UNIVERSITA' CATTOLICA DEL SACRO CUORE
MILANO

Dottorato di ricerca in Economia Pubblica
(DEFAP)
ciclo XXIV
S.S.D: SECS-P/06-P/05-P/01

SOCIAL INTERACTIONS AT SCHOOL

Tesi di Dottorato di:
Marco TONELLO
Matr. No. 3703677

Anno Accademico 2011/12

UNIVERSITA' CATTOLICA DEL SACRO CUORE

MILANO

Dottorato di ricerca in Economia Pubblica

(DEFAP)

ciclo XXIV

S.S.D: SECS-P/06-P/05-P/01

SOCIAL INTERACTIONS AT SCHOOL

Coordinatore: Ch.mo Prof. Massimo Bordignon

Relatore: Ch.mo Prof. Claudio Lucifora

Tesi di Dottorato di:

Marco TONELLO

Matr. No. 3703677

Anno Accademico 2011/12

Acknowledgments

Desidero ringraziare le numerose persone che mi hanno sostenuto, ascoltato, supportato (e, a volte, sopportato…) in questi anni.

Claudio mi ha trasmesso la sua passione e la sua curiosità, la sua instancabile precisone, il suo metodo. Lo rigrazio per l'entusiasmo che mi ha sempre dimostrato, per avermi ascoltato, per aver letto, riletto e corretto le pagine di questa tesi, per avermi accompagnato in questo viaggio tra le "social interactions" ed avermi incoraggiato e consigliato, sempre. Il suo sostegno e i suoi consigli non mi sono mai mancati. Lo ringrazio moltissimo…e so, che ho già scritto troppo per i suoi gusti!!

Ringrazio Paolo Sestito per il suo costante supporto nella raccolta dei dati, e per i pomeriggi passati nel suo studio e gli infiniti consigli che mi ha regalato. Ringrazio Piero Cipollone: i pranzi in sua compagnia sono stati fonte d'idee e intuizioni. Ringrazio Graziella Marzi, che mi ha accolto molte volte nel suo ufficio nella "mia Bicocca", incoraggiandomi, per prima, a intraprendere questa strada e sostenendomi sempre. Ringrazio Magda Bianco e tutti i colleghi (Roberta, Giuliana, Silvia, Giacomo, Cristina P., Stefano, Elisa, Marco e Cristina G.) che mi hanno accolto e sostenuto ormai da quasi un anno. Ringrazio i compagni di dottorato (David, Li Yuan, Han Wei) con cui ho condiviso molti dei mesi "milanesi". A Ylenia spetta un ringraziamento particolare per gli infiniti consigli, e per aver letto e commentato questi capitoli! Ringrazio, poi, le molte persone – colleghi, professori, studenti – che mi hanno dato utili suggerimenti: Simona Comi, Paolo Ghinetti, Gabriele Mazzolini, Elena Cottini, Simone Moriconi, Mauro Testaverde, Daniele Checchi, Lorenzo Cappellari, Piergiovanna Natale, Massimo Bordingnon, Luigi Prosperetti.

Ringrazio Claudio Rossetti (Luiss), Patrizia Falzetti (Invalsi), Gianna Barbieri (MIUR) e Marco Mignani perché il loro lavoro nel rendere disponibile e nell'assemblare i dati è stato semplicemente indispensabile.

Alla mia famiglia va il ringraziamento per il sostegno in tutti questi anni, per aver sostenuto "la ricerca"…di molte strade, per aver condiviso le molte decisioni e indecisioni. A loro devo una gran parte di tutto questo. A Simone va anche un ringraziamento speciale per il supporto tecnologico ed editoriale.

A Ylenia,

ai nostri passi e alle nostre corse,

a questa lunga strada che ha portato a noi…

SUMMARY

CO-AUTHORSHIP DISCLAIMER AND WORKING PAPERS

The third chapter is a joint work with Claudio Lucifora. We equally contribute to the paper. From this chapter we obtained the working paper circulated as: Lucifora, C. and M. Tonello (2012), "*Students' cheating as a social interaction: evidence from a randomized experiment in a national evaluation program"* IZA D.P. No. 6967, October 2012.

The first chapter is a substantially revised version of two working papers previously circulated: Tonello, M. (2011), "*Mechanisms of peer interactions between native and non-native students: rejection or integration?*" Document de Treball IEB 2011/21, University of Barcelona; Tonello, M. (2012), "*Social interactions between native and non-native students: mechanisms and evidence*", Quaderni IEIL, Istituto di Economia dell'Impresa e del Lavoro No. 65/2012, Università Cattolica di Milano.

INTRODUCTION [1]

Economic analysis has always focused on how individual decisions are interconnected through price interactions taking place in the markets. Non-market interactions were phenomena of smaller relevance, without an intrinsic importance and typically conceptualized as problems of 'incomplete markets', preventing the economy to reach a social optimum through the simple clearing markets mechanisms (Manski, 2000). Only starting from the late Nineties, a growing interest developed in understanding how social interactions beyond the marketplace affect individual decisions and outcomes (Blume and Durlauf, 2005). In the last decade, the importance of social interactions in shaping individual behaviour has been widely recognized in both the economic and the sociological literature (Jackson 2006). A number of studies produced empirical evidence documenting the influence of social interactions in many areas: consumption, criminal behaviour, on-the-job productivity and satisfaction, financial decisions (among others: Zimmerman, 2003, Hoxby 2000, Sacerdote 2001, Cipollone and Rosolia 2007, Falk and Ichino 2006, Mas and Moretti 2011, Moretti 2011)[2]. In the thesis, I refer to 'social interactions' as all forms of interdependencies among individuals in which preferences, beliefs and constraints faced by one socioeconomic actor are directly influenced by the characteristics and choices of others (Durlauf and Ioannides, 2009; Zanella 2004). These interactions do not occur because individuals are affected through the effects of the choices of others on prices, rather, social interactions typically have features that render them forms of externalities (Scheinkman, 2008)[3].

Pupils attending school may receive direct and indirect inputs for the development of their cognitive and non-cognitive skills, coming from a variety of sources: teachers, school facilities, parental investments, environment and neighbourhood, and, most importantly, schoolmates. Starting from the Nineties, a large and multidisciplinary literature has focused on the impact and the effects of a pupil's schoolmate's background characteristics and abilities on achievement at school (Gibbons and Telhaj, 2006). My research focuses on social

---

[1] The views expressed in the thesis are those of the author and do not necessarily reflect those of the institutions he belongs to. The usual disclaimers apply.

[2] Concerning the Italian context, evidence on educational peer effects can be found only in few and recent papers: Brunello, De Paola and Scoppa (2010), De Paola and Scoppa (2010), De Giorgi, Pellizzari and Redaelli (2010), De Giorgi and Pellizzari (2011); Cipollone and Rosolia (2007). Few other studies involve the analysis of peer effects at work (De Paola, 2010; Falk and Ichino, 2006), criminal behaviour (Corno, 2009), and hospitalization choices (Moscone, Tosetti and Vittadini, 2011).

[3] In fact, social interactions are sometimes called 'non-market interactions' to emphasize the fact that these interactions are not regulated through the price mechanism.

interactions among pupils attending the same class or the same school. In the existing literature social interactions among schoolmates are commonly referred to as ‘peer effects’ or ‘peer-groups effects’. This term usually indicates social interactions of children or young adults with people of similar age, in order to make a distinction from the broader ‘neighbourhood effects’ stemming from interactions with superiors, family or teachers (Gibbons and Telhaj, 2006). Spanning the economics, education, sociological and psychological field, a rich literature has focused on these aspects, trying to model and measure the effects of social interactions on pupils’ attainment. The first empirical study on peer effects at school dates back to the ‘Coleman Report’ (1966), while, some years later, Becker (1974) was the first to provide a theoretical framework to social interactions. It is only starting from the Nineties that a vast literature has flourished, though a clear consensus on the issue has not been found yet. Indeed, educational peer effects are a complex phenomenon. First of all, peers may affect different outcomes (such as teen pregnancy, drug use, high school attrition, attitudes toward minorities, college choice), and they may have an effect in the accumulation and development of both cognitive and non-cognitive skills (Neidell and Waldfogel, 2010). Then, peer effects may work through multiple channels: a student’s ability can affect his peers through knowledge spillovers and direct peer instruction (i.e. students teaching one another), but also a student’s behaviour may affect his peers (Hoxby, 2000). Finally, peer influence in a classroom may follow a variety of lines, such as race, disability, gender, family income. This complexity is mirrored by the difficulties in the identification of these effects and of the underlying social mechanisms, which, in Manski’s words, “[…] is difficult to impossible” (Manski, 1993, p. 532).

Albeit research on peer effects is a hard empirical challenge, in the end, it provides powerful insights for policy makers. Social scientists have been studying for a long time peer effects, because, if they exist, they potentially affect the optimal organization of any structure where individuals interact (schools, jobs, neighbourhoods, etc.). As outlined above, peer effects constitute a particular class of social interactions and may be considered as some kind of ‘externalities’ in education production activities, or in the accumulation process of human capital. This interpretation justifies the ‘public hand’ intervention to correct them: externalities due to peer effects can be both positive and negative, and in both cases may prevent to reach a social optimum. As a consequence, there is room for policy interventions so to enhance social welfare. Some practical examples in the school context are concerned with tracking (under which students are exposed only with peers with similar achievement), grouping, desegregation policies (whether desegregation plans should assign students to schools outside the neighbourhood or district), or even school choice policy issues

(Ammermueller and Pischke, 2009). If peer effects at school exist and are sufficiently high, then the policy maker should care about a school system design that encourages an efficient distribution of peers because it will make human capital investments more efficient, and will also enhance macroeconomic growth (Hoxby, 2000). From a general policy making perspective, peer effects are also important because they may reinforce the effects of changes in private incentives. This amplification is known as 'social multiplier effect', and the presence of social multipliers has important implications for the policy design that is still greatly unexplored (Glaeser, Sacerdore and Scheinkman, 2003; Durlauf, 2004).

Thesis outline - The thesis contributes to the existing literature in proposing different empirical strategy to identify social interactions parameters and linking the results to simple theoretical frameworks to shed light on the possible social mechanisms driving the estimated effects. I focus on junior high school students. Junior high school is generally considered by educational psychologists as a critical period in the students' educational path, corresponding to students' early adolescence and to the period in which friendships ties are usually formed and interactions with school mates take a relevant part of students' time at school and outside school. The three chapters exploit rich and newly available datasets combining test score results in Math and Language from Invalsi[4] (First Cycle Final Exam and National Evaluation Program), administrative records from Ministry of Education Statistical Office, and the Italian Population Census Survey 2001. The census dimension of Invalsi data allows to overcome problems of underrepresentation and measurement errors typical of international surveys on students' attainment (such as PISA, TIMMS and PIRLS) substantially improving the originality and contributions of the research.

