Influences of School Climate and Teacher's Behavior on Student's Competencies in Mathematics and the Territorial Gap between Italian Macro-areas in PISA 2012

Giuseppe Bove - Daniela Marella - Vincenzina Vitale

Università degli Studi Roma Tre - Department of Education (Italy)

doi: 10.7358/ecps-2016-013-bove

giuseppe.bove@uniroma3.it
daniela.marella@uniroma3.it
vincenzina.vitale@uniroma3.it

EFFETTI DEL CLIMA SCOLASTICO E DEL COMPORTAMENTO DELL'INSEGNANTE SULLE COMPETENZE DEGLI STUDENTI IN MATEMATICA E DIVARIO TERRITORIALE TRA LE MACRO-AREE ITALIANE IN PISA 2012

Abstract

In this study the effects of school and classroom climate and teacher's behavior on Italian students' mathematical achievement score in PISA 2012 were investigated. Simple and scale indices provided by the PISA database, constructed by responses from the students' and principals' background questionnaires, were considered as predictive variables of the math achievement scores. Multilevel models including all the predictive variables, controlling for some relevant student, family background and school variables, confirmed that perceptions of school and classroom climate and teachers' behavior influence mathematics performance in PISA. In particular, the effect of the teacher's use of cognitive activation strategies had the strongest positive effect, followed by the school and classroom climate indicators. Thus, more cognitively activating instruction and an orderly and peaceful atmosphere in schools and classrooms encourage students and help to transform existing interests into mathematic achievement. Our *analyses show that these factors can also influence the gap between Northern and Southern Italian macro-areas. When the predictive variables are added to the control variables in our multilevel models including macro-area indicators, the gap between Northern macro-areas and the Southern-Islands decreases by over twenty per cent, and the gap between Northern macro-areas and the Southern by over fourteen per cent. On the basis of these results, we have provided some useful indications for Italian educators and policy makers.*

Keywords: Mathematical achievement, Multilevel models, PISA, School climate, Territorial gap.

1. Introduction

Several types of educational, psychological, and social factors can influence students' learning and performances at school. In an important comparative study, Wang *et al.* (1993) show that proximal factors (e.g., psychological, instructional, and home/school environment) exert more influence than distal factors (e.g., demographic, policy, and organizational). They emphasize the importance of the effects of classroom management (e.g., promoting a classroom atmosphere conducive to learning) and the quality of student-teacher interaction (e.g., cognitive level of questions), considered as strong as cognitive student abilities and family background. More recently, several studies have suggested that problems such as violence, bullying and similar behaviors cause weakened instruction and that learning requires an orderly and co-operative environment both in and outside of the classroom (e.g., Luiselli *et al.*, 2005; Jennings & Greenberg, 2009).

In particular, with respect to mathematical learning, Kunter *et al.* (2013) have identified three basic dimensions that link teaching and students' outcomes in mathematics classrooms: cognitive activation, supportive climate, and classroom management. The cognitive level of students' activities is a key feature of mathematical instruction promoting conceptual understanding. Besides, the degree of cognitive involvement usually depends not only on the demands of the tasks but also on a supportive classroom climate. Students in classes with a more supportive climate are cognitively more engaged and show more involvement than students in classes with a less supportive climate (Turner *et al.*, 1998). On the other hand, learning depends on the quality of classroom management. Students have more opportunity to engage by learning content when they can spend more time on tasks, and therefore students in less effectively managed classrooms are usually disadvantaged.

For these reasons, school climate and teacher's behavior are considered important factors in international educational surveys, and concerning information gathered from student and school background questionnaires, can be related to the performance in achievement tests. In particular, in the questionnaires administered in the Programme for International Student Assessment (PISA) 2012, promoted by the Organisation for Economic Co-operation and Development (OECD), students' and principals' perceptions of three related aspects were considered: school climate, student-teacher relation, teacher didactic behavior.

Unfortunately, despite its importance, the impact of these aspects on student outcomes have not been rigorously investigated in the Italian educational system. So, to cover this lack, this study aims to examine the influences of school climate and teacher's behavior on student's results in mathematics in the PISA 2012 Italian sample. Another important aspect considered is how these influences are related to the territorial gap between Italian macro-areas. Italian results in many of the international surveys on students' competencies locate the younger Italian generations to the bottom places in the international rankings, and highlight a dramatic divide between the country's northern and southern regions. Studies have examined the influence on this territorial gap of family background (Checchi, 2004), environmental factors like local availability of social capital or other local resources (Bratti *et al.*, 2007), and even the psychological motivation of students to succeed in the PISA test (Quintano *et al.*, 2010), but no studies have analyzed the influence of factors related to climate and student-teacher interaction.

In summary, the present study addressed (a) whether and how Italian students' math achievement scores in PISA 2012 are related to some predictive variables regarding students' and principals' perceptions of school and classroom climate and teachers' behavior, and (b) whether and how these predictive variables influence the territorial gap between Italian macro-areas.

To take into account the hierarchical nature and the complex sampling features of the PISA data set, a multilevel modeling analysis, which allows for the examination of variables at several levels simultaneously (student and school levels) is performed.

2. Overview of the literature

Schools are social places where students do not learn alone but rather in collaboration with their teachers and peers. In many studies researchers have increasingly acknowledged that a safe and healthy school environment is important for promoting students' academic achievement (e.g., Zins *et al.*, 2004). Factors like student truancy or arriving late for school, students lacking respect for teachers or using alcohol or illegal drugs, and students intimidating or bullying other students, have been identified as contributing to students' sense of safety and belonging at school, but less research has examined the degree to which these factors are predictive of performance on standardized achievement tests. At the same time, when teachers and students have closer relationships and better communication, students may be better able to seek help when they need it. Perceptions of teacher support have been associated with greater school liking, greater self-direction, and better academic performance (e.g., Birch & Ladd, 1998). From this perspective, many empirical studies have also confirmed that a better climate in a classroom has positive effects on student learning, because it provides students with sufficient time and an orderly atmosphere in which to be engaged in study activities (e.g., Goh & Fraser, 1998; Seidel & Shavelson, 2007). Thus, climate and student-teacher interactions are analyzed in international educational surveys, and sometimes considered as predictive variables of students' performances.

2.1. *School climate and teacher's behavior in international surveys*

International surveys usually produce complex datasets, combining information from students, teachers and school principals, for many different countries. Items concerning school and classroom climate and student-teacher relations are available at each level (school principals, teachers and students).

Claes *et al.* (2009) investigated causes and consequences of the occurrence of truancy in schools, in the international survey on Civic Education (Cived), carried out in 1999 by the International Association for the Evaluation of Educational Achievement – IEA (Torney-Purta *et al.*, 2001). In the study, regarding primarily fourteen year old students, it was assumed that endemic truancy impacts the general school climate and disturbs the entire pedagogical process within the class, negatively affecting even the pupils who are present in class. Differentiating between the individual, school and country levels by multilevel regression analysis, they found schools that encourage participation, that offer a supporting climate, that are seen as open participating environments, and where parents are strongly involved, record lower truancy levels. Moreover, test scores for civic knowledge are lower in schools with high truancy levels, and this effect remains significant, even taking into account various strong control variables, like socio-economic status.

To determine how negative school factors, such as aggression, are related to the mathematical and scientific performances of students, Perse *et al.* (2011) conducted an analysis of the TIMSS (Trends in International Mathematics and Science Study) 2003 database, for the Slovenian and the international data. Comparisons of the different levels of an index of students' perception of school safety showed that whoever reported not being exposed to aggression at school, on average, got higher (math and science) scores than those who did. Simple regression analysis showed that such aggressive behavior is a good predictor of educational achievement, both for Slovenia and for some high and low-achieving countries selected from the whole TIMSS population, both in the fourth and in the eighth grade.

These phenomena have also been analyzed in the triennial survey of fifteen-year-old students around the world known as PISA. Information concerning school climate and teacher's behavior has been gathered by students' and principals' responses to several questions contained in the background and school context questionnaires. Scale indices are constructed through the scaling of multiple items concerning school and classroom climate, classroom management, teacher's cognitive activation, etc. These indices are available along with raw data in the PISA OECD web site, and they can be used to investigate in detail educational context in the participating countries.

