. The same can be said for second-generation foreign students, although their differences with Italians are not so large. The percentages according to regularity of studies show that repeating students are the most exposed to the risk of bullying. For example, postponed students declare more frequently to being isolated if compared with regular students (6.3% against 4.2%, and 5.7% against 3.9% every week and every day, respectively) and beaten (3.0% against 2.0%, and 2.8% against 1.6% each week and every day, respectively) during the reference school year. Finally, by dividing students into four groups based on the quartiles of the distribution of the ESCS, it is also possible to see that the differences are visible mainly among students who claim to be bullied every day. For instance, the percentage of students who claim to be teased every day by other students gradually decreases from 9.4% in the lowest socioeconomic status class to 6.2% in the highest socioeconomic status class. These differences are much smaller among students who not suffer bullying so frequently, independently on the form of bullying suffered.

ISBN 9788835113850

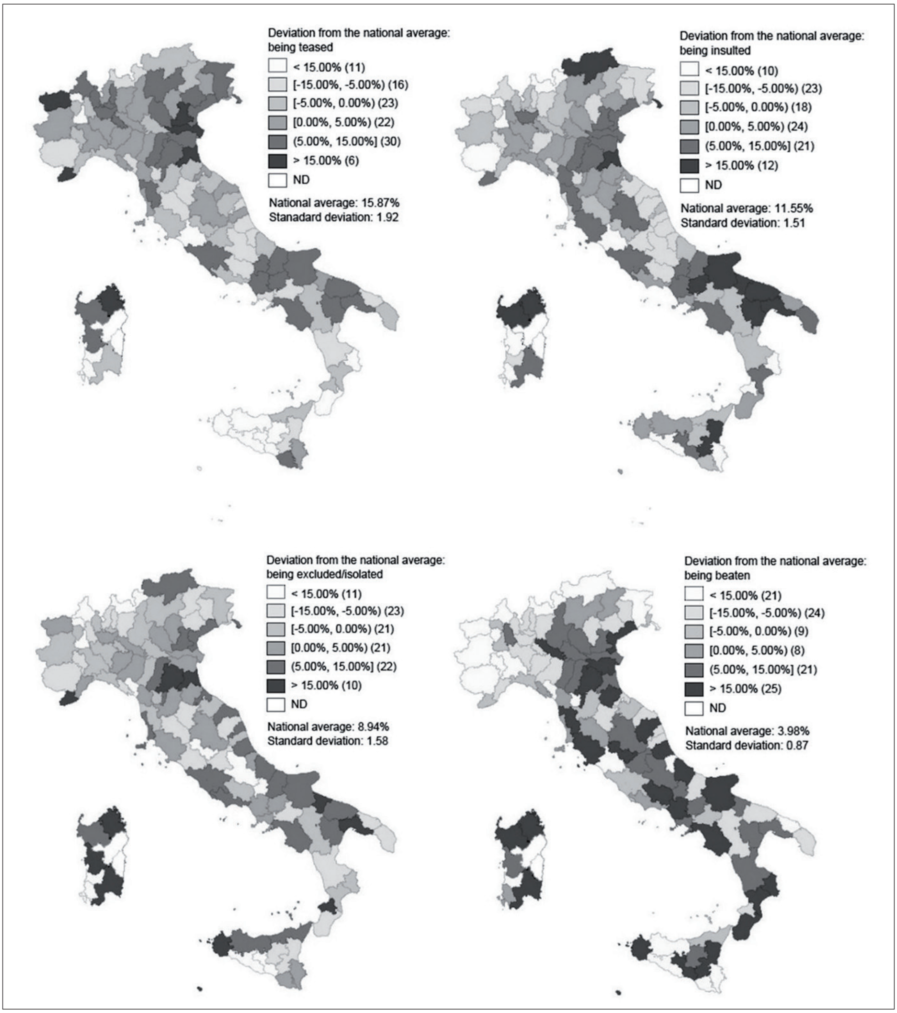

Fig. 1 – Percentage of children enrolled in 5th grade classes in the Italian provinces who claim to have been teased, insulted, isolated/excluded or beaten during the 2014/2015 academic year on a weekly and daily basis. Data referable to the entire population

N.B. Data of the student questionnaire for the provinces of Nuoro and Ogliastra are not available for the 2014/2015 academic year.

Source: our calculations based on INVALSI data

The results of the descriptive analysis investigating the differences in academic performance between students who suffered bullying more or less frequently in the 2014/2015 school year reveal – for all school grades in-

ISBN 9788835113850

cluded in the dataset – the existence of a gradient with scores obtained in the INVALSI tests always higher among those who have never experienced bullying and decreasing with the frequency of such episodes until they reach a minimum value among those students who have declared to be bullied every day. The differences observed in WLE scores, both for Italian and Mathematics, are generally around 10 points (data not shown). However, the differences in scores are also evident by gender, citizenship, regularity and ESCS (Index of Economic, Social and Cultural Status) classes (data not shown), confirming the rich literature on the subject and the need to include these variables in subsequent analyses (Dustmann *et al.*, 2012; Tomul and Savasci, 2012; Agasisti and Longobardi, 2016).

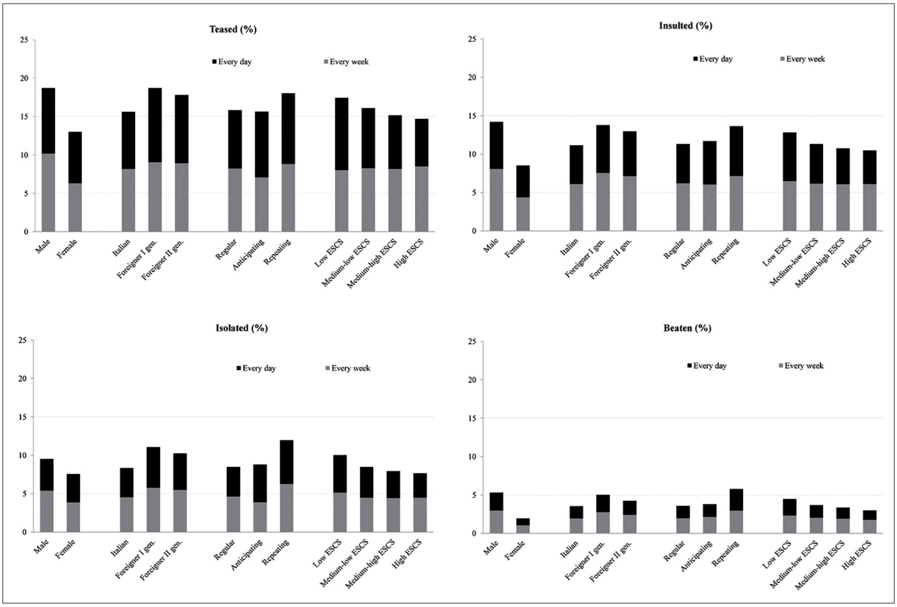

Fig. 2 – Percentage of children enrolled in 5th grade classes who claim to being teased, insulted, isolated and beaten during the 2014/2015 academic year on a weekly and daily basis. Data referable to the entire population*

* Missing and invalid answers are excluded from the calculation of the percentages.

Source: our calculations based on INVALSI data

An overview of the profile of bullied students (including their academic performances) is possible using multiple correspondences analysis, which allows us to observe in a single graphical representation the main patterns of the association between all the study variables.