The first chapter and the second chapter deal with social interactions between native and non-native students. In the last two decades, a lot of Western countries have experienced massive immigration waves. While there is a vast literature on the effects of immigration on natives' labour market outcomes, economic literature on the effects of non-native students on native peers' attainment levels is quite limited and presents mixed evidence. Although it is widely accepted that non-native students typically face more problems at school and have lower scores in standardized tests, causes, consequences and policy implications are still unclear. There is not clear evidence on possible consequences of social interactions between natives and non-natives in educational settings, and it might happen that such interactions (if they exist) could tend either to increase or decrease the existing attainment gaps. For instance, Jensen and Rasmussen (2011) find a negative effect of school ethnic concentration on

[4] Invalsi is the National Institute that carries out the evaluation of students' attainment and schools in Italy.

cognitive outcomes for Danish native students. Brunello and Rocco (2011) provide cross-country evidence of a negative but small effect of the share of immigrants on natives' educational attainment. On top of that, even less is known on the possible underlying mechanisms that such peer interactions may follow. The study of peer interactions between native and non-native students has also important policy implications ranging from the implementation of re-allocation programs (e.g. the 'Boston Moving To Opportunity Program', Angrist and Lang 2004), to non-native students allocation rules across classes or schools, or even 'share-cap' rules that fix a maximum level to non-native students concentration in each school.

The **first chapter** ("*Social interactions between native and non-native students: mechanisms and evidence*") addresses the question of whether social interactions between native and non-native students affect natives' attainment measured by test scores in Math and Language. The chapter proposes a theoretical framework to stylize the possible mechanisms of peer interactions between native and non-native students based on a 'disruption' versus 'integration model' of education production (Lazear, 2001), and tests the theoretical predictions identifying the causal link between non-natives' school share and native students' educational outcomes. The identification strategy hinges upon school fixed effects and selection on observables. Results show that non-native school share has small and negative impacts on test scores of natives' peers. Negative effects on natives' test scores are significantly different from zero only for sufficiently high values of non-native school-share and characterized by a convex relation (i.e. marginally increasing with respect to non-native school share). The empirical evidence is consistent with an integration model of peer interactions.

The **second chapter** ("*Acting-white? Social interactions among non-native students*") addresses the issue of the effects of social interactions within non-native students on their own attainment. The empirical analysis tests the existence and relevance of two potential behavioural channels that might help to explain the underlying mechanisms: 'acting-white' and 'assimilation'. I label 'acting-white' the evidence that within-group negative social interactions are greater the greater is the segregation of minority students within each school. This is a sort of reinterpretation of the 'oppositional culture behaviours' sociological theory (Fordham and Ogbu, 1986). The 'assimilation channel' is tested restricting the analysis on the sub-group of first-generation non-native students who plausibly experience more difficulties to assimilate to the hosting country language and culture with respect to second generation peers. The sources of endogeneity are tackled with an instrumental variable approach. I find negative within-group social interaction effects increasing with respect to the degree of school segregation and decreasing with respect to non-natives' assimilation. These findings support the existence of 'oppositional culture' mechanisms (or 'acting-white behaviours') that exacerbate the negative social interactions effects within the non-native peer group.

The **third chapter** ("*Students' cheating as a social interaction: evidence from a randomized experiment in a national evaluation program*") focuses on students' cheating as a form of social interaction among classmates taking an official exam. Large-scale cheating has been uncovered over the last year at some of the US most competitive schools, and surveys conducted in US and Canada document that a relevant part of high school and college students admit to have cheated in official exams (McCabe, 2005). Economic literature suggests that students' and teachers' cheating activities has been growing hand in hand with the more extensive use of high-stake testing systems with detrimental consequences on the signalling value of the education on the labour market and on the incentives to invest in human capital accumulation. There is little evidence on the effects of cheating behavior for educational outcomes, as well as on the measures taken to contrast its diffusion. This chapter is one of the few work which analyses students' cheating behavior in a social interaction framework. In fact, when a student cheats during an exam, many others – who might otherwise have behaved honestly - end up being influenced thus reacting to such behavior. In this context, even an isolate cheating behavior may propagate and become larger through social interactions. We provide a measure of the social interactions due do students' cheating while taking an exam in terms of 'cheating social multiplier'. First, we build a theoretical model which defines the mechanisms that may drive social interactions in cheating behavior showing that students may optimally decide whether to engage in cooperative effort exchanging information and do so taking into account other students' best response. Then, we estimate the structural parameter corresponding to the social multiplier in cheating behaviors using the Excess-Variance approach developed by Graham (2008) which is based on contrasts in the (excess) variability of conditional between- and within-variance at different levels of aggregation. The natural experiment in Invalsi SNV data is the perfect environment to retrieve the structural parameter as it provides a perfect randomization in the external monitoring technology to which classrooms were subject to during the evaluation. This chapter contributes to several strands of the literature. First, it is one of the few works studying the unintended behavioral consequences due to the introduction of testing systems in education. Second, it contributes to the general literature on social interactions as it is one of the few papers estimating the endogenous part of the peer effects parameters à la Manski. Third, it develops a stylized theoretical model and directly estimates the structural parameters of the social interactions due to students' cheating exploiting a relatively new identification strategy based on the Excess-Variance approach developed by Graham (2008). We find that cheating interactions play a substantial role so that tolerating this practice can have detrimental consequences which are substantially amplified by the effects of the social multiplier: cooperative behaviors, when a strict external monitoring is missing, may generate a change in the equilibrium of students' achievements that is twice as big as the class average achievement.

CHAPTER 1.
SOCIAL INTERACTIONS BETWEEN NATIVE AND NON-NATIVE STUDENTS: MECHANISMS AND EVIDENCE

ABSTRACT

I present an education production function with peers' spillovers encompassing two alternative mechanisms of peer effects between native and non-native students: disruption and integration. The identification strategy exploits the idiosyncratic variation in non-native school share between adjacent cohorts and school fixed-effects to estimate peer effects on natives' school mean test scores in Language and Math. I test the theoretical predictions exploiting a dataset covering the entire population of native and non-native students enrolled in Italian junior high schools. I find that non-native school share has negative impacts on natives' school mean test score especially concerning Language skills. Effects are highly non-linear: non-native school share below 15% does not affect natives' outcomes. The disruptive mechanism of peer interactions is partially rejected by the empirical analysis which rather supports the integration mechanism.

JEL Classification: J15, I21, I28

1.1. INTRODUCTION

Over the last decade, most OECD countries have experienced increased migration, much of it of people whose home language is not the language of instruction in the schools that their children attend: according to PISA 2009 survey, the proportion of 15-years-old students with an immigrant background in the European Union countries was around 9% (PISA, 2009). While there is a vast literature on the effects of immigration on natives' labour market outcomes, economic studies on the effects of non-native students on native peers' attainment levels is quite limited (Gould *et al.*, 2009). These students may be academically disadvantaged either because they are immigrants entering a new education system or because they need to learn a new language in a home environment that may not facilitate this learning. In both cases, they plausibly need special or extra attention from teachers and educators. Moreover, the educational disadvantage experienced by non-native students is substantially influenced by the new environment they face in the hosting country (Dustmann and Glitz, 2011; Dustmann *et al.*, 2011) and peer effects may play a crucial role in narrowing the existing gap - integrating non-native with native students - or exacerbating it - if self-clustering and rejection behaviours are in place (Patacchini and Zenou, 2006).

Starting from Coleman (1966), scholars in the sociology of education have long argued that peers' influence and class ethnic composition are important determinants of students' achievement. Nevertheless, the specific question of whether non-natives affect natives' educational outcomes through social interactions has received relatively little attention and presents mixed evidence (Brunello and Rocco, 2011; Gould *et al.*, 2009). Even less is known on the possible underlying mechanisms that such peer interactions may follow (De Giorgi and Pellizzari, 2011). The aim of the paper is twofold. On the one hand, I propose a theoretical framework to stylize the two mechanisms of peer interactions between native and non-native students based on 'disruption' vs. 'integration' models of educational production. On the other hand, I test the theoretical predictions identifying the causal link between non-natives' school share and native students' educational outcomes. Assuming that non-native students have a higher propensity to disrupt compared to natives and grounding on Lazear (2001) educational production framework, the 'disruptive model' of peer interaction predicts that, in mixed schools, the presence of non-native students generates negative externalities on natives' attainment which are marginally decreasing with respect to non-native share. This model embeds the classical 'bad apple principle' so that one 'disruptive student' is enough to generate bad spill-overs on all the students (Hoxby and Weinghart, 2006; Epple and Romano, 2011; Sacerdote, 2010). The 'integration model' predicts that, for 'sufficiently low' values of non-native school share, non-natives students' disruption does not hurt the educational production process. This is because non-native students are more easily integrated with native peers and engage less frequently in disruptive behaviours when they are not enough to constitute an independent cluster in the school. Exploiting this framework, I want to answer the following research questions: does non-native school share induce negative peer effects on natives' attainment? Is the 'disruption mechanism' sufficient to explain peer effects between native and non-native students? Do different levels of non-native school share have different impacts on natives' so that an 'integration mechanism' might be at work?

From the empirical point of view, I identify peer effects exploiting the within school idiosyncratic variation in non-native share between adjacent cohorts. Our estimation strategy relies on the assumption that changes in non-native school shares between adjacent cohorts in the same school are not correlated with pupils' unobservable characteristics that may be relevant in the educational production process. Solving problems of sorting and omitted variables bias is crucial in the correct identification of the effects. Sorting takes place within schools - as non-native students are non-randomly allocated across classes - and between schools because of non-natives' families residential decisions. I side-step the within school non-random allocation of non-natives across classes taking school-level averages, while school fixed effects control for across school sorting and non-native students endogenous placement. Selection on observables and school fixed-effects limit omitted variable bias in correlated effects (Hoxby, 2000; Gould *et al.* 2009; Brunello and Rocco, 2011). I use as outcome measure Language and Math test scores from the standardised exam taken by all 8th grade students enrolled in Italian junior high schools (*Invalsi First Cycle Exams*[5]).

This paper contributes both to the general literature on peer effects in education and to the specific stream of the literature concerning social interactions between native and non-native students. First, I contribute to the general literature on peer effects as, to the best of our knowledge, our study is among the firsts linking the empirical estimation of peer effects parameters to a stylized theoretical framework[6]. In fact, although there is a large empirical literature on social interactions, still little is known about the economic mechanisms leading to the high level of clustering in behaviour that is so commonly observed in the data (De Giorgi and Pellizzari, 2011). Indeed, the use of a simple theoretical framework does allow to shed light on the interpretation of the results. On top of that, allowing for non-linearities in peer effects is crucial to test which possible mechanisms and which channels peer effects are following (Hoxby and Weinghart, 2006; Imberman *et al.* 2012). Second, I contribute to the

[5] 8th grade students are attending the third year of junior high school. The Italian '*Junior high School Diploma*' corresponds to ISCED level 2.