In OECD, 2013a, p. 62, it is underlined as «Disciplinary climate is also consistently related to higher average performance at the school level. In 48 participating countries and economies, schools with better average performance tend to have a more positive disciplinary climate, even after accounting for the socio-economic status and demographic background of students and schools and various other school characteristics», and also: «The school climate encompasses not only norms and values but also the quality of teacher-student relations and the general atmosphere» (OECD, 2013b, p. 178). The relationship between mathematics performance and the level of teacher-student relations presents a large variation between countries in PISA. In a multilevel regression analysis with mathematics performance regressed on schools' learning environment, resources, policies and practices, and student and school characteristics (OECD, 2013a, table IV.1.12c, p. 247), a school average index of teacher-student relations presented both large positive regression coefficients (Austria, Korea, USA) and large negative regression coefficients (Turkey, Argentina, Germany, Italy). Moreover, with respect to teacher behavior: «On average across OECD countries, students who reported that their teacher uses cognitive-activation strategies and teacher-directed instruction reported particularly high levels of perseverance and openness to problem solving, are more likely to favour mathematics as a field of study over other subjects, and to see mathematics as more necessary to their careers than other subjects compared with students who perform as well but whose teachers do not use these strategies» (OECD, 2013c, p. 114).

School climate and student-teacher relations have also been considered in local studies regarding PISA data. For instance, Konishi *et al.* (2010) examined the relationship between school bullying, student-teacher connectedness, and academic performance in a Canadian student sample drawn from the 2003 data collection for the PISA project. Using a multilevel analysis they found that math achievement was negatively related to school bullying and positively related to student-teacher connectedness.

With respect to the Italian educational system, school climate and student-teacher relations have not been analyzed in connection with student performances in PISA. Some of the items describing the climate in the class in the student questionnaire were analyzed by INVALSI, 2013, where comparing the student responses in 2003-2012 PISA waves, an improvement was detected for the class climate during lessons. At the same time, school climate and student-teacher relations were never considered in studies of the territorial gap in the level of skills achieved by fifteen-year-old students in PISA. For these reasons, in the next pages, after a brief description of the PISA data set, we will study the relation between Italian students' math achievement scores in PISA 2012 and school and classroom climate and teachers' behavior, verifying also if this relation can influence the territorial gap between Italian macro-areas.

3. Methods

3.1. *Source of data*

We report here only the main characteristics of PISA data, because they are largely known among social researchers, and even to the general public. PISA is an aged based survey assessing fifteen-year-old students (in most cases, students approaching the end of compulsory schooling) and taking place in many countries every three years since the year 2000. As emphasized by OECD, 2014, p. 22, «The assessment is forward-looking: rather than focusing on the extent to which these students have mastered a specific school curriculum, it looks at their ability to use their knowledge and skills to meet real-life challenges». Therefore, unlike other large-scale surveys such as TIMSS or PIRLS, PISA does not focus on curricular competencies but on knowledge and skills that can be used in everyday life.

For each assessment, between reading, mathematics and science, one is chosen as the major domain and given greater emphasis. In this paper the last wave of PISA is considered, which refers to data collected in Spring 2012 and whose main focus is on measuring performance in mathematics. In addition to this, a wealth of information on both the students' and schools' characteristics are collected through questionnaires filled in by students, parents and schools' representatives (typically principals). A detailed description of the general characteristics of the survey and the adopted frameworks can be found in OECD, 2013b.

Students' math achievement scores were constructed by PISA from students' performance on standardized paper-and-pencil tests, including multiple-choice and constructed-response items (short-answer, open-ended questions), with high scores indicating high achievement in each subject. Comparability across participating countries is allowed by these standardized achievement measures. An important innovation in PISA 2012 was the rotated design of the student questionnaire to extend the content coverage without increasing the response time for individual students, so that more information could be used in the study. Three distinct forms (A, B and C) were submitted to students. Questions concerning gender, language at home, migrant background, home possessions, parental occupation and education (collected in a common question set present in all forms), were administered to all students, as in previous PISA cycles. Questions regarding attitudinal and other non-cognitive constructs (collected in three distinct question sets rotated in pairs in each form) were administered only to sub-samples of students. Scale indices were computed in order to measure latent constructs that cannot be observed directly. To this aim, student and school questionnaire items were scaled using item response models and weighted likelihood estimation. Parameter estimates, scale reliabilities and other details are provided in OECD, 2014, pp. 316-353. As a consequence of the rotation of the student questionnaire, scale indices based on items in the common question set are defined for all students in the sample, and scale indices based on items in the rotated question sets (present in only two of the three forms) are defined approximately for two thirds of the sample.

3.1.1. Sample

The Italian student sample in PISA 2012 was a stratified two-stage cluster sample. In the first stage schools in which fifteen-year-old students were enrolled, were sampled with systematic probability proportional to size (PPS) sampling in each stratum. The two stratification variables were the geographical area (21 categories) and the study programme (5 categories), and the measure of size was the estimated number of eligible (fifteen-year-old) students enrolled. In the second stage, samples of students were selected within the sampled schools. A frame list of each sampled school's fifteen-year-old students was prepared and thirty-five students were selected with equal probability. All fifteen-year-old students were selected if fewer than thirty-five were enrolled. The original Italian dataset includes 31073 students nested within 1194 schools. In order to reduce the sampling error associated with aggregating micro-level indices to form macro-level indices, we exclude schools with a sample size of less than 15 students, since most of the school indices (macro level indices) involved in the statistical analyses are observed only for two thirds of the original sample. Furthermore, in applying multilevel models we exclude all cases with missing data on the predictive or control variables (*listwise* option), since most of all are missing by design. The final database includes 16709 students within 857 schools: 8276 (49.5%) males, 8433 (50.5%) females. The average school sample size is 19.5.

3.2. *Measures*

In our analyses we are going to study the influence of school and classroom climate and teacher's behavior (predictors) both on student performance in PISA 2012 and on the gap in the mathematical achievement between Northern and Southern Italian macro-areas, controlling for some student, family background and school variables (e.g., gender, attitude towards school and mathematics, the Index of Economic, Social and Cultural Status – ESCS, school study programme, etc.). The predictive and control variables, grouped according to the level (student or school), concern three different aspects: school climate, student-teacher relation, teacher didactic behavior.

All the scale indices used as predictive or control variables, were transformed by the PISA Consortium to an international metric, with an OECD average of zero and an OECD standard deviation of one. So, a positive value of an index means that the respondent answered more favorably, or more positively, than respondents did, on average, across OECD countries.

3.2.1. Mathematics performance

Student performance in mathematics cannot be directly observed but must be inferred from observed item responses. Among several possible approaches to compute literacy scales, PISA uses the methodology referred to as *plausible values*. These values are random draws from opportune posterior distributions defined for each student, representing his/her range of abilities (see

Mislevy, 1991, for details). When established, the overall mathematical literacy scale has the mean set at 500 and the standard deviation set at 100 (for the pooled, equally weighted OECD participating countries). Five plausible values (ranged from 162.2 to 825.8 in our sample) are included for each student in the international PISA database (labeled from pv1math to pv5math). So each analysis has to be undertaken five times, once with each plausible value variable. The five results are then averaged to obtain the final estimates and the significance tests opportunely adjusted for variation between the five sets of results. Details on the scaling procedure in PISA can be found in OECD, 2014, chapter 9, and descriptions of the computations of final standard errors are reported, for instance, in OECD, 2009.

3.3. *Predictive variables*

The main descriptive statistics regarding the predictive variables considered in the following of this paper are reported in *Table 1*.

Table 1. – Descriptive statistics for predictive variables (national values) (n = 16709).