ISBN 9788835113850

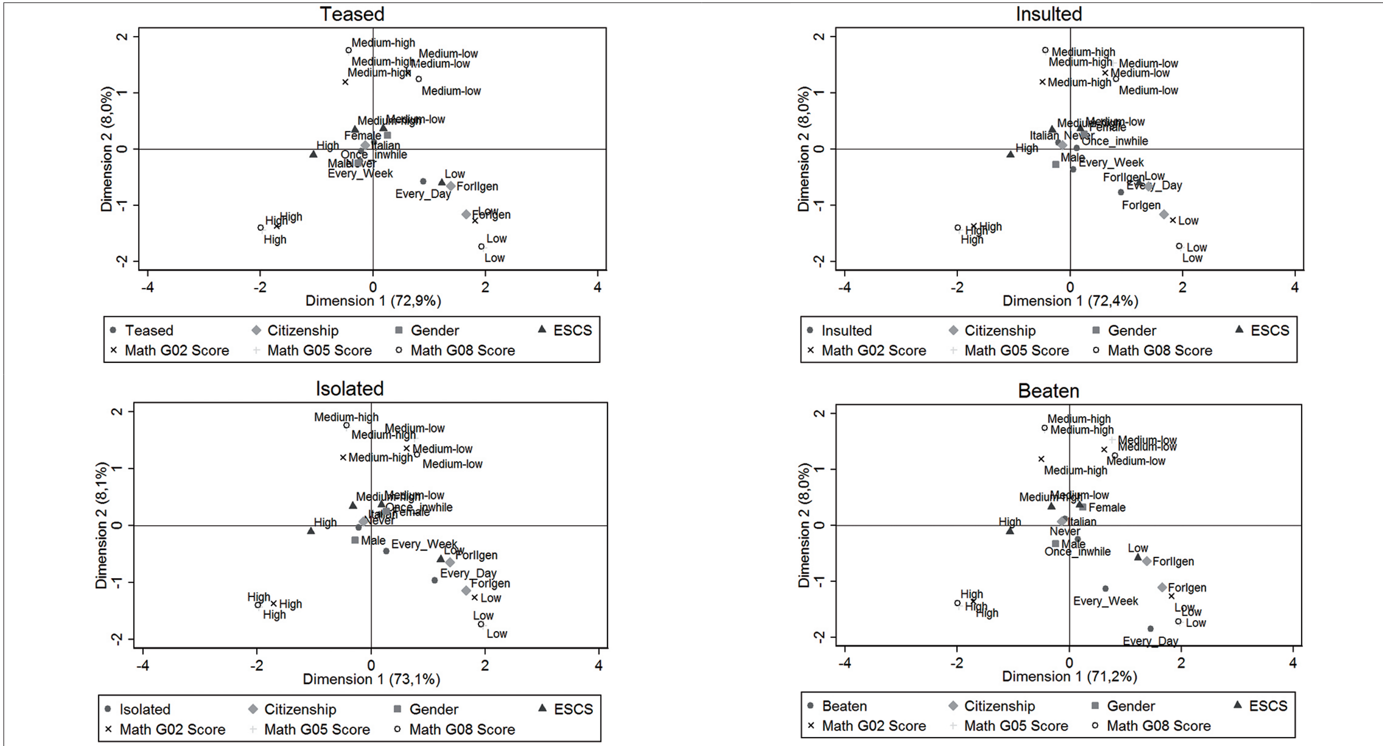

Fig. 3 – Multiple correspondences analysis: profiles of victims of bullying among children enrolled in 5th grade classes in the 2014/2015 academic year, depending on the type of acts suffered (teasing, insults, isolation, physical violence). Data referable to the entire population

Source: our calculations based on INVALSI data

ISBN 9788835113850

In Figure 3, we can observe highly overlapping clusters for all four manifestation of bullying: situations in which a victim is bullied with a greater frequency (daily and weekly), are permanently associated a disadvantageous socio-economic situation, being a foreigner and – although to a lesser extent – being a male, and they are always enclosed in the same quadrant. Academic performance is also highly associated with bullying, with victims who are exposed more frequently to violent behaviours characterized by the lowest achievement in all school grades. The associations are very similar between the four outcome variables related to bullying, however it is evident that for the "physical violence" there is a greater polarization on the vertical axis of the categories referred to the frequency with which these actions are subjected, and an association even more pronounced with the "low" performance in the short and medium term.

Tables 2 and 3 report the results of the regression models estimating the impact of different forms of bullying on the academic achievement in the short (5th grade) and medium-term (8th grade), i.e. in the same year and three years after episodes of violence suffered. In all models the 2nd grade score of students is also included, as a control variable for student's basic skills. Results are highly comparable for Italian and Mathematics tests, thus, only those related to Mathematics are shown. This allows to have more robust estimates for foreigners, whose knowledge of the Italian language could be lower.

The negative association between bullying and academic performance, assessed controlling for basic competence of students and all socio-demographic variables included in the models (ESCS, gender, migration status and regularity of studies) proves that school results worsen with increasing frequency of the acts of violence suffered. This is the case for all the types of bullying episodes and for both time-based perspectives (short and medium term). Nevertheless, the effects have different intensities related to how much time had passed since the events and on the forms of bullying experienced. In general, the drop in academic performance associated to bullying is lower in 8th grade than in 5th grade. Medium-term effects are always about 2 points lower than the short-term ones, for all types of bullying considered. However, even among the same types of bullying, there is a variability of effects on performance: if we observe the effects for those who suffer the different types of bullying "every day", the smallest reduction in score occurs for "being teased" (-5.8 WLE points, for 8th grade), while the largest reduction is observed for "being beaten" (-13.4 WLE points, for 5th grade).

ISBN 9788835113850

Tab. 2 – Association between bullying and academic performance expressed in WLE 200 scores obtained by students in the Mathematics test: coefficients estimated by linear regression models for the four types of bullying considered in the analysis (being teased, being insulted, isolation, physical violence)

Outcome: Math G05 Score				
	Teased	*Insulted*	*Isolated*	*Beaten*
	Coef.	*Coef.*	*Coef.*	*Coef.*
Never (ref.)	0	0	0	0
Every once in a while	-1.39***	-2.48***	-1.50***	-3.14***
Every week	-0.69**	-3.14***	-2.94***	-7.89***
Every day	-7.33***	-8.37***	-9.05***	-13.43***
Outcome: Math G08 Score				
	Teased	*Insulted*	*Isolated*	*Beaten*
	Coef.	*Coef.*	*Coef.*	*Coef.*
Never (ref.)	0	0	0	0
Every once in a while	-0.18n.s.	-2.13***	-0.02n.s.	-1.83***
Every week	1.85***	-1.70***	-0.63n.s.	-5.84***
Every day	-5.76***	-7.49***	-7.19***	-11.69***

*** p-value < 0.001, ** p-value < 0.005, * p-value < 0.05, n.s. not significant

Note: two separate models have been estimated for each type of bullying, one with WLE 200 Math G05 Score and one with WLE 200 Math G05 Score as outcome variable. All models are adjusted for gender, citizenship, regularity of studies, ESCS and Math G02 Score.

Source: our calculations based on INVALSI data

The same type of relationship can be observed from a complementary perspective, considering as a study variable not the numerical WLE score, but rather Learning levels, obtained by recoding quantitative WLE scores into 5 categories corresponding to different levels of competencies. The experience of being bullied, especially with a daily frequency, seems to increase the risk of being in the lowest levels of performance (Level 1 and Level 2), varying from +38% for teasing to +96% for physical violence. It should be noted, though, that the gradient associated with the frequency with which the events take place shows a fluctuating trend for "being teased", while it confirms a decreasing trend for all the other forms of violence suffered.