[6] De Giorgi and Pellizzari (2011) and Duflo et al. (2011) are two notable exceptions.

specific stream of the social interactions literature which examines the effects of non-natives on natives' attainment providing new evidence on the existence and size of peer effects, and on the underlying mechanisms. Our work considerably improves the existing empirical findings on this issue as I exploit a dataset that contains census information on *all* native and non-native 8th grade students. Thanks to this characteristic, the population of non-native students in Italian junior high schools (10 -14 years-old students) is similar to the non-native population of students of many European countries, and especially those which experienced sharp immigration waves in the last decades (e.g. U.K., Spain, Portugal, Ireland) (Eurydice, 2012). Moreover, the census dimension allows to overcome serious problems of under-representation and attenuation bias arising when immigrant shares are included in reduced form estimations providing more accurate results compared to existing studies exploiting survey data on students' attainments (e.g. PISA, PIRLS and TIMMS) (Aydemir and Borjas, 2010).

Our results show that non-native school share has a negative impact on test scores of natives' peers and that the effect is stronger for Language skills. However, negative effects on natives' test scores are significantly different from zero only for sufficiently high values of non-native school-share and marginally increasing with respect to non-native school share. Thus, once I allow for non-linear effects in non-native school share, the general pattern of the results is more consistent with an 'integration mechanism' of peer interactions as negative peer effects are not at work for non-native school shares below 15% for Language test and below 20% for Math.

The rest of the paper is organized as follows. Section 1.2 presents a review of the literature and Section 1.3 explains the theoretical framework. Section 1.4 discusses the main characteristics of the dataset and provides general descriptive evidence while Section 1.5 describes the identification strategy. Section 1.6 and Section 1.7 discuss the results and conduct sensitivity checks. Section 1.8 concludes and derives policy implications.

1.2. LITERATURE

Despite the growing relevance of the immigration phenomenon in Europe and the well-established desegregation literature in the U.S., works investigating peer interactions between native and non-native students in European schools are just a few and present mixed findings. There is no consensus on the effects of social interactions between natives and non-natives in educational settings: existing studies find both that the presence of non-native students negatively influence natives' attainment (Jensen and Rasmussen, 2011; Brunello and Rocco, 2011; Gould *et al.* 2009; Contini, 2011) or do not find sizeable effects (Ohinata and Van Ours, 2011; Geay *et al.* 2012) so that it might happen that such interactions (if they exist) could either increase or decrease the existing attainment gaps (Schnepf, 2007).

Jensen and Rasmussen (2011) analyse the effect of school ethnic concentration on PISA test scores of Danish 9th grade students. To correct for the endogeneity in school ethnic concentration across schools, the authors apply school fixed-effects and IV. The instrumental variable is the ethnic concentration in the geographical area where each school is located. Results show that there is a negative effect of ethnic concentration on students' outcomes which is statistically significant only for native Danish children. Brunello and Rocco (2011) study whether the share of immigrant pupils affects the school performance of natives using aggregate multi-country data from PISA. Aggregation at the country level is exploited to avoid sorting problems of immigrant students within each country, while country fixed effects and socio-economic indicators are used to control for across countries sorting and time trends in immigrants' residential choices. They find that immigrant share has small negative effects on natives' mean test and that a reduction of the dispersion of this share between schools would determine only a small increase in natives' test scores. Gould *et al.* (2009) exploit the variation in the number of immigrants in 5th grade conditional on the total number of immigrant students in grades 4 to 6 to identify the causal link of the immigrant concentration on the outcomes of native students in Israeli schools. The approach is interesting under two main aspects: first, they use quasi-experimental evidence as early '90 immigration waves to Israel are used as an exogenous variation in immigrants' flows; second, they focus on long-term outcomes. Their results point to a strong adverse effect of immigrant concentration on native outcomes. Contini (2011) exploits data from 5th and 6th grade students in Italy to study to which extent immigrant class share influences the attainment of both native and immigrant students. Her identification strategy is based on within school (across-classes) variation in the exposure to non-native peers. She conducts the analysis on a sub-sample of the original census population of schools for which a statistical test of students random allocation across classes is accepted. She finds that the immigrant class-share has weak negative effects on children test scores and that the effects are larger for pupils from disadvantaged backgrounds and negligible for native pupils from richer families.

In opposition to the aforementioned studies, Geay *et al.* (2012) and Ohinata and Van Ours (2011) do not find sizable effects. Geay *et al.* (2012) look at the association between the share of non-native English speakers in the year group and the educational attainment of native English speakers at the end of primary school in England. They use two different approaches. First, they analyse how the effect changes using selection on observables in a standard valueadded OLS model for the education production function. Then, similarly to Gould *et al.* (2009), they implement an IV approach exploiting the exogenous variation given by the influx of white non-native English speakers that happened after 2005, on account of E.U. enlargement to Eastern European countries. Using both approaches they do not find sizable negative effects. Finally, Ohinata and Van Ours (2011) analyse to what extent immigrant class share has effects on native Dutch attainments using individual level data from PIRLS and TIMSS surveys. Their estimation strategy exploits school fixed effects and does not retrieve statistically significant effects.

Differently from European studies which primarily concern first-generation immigrants, U.S. literature traditionally focused on achievement gaps between ethnic minority students (blacks and Hispanics) and white students. Only in the last decades peers' interactions have started to be seen as one the most important determinant of many observed different behaviours and outcomes between white and black students (Heckman, 2011). Early contributions were given by Evans, Oates and Schwab (1992) and Cutler and Glaeser (1997), while Hoxby (2000), Hanushek *et al.* (2009) and Hanushek and Rivkin (2009) are the first to define 'racial peer effects' as a particular group of social interactions taking place between students belonging to different ethnic groups. These works generally point to weak effects of immigrant school share on students' achievement. The effects are higher within students of the same ethnic group than between students belonging to different ethnic groups.

A common characteristic of both European and U.S. literature on peer effects is the main interest in empirical analysis. Theoretical investigation of the mechanisms of educational social interactions is usually neglected although it could help to provide consistent interpretations for the empirical evidence (De Giorgi and Pellizzari, 2011). One notable exception is Cooley (2010) who defines a structural model to explain the achievement gap between black and white students and estimates it using data from North Carolina elementary schools. In line with the reduced-form estimations of racial peer effects, she finds that endogenous peer effects within the non-native peer group are much stronger than effects between natives and non-natives.

1.3. THEORETICAL FRAMEWORK

In this Section, I propose two models of peer effects between native and non-native students. First, I adapt the 'disruption' model à la Lazear (2001) identifying the two types of students who interact in a mixed school with native and non-native students. Then, I propose the 'integration model' which extends Lazear (2001) to allow for heterogeneity in the externalities according to different intensity of the exposure of non-natives to native peers.

Disruption is a possible mechanism of peer interactions that directly influences the learning process and the attainment levels through externalities caused by peers' behaviour[7]. The basic assumptions I made are two: (*i*) one child's disruption hurts the learning process of all students (including the disruptive one); (*ii*) non-native students have a higher propensity to cause interruptions during the learning process. They are stylized assumptions on different 'behaviours' that distinguish the two types. Indeed, the 'disruption mechanism' may follow many different channels and should not be associated to non-native students' 'bad' behaviour: it could be thought as non-native students' need of additional help which causes the teacher to slow down the teaching activity, as well as non-native students' higher propensity to interrupt because of more difficulties to understand due to insufficient language skills. Descriptive evidence and discussions in Appendix A corroborate these hypotheses.

1.3.1 The 'Disruption' Model À La Lazear

Interactions between native and non-native students in the school are such that the misbehaviour of the 'more disruptive' type determines negative externalities on the learning production process which are captured by negative peer effects on per student outcome. Following Lazear (2001), I define p as the probability that any student is not hurting his own learning or other's learning at any moment in the time spent at school. Given a class size of n, the probability that disruption occurs at any moment in time t is $(1 - p^n)$. Define V as the value of a unit of learning, which is influenced by the likelihood that a student is not engaged in a disruptive behaviour in the given instant t, and Z the total number of students in the school. Then, the total output for each school is given by $Y=ZVp^n$, and the output per student by $y=Vp^n$. As discussed above, I assume that non-native students ($j=F$) tend to interrupt more frequently (on average) with respect to native peers ($j=N$), so that I can identify to types of students ($j=N$, F) according to different values of p_j ($p_N > p_F$). Finally, define $\theta<0.5$ the proportion of non-native students in each school so that type F is the 'minority type'. This is consistent with the institutional setting and data used as in Italian schools - as well as in the majority of European schools - non-native students rarely constitute the majority of the school population (Table 1, MIUR, 2009a, 2010, 2011; Eurydice, 2012). Normalizing V to 1 and holding n constant, per-student output in mixed schools is:

[7] Hoxby and Weinghart (2006), De Giorgi and Pellizzari (2011), Epple and Romano (2011), Sacerdote (2010) and McKee *et al.* (2010) point to Lazear (2001) model as one of the potential model of peer interactions in the school.

$$y^D = p_N^{(1-\theta)} p_F^{\theta} \quad [1]$$

where $y = y_N = y_F$ under assumption (*i*). The 'disruptive model' predicts that per-student output (y) is a decreasing and concave function of non-native school share (θ):

$$\frac{\partial y^D}{\partial \theta} = \left[p_N^{(1-\theta)} p_F^{\theta} \right] \ln\left(\frac{p_F}{p_N} \right) = y^D \ln\left(\frac{p_F}{p_N} \right) < 0 \quad [2]$$

$$\frac{\partial^2 y^D}{\partial \theta^2} = \frac{\partial y^D}{\partial \theta} \ln\left(\frac{p_F}{p_N} \right) > 0 \quad [3]$$

Expressions [2] and [3] describe the main characteristics of the 'disruptive model' of educational production which embeds the so-called 'bad-apple principle'. Eq. [2] states that increasing non-native school share always determines negative externalities on students' outcome so that one disruptive student (i.e. 'one bad apple') is enough to generate negative peer effects on all schoolmates. The concave relation described by eq. [2] and [3] shows that as the share of non-native students increases, the school becomes more segregated and the negative effects on per student attainment marginally decreases. The decreasing and strictly concave relation between non-native share and per-student output can be represented as in Figure 1.