Variables	Mean	SD	Min.	Max.
DISCLIMA	0.01626	0.988430	-2.451	1.879
STUDREL	0.02132	0.983734	-2.906	2.364
COGACT	0.01318	0.893115	-3.767	3.319
TCHBEHFA	0.02206	0.951896	-2.504	2.517
TCHBEHSO	0.01707	0.888185	-1.516	3.395
TCHBEHTD	0.01630	0.969952	-3.463	2.753
STUDCLIM	-0.01639	0.940618	-2.428	2.717
TCMORALE	0.00252	0.904310	-2.818	2.054
TEACCLIM	0.00799	0.932255	-2.181	3.176
MEANDISCLIMA	-0.01433	0.434047	-1.386	1.473
MEANSTUDREL	-0.18375	0.323138	-1.140	1.028
MEANTCHBEHFA	0.13206	0..354346	-0.876	1.203
MEANTCHBEHSO	-0.06557	0.371510	-0.950	1.264
MEANTCHBEHTD	-0.17716	0.357401	-1.621	0.869
MEANCOGACT	-0.10319	0.338719	-1.412	1.047

3.3.1. Student-level

The following scale indices were constructed through the scaling of multiple items concerning school climate and teacher's behavior, gathered by students' responses to several questions contained in the background questionnaires. This group includes an index concerning classroom climate, that is the *index of disciplinary climate* (DISCLIMA, based on student answers to items like: «Students don't listen to what the teacher says», with: M = 0.016, SD = 0.988); an index concerning student-teacher relations, that is the *index of teacher-student relations* (STUDREL, based on student answers to items like: «Students get along well with most teachers», with: M = 0.021, SD = 0.983); and four indices concerning teachers' didactic behavior in mathematics lessons, that are: the *index of teacher-directed instruction* (TCHBEHTD, based on student answers to items like: «The teacher sets clear goals for student learning», with: M = 0.016, SD = 0.969), the *index of teachers' student orientation* (TCHBEHSO, based on student answers to items like: «The teacher gives students different work classmates who have difficulties learning and/or to those who can advance faster», with: M = 0.017, SD = 0.888), the *index of teachers' use of formative assessment* (TCHBEHFA, based on student answers to items like: «The teacher tells students how well they are doing in mathematics class», with: M = 0.022, SD = 0.951), the *index of teacher's use of cognitive activation strategies* (COGACT, based on student answers to items like: «The teacher asks questions that make students reflect on the problem», with: M = 0.013, SD = 0.893).

3.3.2. School-level

In this section, we distinguish between indices obtained by principals' responses to questions of the school questionnaire, and indices obtained as school means of some students' level indices presented in the previous section.

The following scale indices were constructed through the scaling of multiple items concerning school climate and teacher's behavior, gathered by school principals' responses to several questions contained in the background school questionnaire. This group includes: the *index on teacher-related factors affecting school climate* (TEACCLIM, derived from principals' report on the extent to which the learning of students was hindered by the following factors in their schools: (i) students not being encouraged to achieve their full potential; (ii) poor student-teacher relations; etc., with: M = 0.007, SD = 0.932), the *index of student-related factors affecting school climate* (STUDCLIM, derived from principals' report on the extent to which the learning of students was hindered by the following factors in their schools: (i) students truancy; (ii) students skipping classes; etc., with: M = -0.0163, SD = 0.940), the *index of teacher morale* (TCMORALE, derived from principals' report on the extent to which they agree with the following statements considering teachers in their schools: (i) the morale of the teachers in this school is high; (ii) teachers work with enthusiasm; etc., with: M = 0.002, SD = 0.904).

To the previous three indices, that are directly defined at the school-level, we also added six other variables (*contextual variables*) that arise as school means of the six indices defined as predictive variables at the student-level, namely: MEANDISCLIMA, MEANSTUDREL, MEANTCHBEHTD, MEANTCHBEHSO, MEANTCHBEHFA and MEANCOGACT. As averages of the opinions expressed by several students in the same school, these contextual variables can be assumed as proxies of the school situation regarding respectively the school climate, the student-teacher relations and the teacher didactic behaviors. Besides, introducing also these six variables as predictive variables at the school-level, allow us to express differences between within-school and between-school regressions in the multilevel models that will be presented in the following of this paper.

3.4. *Control variables*

To explain the dependence of mathematics performance by the predictors defined in the previous sections, other variables were included as control variables both at student and school-level.

3.4.1. Student-level

In this group we have: *gender*, the *index of economic, social and cultural status* (ESCS, with: M = -0.018, SD = 0.962) and its squared values (ESCS2), the *index of attitude towards school* (learning outcomes) (ATSCHL, based on items like: «School has done little to prepare me for adult life», with: M = 0.026, SD = 0.933), the *index of attitude towards school* (learning activities) (ATTLNACT, based on items like: «Trying hard at school will help me get a good job», with: M = 0.017, SD = 0.942), the *index of mathematics self-concept* (SCMAT, based on items like: «I am just not good at mathematics», with: M = 0.024, SD = 0.973»), the *index of mathematics anxiety* (ANXMAT, based on items like: «I often worry that it will be difficult for me in mathematics classes», with: M = 0.032, SD = 0.852), the *index of home educational resources* (HEDRES, based on items measuring the existence of educational resources at home including a desk and a quiet place to study, a computer that students can use for schoolwork, etc., with: M = 0.004, SD = 0.856).

3.4.2. School-level

In this group we have: a scale index concerning school resources, that is the *index of quality of school educational resources* (SCMATEDU, derived from items measuring principals' perceptions of potential factors hindering instruction at their school, like: «Shortage or inadequacy of science laboratory equipment» or: «Shortage or inadequacy of computers for instruction», with: M = 0.015, SD = 0.884); and two simple indices regarding school activities, that are: the *index of creative extracurricular activities at school* (CREACTIV, derived from principals' report on whether their schools offered the following activities to students in the national modal grade for fifteen 15-year-olds in the academic year of the PISA assessment: (i) band, orchestra or choir; (ii) school play or school musical; and (iii) art club or art activities, with: M = 0.043, SD = 0.911); the *index of mathematics extracurricular activities at school* (MACTIV, derived from principals' report on whether their schools offered the following activities to students in the national modal grade for 15-year-olds in the academic year of the PISA assessment: (i) mathematics club; (ii) mathematics competition; (iii) clubs with a focus on computers/ ITC; and (iv) additional mathematics lessons, with: M = 0.081, SD = 0.971).

In addition, we included the school mean of the *index of economic, social and cultural status* (MEANESCS) and two variables concerning the sample stratification, that are: the *macro-area* (with categories: NW North-West, NE North-East, C Centre, S South, SI South-Islands); the *study programme* (with categories: Academic track, Technical, Vocational and others). The former allows us to analyze the influence of the predictors on the geographical gap of performances, the latter is necessary to compare the national study programmes and to control for differences in their geographical distribution.

4. Data analysis

4.1. *Data analysis plan*

Data in educational studies have a hierarchical structure; in a two-level model, students are nested within schools and variables are measured at each level of the hierarchy. Each level is (potentially) a source of unexplained variability. It is clear that students belonging to the same school are often more alike than two randomly chosen students, so the average correlation between variables measured on students from the same school is higher than that measured on students from different schools. The problem of no *independent identically distributed* (*i.i.d.*) observations is also common in surveys employing multistage sampling designs, with often unequal selection probability at some stage, as in PISA 2012. Traditional regression methods ignore dependence of observations leading to underestimate standard errors. Smaller standard errors increase the significance of hypothesis tests assuming there is more information in the data than there really is.

It is possible to model such a complex population and sampling structure as a hierachical system of regression equations, within the class of multilevel models (for an overview see, for instance, Hox, 2010, or Snijders & Bosker, 2012).

In this paper, an approach integrating model and design-based frameworks is used to take into account for the particular complex sampling features of PISA 2012. The main characteristics of this approach are the multilevel model specification and the use of sampling weights at both student and school levels, to adjust the estimation for the unequal probability of selection. Moreover, the methodology of *plausible values* suggested in OECD, 2014, was incorporated in the estimation and hypotheses testing procedures.

4.2. *Data analysis procedure*

We apply several multilevel regression models with two levels and a random intercept, using *Mplus* software version 6 (Muthén & Muthén, 1998-2010). The dependent variable is the mathematics performance. Predictors and control variables at school and student levels are described in the «Measures» section. The estimation method uses the five plausible values for each student's performance on the mathematics scale and the final student sampling weights provided with the PISA 2012 database. Students weights are rescaled to sum up within each school to the school sample size, according to Pfeffermann *et al.* (1998). School weights, corresponding to the sum of student final weights within each school, are rescaled using the default method implemented in *Mplus*. The new composite weights, which are the product of the rescaled student and school weights, sum up to the total sample size. Models estimation uses the five imputed datasets according to PISA procedure, combining results from these analyses. In the presence of sampling weights, in *Mplus*, parameters are estimated by maximizing a weighted loglikelihood function (Asparouhov, 2006; Asparouhov & Muthen, 2006), and standard error computations use a sandwich estimator. Estimates are robust to non – normality and non – independence of observations. For details, see Muthén and Muthén, 1998-2010. Data analysis was conducted in three phases.