ISBN 9788835113850

Tab. 3 – Association between bullying and academic performance expressed in Mathematics learning levels: Odds Ratios (OR) of being in the lowest levels of performance (Level 1 and 2) estimated by logistic regression models for the four types of bullying considered in the analysis (being teased, being insulted, isolation, physical violence)

	Outcome: Math G08 Competence Levels			
	Teased	*Insulted*	*Isolated*	*Beaten*
	Coef.	*Coef.*	*Coef.*	*Coef.*
Never (ref.)	1	1	1	1
Every once in a while	0.99n.s.	1.12***	0.99n.s.	1.10***
Every week	0.91***	1.10***	1.06*	1.44***
Every day	1.38***	1.56***	1.56***	1.96***

Note: one model has been estimated for each type of bullying. All models are adjusted for gender, citizenship, regularity of studies, ESCS and Math G02 Score.

*** p-value < 0.001, ** p-value < 0.005, * p-value < 0.05, n.s. not significant

Source: our calculations based on INVALSI data

5. Discussion and conclusions

Results presented in this study offer important insights about bullying and arise some issues that would be interesting to further investigate. First of all, they show that in Italy bullying is a widespread phenomenon already in primary school: most of the students who attend 5th grade declare to suffer physical or verbal violence by their schoolmates, and an important proportion of them claim to be bullied on a weekly or even daily basis. We also demonstrate the existence of a different distribution of different types of bullying in the national territory. Closely related are the differences in prevalence, that produce significant variation in terms of effects of bullying on the academic achievement. Being teased, for example, seems to have a negative impact on academic performance only when it occurs with a daily frequency, whereas, probably due to its being a "mild" form of bullying, it could be sufficiently tolerated and/or contrasted when it occurs only sporadically. On the other hand, the most serious forms of bullying (insults, isolation and physical violence) have a significant association with the academic performance even when they occur only occasionally. The results of our analysis also confirm the fundamental characteristics of the bullied victim. First and second generation students and children of low socio-economic status are particularly vulnerable to bullying.

ISBN 9788835113850

Another important result is that bullying victims generally obtain lower scores in standardized tests, not only at the end of the academic year or when the phenomenon is suffered but also three years later, with a gradient often proportional to the frequency of the episodes. This would actually support the hypothesis that suffering frequently acts of bullying may adversely affect the academic performance, having prolonged effect also over time. However, the relationship between the two phenomena (bullying and performance) is certainly complex: the fact that in the correspondence analysis 2nd grade score (prior to the detection of having suffered bullying) is also highly correlated with bullying and with all the "typical" characteristics of bullying victims, suggests the possibility that poor performance is primarily associated with a social, economic "condition of weakness" that is also the "fertile ground" for bullying. For this reason, further analyses are needed to better clarify the complex relationships between bullying, social disadvantage and academic performance.

References

Agasisti T., Longobardi S. (2016), "Come gli studenti delle regioni italiane superano il gap socio-economico: evidenze da PISA 2012", in L. Palmerio (a cura di), *OCSE PISA 2012. Contributi di approfondimento*, FrancoAngeli, Milano, pp. 9-27.

Alivernini F., Manganelli S., Cavicchiolo E., Lucidi F. (2019), "Measuring bullying and victimization among immigrant and native primary school students: evidence from Italy", *Journal of Psychoeducational assessment*, 37 (2), pp. 226-238.

Barker E.D., Boivin M., Brendgren M., Fontaine N., Arseneault L., Vitato F., Bissonnette C., Trembley R.E. (2008), "Predictive validity and early predictors of peer-victimization trajectories in preschool", *Archives of General Psychiatry*, 65, pp. 1185-1192.

Beran T.N., Hughes G., Lupert J. (2008), "A model of achievement and bullying: analysis of the Canadian National Longitudinal Survey of Children and Youth Data", *Educational Research*, 50 (1), pp. 25-39.

Brown S., Taylor K. (2007), "Bullying, education and earnings: evidence from the National Child Development Study", *Economics and Education Review*, 27, pp. 387-401.

Dustmann C., Frattini T., Lanzara G. (2012), "Educational achievement of second-generation immigrants: an international comparison", *Economic Policy*, 27 (69), pp. 143-185.

Genta M.L., Menesini E., Fonzi A., Costabile A., Smith P.K. (1996), "Bullies and victims in schools in Central and Northern Italy", *European Journal of Psychology of Education*, 11 (1), pp. 97-110.

ISBN 9788835113850

Gini G. (2004), "Bullying in Italian Schools. An overview of intervention programs", *School Psychology International*, 25 (1), pp. 106-116.
INVALSI (2018), INVALSI Rapporto nazionale 2018.
Lacey A., Cornell D. (2013), "The Impact of Teasing and Bullying on Schoolwide Academic Performance", *Journal of Applied School Psychology*, 29, 3, pp. 262-283.
Ponzo M. (2013), "Does bullying reduce educational achievement? An evaluation using matching estimators ", *Journal of Policy Modeling*, 35 (6), pp. 1057-78.
Oliveira F.P., Almeida de Menezes T., Irffi G. Resende Oliveira G. (2018), "Bullying effect on student's performance", *Economia*, 19, pp. 57-73.
Smith P., Talameli L., Cowie H., Naylor P., Chauhan P. (2004), "Profile of non-victims, escaped victims and new victims of school bullying", *British Journal of Psychological Society*, 74, pp. 565-81.
Tomul E., Savasci H.S. (2012), "Socioeconomic determinants of academic achievement", *Educational Assessment, Evaluation and Accountability*, 24 (3), pp.175-187.
Woods S., Wolke D. (2004), "Direct and relational bullying among primary school children and academic achievement", *Journal of School Psychology*, 42, pp. 135-55.

ISBN 9788835113850

4. A comparison of regression tree-based features selection methods for the prediction of academic performances

by Lorenzo Mancini, Chiara Sacco

Students' academic achievements are the result of the influence of several different factors: socio-economic, socio-emotional and environmental factors as student's own characteristics, the characteristics of their family, the network of their social relationships as well as the characteristics of the schools, the teachers or the class. One of the main research topics in the educational field is the identification of the factors mostly influencing the student academic performance.

In recent years, the introduction of automated methods of data collection such as the computer based test has made available a large amount of data but, usually, only a limited number of variables are considered in the prediction models, selected on the base of theoretical knowledge and literature review. Modelling the relationships between large set of variables can be cumbersome in classical statistical methods, which have to face with critical issue as overfitting and multicollinearity. Variable selection methods allow to overcome these problems by removing all the redundant information from the model, thus obtaining an easier model to interpret. These methods result both into models with better performance and less biased estimates.

The aim of this study is to compare two tree-based variable selection methods to identify the most relevant predictors of 8th grade students' performances at INVALSI test in Italian language and to rank the selected variables accordingly with their importance for prediction. This approach has the benefit to account all the variables in one model simultaneously, allowing to retain all the possible predictors. The analysis of the selected variables and their importance rank give new insights in understanding the mechanism underlying the student's academic performance.

ISBN 9788835113850

Il successo accademico degli studenti è spesso il risultato dell'influenza di diversi fattori concomitanti. Ne sono un esempio i fattori socio-economici, socio-emotivi e ambientali, così come le caratteristiche proprie dello studente, della famiglia, della rete delle loro relazioni sociali, nonché le caratteristiche delle scuole, degli insegnanti o della classe. Uno dei principali topic di ricerca in campo educativo è l'identificazione dei fattori che, più degli altri, influenzano il rendimento scolastico degli studenti.