[Figure 1 here]

1.3.2 The 'integration'model

The 'integration model' embeds the 'subcultural model' proposed in the U.S. sociological literature (Fordham and Ogbu, 1986; Steele and Aronson, 1998) and exploited to ground the evidence of 'acting-white' behaviours in U.S. schools (Fryer and Torelli, 2010). Anytime an integration mechanism is at work, native students exert positive externalities on non-native peers so that the difference between native and non-native students, in terms of propensity to disrupt, tend to be attenuated ($p_F \rightarrow p_N$). Integration, however, has some cost which I assume to be the effort made by native students to integrate non-native peers. Intuitively, if non-native students are relatively isolated, integration is less costly for native students (Lazear, 1999). On the contrary, anytime non-native students become prevalent enough to form a 'critical mass', the native type rejects them because the effort of integration becomes too high. The rejection may be due to different reasons: natives may be willing to make sufficient effort to include a few minority members but unwilling to make the effort to include numerous non-native schoolmates, but also unwilling to include some non-native students while rejecting others (Hoxby and Weingarth, 2006).

The ‘integration mechanism’ is formalized transforming the non-native propensity of disrupt (p_F) into a decreasing function of the proportion of non-native students θ (i.e. $p_F(\theta)$):

$$y^I = p_N^{(1-\theta)}\left[p_F(\theta)\right]^{\theta} \qquad [4]$$

Thus, define $p_F(\theta)$ as a continuous, twice differentiable function satisfying the following properties[8]:

$$p_F(\theta) = \begin{cases} p_N \text{ if } \theta = 0 \\ \overline{p}_F < p_F(\theta) < p_N \text{ if } \theta \in (0;0.5) \\ \overline{p}_F < p_N \text{ if } \theta = 0.5 \end{cases} \qquad [5.1]$$

$$p_F^{'}(\theta) = \frac{\partial p_F(\theta)}{\partial \theta} = \begin{cases} 0 \text{ if } \theta = 0 \\ \overline{p}_F^{'} < p_F^{'}(\theta) < 0 \text{ if } \theta \in (0;0.5) \\ \overline{p}_F^{'} < 0 \text{ if } \theta = 0.5 \end{cases} \qquad [5.2]$$

where $\overline{p}_F = p_F(0.5)$ and $\overline{p}_F^{'} = \left.\frac{\partial p_F(\theta)}{\partial \theta}\right|_{\theta=0.5}$.

In particular, notice that if non-native share is sufficiently small, the propensity of non-disruption of non-natives approaches the natives’ one (i.e. if $\theta \to 0^+$ then $p_F(\theta) \to p_N$ and $p_F^{'}(\theta) \to 0$) as a result of the integration process. On the contrary, if non-native share increases, the gap in the propensity of non-disruption between the two types grows (i.e. if $\theta \to 0.5^-$ then $p_F(\theta) \to \overline{p}_F$ and $p_F^{'}(\theta) \to \overline{p}_F^{'}$).

As a result, the ‘integration mechanism’ determines important differences in the predicted effects due to non-native students’ school share with respect to the simple ‘disruption model’. In fact, contrary to the ‘disruption model’ which generates strictly negative externalities (eq. [2]), the ‘integration mechanism’ allows for non-negative externalities on students’ outcomes:

[8] For example, the function $p_F(\theta)$ can be defined according to an integration index $I(\theta) = \theta / (1-\theta)$ representing the ratio between the number of non-native and native students.

$$\frac{\partial y^I}{\partial \theta} = y^I \left\{ \ln \left[\frac{p_F(\theta)}{p_N} \right] + \frac{\theta}{p_F(\theta)} p_F^{'}(\theta) \right\} \leq 0 \qquad [6]$$

In particular, the 'integration mechanism' makes the non-native peers' negative spillovers - due to the disruption mechanism - decrease for sufficiently low non-native school shares[9]:

$$\begin{aligned} &\text{if } \theta \to 0^+ \Rightarrow \frac{\partial y^I}{\partial \theta} \to 0 \\ &\text{if } \theta \to 0.5^- \Rightarrow \frac{\partial y^I}{\partial \theta} \to k < 0 \end{aligned} \qquad [7]$$

where k is a negative real number.

[Figure 2 here]

Figure 2 represents the basic intuitions from the 'integration mechanism' in eq. [6] and [7]: the dotted and dashed lines show two possible shapes of the relation between non-native school share (θ) and per student outcome (y) consistent with the 'integration model'. The dashed line is globally convex, the dotted line is convex for θ approaching zero. This implies that, even without specifying a precise form of the function $p_F(\theta)$, the externalities from 'disruptive non-native peers' are close to zero for sufficiently low values of the share of non-natives. This is because the education production function with the integration mechanism follows the predictions of the 'subcultural model' showing that the minority type can be integrated by the majority type as long as this does not entail high cost. As demonstrated by Lazear (1999), this 'integration mechanism' acts as a 'cultural acquisition' behaviour that cancels out the distinction between the two types ($p_F \to p_N$) and it is more likely to occur when the presence of non-native students in each school is below a certain 'critical mass value'[10].

To sum up, the 'disruption mechanism' predicts negative and marginally decreasing peer effects of non-native school share on per student outcome. The 'integration mechanism' mitigates these effects and predicts 'non-linear effects' with respect to non-native school share which are close to zero when non-native school share is 'sufficiently low'. I test these theoretical predictions in the empirical application exploiting a rich dataset containing census

[9] Analytical derivations in Appendix B.

[10] Lazear (1999) presents a model of 'cultural acquisition' and shows that "[…] incentives to be assimilated into the majority culture depend on the size of the relevant groups. The smaller is the minority relative to the majority, the greater is the incentive of a minority member to acquire the culture of the majority" (Lazear, 2001, p. 791).

information on test scores and administrative records on all 8th grade native and non-native students for three subsequent school years.

1.4. DATA AND DESCRIPTIVE STATISTICS

Many European countries - including Italy, Spain, U.K., Portugal and Ireland - experienced massive migration waves starting from the late Nineties. The share of foreign population has risen rapidly: from 1997 to 2007 Italy records an increase of 242 %, Spain 627%, the U.K. 92%, Portugal 147% (OECD, 2010), and, as a consequence, non-native students are nowadays a relevant part of the school population and generate a wide range of occasions for peer interactions between students of different ethnic origins.

[Figure 3 here]

This gives rise to a quantitatively large, but relatively unknown, phenomenon. For instance, in school year 1996-97 only 0.7% of students in the Italian school system had a non-Italian citizenship, while in 2010-11 the share has grown up to 7.0%, with peaks of more than 9% in primary and junior high schools, in line with average trends in most European countries (Figure 3; Eurydice, 2012). This characteristic of rapid and sheer increase in the non-native school population also improves the identification of peer effects. In fact, the students' populations used in this works are the first cohorts who have been exposed to the immigration externalities of non-native students on natives' outcomes and this help to limit confounding long-term effects (Gould *et al.*, 2009).

We exploit a unique dataset that combines the *Invalsi First Cycle Final Exam* data[11], administrative records from *Ministry of Education Statistical Office*, and the *Italian Population Census Survey 2001*. *Invalsi First Cycle Exam* (from now on 'Invalsi IC') data contain Math and Language test scores and administrative records on all students enrolled in Italian junior high schools. The census dimension of Invalsi IC tests allows to overcome problems of underrepresentation of immigrant individuals in sample surveys which lead to a substantial attenuation bias (Aydemir and Borjas, 2010). Additional information about socio-economic family background are obtained as school-level averages of Census variables linked to each school using an original matching technique that identifies for each junior high school its 'catchment area'.

Test scores in Invalsi IC data range from 0 to 100 and refer to the fraction of right answers for each of the two subjects. Three waves are available, corresponding to 2007-08, 2008-09 and

[11] INVALSI (*Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione*) is the independent public institute carrying out the evaluation of Italian school system and test students' attainment levels.

2009-10 school years final exams (about 500,000 students per wave). Individual information cover year of birth, gender, citizenship (Italian, non-Italian); how long the student is in Italy if born abroad (from primary school, for 1-3 years, less than 1 year); mother's and father's place of birth (Italy, EU, European but non-EU, other non-European country), grade retention (if the student is 'regular' i.e. if he/she is 14 years old at the end of the school year; 'in advance' i.e. younger than 'regular' students, or 'retained' i.e. older than 'regular' students). Administrative records from Ministry of Education Statistical Office provide general information about school characteristics (i.e. type of school, public vs. private, number of students enrolled and number of teachers, average class size) matched to Invalsi First Cycle data through an anonymous school identifier. Finally, Census 2001 contains information about resident population in Italy in 2001. Each school is matched to a group of census divisions through an original matching technique designed to associate to each junior high school a group of census cells constituting its 'catchment area' (Barbieri *et al.* 2011)[12]. This procedure allows matching to each junior high school variables from 2001 Population Census Survey covering a great variety of demographic and socio-economic information on resident population (gender, age, ethnic origins, education, labour force participation, occupation, households' composition and houses characteristics).

1.4.1. Descriptive statistics

We exploit the panel dimension of the dataset which is constituted by 5616 junior high schools (*s=1…5616*) and three school years (*t=2008, 2009, 2010*)[13]. Mean test scores and mean individual characteristics are obtained averaging at the school level individual level information. School characteristics are matched from Census and administrative school records as explained above. I define to as 'non-native' student an individual enrolled in the Italian school system and having both parents without Italian citizenship[14].

[Table 1 here]

Panel A in Table 1 describes the general characteristics of the schools in the dataset (number of students and schools, non-native school share, average school size, average class size) with respect to geographical macro-area and Invalsi IC waves while Panel B provides descriptive statistics concerning non-native school share, school size and average class size. The distribution of non-native students across the territory is not homogenous: in Northern and

[12] See Appendix B for a detailed description of the dataset and the matching techniques used.
[13] From the original population of 6290 schools, almost 5% are dropped because they appear in only one wave. To be consistent with the theoretical framework, we also exclude 22 schools with non-native school share in 8^{th} grade greater than 0.5. Robustness checks show that results do not change even without dropping these schools. *t=2008* stands for school year 2007-08 and so on.
[14] This definition coincides with the definition of the Ministry of Education Statistical Office.

Centre regions the average non-native school share is around 11-10% while it dramatically falls in the South (2%). Moreover, in the Appendix A I provide a short description of the regulatory framework concerning allocation rules of non-native students across classes and schools and show that non-native students are not randomly allocated across classes in the same school. Average school and class size are generally equally distributed across the IC waves although sensibly smaller in the South of the country. This suggested appropriate robustness checks performed in Section 1.7.1.