*Table 2. – Multilevel regression analyses on mathematical performances in PISA 2012 Italian sample (Overall models) – (*p < 0.05 and **p < 0.01 and ***p < 0.001).*

		Empty-model (1)			Empty-model plus student level variables (2)			Empty-model plus student and school level variables (3)		
	FIXED PART	*Coefficient*	*S.E.*	*t-value*	*Coefficient*	*S.E.*	*t-value*	*Coefficient*	*S.E.*	*t-value*
	Intercept	493.5***	3.049	161.878	457.618***	3.103	147.49	457.224***	3.057	149.55
Student-level										
	Predictive variables									
	DISCLIMA				2.770**	0.799	3.468	2.762**	0.798	3.461
	STUDREL				-3.083**	1.039	-2.966	-2.951**	1.047	-2.819
	TCHBEHFA				-2.498*	1.040	-2.401	-2.511*	1.040	-2.414
	TCHBEHSO				-12.857***	1.087	-11.824	-12.867***	1.087	-11.837
	TCHBEHTD				-3.030**	1.030	-2.942	-3.052**	1.030	-2.964
	COGACT				11.056***	0.979	11.297	11.045***	0.978	11.297
	Control variables									
	GENDER				23.023***	1.872	12.297	23.087***	1.870	12.349
	ESCS				4.668***	0.841	5.551	4.667***	0.841	5.546
	ESCS^2				–	–	–	–	–	–
	HEDRES				–	–	–	–	–	–
	ATSCHL				–	–	–	–	–	–
	ATTLNACT				–	–	–	–	–	–
	ANXMAT				-12.169***	0.987	-12.329	-12.211***	0.988	-12.364
	SCMAT				20.469***	0.938	21.829	20.445***	0.937	21.809
School-level										
	Predictive variables									
	MEANDISCLIMA				13.823**	5.089	2.716	11.366*	4.918	2.311
	MEANSTUDREL				–	–	–	-12.278*	5.716	-2.148

	Empty-model (1)			Empty-model plus student level variables (2)			Empty-model plus student and school level variables (3)		
FIXED PART	*Coefficient*	*S.E.*	*t-value*	*Coefficient*	*S.E.*	*t-value*	*Coefficient*	*S.E.*	*t-value*
MEANTCHBEHFA	(1)			-37.823***	7.281	-5.195	-33.114***	6.866	-4.823
MEANTCHBEHSO				-33.168***	7.367	-4.502	-28.411***	7.121	-3.990
MEANTCHBEHTD				-26.295**	8.794	-2.990	-24.712**	9.085	-2.720
MEANCOGACT				57.877***	6.988	8.283	54.010***	6.968	7.751
STUDCLIM							9.017***	1.749	5.156
TCMORALE							–	–	–
TEACCLIM							–	–	–
Control variables									
CREACTIV							–	–	–
MACTIV							3.722*	1.679	2.216
SCMATEDU							–	–	–
MEANESCS				34.28***	4.512	7.597	27.669***	4.400	6.289
RANDOM PART									
School-level variance	4048.274***	233.35	17.349	1085.065***	92.452	11.736	1005.287***	86.323	11.646
Student-level variance	4039.654***	77.711	51.983	3042.394***	61.489	49.479	3042.472***	61.507	49.465
(Residual) intraclass correlation	0.505			0.263			0.248		
Deviance	188777.680			183170.964			183115.862		
R-square	–			0.489			0.019		

*Table 3. – Multilevel regression analyses on mathematical performances in PISA 2012 Italian sample (GAP analysis) – (*p < 0.05 and **p < 0.01 and ***p < 0.001).*

		Model 1			Model 2			Model 3		
		(1)			(2)			(3)		
	FIXED PART	*Coefficient*	*S.E.*	*t-value*	*Coefficient*	*S.E.*	*t-value*	*Coefficient*	*S.E.*	*t-value*
	Intercept	419.383***	5.912	70.934	408.739***	6.857	59.605	433.096***	6.267	69.104
	Macro-area and Study programme									
	NORTH WEST	20.024**	7.455	2.686	28.211***	6.038	4.672	19.562***	4.624	4.231
	NORTH EAST	28.257***	7.320	3.860	35.339***	6.071	5.821	23.861**	5.248	4.547
	SOUTH	-29.126***	6.537	-4.455	-13.508*	5.905	-2.287	-16.322**	4.867	-3.354
	SOUTH-ISLES	-44.278***	6.441	-6.874	-31.687***	5.796	-5.467	-26.479***	4.804	-5.512
	ACADEMIC TR.	111.443***	4.795	23.243	64.112***	7.410	8.652	29.906***	7.100	4.212
	TECHNICAL	65.520***	5.074	12.914	41.161***	5.701	7.219	24.581***	5.293	4.644
	N. WEST-A. TR.	–	–	–	–	–	–	–	–	–
	N. WEST-TECH..	–	–	–	–	–	–	–	–	–
	N. EAST-A. TR.	–	–	–	–	–	–	–	–	–
	N. EAST-TECH.	23.648*	9.174	2.578	16.664*	7.361	2.264	10.074	6.887	1.463
	SOUTH- A. TR.	–	–	–	–	–	–	–	–	–
	SOUTH-TECH.	–	–	–	–	–	–	–	–	–
	S. ISLES-A. TR.	–	–	–	–	–	–	–	–	–
	S. ISLES-TECH.	–	–	–	–	–	–	–	–	–
Control variables										
	Student-level									
	GENDER				21.756***	1.892	11.499	23.027***	1.887	12.200
	ESCS				4.717***	0.853	5.531	4.664***	0.843	5.532
	ESCS^2				–	–	–	–	–	–

		Model 1			Model 2			Model 3		
		(1)			(2)			(3)		
	FIXED PART	*Coefficient*	*S.E.*	*t-value*	*Coefficient*	*S.E.*	*t-value*	*Coefficient*	*S.E.*	*t-value*
	HEDRES				–	–	–	–	–	–
	ATSCHL				–	–	–	–	–	–
	ATTLNACT				–	–	–	–	–	–
	ANXMAT				-12.092***	1.012	-11.945	-12.050***	0.989	-12.183
	SCMAT				20.299***	0.941	21.571	20.615***	0.936	22.025
	School-level									
	CREACTIV				–	–	–	–	–	–
	MACTIV				5.883**	1.949	3.019	3.733*	1.502	2.485
	SCMATEDU				-	-	-	-	-	-
	MEANESCS				41.589***	5.244	7.931	23.333***	4.551	5.128
Predictive variables										
	Student-level									
	DISCLIMA							2.758**	0.798	3.458
	STUDREL							-3.002**	1.038	-2.892
	TCHBEHFA							-2.597*	1.034	-2.512
	TCHBEHSO							-12.847***	1.085	-11.841
	TCHBEHTD							-3.024**	1.032	-2.930
	COGACT							11.056***	0.981	11.267
	School-level									
	MEANDISCLIMA							13.744**	4.470	3.075
	MEANSTUDREL							–	–	–
	MEANTCHBEHFA							–	–	–
	MEANTCHBEHSO							-29.464***	6.247	-4.716

	Model 1			Model 2			Model 3		
	(1)			(2)			(3)		
FIXED PART	*Coefficient*	*S.E.*	*t-value*	*Coefficient*	*S.E.*	*t-value*	*Coefficient*	*S.E.*	*t-value*
MEANTCHBEHTD							-19.395*	7.814	-2.482
MEANCOGACT							36.172***	6.238	5.798
STUDCLIM							6.573***	1.536	4.280
TCMORALE							–	–	–
TEACCLIM							–	–	–
RANDOM PART									
School-level variance	1691.444***	129.491	13.062	1081.739***	81.958	13.199	738.598***	65.022	11.359
Student-level variance	4049.324***	72.950	55.508	3198.13***	62.697	51.010	3042.569***	61.651	49.351
(Residual) intraclass correlation	0.295			0.253			0.195		
Deviance	188083.200			183967.376			182899.010		
R-square	0.292			0.254			0.116		

Phase 1 – As generally recommended for multilevel analysis, we start with computation of *student* and *school level* variances estimates for the dependent variable (mathematical performance), by a one-way ANOVA type model with a random intercept (named *empty model*). Results are shown in *Table 2*, columns under (1). This variance decomposition allows us to estimate the *intraclass correlation coefficient* (ICC, proportion of total variance of the dependent variable that is due to the grouping at school level). A low intraclass correlation indicates that schools are performing similarly while higher values point towards large differences between school performance. A significant between school variance and a substantial ICC are essential conditions for application of multilevel modeling.