Negli ultimi anni si è resa disponibile una mole di dati sempre maggiore, anche grazie all'introduzione di metodi di raccolta automatizzati, come, per esempio, le prove computer-based*. Nonostante la grande disponibilità di dati, solitamente, nei modelli di predizione, sono considerate solo un numero limitato di variabili, selezionate sulla base di conoscenze teoriche pregresse o dell'analisi della letteratura esistente. Infatti, i metodi statistici classici spesso incontrano difficoltà nel trattare un numero elevato di variabili in quanto è probabile incorrere in problemi come* overfitting *e multicollineareità. I metodi di variable selection consentono di superare questi problemi, rimuovendo tutte le variabili ridondanti dal modello, e ottenendo al contempo un modello di più semplice interpretazione. Questi metodi, inoltre, consentono di ottenere modelli con performance migliori e stime meno distorte.*

Lo scopo di questo studio è quello di confrontare due metodi di variable selection tree-based*, con l'obiettivo di identificare i predittori più rilevanti dei risultati al test INVALSI di Italiano degli studenti all'ultimo anno della scuola secondaria di primo grado e, contestualmente, di individuare l'ordine di importanza rispetto alla predizione delle variabili selezionate. Questo approccio ha il vantaggio di consentire l'inclusione nel modello di tutte le variabili simultaneamente. Lo studio delle variabili selezionate dai due metodi, e del relativo grado di importanza per la predizione, permette una comprensione più profonda dei fattori determinanti il successo accademico degli studenti.*

1. Introduction

The prediction accuracy of academic performances is one of the most challenging research topics and, usually, the main task consists in identifying the factors which most influence the learning process. It is well known, in the educational field, that student's academic achievements are influenced by several socio-economic, socio-emotional and environmental factors as student's own characteristic, the characteristic of their family, the network of their social relationships as well as the characteristic of the schools, the

ISBN 9788835113850

teachers or the class. It turns out that the educational system is a very complex phenomenon as it includes a huge amount of concurrent factors, which influence academic performances. Various studies over the years have attempted to describe this process from different points of view (Gallina, 2006; Passow *et al.*, 1976).

Despite the current availability of a very large amount of data on student, teacher and school characteristics, the classical statistical methods are hindered by several difficulties while handling high dimensional data. This could be one of the reasons why usually prediction models considered a limited number of variables, selected according to theoretical knowledge and literature review. Indeed, in the context of classical statistical models, dealing with high dimensional data could lead to crucial issues such as overfitting, non-convergence and multicollinearity.

In the last decades, several feature selection methods (Miao and Niu, 2016; Chandrashekar and Sahin, 2014) have been proposed to overcome the issues of modelling a large amount of variables by reducing the data dimensionality and identifying the relevant variables. These methods have become widespread because they allow to obtain an easier model to interpret, free from redundant information. Variable selection allows to identify a subset of relevant predictors and, at the same time, to remove from the model all the irrelevant variables for the prediction of the outcome. Excluding the redundant and noisy variables can improve the performance of the model, avoiding possible bias in the estimates (Chandrashekar and Sahin, 2014) and result into faster computational times.

In the educational field, several works applied variables selection methods to improve the accuracy of prediction models for student performances (see, among others, Acharya and Sinha, 2014; Ramaswami and Bhaskaran, 2009; Cortez and Silva, 2008). However, as far as we know, few studies have focused on the Italian educational system with the aim of identifying among the various student's and school's characteristics, those who are mainly associated with student's academic achievements.

In particular, this study focused on identifying the relevant predictors of the students' performance at INVALSI test in Italian language. We compared the performance of regression trees (Breiman *et al.*, 1984) and multilevel regression trees (Sela and Simonoff, 2012) to evaluate which one maximizes the prediction accuracy of the students' performance.

The main advantage of tree-based methods is that the algorithms are not based on strong assumptions about the functional form that describes the relation between the outcome and the covariates. Not making any assumptions leads to a model that is free to learn the functional form from the data

ISBN 9788835113850

and makes nonparametric learning methods a more flexible tool, which is able to analyse and represent the complexity of the studied phenomenon (Hollander *et al.*, 2013). One issue of the educational data is represented by the hierarchical structure, in which students (level 1) are nested within schools (level 2). For tree-based methods, the impact of multilevel data structure on the performance is controversial. In particular, some studies have highlighted that for regression trees the differences in prediction accuracy can be negligible whether data are treated as multilevel or single level only (Fu and Simonoff, 2015). Nevertheless, it appears that multilevel data can deleteriously affect the computation of the variable importance (Martin and Von Oertzen, 2015; Loh and Shih, 1997). For this reason, regression trees should be compared with an alternative method that considers the data structure. The machine learning techniques used in this work take into account all the variables in one model simultaneously, allowing to retain all the possible predictors. Although many predictors could have a significant impact on student's performance, not all of them contribute in the same way to explain variability in student's results. A further advantage of tree-based methods is that they allow to rank the selected variables accordingly with their importance for prediction. For each variable, the corresponding relative importance index is evaluated by the model, which automatically excludes the variables with null importance, i.e. the variables not useful for prediction. Thus, tree-based methods allow to perform, at the same time, variable selection and ranking.

These characteristics make tree-based methods a valid alternative to traditional regression models. Regression models are not able to identifying automatically the best subset among many variables and usually require additional steps for variable selection as stepwise selection (Efroymson, 1966; Draper and Smith, 1966) or the best subset selection method (Beale *et al.*, 1967; Hocking and Leslie, 1967). However, in presence of the dependence structure introduced by multilevel data, the use of more widespread parametric variable selection methods as stepwise selection is strongly discouraged (Pinheiro and Bates, 2000), whereas is recommended the use of information-theoretic tools to select the model with the best subset of predictors (Burnham and Anderson, 2002). This approach selects the model with the highest predictive power estimating a penalty term to account for the model complexity (Vaida and Blanchard, 2005). On the other hand, this approach shows several weaknesses with respect to tree-based methods: first of all, it does not return a rank of the variables in function of their importance for prediction; secondly, it could be computationally demanding since the number of the possible models increases with the number of predictors.

ISBN 9788835113850

In this work, we focused on the comparison of the performances of two tree-based selection methods, namely, Regression Tree (RT) and Random Effects Expectation Minimization Recursive Partitioning (RE-EM) tree, in order to identify which of them improves the accuracy of the prediction of the students' performances and which is the ranking of the selected variables according with their importance for prediction. In addition, we compared the prediction accuracy of the tree-based methods with the standard linear mixed effect model, assumed as reference model.

The variables selected through tree-based methods could shed light on the existing theories from a different prospective or could give a deeper insight into the mechanism underlying the student's academic performances.

2. Data

The Italian National Institute for the Evaluation of the Education and Training Educational System (INVALSI), annually, carries out standardized tests to assess the performance of all Italian students at the end of the second and the fifth years of primary school, at the end of lower secondary school, and at the end of the second year of higher secondary school. This study exploits the standardized test administered by the INVALSI for the school year 2017/2018 focusing on the students in the 3^{rd} grade of the lower secondary school. INVALSI required to the students to compile a questionnaire after the standardized tests in Mathematics and Italian. The student questionnaire collected information about home background, including parent's country birth, parent's occupational status and educational qualification, language and dialect spoken at home and home resources. In addition, the student questionnaire contains more than 50 multiple-choice questions about student's anxiety during the test, student's motivation and interest in study, student's view about the school life (perception of school environment and relationships with peers), parent's support, student's self-efficacy and student's future expectations.

In our analysis we focused on the INVALSI representative sample, composed by 29,568 students of 940 Italian secondary schools that participated to the INVALSI test to assess achievement in Italian language. All the exploratory variables included in the variable selection models are illustrated in the Tab. 1.