[Table 2 here]

[Table 3 here]

Table 2 shows school mean and standard deviation of test scores according to the native/non-native status: gaps between mean test scores for natives and non-natives are large and statistically significant. In Table 3 I report the coefficient of the dummy variable '*being non-native*' obtained running individual-level pooled OLS regressions on the whole sample of IC 2009 and 2010 students. I first show the raw coefficient of the unconditional attainment gap where I control only for cohort fixed effects to capture possible trends in the two IC waves: non-native students have test score lower than native peers by 11.65 points in Language, and 8.36 points in Math. Then, I progressively add controls for individual characteristics (gender, retention, parents' origins, time spent in the host country since birth), time-variant school characteristics (school size, average class size, pupil-teacher ratio) and school fixed-effects. The conditioned gaps turn out to be smaller than the unconditioned one, but still significantly different from zero: *coeteris paribus*, being non-native implies a lower test score in Language (-3.44 points) and in Math (-1.79 points).

Two main results can be drawn from general descriptive evidence. First, there exists a sizable gap in test scores results between native and non-native students which is greater in Language than in Math skills. Second, even after taking into account individual characteristics, parental background, school characteristics and territorial differences, the attainment gap is reduced but still persists. Given that the gap does not disappear controlling for usual school and family background inputs, it is plausible to think that 'social' inputs and peers' externalities may play a crucial role in explaining these gaps (Akerlof and Kranton 2002; Heckman 2011; Freyer 2011; Patacchini and Zenou, 2006 among others).

1.5. EMPIRICAL STRATEGY

The estimation of peer effects between native and non-native students must address several empirical difficulties concerning different types of students' sorting and omitted variable bias in correlated effects. First of all, non-natives' concentration in the schools may be endogenous because of households' housing decisions. One must take into account the endogenous placement of immigrants into some geographical areas that are usually more likely to be populated also by lower-achieving native students, regardless of the local level of immigrant concentration (Gould *et al*., 2009). Second, the peer group can be the result of individual and families choices: for example, given the residential choice of the household, individuals might choose a certain school on the basis of some (perceived) school quality. Third, given the school choice, the allocation of non-native students among classes is not random, but usually depends on school staff choices, previous school path, law or regulations (Ammermueller and Pischke, 2009). Besides self-selection issues, the estimation of a reduced form model retrieving the peer effect parameters is also difficult because of the problems arising from the presence of the correlated effects that will give rise to bias if they are correlated with peer group composition (Manski, 1993). The sorting processes described and the difficulty to control for all possible correlated effects may lead to a negative spurious correlation between attainments levels of native students and non-native school share, independently from the fact that non-native students actually cause some externalities on natives' outcomes.

Our estimation strategy relies on the assumption that changes in non-native school shares between adjacent cohorts within the same school are not correlated with pupils' unobservable characteristics that may be relevant in the educational production process. The strategy implemented rests on averaging procedures and selection on observables to solve the sorting mechanisms described above (sorting across classes in the same school, sorting across schools in the same areas and endogenous placement across areas) and school fixed-effects to limit possible bias due to omitted variables in correlated effects. In the empirical specification I use as outcome variable natives' per student outcome (y^N) as the focus of this work is on peer effects on natives' attainment due to non-native peers' spill-overs.

1.5.1. Baseline empirical model

We solve sorting of non-native students across classes within the same school using school level averages (Card and Rothstein, 2007) and I identify the effect of non-native school share on natives' attainment exploiting school by time variations in the data. In fact, any non-randomness due to across classes sorting would give rise to a class-specific error term correlated with the observed variables which potentially bias OLS estimates from individual-level data15. Conducting our analysis at the school level solves the sorting of non-natives across classes in the same school as long as I assume that: (i) the class-specific error component averages to zero across all classes in the school; (ii) the individual-specific error component are mean zero for all natives in each school (Card and Rothstein, 2007). Thus, I start from the following empirical specification at the school level:

$$y_{st}^{N} = \beta P_{st}^{F} + X_{st}^{N}\alpha + \varphi_s + \varphi_t + \eta_{st}^{N} \qquad [8]$$

where:

$$P_{st}^{F} = \left.\frac{non\text{-}natives_{st}}{(natives_{st} + non\text{-}natives_{st})}\right|_{grade=8} \qquad [8.1]$$

y_{st}^{N} is the school mean test score of all 8th grade native students ($j=N$) in school s and school year t; X_{st}^{N} is a vector containing mean characteristics of native students in school s and year t; φ_S are school fixed effects and the term φ_t includes cohort and territorial fixed-effects. P_{st}^{F} is the variable of interest and it is defined as the share of 8th grade non-native students in school s and year t (henceforth '*non-native school share*'). School fixed effects solve omitted variable bias in individual mean characteristics and school mean characteristics which may influence natives' attainment (i.e. the correlated effects). Notice that P_{st}^{F} is used as a proxy for average non-native peers' characteristics (X_{st}^{F}). The rationale for this stands in the fact that the share of non-native peers in the school is a good proxy of peers' characteristics but it is also predetermined with respect to the outcome measure and thus not affected by common school-level shocks (i.e. the correlated effects) (Ammermüller and Pischke, 2009; Angrist and Lang, 2004). Indeed, the use of peers average characteristics (X_{st}^{F}) may determine serious problems of collinearity with respect to individual characteristics and may be correlated with common shared variables. Moreover, because of the well-known 'Reflection Problem' (Manski, 1993), I cannot distinguish whether β reflects the exogenous effects of peers' characteristics or the endogenous effects operating through peers' achievement. Anyway, finding evidence of the 'social effects' (i.e. both endogenous and exogenous peer effects) is still of substantial policy interest (Ammermüller and Pischke, 2009).

[15] Non-random allocation of non-native students across classes is common in many European countries. Ammermueller and Pischke (2009) provide evidence for primary schools in France and Sweden. Contini (2011) finds that for approximately 22% of junior high schools in Italy the null of non-random allocation of immigrant students across classes within the same school cannot be rejected (at 10% confidence level, 6th grade).

Another important source of endogeneity that must be addressed in our empirical model is across schools sorting of non-native students. School fixed-effects and geographical area fixed effects already capture this sorting[16]. However, I also exploit the original features of our dataset and add to the specification in eq. [8] a set of school by year variables (W_{st}) which capture the socio-economic characteristics of each school catchment-area and help to control for non-natives' sorting across schools. The socio-economic variables are chosen to capture catchment-area characteristics that could have attracted immigrant families in the past and thus influence the actual non-native school share. For example, I include male and female occupation rates, population density, indicators for poor housing conditions which are considered relevant determinant of immigrants' residential choices (Boeri *et al.*, 2011). I also include the number of non-Italian residents in each school catchment area in 2001 (i.e. at the beginning of the sharp increase in the Italian immigration trend) which can be shown to be a strong predictor of the actual non-native school shares.

A final concern may arise in cases in which the variation of non-native shares across subsequent cohorts is potentially endogenous because of some sort of 'native flight' or underlying time trends (Betts and Fairlie 2003, Hoxby 2000 among others). In this case, I apply the same strategy used by Gould *et al.* (2009) and Brunello and Rocco (2011) conditioning on the total stock of non-native students in the school (i.e. the total number of non-native students in 6th, 7th and 8th grades) and on total school size (i.e. the total number of students enrolled in the school) (S_{st}). Conditioning on these variables, the share of non-native students who attend 8th grade in each school can be considered as good as random, while any residual correlation between non-native shares and school characteristics is captured by the school fixed effects. I also include in vector S_{st} relevant time-variant school characteristics such as average class-size, pupil-to-teacher ratio and 'cheating dummies'[17]. Equation [9] represents our baseline empirical specification:

$$y_{st}^{N} = \beta P_{st}^{F} + X_{st}^{N}\alpha + W_{st}\delta + S_{st}\gamma + \varphi_s + \varphi_t + \eta_{st}^{N} \qquad [9]$$

[Table 4 here]

Table 4 contains the complete list and description of the variables included in the X_{st}, W_{st} and S_{st} vectors. The estimation of β in eq. [9] allows a causal interpretation of the effect of non-

[16] Geographical area fixed-effects are in the form of interaction variables between five territorial dummies (North West, North East, Centre, South, Islands) and year dummies. In the sensitivity analysis we show that results are robust introducing up to 103 territorial dummies corresponding to school-districts and province level.
[17] See Appendix C for detailed description on how catchment-areas are built and school variables are constructed.

native school share on natives' attainment which I interpret as non-natives' peer effects on natives' attainment. If $\beta < 0$ I might conclude that the presence of non-native students cause negative externalities on the attainment of native peers and that a possible 'disruption mechanism' is at work.

1.6. RESULTS

In this section, I first present the results from the baseline specification and then I seek for non-linear effects in order to test the mechanisms illustrated in Section 1.3.

[Table 5 here]

Table 5 contains the results for the baseline model. The dependent variable is the Invalsi IC school mean test score for native students. I conduct our analysis separating Language from Math test score. The rationale for doing it being that I expect peer effects to have greater impact on Language tests because language skills are directly influenced by the use of Italian language with native peers. I progressively add school variables controls (S_{st}) in columns (II) and catchment-area socio economic variables (W_{st}) in columns (III). Thus, the coefficients estimated in columns (I) correspond to eq. [8], while the ones estimated in columns (III) to eq. [9]. Adding school and catchment-area controls significantly influences the estimates improving the school fixed-effects basic framework and limiting the possible biases due to across school sorting.

Focusing on the estimates of β from eq. [9] (columns III), I find negative and statistically significant effects and that natives' Language skills are more influenced by peers effects compared to Math. Increasing non-native school share by 1% determines a decrease of -4.85 points in native peers' Language school mean test score and -3.53 in Math, corresponding to a decrease of 0.66 standard deviations for Language and 0.35 for Math. Given that school composition usually changes a lot from primary to junior high schools (MIUR 2009, 2010), these effects can be interpreted as the result of the cumulated externalities experienced by native peers in the exposure to non-natives during the three years of junior high school. Notice also that the cohorts used in this study are actually the first to be exposed to a 'relevant' presence of immigrant students in the schools so that long-term confounding effects are limited in our setting thanks to the characteristics of the immigration waves (see Section 1.4 and Appendix A). However, non-native students can be enrolled *during* the school year or suffer higher grade retention compared to natives so that school composition could be subject to sensible changes from grade 6 to 8[18]. Because of the possibility of these changes in the non-native school composition, I can interpret the estimated peer effects as an *upper bound* of the cumulated externalities.

Our results are in line with Brunello and Rocco (2011), Jansen and Rasmussen (2011), Contini (2011) and Gould *et al.* (2009). Brunello and Rocco (2011) find that a one percentage point increase in the share of immigrant students is expected to decrease by 1.38 points the average test scores for native students (0.018 standard deviations). Contini (2011) finds that the class share of immigrant students decreases 6th grade individual test score by 0.66 standard deviations for Language and 0.14 for Math[19]. Jansen and Rasmussen (2011) find negative and significant effects only for Math: a 1% increase in immigrant school concentration reduces individual Math score by 1.05 points (0.011 standard deviations).