Phase 2 – We analyze the influence of predictors on mathematical performances by a multilevel model including all the predictive variables and all the control variables (*overall model*), excluding only the macro-area and study programme indicators. Student and school level variables are included in two successive steps, respectively. Results are shown in *Table 2*, columns under (2) and (3).

Phase 3 – A sequence of three *nested multilevel models* is applied in order to analyze the influence of predictors on the gap in mathematical performances between Northern and Southern Italian macro-areas. Below, we briefly describe these three nested models:

Model 1 adds to the *empty model* the two variables *macro-area* and *study programme* and their interactions, respectively. This model provides estimates of mathematical literacy scales for the Italian macro-areas, taking into account the complex sampling structure of PISA. Baseline estimates of the gap between Northern and Southern Italian macro-areas are obtained. Results are shown in *Table 3* (columns under Model 1).

Model 2 adds to Model 1 all the control variables at student and school level allowing to analyze the influence of such variables on the gap between Northern and Southern Italian macro-areas. Results are shown in *Table 3* (columns under Model 2).

Model 3 adds to Model 2 all the predictive variables at student and school level (school and class climate and teacher behavior). This allows us to analyze the influence of such variables on the gap between Northern and Southern Italian macro-areas, conditioning on the control variables. Results are shown in *Table 3* (columns under Model 3). All the scale indices are centered on their national grand mean. In order to obtain reliable mean estimates, centering was carried out before deletion of cases with any missing value.

For each model, we use a *backward selection method*, removing the variable with highest *p-value* and rerunning the model until all remaining variables have *p-values* less than 0.05. Results are shown in *Tables 2* and *3*, they will be presented in the next section. In *Tables 2* and *3*, variables not included in the final model, because they were removed by the backward procedure, are indicated by the minus sign in each column. We compare nested models by means of a chi-square difference test equivalent to a rescaled likelihood ratio test (*r-LRT,* for short). For details, see Satorra & Bentler, 2001.

5. Results

5.1. *Empty model*

To know if there exists a significant variance in mathematical performances across schools to be analyzed by multilevel modeling, a simple linear model with no predictors and with a random intercept was considered (*empty model*). Results are shown in *Table 2*, columns under (1). The estimated mean of the mathematical literacy score for all schools is 493.5 (due to the exclusion of schools with less than 15 students and missing cases treatment, this mean score is a value between the Italian national PISA mean score of 485 and the Italian modal grade PISA mean score of 499). The outputs also provide a partitioning of the total variance between student and school levels: the estimated student level variance was 4039.7 and the school level variance was 4048.3. Therefore, ICC results equal to 0.505, that is 50.5% of variance in the outcome is due to differences among schools, and intercepts vary significantly across schools (*test* $t = 17.3$, $p < .001$). The ICC value provides an indication of high variability between Italian school performances in mathematics, considering that PISA 2012 mean for ICC for participating countries is 0.37, with low values for Iceland (0.12) Finland (0.13) Sweden (0.15) and Norway (0.15) and high values for the Netherlands (0.66) Hungary (0.65) Turkey (0.62) and Slovenia (0.59). The deviance reported in *Table 2* is a goodness of fit measure expecting to go down when predictive or control variables are added to the model.

The empty model does not explain any variance of the dependent variable, and results suggest that the development of multilevel models to explain variability within and between schools is warranted.

5.2. *Overall model construction*

5.2.1. Empty-model plus student level variables

We gradually built our model, starting by including only information obtained from the students' questionnaires (student-level variables). Therefore, this model includes all the fourteen student level variables (predictive and control) listed in *Table 2*. As well as this, school means of the indices ESCS, DISCLIMA, STUDREL, TCHBEHFA, TCHBEHSO, TCHBEHTD and COGACT, were also included to express differences in individual and contextual effects of these variables on mathematical performance. Results are shown in *Table 2*, columns under (2). Four variables at student level (ESCS2, HEDRES, ATSCHL, ATTLNACT) and one variable at school level (MEANSTUDREL) were removed by the backward procedure.

In the final model, the student level variance (3042.4) and the school level variance (1085.1) are much lower than in the empty model, so the added student variables remaining after the backward procedure explain part of the student and school variability of the dependent variable. The proportional reduction in the estimated total variance (R^2 as proposed in Snijders & Bosker, 2012, p. 112) was 0.489 (or 48.9%) in respect to the empty model. The residual intraclass correlation is 0.263 (26.3% of the residual total variance is now explained by school differences). The deviance also goes down, indicating a better fit than the empty model (*r-LRT* = 2695.601 with *p-value* < 0.0001). The new intercept estimated value of 457.6 represents the expected mathematical literacy score for a female student having national mean values for all the student-level variables included in the regression model, and attending a school with national mean values for all the school level variables. According to the regression coefficients presented in *Table 2* for this model, we found that the two predictive variables regarding the climate in a classroom and student-teacher relations (i.e. DISCLIMA and STUDREL) are both significant at the student level, after controlling for gender, some relevant attitudes of students towards school and mathematics, and family background. However, they have opposite signs. The index of disciplinary climate positively affects mathematical performances (DISCLIMA and the corresponding school mean – MEANDISCLIMA, have regression coefficients respectively of 2.77 and 13.82). For this variable, the contextual effect of the school mean gives an additional contribution over the individual effect, so that a student with a given perception of his/her class climate obtains, on average, a higher mathematical score if he/she is also in a school with a better perceived mean climate. The effect of MEANDISCLIMA is about twice that of DISCLIMA, when they are compared with respect to one standard deviation increase. The index of teacher-student relations has a negative impact on mathematical performances (STUDREL and the corresponding school mean – MEANSTUDREL, have regression coefficients -3.1 and -11.3, respectively. Only the first coefficient is significant).

The regression coefficients for the four indices concerning teacher didactic behavior (i.e. TCHBEHFA, TCHBEHSO, TCHBEHTD and COGACT) are significant. The index of teacher's use of cognitive activation strategies (COGACT) presents a positive and relevant influence on mathematical performance, with a regression coefficient of 11.06 for the student level and 57.9 for mean school level (MEANCOGACT). Again, the contextual effect of the school mean gives an additional contribution over the individual effect, so that a student with a given perception of his/her teacher's use of cognitive activation strategies in mathematical lessons, on average, has a higher mathematical score if he/she is also in a school where use of activation strategies is better perceived. Also for this variable, the school level effect is about twice that of the student level, when they are compared with respect to one standard deviation increase. The index of the teachers' use of formative assessment (TCHBEHFA), the index of the teachers' student orientation (TCHBEHSO) and the index of the teacher-directed instruction (TCHBEHTD) have negative regression coefficients, particularly relevant in the case of TCHBEHSO, that has a regression coefficient of -12.9 for the student level and -33.2 for mean school level (MEANTCHBEHSO). Possible interpretations of these negative influences are provided in the «Discussion» section.

The analysis of results concerning individual control variables for this model confirms that girls score lower on mathematics performance, after controlling for the cited student and family characteristics. The family background, represented by the index of economic, social and cultural status (ESCS) and the index of home educational resources (HEDRES), has a significant and positive influence only with respect to the former. Student aptitude towards school (ATTSCHL, ATTLNACT) does not appear influential for mathematical performances, while mathematics anxiety (ANXMAT) and, above all, mathematics self-concept (SCMAT) play an important role.