ISBN 9788835113850

Tab. 1 – Variables included in the model

Variable name	*Variable description*
Score INVALSI 14/15	Student score at INVALSI test in Italian language in s.y. 2014/2015
Oral score	Student oral exam score in Italian language attributed by the teacher
ESCS	Student economic, social and cultural status indicator
Kindergarten	Student attendance at kindergarten (1 = “yes”, 0 = “no”)
Gender	Student gender (1 = “female”, 0 = “male”)
Foreign 1st gen	First generation foreign student (1 = “yes”, 0 = “no”)
Foreign 2nd gen	Second generation foreign student (1 = “yes”, 0 = “no”)
Retaining student	Student retaining (1 = “yes”, 0 = “no”)
Int_Foreign1st_Retainig	Interaction term between first generation foreign student and retaining student
ESCS School	School mean economic, social and cultural status
School score INVALSI 14/15	School mean score at INVALSI test in Italian language in a.y. 2014/2015
Percentage 1st gen foreign	School percentage of first generation foreign student
Percentage of retaining students	School percentage of students not attending academic year as provided from scholastic program
School dimension	School total number of students
School missing %	School percentage of mean missing answer to INVALSI test
Language	Student language (1= “No Italian”, 0 = “Italian”)
Dialect	Speak regularly dialect at home (1= “yes”, 0 = “no”)
Qualification Expectation	Student expected educational qualification
Q01_ITA	Indicator of student anxiety during INVALSI test in Italian Language
Q02_MAT	Indicator of student anxiety during INVALSI test in Math
Q06_ITA	Indicator of student home resources
Q10_ITA	Indicator of student interest in studying Italian language
Q10_MAT	Indicator of student scholastic experience
Q11_MAT	Indicator of parents sensitivity and support
Q12_MAT	Indicator of student self-efficacy
Q13_MAT	Indicator of student relationships with peers
Q14_MAT	Indicator of student’s future expectations
Q15_MAT	Indicator of student’s interest in studying Math

Note: Suffix “Q” identifies questions from Student’s INVALSI questionnaire. The items associated to each question have been reported on a continuous scale through Graded Response Model (Samejima, 1969).

ISBN 9788835113850

3. Methods

3.1. Tree-based Methods

To identify a prediction model for student's achievement out of several INVALSI student variables we exploited two tree-based methods: Regression Tree (RT) and RE-EM tree (Sela and Simonoff, 2012; Hajjem *et al.*, 2011). Classification and Regression Trees (CART) is a machine learning technique introduced by Breiman (Breiman *et al.*, 1984), which consists of a set of rules used for prediction or classification. The main idea behind these methods is to partition the sample space recursively into sub-groups, which are smaller in number and more homogeneous at each iteration of the algorithm.

In this study, a pruning strategy, based on cross-validation analysis of the complexity parameter (cp), has been applied to avoid problem of overfitting that might lead to poor performance of the model. The optimal value of cp found by 10-fold cross-validation has then been used to prune the tree and the resultant model has been used to verify the prediction performance and to identify the selected variables.

RE-EM tree is an extension of CART method for multilevel data, including subject-specific random effects in the tree structure. The inclusion of the subject-specific random effects is based on the idea of estimating the random effects iteratively, removing them from the response and then computing the regression tree.

3.2. Methods' evaluation

As in this study nonparametric methods are considered, it is not possible to compare their performances with classical statistical indicators as the Akaike Infromation Criterion (Akaike, 1974) or the Bayesian Information Criterion (Schwarz, 1978). For this reason, to compare the models in terms of goodness of fit, we considered the prediction accuracy and model complexity (Sanchez-Pinto *et al.*, 2018; Lim *et al.*, 2000). Prediction accuracy has been measured using mean square error (MSE), i.e. the sum of the differences between the predicted and the observed values, whereas the model complexity has been evaluated in terms of number of variables included in the model. The MSE of the two methods has been subsequently compared with the MSE obtained with a standard Linear Mixed Model (LME). Indeed, this model is a widespread parametric solution for modeling the relationship between variables when data show a multilevel structure (Pinheiro and Bates, 2000; Wolfinger and O'Connell, 1993).

ISBN 9788835113850

The variable importance index, measured by the two tree-based methods, has been categorized basing on the quintiles for an immediate interpretation of the ranking of the variables. The variable importance for both methods corresponds to the sum of goodness of fit increment whenever the variable is selected for a split.

We have implemented all the analyses using R software (version 3.6). In particular, we used the *rpart* package (Therneau and Atkinson, 2018) for RT and the *REEMtree* package (Sela and Simonoff, 2011) for RE-EM. For the estimation of the LME we used the *nlme* package (Pinheiro *et al.*, 2020).

4. Results

RE-EM outperformed the RT in terms of prediction accuracy, with a lower MSE value (577.25 for RE-EM and 632.64 for RT), and in terms of model parsimony, with a total number of predictors equal to 21 (24 predictors selected by RT). It is interesting to notice that taking into account the multilevel structure results to be an advantage for regression tree methods. Indeed, the RE-EM method has achieved better performance with a smaller number of variables with respect to RT.

The MSE of the standard LME model was equal to 589.43, thus, the model reached a better performance on the data with respect to RT, although it does not perform variable selection and the estimates were computed on the whole set of 28 variables. The prediction accuracy of the LME has been outperformed by RE-EM, which resulted to be a better alternative.

Tab. 2 presents the selected variables and the corresponding importance quintiles for each selection method. The color gradation denotes the quintiles of variable importance, with grayest cells representing the most important variables and white cells representing the not selected variables. It can be notice that the most important variables have been identified in accordance by both methods. In particular, student's previous scores at INVALSI test in Italian language (*Score INVALSI 14/15*), student's oral exam score in Italian language (*Oral score*), student's qualification expectation and student's self-efficacy (*Q12_MAT*) were ranked at 5th quintile by both the methods. Despite student's qualification expectation is selected as an important factor in determining students' academic success, the student's future expectations (*Q14_MAT*) is ranked at 2nd quintile by RT and 3rd quintile by RE-EM. The students' socio-economic background (ESCS) is ranked at the 3rd quintile by both algorithms. According with both methods, the students' score at INVALSI test 2014-2015, the anxiety during the test (*Q01_ITA*), the interest in

ISBN 9788835113850

the study of the Italian language (*Q10_ITA*) and the student's perception of the scholastic experience (*Q10_MAT*) showed a greater importance than the student's socio-economic background. The other variables ranked at the 3rd quintile by the RE-EM are the student's interest in studying Math (*Q15_MAT*) and the anxiety during INVALSI Math test (*Q02_MAT*). The RT ranked at the 3rd quintile the indicator of student anxiety during INVALSI Math test (*Q02_MAT*), the home resources (*Q06_ITA*) and the school dimension.

Tab. 2 – Heat map of variables selected and their importance quintile for each method

	Rank	
	RE-EM	*RT*
Score INVALSI 14/15		
Oral score		
Qualification Expectation		
Q12_MAT		
Q01_ITA		
Q10_ITA		
Q10_MAT		
School score INVALSI 14/15		
ESCS		
Q02_MAT		
Q14_MAT		
Q15_MAT		
School dimension		
School ESCS		
Int_Foreign1st_Retainig		
Percentage 1st gen foreign		
Percentage of retaining students		
Retaining student		
Q06_ITA		
Q11_MAT		
Q13_MAT		
Dialect		
Gender		
Kindergarten		
Language		
School missing %		
Foreign 2nd gen		
Foreign 1st gen		

Source: our elaboration on INVALSI data, grade 8th, 2017/2018

ISBN 9788835113850

Both RT and RE-EM agree in excluding the following variables: the use of dialect at home, the language spoken at home, the gender and the indicator of first generation immigrant. Moreover, RE-EM excludes the student attendance at kindergarten, the indicator of second generation immigrant, ranked at the 1st quintile by RT, and the mean school percentage of missing answer at INVALSI test, ranked at the 4th quintile by RT. It is interesting to notice that the school percentage of mean missing answers at the INVALSI test (*School missing %*) has been treated in a diametrically opposite way by the two methods: the RE-EM excluded this variable, while the RT raked it at the 4th quintile.

5. Discussion

Despite substantial differences, the presented methods agree over many of the predictors considered in the analyses, in particular, on the variables excluded from the analysis and on the top ranked ones. Ranking the variables on the base of the importance for prediction has given new interesting insights on the intercurrent relations between the socio-economic, socio-emotional and environmental factors.