The baseline model estimates improve existing empirical studies under, at least, two main aspects. First, I use the universe of native and non-native students and thus do not suffer from attenuation bias (Aydemir and Borjas, 2010). This can partially explain the fact that the effects I find are greater compared to Brunello and Rocco (2011) and Jensen and Rasmussen (2011), but in line with Contini (2011). Second, aggregation at the school level ensures consistent estimates for the peer effects parameter because estimations from individual-level OLS with school fixed effects are inconsistent as long as non-native students are non-randomly allocated across classes (Ammermüller and Pischke, 2009)[20]. In any case, the magnitude of the estimated effects are only partially comparable to the aforementioned studies: Brunello and Rocco (2011) include in the sample several countries with immigration histories different from the Italy and, in general, continental Europe (such as the U.S., New Zealand, Mexico, Russia, Canada), while Jansen and Rasmussen (2011) and Contini (2011) exploit individual level data and within school variation in immigrant class shares.

1.6.1. Non linear effects: 'disruption' vs 'integration' mechanism

The theoretical framework predicts that in case the 'integration mechanism' plays a substantive role, the effects of non-native share vary substantially with respect to different levels of P_{st}^{F}. Therefore, it is crucial to test for possible non-linearity in the peer effects to

[18] In the robustness checks (Section 3.7.1) we show that grade retention does not to induce bias in the results. The presence of non-native students enrolled since less than one year is actually negligible in our sample (about 0.004%).

[19] The effects are taken from the sum of the peer variables of the share of first and second generation immigrants (Contini, 2011, Table 6).

[20] An alternative method to solve this problem would be to exclude all schools where the null of random-allocation across classes is rejected (see Contini, 2011). However, given that this is more likely to happen for schools with a limited number of non-natives, it would be more difficult to test for non-linearity in the effects and underlying social mechanisms.

distinguish which of the two mechanisms is at work. To this purpose, I introduce a linear spline functional form in non-native school share dividing the non-native school share range into two intervals with the breaking point $T \in (0; 0.5)$:

$$y_{st}^{N} = \beta_1 P_{st}^{F1} + \beta_2 P_{st}^{F2} + X_{st}^{N}\alpha + W_{st}\delta + S_{st}\gamma + \varphi_s + \varphi_t + \eta_{st}^{N}$$
$$P_{st}^{Fi} = \begin{cases} \theta_{st} & \text{if } 0 \le \theta_{st} < T \\ 1-\theta_{st} & \text{if T} \le \theta_{st} < 0.5 \end{cases} \qquad [10]$$

We accept the hypothesis that a simple ‘disruption mechanism’ is at work if two conditions hold: (*i*) peer effects are negative and statistically significant for every value of the non-native school share range (i.e. $\beta_1 < 0$ and $\beta_2 < 0$ for every value of T); (*ii*) a strictly concave relation exists between non-native school share and native educational outcome. This is because the ‘disruption mechanism’ implies that the estimated peer effects (β) should be greater for lower values of non-native school shares (i.e. $|\beta_1| > |\beta_2|$, the ‘bad apple principle’). On the other hand, I accept the hypothesis that an ‘integration mechanism’ is at work if: (*i*) peer effects are negative and statistically significant only for ‘sufficiently high’ values of *T*. (i.e. $\beta_1 = 0$ and $\beta_2 < 0$); (*ii*) the ‘integration mechanism’ entails a convex relation between non-native school share and natives’ educational outcome (at least) as $\theta \to 0$ (i.e. $|\beta_1| < |\beta_2|$ for ‘sufficiently low’ values of *T*).

To seek for structural changes in the effects I implement different values of the break point (*T*) and report the results in Table 6. This allows showing in a flexible way how effects vary above and below any given threshold. In the sensitivity analysis (Section 1.7) I present additional tests for non-linear effects in non-native school share implementing different methods.

[Table 6 here]

The effects are highly non-linear: I always reject the null that $\beta_1 - \beta_2 = 0$. Setting the threshold at the mean of the non-native school share distribution (*T*=0.068) I obtain that increasing by 1% the non-native share has not statistically significant effects if the non-native school share is below the threshold, while it decreases natives’ Language and Math test scores by almost 5 points if the share is above 6.8%. The general pattern of the results shows that the increase of non-native share has negative and statistically significant effects only for sufficiently large values of *T*. I cannot reject the null that $\beta_1 = 0$ and $\beta_2 < 0$ for *T*<0.10, while, if *T*>0.10, β_1 and β_2 are both negative and statistically significant for Language. Concerning the magnitude, effects are greater for higher values of non-native school share, thus rejecting

$|\beta_1|>|\beta_2|$ both for Language and Math. The concave relation of the 'disruption model' in is not found in the empirical estimation of the effects, which are consistent, on the contrary, with a non-linear convex relation: negative marginal effects are present only for high levels of non-native school share and are generally increasing with respect to non-native school share.

[Table 7 here]

In Table 7 I introduce a spline function with two break points, where the first one is fixed at 10% (T_1=0.10) and I set different values for the second (T_2=0.15; 0.20; 0.30)[21]. The rationale is the following: with one break point I exclude that the structural break (T) is smaller than the threshold of 10%, indeed the effects above 10% are still unclear. Effects below the threshold of 10% and between 10 and 15% are never statistically significant. Results for Language show negative and significant effects between 10 and 20%, 10 and 30% levels of non-native school share, while results for Math are negative and statistically significant only for high shares (above 20%).

Summing up and interpreting together the results from Table 6 and Table 7 I find that non-linear effects reject the hypothesis of concave relation. For Language, I cannot reject the null that $\beta_1 = 0$ and $\beta_2 < 0$ for T<0.15, while for Math the same result holds for T<0.20. Thus, our findings are more consistent with the theoretical predictions of the 'integration model' of peer effects rather than with the simple 'disruption model'. Interestingly, effects are stronger for language skills where the 'disruption' plausibly occurs more frequently given the greater difficulties to learn a non-mother-tongue for non-natives.

1.7. ROBUSTNESS CHECKS

We test the robustness of our results under three main dimensions. First, I test the robustness with respect to class-size effects and grade retention. Then, I test for possible concerns due to the main source of endogeneity (across school sorting). Finally, I show further evidence on non-linear effects.

1.7.1. Robustness to class-size effects and grade retention

Existing literature on class size effects in compulsory school reports controversial results exploiting both experimental and non-experimental data (McKee *et al.* 2010). The joint estimation of peer effects and class-size effects is hard in practice because class size could, in principle, both amplify or reduce existing social interactions if conformity type peer effects or

[21] Robustness checks for other thresholds between 0.10 and 0.30 are always consistent with these results.

oppositional behaviour are, respectively, assumed (Brock and Durlauf, 2001; Sacerdote, 2010). For instance, Graham (2008) exploit random allocation of students to large and small classes provided by the Tennessee STAR Project to estimate the intensity of peer effects in education attainments under the assumption that smaller classes intensify conformity types of social interactions among classmates. In our framework it is unlikely to disentangle the single contribution of these two channels to each student's attainment in the absence of an exogenous variation in class size. For this reason, the theoretical framework is developed assuming no class-size effects (i.e. holding n constant). In the identification strategy, given that class size effects are part of the class-specific error component, they are assumed to be averaged out in the aggregation from individual to school-level data (Section 1.5). The estimation of the empirical model takes this into account and controls for average class size and its square in all the specifications.

We perform robustness checks to ensure that class size effects do not play a substantive role in the data used. First, I noticed that the coefficients of the class size variables are never statistically significant. I also try a different specification for class size variable using a categorical variable instead of the continuous one implemented in the main specifications, but results never show differences. Then, given that average class size is relatively smaller in the South, I repeat the analysis adding an interaction between class size variables and a South dummy. Results never show statistically significant coefficients (Table 8, column I). Finally, I interact the share of non-native students with a dummy variable equal to 1 when average class-size is greater of the median value (for each cohort). If some form of class size effects are at work, I would expect a statistically significant coefficient, so that the general peer effects captured by β would be either reinforced or mitigated. Table 8 shows that this is not the case so that I can conclude that class size effects are adequately controlled for in the empirical specification and do not play a substantive role in the data at hand[22].

[Table 8 here]

Contini (2011) underlines that non-native students in Italian schools typically face higher grade retention with respect to native. Data from Ministry of Education and Invalsi define to as 'retained' a student enrolled in a lower grade with respect to his/her age. However, grade retention for non-natives may occur for three different reasons. Non-native students are enrolled in a lower grade because (*i*) they are enrolled when the school year is already started and their language proficiency is insufficient to face the grade corresponding to their age; (*ii*) because of differences due to the previous school path in a school system which does not

[22] Additional robustness checks have been made concerning class size effects and are available from the authors upon request. However, we conclude that these effects, although theoretically relevant, are reasonably controlled for in the estimation strategy and do not induce bias in the results.

overlap with the Italian one; (*iii*) because they are held back at the end of the school year for insufficient proficiency and forced to repeat the grade (see Appendix A for details on the institutional setting). In our dataset 37.33% of non-native students are retained (compared to 6.25% of natives), but I cannot distinguish which form of grade retention each student experiences. This is because data from Invalsi and official statistics from Ministry of Education do not distinguish among these different forms of grade retention for non-native students. Possible threats to the identification strategy arise only from the third type of grade retention as non-native students held back at the end of the school year (because of insufficient marks) and repeating the same grade the following year may undermine the idiosyncratic variation in non-native school shares between two adjacent cohorts.

Although I cannot disentangle the fraction of retained because held back at the end of the 8^{th} grade, surveys on immigrant students in the Italian school system (CNEL, 2011; CENSIS, 2008) show that the first and second types of grade retention are widely used from teachers and School Heads as a tool to facilitate non-natives' integration and language proficiency. More than one third of the junior high school teachers interviewed confirm that non-native students are usually allocated to a lower grade, especially if language proficiency is poor. Evidence from *ad hoc* elaborations from Ministry of Education and *ISMU Foundation* show that in the 2010-11 school year 47.9% of non-native students enrolled in junior high schools are classified as 'retained', but only 9.1% are 'grade-repeaters' because held back for insufficient proficiency at the end of the school year (MIUR, 2011). To test the robustness of our results, I perform the analysis exploiting as source of variation the difference between non-native shares in 2008 and 2010 Invalsi IC waves (i.e. dropping the observations for 2009). In this way I exclude the possibility that a fraction of the non-native school share is composed by non-native grade-repeaters held back at the end of the 8^{th} grade.

[Table 9 here]

Albeit less precisely estimated, results in Table 9 (column I) show that there are not significant differences in the effects. In column II I separate the fraction of the school share constituted by 'retained' and 'non-retained' non-native students. Retained non-natives include all three types of grade-retention. The externalities caused by 'retained' non-native students are not statistically significant, both for Language and Math test scores. On the contrary, 'non-retained' students determine negative externalities on native peers. These results confirm the robustness of the analysis as the relevant part of negative peer effects seem to be driven by the non-retained fraction of non-natives students. Moreover, the negative externalities are statistically significant only for Language test scores suggesting that initial allocation to lower grades directly befits non-natives in the improvement of language skills and indirectly benefit natives that do not receive negative externalities.