5.2.2. Empty-model plus student and school level variables

Now we add the school level variables obtained from the principals' questionnaire (school-level variables) to the previous model, to examine their influence on mathematical performance. Results are shown in *Table 2*, columns under (3). Four variables at student level (ESCS2, HEDRES, ATSCHL, ATTLNACT) and four variables at school level (TCMORALE, TEACCLIM, CREACTIV, SCMATEDU) were removed by the backward procedure. In the final model, the student-level variance is 3042.4 and school-level variance is 1005.3. We note that the student level variance is unchanged with respect to the previous model, as a result of having added all school level variables. The proportional reduction in the estimated total variance is 0.019 (or 1.9%) with respect to the previous model. However, if we measure the proportional reduction only with respect to the school-level variance, we obtain a 7.4% reduction, so the new school level variables remaining after the backward procedure (MACTIV and STUDCLIM) explain part of the school variability of the dependent variable. The residual intraclass correlation is 0.248 (24.8% of the residual total variance is now explained by school differences). The deviance also goes down, which indicates that the model fits better than the previous model (*r-LRT* = 36.717 with *p-value* < 0.0001).

The intercept estimated value of 457.2 is substantially unchanged with respect to the previous model and has a similar interpretation. Among the school-level predictive variables included in the model, the *principal's report on student-related factors affecting school climate* (STUDCLIM) has a significant regression coefficient of 9.02. Therefore, after controlling for other variables, students are more likely to perform better in mathematics, when attending schools where principals believe that students' behavior (e.g., truancy, skipping classes, arriving late for school, etc.) hinders learning to a lesser extent (e.g., a better disciplinary climate at school). This result confirms what was obtained with the previous model at the student level, with the *index of disciplinary climate* (DISCLIMA). On the contrary, neither regression coefficients of the *principal's report on teacher-related factors affecting school climate* (TEACCLIM) or the *principal's report on teacher morale* (TCMORALE) are significant (a common result for most of the OECD participating countries (see OECD, 2013a, table IV.1.12b, p. 243), as a consequence of their small correlations with performances in mathematics.

Amongst the control variables included at the school-level, only the *index of mathematics extracurricular activities at school* (MACTIV) has a significant regression coefficient of 3.7, meaning students are more likely to perform better in mathematics, when attending schools where principals report high levels of mathematics extracurricular activities. The influences of the *index of creative extracurricular activities at school* (CREACTIV) and the *index of quality of school educational resources* (SCMATEDU) result as not significant.

6. Comparison of predictor variables

In this section we compare the effects of school and class climate and teacher's behavior on student performance in mathematics, by using the regression coefficients obtained in the previous section. Despite the fact that most predictors are scale indices transformed to an international metric with an OECD average of zero and an OECD standard deviation of one, in the comparison of the effects we prefer to take into account the small differences in their standard deviations (due to the selection of the Italian subsample). To this aim, the effect of each student and school level predictor on the mathematical literacy score is computed as the product of the corresponding unstandardized regression coefficient (in *Table 2*) with the standard deviation (in *Table 1*). We point out that all these effects can be considered as *net* effects, because they represent changes in the dependent variable that can be associated with an increase of one standard deviation of the given predictor, for fixed values of the other predictors and control variables included in the regression model. Moreover, when for a student level predictor a school mean is also included as a school level predictor in the model, the school level effect is added to the student level effect. For the climate indicators, the effect of STUDCLIM index is added to DISCLIMA and MEANDISCLIMA.

Accordingly, the total (student and school level) effect of teacher's use of cognitive activation strategies results as the strongest positive effect (one standard deviation increase determines, on average, an increase of 28 points on the mathematical score), followed by the school and classroom climate (16 point increase on the mathematical score, evenly distributed between DISCLIMA and STUDCLIM). Among the negative effects, that of the teachers' student orientation seems remarkable (one standard deviation increase determines, on average, a -22 point decrease), followed by the teachers' use of formative assessment (-14 point decrease) and the teacher-directed instruction (-12 point decrease). The negative effect of the teacher-student relations results as relatively weak (-7 point decrease).

6.1. *Nested models to analyze the geographical gap in mathematics performances*

In the presentation of the following three models, we focalize on the analysis of changings concerning the regression coefficients of the macro-area indicators, disregarding the analysis of the effects of predictors and control variables on the dependent variable, already analyzed in the previous sections.

In Model 1 the fixed effects of the variables *macro-area* and *study programme* (see «Measures» section for a description of the categories) and their interactions are added to the random intercept in this multilevel model. Results are shown in *Table 3* (column Model 1). All the interaction dummies except one (NE-Technical) were removed by the backward selection method. As expected, by entering only school-level variables in the model, the student level variance (4049.3) remains essentially unchanged with respect to the empty model, while the school-level variance (1691.4) is much lower. The proportional reduction in the estimated total variance was 0.292 (or 29.2%) with respect to the empty model. The residual intraclass correlation is 0.295 (29.5% of the residual total variance). The deviance also goes down, which indicates that the model fits better than the empty model (*r-LRT* = 421.0501 with *p-value* < 0.0001).

The intercept value of 419.4 shown in *Table 3*, is the estimate of the mean mathematical literacy score for students having the two reference categories C-Centre (for macro-area) and Vocational and others (for study programme). As it is shown in the table, this baseline literacy score increases when we consider the northern macro-areas NW-North/West (+20.02) and NE-North/East (+28.26), while it decreases for the southern macro-areas S-South (-29.13) and SI-South/Islands (-44.28). For the other two categories of the study programme, the literacy score increases respectively by 111.4 points (Academic track) and 65.52 (Technical). A further increase of 23.65 points regards the interactions between categories NE-North/East and Technical.

The macro-areas regression coefficients estimates allow us to quantify the gap in mathematics performance between students from different geographical areas. Therefore, for instance, we see that the gap NE-SI amounts to 72.54 points (28.26 + 44.28) and the gap NW-SI amounts to 64.3 points (20.02 + 44.28). We can investigate the reasons of these gaps adding new information (control and predictive variables) to this model and studying corresponding changes in the gap measures obtained. We start adding control variables to this model in the following section.

In Model 2 all the control variables at student and school level are added to the previous Model 1, to analyze their influence on the gap between Northern and Southern Italian macro-areas. Results are shown in *Table 3* (column Model 2).

Also for this model the only interaction dummy not removed by the backward selection method was NE-Technical. The school-level variance (1081.7) and the student-level variance (3198.1) were lower than in Model 1, and the proportional reduction in the estimated total variance was 0.254 (or 25.4%). The residual intraclass correlation is 0.253 (25.3% of the residual total variance). The deviance also goes down, which indicates that the model fits better than the previous model (*r-LRT* = 1876.344 with *p-value* < 0.0001).

The intercept value of 408.7 shown in *Table 3*, is the expected mean mathematical literacy score for female students having the two reference categories C-Centre (for macro-area) and Vocational and others (for study programme), and having national mean values for all the student-level control variables included in the regression model, and attending a school with national mean values for all the school-level control variables. Again, this baseline literacy score increases when we consider the northern macro-areas NW-North/West (+28.21) and NE-North/East (+35.34), while decreases for the southern macro-areas S-South (-13.51) and SI-South/Islands (-31.69). For the other two categories of the study programme, the literacy score increases respectively by 64.11 points (Academic track) and 41.16 (Technical). A further increase of 16.66 points regards the interactions between categories NE-North-East and Technical.

We confine our analyses to the comparison of the territorial gaps provided by Model 1 and Model 2, respectively. We see that the gap NE-SI amounts to 67.03 (35.34 + 31.69), a 7.6% reduction with respect to the gap estimated by Model 1. Similarly, a 6.8% reduction is obtained for the gap NW-SI, and stronger reductions (approximately 15%) regard the gaps NE-S and NW-S. Therefore, we can observe that the gap between Northern and Southern Italian macro-areas reduces when controlled for some personal student characteristics (anxiety and self-concept) linked with mathematics, the economic, social and cultural status of families and schools, and the propensity to mathematical extracurricular activities of the schools.

We can conclude our analysis of the gap in the following section, adding all the predictive variables to Model 2. In Model 3 all the predictors at student and school level are added to the previous Model 2, to analyze their influence on the gap between Northern and Southern Italian macro-areas. Results are shown in *Table 3* (column Model 3).