Students' individual characteristics such as self-efficacy or qualification expectation resulted of primary importance in determining students' academic performances, and they overcome other factors such as ESCS. Despite it is well known the significant impact of ESCS on student's performance (OECD, 2016; INVALSI, 2018), in this study its importance is not among the highest but it is only ranked at the 3rd quintile. The effect of the socio-economic background on educational achievement could be only moderate when taking into account students' characteristics as self-efficacy or interest in study (Marks, 2017). Indeed, the key role of self-efficacy and qualification expectation on student's academic performance has been widely studied (Caprara *et al.*, 2008; Bandura *et al.*, 2001; Caprara, 2001) and the results of PISA tests demonstrated that, in some countries, including Italy, self-efficacy is a stronger predictor of academic achievement than student or school socio-economic background (OECD, 2004). Furthermore, Artlet *et al.* (2003) shown that self-efficacy is highly correlated with student ESCS and students with high levels of ESCS result into higher levels of self-efficacy. According to the literature, students with high self-efficacy have greater academic expectations than students with low self-efficacy and they achieve better academic performance (Schunk, 2012; Zimmerman *et al.*, 1992). Numerous studies have found that the students' education expectations are positively

ISBN 9788835113850

correlated with good academic performance at all grades (see, among others: Okagaki and Frensch, 1998; Ainley *et al.*, 1991; Marjoribanks, 1987). In accordance with previous studies (Chapell *et al.*, 2005; Eysenck and Calvo, 1992; Meece *et al.*, 1990), another socio-emotional variable that results to be important for the prediction of the outcome is the student anxiety during INVALSI test in Italian language, ranked at the higher importance quintiles by both methods. This variable has been found to have a negative effect on students' performances. Pomerantz *et al.* (2002) analysed the gender differences in academic performance and internal distress, demonstrating that girls outperformed boys across four subjects but were also more prone to anxiety. Gottfried (1985) demonstrated the relation between academic intrinsic motivations (i.e. the genuine interest in an activity), academic anxiety and school achievement. In our study, the student's intrinsic motivation is represented by the student's interest in studying the Italian language. In general, highly motivated students achieve better academic results than less motivated students (Tella, 2007; Tavani and Losh, 2003) and Sikhwari (2014) showed that females are highly motivated compared to their male peers.

It is interesting to notice that gender and the indicator of second generation immigrant, two variables that are widely associated with inequalities in education achievement, are excluded by both methods. Girls usually result to outperform boys in reading tests (INVALSI, 2019; Legewie and Di Prete, 2012). However, the effect of the student's gender could be moderated by the inclusion in the models of student's individual characteristics. Spinath *et al.* (2014) demonstrated that gender differences in students' individual characteristics contribute to a significant extent to gender difference in school performance.

Concerning the role of the immigrant status in the academic achievement, several studies suggested a non-complete integration of the immigrant into the society (Schnell and Azzolini, 2015; Azzolini *et al.*, 2012). On the other hand, since second generation immigrants attend the entire school cycle in the same country, their integration should be higher than that of first-generation ones (Schneeweis, 2011; Schnepf, 2004). Meunier (2011) highlighted that the poor performance of students is the outcome of a set of characteristics (such as lower language skills, lower socio-economic and cultural status) rather than the immigrant status itself. This consideration suggests that the effect of the immigrant status on performance could be mitigated by the inclusion of other students' variables in the model. This hypothesis will be taken into account in future researches. Also, the dialect and the language spoken at home, excluded by both methods, could be strongly correlated with the immigrant status and their influence on students' achievement could be

ISBN 9788835113850

mitigated with the inclusion in the model of other students' individual characteristics. This hypothesis should be further investigated in future studies.

In addition, it is important to notice that almost all the variables excluded by the two methods are binary variables: the use of the dialect at home, the language spoken at home, the attendance at kindergarten, the gender and the immigrant status. One well-known limitation of the tree methods is that the unrestricted search approach of the best variable to split the sample space induces a bias in variable selection (Loh and Shih, 1997; Doyle, 1973). Specifically, regression tree tends to select variables that have more categories because those variables provide more potential splits.

5.1. Strengths and limitations of the study

The main advantage of CART and RE-EM is that they are based on nonparametric machine learning algorithms, i.e. the algorithms are not based on strong assumptions about the functional form which describes the relation between the outcome variable and the covariates. This results into a model free to learn any functional form from the data and makes nonparametric learning methods a more flexible tool – if compared with parametric method as standard linear regression – able to analyze and represent the complexity of the educational phenomenon (Hollander *et al.*, 2013).

An additional advantage of the tree-based methods is that the variables selected in each model are ranked by importance. The computation of the variable importance is embedded in the tree-based algorithm, allowing performing the selection and the ranking of the variables at once. On the contrary, in the classic regression model, it is necessary to perform further steps of variable selection, as the best subset selection method (Beale *et al.*, 1967; Hocking and Leslie, 1967), which is also used in the framework of linear mixed model (Burnham and Anderson, 2002), i.e. in presence of multilevel data. However, this method could be more computationally demanding than tree-based methods and it does not allow ranking the variables in function of their importance for prediction. The advantages of CART and RE-EM make them an attractive alternative to traditional regression models, in particular in presence of a large number of predictors.

A major drawback of tree-based methods is their sensitivity to the characteristics of the sample, i.e. a small change in the sample could cause a relevant change in the results of the decision tree, causing instability (Timofeev, 2004). We point out that RE-EM is a relatively new method, which specifically addresses the problem of adapting CART to the case of multilevel data, and,

ISBN 9788835113850

as far as we know, there are not studies which have investigated the differences in the sensitivity to the sample characteristics of the two methods. To overcome this problem, in this study we tested the optimal parameters of the methods by means of a 10-fold cross-validation. Nevertheless, specific methods as Random Forest (Breiman, 2001) have been developed over the years to address the problem of overfitting in CART. Moreover, Random Forest could also help to overcome the variable selection bias of CART and RE-EM methods, i.e. the tendency to select variables that have more possible categories. It is our intention, in future studies, to include these models in the comparison.

6. Conclusions

The aim of the present study was to compare the performance of two tree-based variable selection methods to identify the predictors of the students' performance at INVALSI test in Italian and rank them on the base of their importance.

The RE-EM has resulted to be the best method for our data as it has better prediction accuracy with a more parsimonious model compared to RT. This result suggests that considering the multilevel structure of the data in the RE-EM could be a further advantage for the nonparametric tree-based methods as it resulted in an improvement in the predictive power. In addition, RE-EM results to be a better alternative in terms of prediction accuracy if compared to LME model.

A crucial advantage of the tree-based methods is the possibility to rank the variables according their importance. The variable selection and the variable importance ranking of the RE-EM suggested the relevance of the individual socio-emotional characteristics as predictors of the academic achievement. The students' attitude, expectations and motivations seem to influence the students' performances more than factors usually considered of paramount importance in the context of the educational inequalities, as the socio-economic background, the gender and the foreign status of the students. This suggests students' individual characteristic as fundamental factors to take into account when investigating the students' academic success.

This result could be a useful guideline for schools and policy makers, which should be aware about the key role played by of students' individual socio-emotional variables in order to design effective interventions to improve the students' achievements.

As previously discussed, the tree-based methods suffer from specific drawbacks as the variable selection bias, i.e. they tend to exclude mostly the binary variables in favour of the continuous variables, and the sensitivity to the sam-

ISBN 9788835113850

ple characteristics. In future studies, we aim to investigate more in depth the selection bias and the influence on the predictive performance of the model and to include in the comparison other models less prone to overfitting, as Random Forest (Hastie *et al.*, 2009) or Lasso (Tibshirani, 1996). Furthermore, in this study we have focused on only thirty student's variables. The integration in the model of teacher's and school's variables will be the base of future researches.