1.7.2 Robustness to across schools sorting

The identification is designed to control for across school sorting through school fixed-effects, territorial by year fixed effects and school specific catchment-area socio economic variables. To test that the identification strategy is suitable to capture this main source of endogeneity, I split the sample of schools into two groups according to school location in big or small municipalities. I define 'big municipalities' those with three or more junior high schools in their territory, while 'small municipalities' have one or two junior high schools[23]. The enrolment rules are based on residency criteria. Students have to attend the junior high school in the same municipality where they live with their family. If there is more than one school, families have to enrol their child to the school of the area where they reside. They are allowed to enrol the child to another junior high school of the municipality only if free slots are available.

Thus, the enrolment institutional framework limits *per se* across school sorting. However, 'cream-skimming' and self-selection processes are still possible and more likely to happen in big municipalities where there is a sufficiently large number of schools and families have some degree of 'choice'. On top of that, 'big municipalities' are located in more urbanized areas and benefit from higher public transportation means that could favour the commuting process to a distant junior high school, alternative to the one nearby home. Thus, I estimate separately eq. [9] on the subsample of small and big municipalities. If across school sorting is at work, the estimations should differ substantially in the two groups of schools inducing a negative spurious correlation between natives' mean test scores and non-native shares, and downward bias in the estimation of β. Given that across school sorting is more likely to happen in urban areas (i.e. big municipalities group), concerns for across school sorting would then arise if I systematically find that $\left|\beta_{big\_municip}\right| > \left|\beta_{small\_municip}\right|$.

[Table 10 here]

Estimations in Table 10 reject this hypothesis: effects are similar in the two subsamples, though slightly larger, in absolute terms, in small municipalities. An additional sensitivity check was carried out using 103 territorial dummies corresponding to junior high school districts (which also correspond to Italian provinces, NUTS5) instead of the five areas territorial dummies (North East, North West, Centre, South, Islands). School districts by year

[23] This distinction comes from Invalsi IC data (see Appendix C for details).

fixed-effects and school fixed-effects would capture any kind of across-school sorting within each school district. Results in Table 10 do not show significant differences from the baseline estimates, confirming the goodness of the baseline model estimates.

1.7.3 Tests for non linear effects in non-native school share

We progressively add to the baseline model higher order terms of the non-native school share variable to test the possible concave relation predicted by the 'disruptive model' or even any cubic or quadratic relevant relationship.

[Table 11 here]

Table 11 shows that higher order terms do not have statistically significant coefficients neither for Language nor for Math. The negative sign for the coefficient of the quadratic term in the first column further rejects the hypothetical concave relation predicted by the 'bad apple principle' in the 'disruptive model'. Then, I use the spline functional form to test whether it is possible to find evidence of statistically significant effects for some thinner intervals of the distribution of non-native school share.

[Table 12 here]

In Table 12 I let spline thresholds coincide with the 20^{th}, 40^{th}, 60^{th} and 80^{th} percentiles. This test increments the robustness of the findings concerning the use of only one threshold exogenously determined. Results show once more that negative and significant effects are concentrated in the upper deciles of the distribution of non-native school share.

1.8 CONCLUSIONS

This paper sheds light on peer effects between native and non-native students. Results are of substantial interest given the limited evidence of peer effects between natives and immigrants in European settings, and given the growing relevance on the immigration phenomenon and its impacts, not only on the labour markets, but also on the education systems. Our results contribute to the existing literature in three main aspects. First, I provide a theoretical framework to interpret the underlying social mechanisms that determine evidence of peer effects; second, I estimate the effect of non-native school share on natives' attainments identifying the peer effects parameter (β) exploiting a rich dataset covering the entire 8^{th} grade students population of native and non-native students; third, allowing for non-linear effects, I provide empirical evidence to test the stylized predictions of the theoretical framework. Increasing non-native school share by 1% determines a decrease of -4.85 points in native peers' Language test score and -3.53 in Math. These results are in line with a part of the evidence from European literature on peer effects between immigrant and native students (Brunello and Rocco, 2011; Jensen and Rasmussen, 2011; Contini, 2011). Differently from Geay *et al.* (2012) and Gould *et al.* (2009) who find long-term effects of the exposure to non-native peers, the effects I estimate can be interpreted as an upper bound of the externalities cumulated by native peers during the three years of junior high school.

Introducing non-linearity and rooting our analysis on the comparison between the 'disruptive' and the 'integration' model of education production proposed in the theoretical framework allows interpreting the results in a more precise way. The overall pattern of our findings is more consistent with the 'integration model' of peer interactions and robust under many dimensions. In fact, negative effects are concentrated only in schools with sufficiently high values of non-native school share and are not marginally increasing with respect to non-native school share (i.e. not characterized by a strictly concave relation). In particular, peer effects are close to zero for non-native school shares below 15% for Language and below 20% for Math.

This work also suggests important policy implications concerning allocation rules of non-native students across classes and across schools. Notice that policy implications would be substantially different according to the mechanism that is at work. The simple 'disruption mechanism' would entail average outcome to be maximized when schools are totally segregated by type of student. On the contrary, the 'integration mechanism' let allocation rules play a substantive role in minimizing the negative externalities and fostering the integration processes. In fact, according to the 'integration mechanism' any allocation rule should be constructed so to avoid any concentration of non-native students in the same school and rather distribute them equally. As our empirical results support this latter mechanism of social interactions between native and non-native students, I can posit that a relative isolation of non-native students from other non-native peers is beneficial for natives as it forces the integration mechanism between the two peer groups. A non-native school share below 15% in each school would help the 'integration mechanism' to be at work. For example, a recent regulation act from the Italian Ministry of Education imposes a cap threshold of 30% to non-native share in each school and class[24]. According to our findings, this threshold would be inefficiently high and may not have any effect to the educational production of Language and Math skills.

[24]See Appendix A for a details.

To conclude, the ‘disruptive mechanism of native/non-native students peer interactions’ is able to explain only a part of the empirical evidence. Once I add non-linearity, this mechanism is partially rejected by the empirical analysis which rather shows that, as long as non-native school share is sufficiently low, non-native students do not generate negative peer effects on native outcomes. Negative effects seem to be concentrated in schools where non-native students are enough to form a ‘critical mass’ so that they tend to cluster and do not integrate with native peers. The ‘integration mechanism’ could be at work where non-native share is ‘sufficiently low’ so that it is not too costly for natives to make effort to interact and integrate non-native peers, and, on the other way round, non-natives are ‘forced’ to interact with native peers. This interpretation is also in line with the general evidence of ‘acting white’ behaviours in the U.S. schools. Interestingly, all the results are stronger for Language test scores, confirming that language is more influenced by peer interactions between natives and non-natives compared to mathematical skills.

APPENDIX A. INSTITUTIONAL FRAMEWORK

Immigration flows and the consequent presence of non-native children in the Italian school system have a relatively recent history. Italy experienced only limited immigration before 1970: until the early Nineties there was a substantial internal migration (from the South to the North) and still relevant external migration. Massive immigration from North Africa first, and Eastern countries then, started in the Nineties, but sharply increased only in the last decade (Mencarini *et al.*, 2009). The foreign resident population has risen rapidly: in 1999 it only accounted for 1.9% of the total resident population, in 2008 the share of foreign residents has grown up until 7.3% (Billari and Dalla Zuanna, 2008). The same pattern can be found in the total number and the share of non-native students enrolled in the school system in the last fifteen years (Table A1). Concerning the general time trends, the variation in non-native students' population is now decreasing, after the peaks at the end of the Nineties and at the beginning of the present decade (MIUR, 2011). Students from European countries (EU and non-EU) and from Africa cover more than two thirds of the non-native students population, while students from Romania, Albania and Morocco contribute for almost 45% of the total non-native students population.

A.1. Non-native students' allocation rules

D.P.R. No. 394/1999 constitutes the reference regulatory framework concerning non-native enrollment in Italian schools. The basic elements to recall here are three: first, the right and the duty for every immigrant individual in school age, to be enrolled in the suitable school institution, independently from their legal or illegal status; second, the duty for every school to accept and enrol immigrant students in every moment of the school year; third, the competence of the School Board and Head to allocate foreign students so to avoid the "[…] constitution of classes where their presence is predominant". Non-native students should be allocated to the grade and class appropriate for their age (so called 'age-rule'). However, the School Board is allowed to allocate non-native incoming students to a lower grade depending on the native country school system, language skills, and type of school path followed in the previous school system.

To provide evidence of the non-random allocation of non-native students across classes in the same school, I calculate a dissimilarity index (D) at the school level. The index was first proposed by Duncan and Duncan (1955) and then extensively used in school and residential segregation analysis (among the others, Clotfelter, 1999; Echenique and Fryer, 2007). It provides a measure of the *evenness* in the distribution of non-native students. Given that in each school there are N_j classes ($c=1...N_j$), the dissimilarity index at the school level (D^s) measures the percentage of non-native students that would have to change class for each class to have the same percentage of non-native students as the one of the whole school (i.e. the non-native school share). In symbols:

$$D^j = \frac{1}{2}\sum_{c=1}^{N_j}\left|\frac{Natives_c}{Natives_j} - \frac{Non-natives_c}{Non-natives_j}\right| \quad \text{[A.1]}$$

where $Natives_c$ and $Non\text{-}natives_c$ represent, respectively, the total number of native and non-native students in class c of school j, and $Natives_j$ and $Non\text{-}natives_j$ represent the total number of native and non-native students in school j. D^j ranges from 0 (perfectly even distribution, meaning 'no segregation') to 1 (perfectly uneven distribution, i.e. 'maximum segregation'). The graph box in Figure A.1 portraits the results distinguishing among three geographical macro areas. The distribution of non-native students across classes cannot be considered even: median value of the dissimilarity index is 15% in the North, 17% in the Centre and 13% in the South, so that, for example, in the North on average 15% of non-native students has to be reallocated from one class to another to obtain an 'even' distribution within the school.

In January 2010, the Italian Ministry of Education introduced a new rule for the allocation of non-native students within classes and schools, establishing that classes and schools should not contain more than 30% of non-native students (i.e. students with non-Italian citizenship)[25]. The idea behind the implementation of such a threshold is to avoid social segregation in the schools and in the classes within schools, especially in areas where immigrant population is particularly high. The rule is enforced starting from the first-grade-classes of primary, lower and upper secondary schools of the 2010-11 school year. Its impact is not huge but still relevant, especially in the North and Centre of Italy: in Lombardy, for example, more than 29% of the classes in the junior high schools have a concentration of more than 30% of non-native students (the percentage decreases to the 27% if I consider only non-native students born abroad) (MIUR, 2010).