For this final model we note that all the interactions were removed by the backward selection method, including the dummy NE-Technical. However, in order to maintain the nested structure with Model 2 and compute the *r-LRT*, we prefer to include the interaction NE-Technical in the model. The school-level variance (738.6) and the student-level variance (3042.57) were lower than in Model 2. The proportional reduction in the estimated total variance at student level was 0.116 (or 11.6%). The residual intraclass correlation is 0.195 (19.5% of the residual total variance). The deviance also goes down, which indicates that the model fits better than Model 2 (*r-LRT* = 557.016 with *p-value* < 0.0001). In this final model, the gap NE-SI amounts to 50.34, corresponding to a further 24.9% reduction with respect to the gap estimated by Model 2. Similarly, a 23.14% reduction is obtained for the gap NW-SI, and lower reductions also regard the gaps NE-S (17.7%) and

NW-S (14%). Most of the predictors analyzed in the final overall model [*Table 2*, columns under (3)] have also significant regression coefficients in Model 3, with similar sizes (only MEANSTUDREL and MEANTCHBEHFA were not significant). Therefore, we can conclude that school and classroom climate and teachers' behavior are not only important to determine the student mathematical literacy level, but play an influence on the gap between Northern and Southern Italian macro-areas. Implications of these results from an educational point of view will be presented in the next section.

7. Discussion

In this section we summarize the major findings of our study, outlining possible interpretations and providing some indications useful for policy makers. In the next section, we will consider the limitations suggesting directions for future research.

The aim of the present study was to investigate the relation between Italian students' math achievement scores in PISA 2012 and school's and classroom climate and teachers' behavior, verifying if this relation can also influence the territorial gap between Italian macro-areas. Scale indices provided by the PISA 2012 database, and constructed by responses in the students' and principals' background questionnaires, were considered as predictive variables of the math achievement scores. These predictive variables concern perceptions with respect to three related aspects; school and classroom climate, student-teacher relations, teacher didactic behavior. We analyzed the influence of predictors on mathematical performances by a multilevel model including all the predictive variables, controlling for some relevant student, family background and school variables.

The data analysis confirmed that perceptions of school and classroom climate and teachers' behavior influence mathematics performance in PISA. In particular, the effect of teacher's use of cognitive activation strategies resulted as the strongest positive effect (one standard deviation increase determines, on average, an increase of 28 points on the mathematical score), followed by the school and classroom climate (16 point increase). Thus, more cognitively activating instructions and an orderly and peaceful atmosphere in schools and classrooms encourage students and help to transform existing interests into mathematics achievement. Significant negative effects characterize the teachers' student orientation (-22 point decrease), followed by the teachers' use of formative assessment (-14 point decrease) and the teacher-directed instruction (-12 point decrease). Thus, teacher behaviors perceived by students as characterized by a frequent use of student orientation, formative assessment and direct instruction are associated with low levels of performances in mathematics. A similar result concerns the student-teacher relations, even though the negative effect on mathematics performance is relatively weak (-7 point decrease). Although partly surprising, these findings are in line with previous results for PISA data at the international level (e.g., OECD, 2013a and 2013c).

Our analyses show that perceptions of school and classroom climate and teachers' behavior are not only important to determine the student mathematical literacy level, but can also influence the gap between Northern and Southern Italian macro-areas. When the predictive variables are added to the control variables in our multilevel model including macro-area indicators, the gaps between Northern macro-areas and the South-Islands decrease over twenty per cent and the gaps between Northern macro-areas and the Southern by over fourteen per cent.

How can these findings be explained? What do our results suggest for professionals in education and policy makers?

Higher cognitive demands that stimulate cognitive functioning and processing, encourage the students' interest to discover and understand mathematical concepts and procedures. The teacher who stimulates students to explain and compare their thoughts and solution methods, increases the likelihood of cognitive activation. Several studies found a significant relationship between higher-order thinking skills and students' mathematics performance (e.g., Stein and Lane, 1996; Wenglinsky, 2002; Shayer & Adhami, 2007), and our study reinforces the indication that cognitive activation is an important means for improving student learning in mathematics. To this aim, teacher training and professional development may be effective ways of increasing cognitive activation in the classroom. Furthermore, Kunter *et al.* (2013) underline how the level of cognitive activation depends on the teacher's pedagogical content knowledge, so this suggests that it is possible to promote cognitively activating instructions by implementing professional development to enhance and advance teachers' pedagogical content knowledge. From this perspective, it seems desirable for the Italian Ministry of Education (MIUR) to strengthen training for future teachers of secondary schools, activating the specific degree programs (biennial second level degree plus one year training) already provided by law (Ministerial Decree nr. 249, September 10, 2010). The positive relationship between school and classroom climate and mathematics outcomes is in line with previous research results (Lee and Bryk, 1989; Power *et al.*, 1989; Wang *et al.*, 1993; Stewart, 2008). Students profit from an orderly school and classroom atmosphere with fewer disruptions and discipline problems. It seems likely that students in these schools have more opportunity to focus on content-related tasks and topics. Research supports the notion that a positive school climate promotes students' abilities to learn and take part in cooperative learning. Respect, group cohesion, and mutual trust have been shown to directly improve the learning environment.

In a broader discussion that concerns school climate, the contradictory small negative effect of the index of teacher-student relations (STUDREL) deserves particular attention. The index is based on students' answers to items like: «Students get along well with most teachers», «Most teachers are interested in students' well-being», «Most of my teachers really listen to what I have to say», etc. etc. These items are not referred in particular to mathematics lessons (like the items regarding the DISCLIMA index), but consider the general relationships with teachers at school and classroom levels. This could be one reason for the weak (or not significant) level of influence of this index on mathematics performance found in Italy and other participating countries in PISA 2012. The interpretation of the negative sign of the effect of STUDREL on mathematics outcomes is more difficult and controversial. We could think that in the Italian educational system, schools and classrooms where a *laissez faire* policy is prevailing, where teachers are permissive and let students spend more time on free school activities or recreation, are those situations where lower mathematics outcomes are frequently associated with a better perception of teacher-student relations, but this hypothesis needs further investigation and appropriate data, not available in the PISA dataset.

Improvements of school and classroom climate can be achieved by acting in various directions. Bryk *et al.* (2010) underscored how their research has shown relational trust (good social relationships among members of the school community) is the «glue» or the essential element that coordinates and supports the processes which are essential to effective school climate improvement. From this perspective, two relevant factors seem to be the frequency and quality of teacher-student social interactions and the encouragement of parents participation at school. Teachers influence their students not only by how and what they teach but also by how they relate to students, engage students in social interactions, model emotional constructs and appropriate behaviors, and manage the classroom. Positive teacher and student social interactions contribute to students' sense of self-esteem and foster a sense of membership in the classroom and school. Schools can have a major impact on truancy and skipping levels by promoting school involvement from parents and by providing a supportive and authoritative environment.

These considerations highlight the need for policies and interventions that can better prepare teachers to develop supportive relationships with all students and promote students' feelings of connectedness towards school.

From this point of view, programs and intervention strategies promoting teachers' social and emotional competence and well-being seems promising (e.g., Jennings & Greenberg, 2009). Finally, the positive effect of the promotion of mathematical extracurricular school activities on mathematical performance highlighted by our multilevel model, reinforces the indication that a prolonged presence at school of students also involved in educational related activities, can improve their performances besides their feeling of connectedness. Especially in those areas of the country where the social environment has negative effects on student learning (this is the case of many southern areas), a prolonged presence at school can be a means to alleviate the territorial divide.

7.1. *Limitations and future research*

Our results also raise some theoretical and methodological questions, however. The first question relates to the type of data source. In most of the cases, Italian fifteen-year-old students have been in their current school for only two or three years, and this means that much of their school experience took place in previous schools. So, the learning environment examined by PISA may only partially reflect students' experience in education, and contextual data collected are an imperfect proxy for students' learning environments up until they reach the age of fifteen.

Moreover, simple and scale indices obtained by PISA questionnaires reflect just partially the school situation, because they are based mainly on students and principal perceptions rather than on direct observations of the phenomena under investigation (e.g., video-based classroom studies, administrative registrations, etc.). Further, we remark that this paper has descriptive aims because we estimate statistical dependence without giving them any causal interpretation.

For all the previous considerations, our results concerning effects of predictive variables on mathematical performances need to be further investigated and supported by appropriate quasi-experimental and longitudinal studies concerning the Italian educational system. From a methodological point of view, further developments include the bias correction stemming from covariate measurement error. A problematic aspect of the contextual analysis is that school-level characteristics are often measured by aggregating student-level characteristics within each school (e.g., school-average of ESCS, climate, etc.). However, the observed school average may not be a reliable measure of the unobserved school average, and results in a biased estimate of the contextual effects. Although, to reduce the problem, we have decided to exclude from the sample schools with too few students, the bias can be also amended by fitting a multilevel structural equation model as suggested in Lüdtke *et al.*, 2008, and Grilli & Rampichini, 2011.