To summarize, in this study we have applied two different variable selection methods to predict the student's scholastic performance and to identify the relevant factors. The RE-EM outperformed the RT in terms of prediction accuracy and parsimony and allowed accounting for the hierarchical structure of the data. Moreover, the RE-EM outperformed the LME, considered in this study as reference model, due to its large use as parametric regression model in the case of multilevel data.

Since tree-based methods are able to identify complex relations among variables, researchers should consider the characteristics of the variable selection methods, in order to choose the most suitable one, and the most important variables, for investigating the research questions.

References

Acharya A., Sinha D. (2014), "Application of feature selection methods in educational data mining", *International Journal of Computer Applications*, 103, 2, pp. 34-38.

Ainley J., Foreman J., Sheret M. (1991), "High school factors that influence students to remain in school", *Journal of Educational Research*, 85, pp. 69-80.

Akaike H. (1974), "A new look at the statistical model identification", *IEEE transactions on automatic control*, 19, 6, pp. 716-723.

Artlet C., Baumert J., Julius-McElvany N., Peschar J. (2003), *Learners for Life. Student Approaches to Learning. Results from PISA 2000*, Paris, OECD.

Azzolini D., Schnell P., Palmer J.R. (2012), "Educational achievement gaps between immigrant and native students in two 'new' immigration countries: Italy and Spain in comparison", *The Annals of the American Academy of Political and Social Science*, 643, 1, pp. 46-77.

Bandura A., Barbaranelli C., Caprara G.V., Pastorelli, C. (2001), "Self efficacy beliefs as shapers of children's aspirations and career trajectories", *Child Development*, 72, 1, pp. 187-206.

Beale E.M.L., Kendall M.G., Mann D.W. (1967), "The discarding of variables in multivariate analysis", *Biometrika*, 54, 3-4, pp. 357-366.

Breiman L., Friedman J., Olshen R., Stone C. (1984), "Classification and regression trees", *Wadsworth Int. Group*, 37, 15, pp. 237-251.

Breiman L. (2001), "Random forests", *Machine Learning*, 45, 1, pp. 5-32.

ISBN 9788835113850

Burnham K.P., Anderson D.R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer-Verlag, New York.

Caprara G. (a cura di) (2001), *La valutazione dell'autoefficacia. Costrutti e strumenti*, Erickson, Trento.

Caprara G., Fida R., Vecchione M., Del Bove, G., Vecchio G., Barbaranelli C., Bandura A. (2008), "Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement", *Journal of Educational Psychology*, 100, 3, pp. 525-534.

Chandrashekar G., Sahin F. (2014), "A survey on feature selection methods", *Computers and Electrical Engineering*, 40, 1, pp. 16-28.

Chapell M.S., Blanding Z.B., Silverstein M.E., Takahashi M., Newman B., Gubi A., McCann N. (2005), "Test anxiety and academic performance in undergraduate and graduate students", *Journal of educational Psychology*, 97, 2, pp. 268-274.

Cortez P., Silva A.M.G. (2008), "Using data mining to predict secondary school student performance", *Proceedings of 5th Annual Future Business Technology Conference*, 39, 7, pp. 5-12.

Doyle P. (1973), "The use of automatic interaction detector and similar search procedures", *Journal of the Operational Research Society*, 24, 3, pp. 465-467.

Draper N., Smith H. (1966), *Applied Regression Analysis*, Wiley, New York.

Efroymson M. (1966), *Stepwise regression – A backward and forward look*, paper presented at the Eastern Regional Meetings of the Institute of Mathematical Statistics, Florham Park (NJ).

Eysenck M.W., Calvo M.G. (1992), "Anxiety and performance: The processing efficiency theory", *Cognition and Emotion*, 6, 6, pp. 409-434.

Fu W., Simonoff J. S. (2015), "Unbiased regression trees for longitudinal and clustered data", *Computational Statistics and Data Analysis*, 88, pp. 53-74.

Gallina V. (a cura di) (2006), *Letteratismo e abilità per la vita: Indagine nazionale sulla popolazione italiana, 16-65 anni*, Armando, Roma.

Gottfried A.E. (1985), "Academic intrinsic motivation in elementary and junior high school students", *Journal of Educational Psychology*, 77, 6, pp. 631-645.

Hastie T., Tibshirani R., Friedman J. (2009), *The elements of statistical learning*, Springer, New York.

Hajjem A., Bellavance F., Larocque D. (2011), "Mixed effects regression trees for clustered data", *Statistics and Probability Letters*, 81, 4, pp. 451-459.

Hocking R.R., Leslie R.N. (1967), "Selection of the best subset in regression analysis", *Technometrics*, 9, 4, pp. 531-540.

Hollander M., Wolfe D.A., Chicken E. (*eds.*) (2013), *Nonparametric statistical methods (Vol. 751)*, John Wiley and Sons, New York.

INVALSI (2018), *Rapporto nazionale Prove INVALSI 2018*, retrieved on September, 15, 2020, from https://www.invalsi.it/invalsi/doc_evidenza/2018/Rapporto_prove_INVALSI_2018.pdf.

INVALSI (2019), *I risultati delle prove INVALSI 2019 a colpo d'occhio*, retrieved on September, 15, 2020, from https://invalsi-areaprove.cineca.it/docs/2019/Uno_sguardo_generale_sui_risultati_delle_prove_INVALSI_2019.pdf.

ISBN 9788835113850

Legewie J., Di Prete T.A. (2012), "School Context and the Gender Gap in Educational Achievement", *American Sociological Review*, 77, 3, pp. 463-85.

Lim T.S., Loh W.Y., Shih Y.S. (2000), "A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms", *Machine Learning*, 40, 3, pp. 203-228.

Loh W.Y., Shih Y.S. (1997), "Split selection methods for classification trees", *Statistica sinica*, 7, 4, pp. 815-840.

Marjoribanks K. (1987), "Ability and attitude correlates of academic achievement: Family-group differences", *Journal of Educational Psychology*, 79, pp. 171-178.

Marks G.N. (2017), "Is SES really that important for educational outcomes in Australia? A review and some recent evidence", *Aust. Educ. Res.*, 44, pp. 191-211.

Martin D.P., Von Oertzen T. (2015), "Growth mixture models outperform simpler clustering algorithms when detecting longitudinal heterogeneity, even with small sample sizes", *Structural Equation Modeling: A Multidisciplinary Journal*, 22, 2, pp. 264-275.

Meunier M. (2011), "Immigration and student achievement: Evidence from Switzerland", *Economics of Education Review*, 30, 1, pp. 16-38.

Miao J., Niu L. (2016), "A survey on feature selection", *Procedia Computer Science*, 91, pp. 919-926.

Meece J.L., Wigfield A., Eccles J.S. (1990), "Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics", *Journal of Educational Psychology*, 82, 1, pp. 60-70.

Okagaki L., Frensch P.A. (1998), "Parenting and children's school achievement: a multiethnic perspective", *American Educational Research Journal*, 35, 1, pp. 123-144.

OECD (2004), *Learning for tomorrow's world: First results from PISA 2003*, OECD, Paris.

OECD (2016), *PISA 2015 Results (Volume I): Excellence and Equity in Education*, OECD Publishing, Paris.

Passow A.H., Eckstein M., Mallea J. (1976), *The National Case Study: An Empirical Comparative Study of Twenty-One Educational Systems*, Wiley, New York.

Pinheiro J.C., Bates D.M. (2000), *Mixed-Effects Models in S and SPLUS*, Springer, New York.

Pinheiro J., Bates D., DebRoy S., Sarkar D., R Core Team (2020), "nlme: Linear and Nonlinear Mixed Effects Models", *R package version*, 3, pp. 1-148, https://CRAN.R-project.org/package=nlme.