Despite the previous limitations, our results are nonetheless interesting for several reasons. To the best of our knowledge, this is the first paper to explore at a national level the influences of school climate and teacher's behavior on student's competencies in mathematics and on the territorial gap between Italian macro-areas in PISA. Moreover, in the literature the majority of studies concerning school climate and teachers' behavior do not examine their effects within multilevel/hierarchical frameworks. The analysis of the territorial dimension is based on the application of a hierarchy of nested multilevel models, going beyond the simple inclusion of macro-area dummies.

Finally, although statistical dependence does not necessarily imply causation, our results contribute to identifying important predictive variables for mathematical student performances, and to providing indications useful for Italian educators and policy makers.

References

Asparouhov, T. (2006). General multilevel modeling with sampling weights. *Communications in Statistics-Theory and Methods*, *35*(3), 439-460.

Asparouhov, T., & Muthen, B. (2006). *Multilevel modeling of complex survey data. Proceedings of the Joint Statistical Meeting in Seattle, August 2006*. ASA section on Survey Research Methods, 2718-2726.

Birch, S., & Ladd, G. W. (1998). Children's interpersonal behaviors and the teacher-child relationship. *Developmental Psychology*, *34*, 934-946.

Bratti, M., Checchi, D., & Filippin, A. (2007). Territorial differences in Italian students' mathematical competencies: Evidence from PISA 2003. *Giornale degli Economisti e Annali di Economia*, *66*, 299-333.

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago: University of Chicago Press.

Checchi, D. (2004). Da dove vengono le competenze scolastiche? L'indagine PISA 2000 in Italia [Where do scolastic skills come from? The PISA 2000 survey in Italy]. *Stato e Mercato*, *72*, 413-453.

Claes, E., Hooghe, M., & Reeskens, T. (2009). Truancy as a contextual and school-related problem: A comparative multilevel analysis of country and school characteristics on civic knowledge among 14 year olds. *Educational Studies*, *35*, 123-142.

Goh, S. C., & Fraser, B. J. (1998). Teacher interpersonal behavior, classroom environment and student outcomes in primary mathematics in Singapore. *Learning Environments Research*, *1*, 199-229.

Grilli, L., & Rampichini, C. (2011). The role of sample cluster means in multilevel models. A view of endogenity and measurement error issues. *Methodology*, *7*(4), 121-133.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York: Routledge.

INVALSI (2013). *OCSE PISA 2012*. Italian national report INVALSI. Frascati (Italy).

Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research*, *79*, 491-525.

Konishi, C., Hymel, S., Zumbo, B. D., & Li, Z. (2010). Do school bullying and student-teacher relationships matter for academic achievement? A multilevel analysis. *Canadian Journal of School Psychology*, *25*, 19-39.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Kraus, S., & Neubrand, M. (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project*. New York: Springer.

Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, *62*, 172-192.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthen, B. (2008). The multilevel latent covariate model: A new reliable approach to group level effects in contextual studies. *Psychological Methods*, *13*, 203-229.

Luiselli, J., Putnam, R., Handler, M., & Feinberg, A. (2005). Whole-school positive behaviour support: Effects on student discipline problems and academic performance. *Educational Psychology*, *25*, 183-198.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, *56*, 177-196.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles: Muthén & Muthén.

OECD (2009). *PISA 2009. Data analysis manual: SPSS and SAS* (2nd ed.). Paris: OECD.

OECD (2013a). *PISA 2012. Results: What makes schools successful? Resources, policies and practices (Volume IV)*. Paris: OECD.

OECD (2013b). *PISA 2012. Assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD.

OECD (2013c). *PISA 2012. Results. Ready to learn: Students' engagement, drive and self-beliefs (Volume III).* Paris: OECD.

OECD (2014). *PISA 2012. Technical report*. Paris: OECD.

Perše, T. V., Kozina, A., & Leban, T. R. (2011). Negative school factors and their influence on math and science achievement in TIMSS 2003. *Educational Studies*, *37*, 265-276.

Pfeffermann, D., Skinner, C., Holmes, D. J., Goldstein, H., & Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society*, s. B, *60*(1), 23-40.

Power, F. C., Higgins, A., & Kohlberg, L. (1989). *Lawrence Kohlberg's approach to moral education*. New York: Columbia University Press.

Quintano, C., Castellano, R., & Longobardi, S. (2010). Gli studenti italiani e il literacy divide: differenziali territoriali delle competenze e del comportamento di risposta [Italian students and literacy divide: Territorial differences of skills and behavior response]. In *INVALSI (2010). PISA 2006. Approfondimenti tematici e metodologici*. Roma: Armando.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507-514.

Seidel, T., & Shavelson, R. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*(4), 454-499.

Shayer, M., & Adhami, M. (2007). Fostering cognitive development through the context of mathematics. Results of the CAME project. *Educational Studies in Mathematics*, *64*(3), 265-291.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis. An introduction to basic and advanced multilevel modeling* (2nd ed.). Los Angeles: Sage.

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, *2*(1), 50-80.

Stewart, E. B. (2008). School structural characteristics, student effort, peer associations, and parental involvement: The influence of school- and individual-level factors on academic achievement. *Education & Urban Society*, *40*, 179-204.

Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age 14*. Amsterdam: IEA.

Turner, J. C., Cox, K. E., DiCintio, M., Meyer, D. K., Logan, C., & Thomas, C. (1998). Creating contexts for involvement in mathematics. *Journal of Educational Psychology*, *90*(4), 730-745.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, *6*(3), 249-294.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives*, *10*(12). Retrieved (29/06/2015) from: http://epaa.asu.edu/epaa/v10n12/

Zins, J. E., Bloodworth, M. R., Weissberg, R. P., & Walberg, H. J. (2004). The scientific base linking social and emotional learning to school success. In J. Zins, R. Weissberg, M. Wang, & H. Walberg (Eds.), *Building academic success on social and emotional learning: What does the research say?* (pp. 3-22). New York: Teachers College Press.

Riassunto

In questo lavoro si analizzano gli effetti che il clima generale della scuola e quello di classe, congiuntamente al comportamento dell'insegnante per come è percepito dagli studenti, esercitano sui risultati nelle prove di matematica dell'indagine PISA 2012. A tale scopo, alcuni degli indicatori di clima e di comportamento del docente forniti dal consorzio internazionale di PISA 2012, costruiti sulla base delle risposte fornite ai questionari da parte di studenti e presidi, vengono considerati come variabili predittive del punteggio in matematica dello studente. L'utilizzo di modelli di regressione multilivello ha quindi consentito di mettere in evidenza l'influenza degli indicatori scelti sui risultati in matematica, tenendo sotto controllo alcune caratteristiche importanti del nucleo familiare e della scuola di appartenenza dello studente. L'utilizzo di strategie di attivazione cognitiva da parte dell'insegnante emerge come l'aspetto a maggior impatto positivo, seguito dal clima di scuola e quello di classe. Le analisi condotte mostrano inoltre che tali aspetti influiscono sul divario territoriale tra le macro-aree italiane settentrionali e meridionali. In particolare, confrontando le variazioni nei valori attesi per i punteggi medi delle diverse aree, ottenute aggiungendo al modello base le variabili predittive rilevanti, si evidenzia come il divario si riduca del venti per cento, quando considerato rispetto al confronto tra Nord e Sud-Isole, e di oltre il quattordici per cento, quando considerato rispetto al confronto tra Nord e Sud. Sulla base dei risultati ottenuti vengono fornite alcune indicazioni utili a coloro che operano in ambito educativo.

Parole chiave: Apprendimento matematico, Clima scolastico, Divario territoriale, Modelli multilivello, PISA.

How to cite this Paper: Bove, G., Marella, D., & Vitale, V. (2016). Influences of school climate and teacher's behavior on student's competencies in mathematics and the territorial gap between Italian macro-areas in PISA 2012 [Effetti del clima scolastico e del comportamento dell'insegnante sulle competenze degli studenti in matematica e divario territoriale tra le macro-aree italiane in PISA 2012]. *Journal of Educational, Cultural and Psychological Studies*, *13*, 63-96. doi: 10.7358/ecps-2016-013-bove