Pomerantz E.M., Altermatt E.R., Saxon J.L. (2002), "Making the grade but feeling distressed: Gender differences in academic performance and internal distress", *Journal of Educational Psychology*, 94, 2, pp. 396-404.

Ramaswami, M., Bhaskaran R. (2009), "A study on feature selection techniques in educational data mining", *Journal of Computing*, 1, 1, pp. 7-11

Sanchez-Pinto L.N., Venable L.R., Fahrenbach J., Churpek M.M. (2018), "Comparison of variable selection methods for clinical predictive modeling", *International Journal of Medical Informatic*s, 116, pp. 10-17.

ISBN 9788835113850

Samejima F. (1969), "Estimation of latent ability using a response pattern of graded scores", *Psychometrika Monograph Supplement*, 34, pp. 1-97.

Schneeweis N. (2011), "Educational institutions and the integration of migrants", *Journal of Population Economics*, 24, 4, pp. 1281-1308.

Schnell P., Azzolini D. (2015), "The academic achievements of immigrant youths in new destination countries: Evidence from southern Europe", *Migration Studies*, 3, 2, pp. 217-240.

Schnepf S.V. (2004), *How different are immigrants? A cross-country and cross-survey analysis of educational achievement*, IZA, Discussion Paper 1398, Bonn.

Schunk D.H. (2012), *Learning theories an educational perspective sixth edition*, Pearson, Upper Saddle River (NJ).

Schwarz G. (1978), "Estimating the dimension of a model", *The Annals of Statistics*, 6, 2, pp. 461-464.

Sela R.J., Simonoff J.S. (2011), "REEMtree: Regression Trees with Random Effects", *R package version 0.90*, 3, pp. 741-749.

Sela R.J., Simonoff J.S. (2012), "RE-EM trees: A data mining approach for longitudinal and clustered data", *Machine Learning*, 86, 2, pp. 169–207.

Sikhwari T.D. (2014), "A study of the relationship between motivation, self-concept and academic achievement of students at a university in Limpopo Province, South Africa", *International Journal of Educational Sciences*, 6, 1, pp. 19-25.

Spinath B., Eckert C., Steinmayr R. (2014), "Gender differences in school success: what are the roles of students' intelligence, personality and motivation?", *Educational Research*, 56, 2, pp. 230-243.

Tavani C.M., Losh S.C. (2003), "Motivation, self-confidence, and expectations as predictors of the academic performances among our high school students", *Child Study Journal*, 33, pp. 141-151.

Tella A. (2007), "The impact of motivation on student's academic achievement and learning outcomes in mathematics among secondary school students in Nigeria", *Eurasia Journal of Mathematics, Science & Technology Education*, 3, 2, pp. 149-156.

Therneau T., Atkinson B. (2018), "rpart: Recursive Partitioning and Regression Trees", *R package version*, 4, pp. 1-13.

Tibshirani R. (1996), "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267-288.

Timofeev R. (2004), Classification and regression trees (CART) theory and applications, Humboldt University, Berlin.

Vaida F., Blanchard S. (2005), "Conditional Akaike information for mixed-effects models", *Biometrika*, 92, pp. 351-370.

Wolfinger R., O'Connell M. (1993), "Generalized linear mixed models a pseudo-likelihood approach", *Journal of Statistical Computation and Simulation*, 48, 3-4, pp. 233-243.

Zimmerman B.J., Bandura A., Martinez-Pons M. (1992), "Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting", *American Educational Research Journal*, 29, 3, pp. 663-676.

ISBN 9788835113850

The authors

Elisa Caponera, PhD, is researcher at INVALSI. Expert on questionnaire construction and validation, she was Italian National Research Coordinator (NRC) for IEA TIMSS and ICILS projects. Her current themes of research are parent involvement at school, gender difference in Mathematics achievement, school effectiveness and equity of school system.

Clelia Cascella held a PhD in Research Methodology and a Ph.D. in Economics before being appointed as researcher in Social Statistics and Psychometrics at INVALSI. Her research interests spread from educational inequalities to research methods, with a focus on Item Response Theory and Multilevel modeling.

Elena Demuru has a degree in Statistics and a PhD in Demography. She has done research in the fields of demography and epidemiology in collaboration with ISTAT and INMP and worked as research assistant at the Statistical Office of INVALSI. She is currently employed as researcher at the Italian National Institute of Health (ISS).

Patrizia Giannantoni holds a PhD in statistics and demography from a multinational program involving University of Rome and Lund and Max Plank Institute in Rostock. She has worked on psychometric evaluation of developmental tests in collaboration with CNR and University of Parma and participated in research projects on migration and integration as research fellow at University of Naples. Since 2017 she has joined the statistical office of INVALSI, keeping her research interests on migrant integration, effects of parental support and educational inequalities.

ISBN 9788835113850

Jana Kopecna, PhD in Demography. She has been involved in research projects on international migration and educational integration of migrants, at the University Sapienza of Rome. She is currently employed at the Statistical Office of INVALSI.

Lorenzo Mancini, PhD in Statistical Science, is a research assistant at the Innovation and Development department at INVALSI. His expertise involves data analysis and elaborations for the Italian National Evaluation System and for the PON Valu.E project. He obtained the Double Bachelor's Degree at the University of Glasgow, Department of Statistics.

Laura Palmerio, PhD, is senior researcher at INVALSI where she is in charge of the International Surveys department. She's the Italian responsible for OECD surveys (PISA, TALIS) and IEA surveys (TIMSS, PIRLS, ICCS). Main research interests: relations between Reading Literacy and Mathematics, civic and citizenship education, gender differences in school achievement.

Chiara Sacco, PhD in Statistical Methodology for Scientific Research, currently works at the INVALSI Research Institute as statistician. The main research interests are in the context of multivariate data analysis for high dimensional data with particular focus on dimension reduction strategies, model based clustering and latent variable models.

Cristina Stringher, Phd, INVALSI researcher, leads a team for the study of Learning to learn (L2L) in Italy, Europe and Latin America, first international project by INVALSI. She also manages the Italian experimentation on preschool quality indicators. Past member of the Informal Advisory Group and the Italian manager for the OECD Study on Social and Emotional Skills and Chair of the Working Group on Data Development at the OECD ECEC Network, she was invited expert by the European Commission in the working groups on ECEC quality indicators, on the EU Key Competencies and on L2L Frameworks.

ISBN 9788835113850

ISBN 9788835113850

ISBN 9788835113850

Over the years, interest in data has always grown and, aware of their centrality, many institutions, both public and private, share their data to facilitate the work of all those who wish to use them to interpret phenomena. In the education field, the data produced by INVALSI undoubtedly have a leading role, both at a sample and census level. The availability of data on learning achievements and living conditions of students (the so-called "context data"), as well as on the professional and operational conditions of teachers and School Managers, collected through specific questionnaires, is a valuable source of information based on which it is possible not only to plan improvement interventions in the didactic field, but also to undertake stimulating paths of educational research.

This volume hosts four research papers, presented within the III Seminar "INVALSI data: a research tool", which took place in Bari from 26 to 28 October 2018. Thanks to the INVALSI data, the authors conducted interesting in-depth analysis of various aspects relating to the Italian education system.

Patrizia Falzetti is Head of the INVALSI Statistical Service, which manages the acquisition, analysis and return of data concerning national and international surveys on learning to individual schools, stakeholders and the scientific community.

FrancoAngeli
La passione per le conoscenze

Unione Europea
FONDI STRUTTURALI EUROPEI
PON 2014-2020
MIUR
PER LA SCUOLA - COMPETENZE E AMBIENTI PER L'APPRENDIMENTO (FSE-FESR)