A.2. Non-native students' behaviour at school

The theoretical framework proposed in Section 1.3 hinges upon two main assumptions concerning the school behaviour of native and non-native students: (*i*) non-native students are

[25] "*Indicazioni e raccomandazioni per l'integrazione di alunni con cittadinanza non italiana*", MIUR, Circolare Ministeriale No. 2/2010 (C.M. 8/1/2010, n. 2).

more disruptive compared to native peers ($p_N > p_F$); (*ii*) disruption has similar effects on all students. In this section, I provide descriptive evidence to corroborate these hypotheses. International studies show that language, culture and previous school path negatively affect non-natives' school performance and behaviour (among others: OECD 2010, Stanat and Christensen 2006, Schnepf 2007, Dustman and Glitz, 2011). There is also direct evidence of the fact that minority students show lower discipline with respect to natives (Kinsler, 2010). Existing studies applying Lazear (2001) model to empirical estimates use the fraction of low-income students in the school (i.e. students eligible for subsidized lunch) as a proxy for the fraction of the students with disruptive behaviour (Mueller, 2011; Mc Kee *et al.*, 2010). Mc Kee *et al.* (2010) justify this assumption grounding on indirect evidence from the Early Childhood Longitudinal Survey where kindergarten teachers were asked to whether the level of child misbehavior interferes with teaching activities. The authors show that higher proportions of teachers agree or strongly agree with the statement are in schools with higher shares of students eligible for subsidized meals.

Concerning the Italian context, evidence on non-native students behaviour at school can be drawn from two surveys about non-native students integration in the school system (CENSIS, 2008) and non-native adolescents integration in society (CNEL, 2011)[26]. CENSIS (2008) survey interviewed a national representative sample of 414 teachers in schools with non-native students and 608 immigrant households. Teachers were asked which kind of problems the presence of non-native students in the school entails on learning processes. Results reported in Table A.2 show that the main difficulties mentioned by teachers concern language difficulties in communicating with non-natives, slowing down the teaching activities and adapting the teaching activities to non-natives. Moreover, 83% of the teachers declare to have difficulties in communication with non-native students because of the language, 73% of the teachers undertake specific activities to help non-native students in catching up native attainment levels. CNEL (2011) survey interviewed a national representative sample of native and non-native students with the aim of assessing non-native adolescents integration in society. The results confirm that non-native have more difficulties at school: shyness, language and discipline are important factors determining these difficulties. The representative sample of non-natives interviewed declared to have had attainment difficulties at school (43.3%), difficulties in interactions with classmates (33.3%) and teachers (24%). In particular, they claim that difficulties in interactions are especially due to language (30.2%),

[26] CNEL (Consiglio Nazionale dell'Economia e del Lavoro) is the National Bureau for Economics and Labour Research; CENSIS (*Centro Studi Investimenti Sociali*) is a foundation carrying out socio-economic research since 1964.

integration (28%) and discipline problems (44.5%)[27] (CNEL, 2011). Thus, descriptive evidence from both surveys is supportive of the first assumption showing that, on average, non-native students cause more disruption compared to natives.

Additional evidence is obtained with the micro-data of the National Evaluation Program carried out by the Invalsi starting from school year 2009-10 on all 5th and 6th grade students enrolled in elementary and junior high schools in Italy (Invalsi, 2010b). I exploit information from 6th grade *Students' Questionnaire* (6th grade students are enrolled in the first year of the junior high school). Table A.3 shows the share of native/non-native, low/high ability students who agree or strongly agree with statements concerning personal difficulties in studying (statements a.1 and a.2) and personal evaluations concerning the slowing down of the learning activities (statements b.1 and b.2). High ability and low ability students are classified according to whether the teachers' mark for each student at the end of the first semester (in late January) is above or below the median. Non-native students suffer more difficulties in learning activities (especially in Language, statement a.2) but experience in the same way as native peers the slowing down of the teaching activity (statements b.1 and b.2). High and low ability students have different feelings about personal difficulties in learning activities (statements a.1 and a.2) but report the same impressions of the slowing down of the teaching activity (b.1 and b.2). The pattern of these answers thus supports both assumptions: non-native students feel greater difficulties in learning activities and plausibly cause more interruptions during the lectures; the consequences of interruptions and disruptive behaviours affect in the same way native and non-native, and high and low ability students.

APPENDIX B. ANALYTICAL DERIVATION

Recall the education production function with 'integration mechanism' (y^I) and its first derivative with respect to non-native school (eq. [4] and [5]) and complete the properties and definitions concerning $p_F(\theta)$ with the properties of the second derivative:

$$p_F''(\theta) = \frac{\partial^2 p_F(\theta)}{\partial \theta^2} = \begin{cases} 0 \text{ if } \theta = 0 \\ < 0 \text{ if } \theta \in (0; 0.5) \\ \overline{p}_F'' < 0 \text{ if } \theta = 0.5 \end{cases} \quad \text{[B.1]}$$

[27] Results are statistically different (at 5 or 10% level) with respect to the same answers given by a representative control group of natives.

where: $\overline{p}_F^{''} = \left.\frac{\partial^2 p_F(\theta)}{\partial\theta^2}\right|_{\theta=0.5}$.

The second derivative of y^I with respect to θ takes the following form:

$$\frac{\partial^2 y^I}{\partial\theta^2} = \frac{\partial y^I}{\partial\theta}\left\{\ln\left[\frac{p_F(\theta)}{p_N}\right] + \frac{\theta}{p_F(\theta)}p_F^{'}(\theta)\right\} + \\ + y^I\left\{\frac{p_F^{'}(\theta)}{p_F(\theta)} + \frac{p_F^{'}(\theta)}{p_F(\theta)} + \theta\frac{p_F(\theta)p_F^{''}(\theta) - p_F^{'}(\theta)p_F^{'}(\theta)}{[p_F(\theta)]^2}\right\} = \\ = \frac{\partial y^I}{\partial\theta}\left\{\ln\left[\frac{p_F(\theta)}{p_N}\right] + \theta\frac{p_F^{'}(\theta)}{p_F(\theta)}\right\} + \frac{y^I}{p_F(\theta)}\left\{2p_F^{'}(\theta) + \theta p_F^{''}(\theta) - \theta\frac{\left[p_F^{'}(\theta)\right]^2}{p_F(\theta)}\right\} \tag{B.2}$$

Then, for $\theta \to 0^+$:

$$\lim_{\theta\to 0^+}\frac{\partial^2 y^I}{\partial\theta^2} = \frac{\partial y^I}{\partial\theta}\left\{\ln\left[\frac{p_F(\theta)}{p_N}\right] + \theta\frac{p_F^{'}(\theta)}{p_F(\theta)}\right\} + \frac{y^I}{p_F(\theta)}\left\{2p_F^{'}(\theta) + \theta p_F^{''}(\theta) - \theta\frac{\left[p_F^{'}(\theta)\right]^2}{p_F(\theta)}\right\} = 0^- \tag{B.3}$$

For $\theta \to 0.5^-$ the second derivative is different from zero, but undetermined as it depends on the values $\overline{p}_F, \overline{p}_F^{'}, \overline{p}_F^{''}$:

$$\lim_{\theta\to 0.5^-}\frac{\partial^2 y^I}{\partial\theta^2} = (p_N\overline{p}_F)^{1/2}\left[\ln\frac{\overline{p}_F}{p_N} + \frac{1}{2}\frac{\overline{p}_F^{'}}{\overline{p}_F}\right]^2 + \left[2\overline{p}_F^{'} + \frac{1}{2}\overline{p}_F^{''} - \frac{1}{2}\frac{(\overline{p}_F^{''})^2}{\overline{p}_F}\right] \neq 0 \tag{B.4}$$

The sign of the second derivative globally depends on $p_F(\theta)$ functional form. However, it is possible to derive its sign for $\theta \to 0^+$ that together with the information on first derivative is sufficient for an horizontal inflection point to exist in a neighbourhood of $\theta = 0^+$ (assuming $\left.p_F^{''}(\theta)\right|_{\theta=0} \neq 0$). These results allow to draw the qualitative graphs in Figure 2 which shows the decreasing slope and the horizontal inflection point in a neighbourhood of $\theta=0^+$, but undetermined concavity or convexity for $\theta>0$.

APPENDIX C. DETAILED DATASET DESCRIPTION AND CONSTRUCTION

We match three datasets. The first contains individual level information on each 8th grade student who attended an Italian junior high school and sit the Invalsi *First Cycle Final Exam* (Invalsi IC) in school years 2007-08, 2008-09 and 2009-10; the second contains school level information from administrative records from Ministry of Education Statistical Office; the third collects information of each school 'catchment-area' from Census 2001. To our knowledge, this is the first time that a dataset with such a variety of information and covering the universe of 8th grade students is made available for the Italian school system.

Individual level information. Invalsi IC data are the first experience of standardized test scores census survey taken on all Italian students. The 'First Cycle Final Exam' was conducted since 2007-08 school year. However, only starting from the 2009-10 s.y. test scores contribute for one sixth of the junior high school final grade. The dataset contains test scores and individual information on about 1,504,286 8th grade students, aged between 13 and 14, who took the Invalsi standardized tests at the end of the 'first cycle' of compulsory education (i.e. after five years of primary education and three years of junior high school). Math and Italian Language tests take place in June. Each part usually lasts one hour and between Language and Math test students have a fifteen minutes break. Data contain separate test scores for Maths and Italian Language ranging from 0 to 100 (percentage of right answers), and individual information is provided by the school administrative staff through school records (thus, not directly asked to students). Because of cheating evidence (Invalsi 2008,2009, 2010a), for each student I have both the raw and cheating-corrected Maths and Language test score. Sensitivity analysis confirms that raw and cheated-corrected results almost coincide once I control for geographical differences (i.e. I introduce in the model macro-area, regional or province dummies). Therefore, I use raw test scores, add geographical controls and a subject and school specific dummy indicating if the school has an high-cheating evidence. The 'high cheating dummy' is calculated starting from cheating coefficients obtained through a fuzzy-logic correction procedure explained in detail in Invalsi (2010a) Appendix 9. The dummy identifies the schools with heavy evidence of cheating behaviours (it takes value 1 if the school is in the lowest decile of the distribution of the subject specific cheating coefficient). Robustness checks replicate the construction of the 'high-cheating dummy' with other percentiles (1-5, 1-15, 1-20) without showing differences in the results.

School level information. Invalsi and Ministry of Education Statistical Office provided us with additional school level information. For each junior high school I know: ownership (i.e. state school or private institution), administrative organization (i.e. whether it is an institute having both elementary and junior high schools, or whether it is a junior high school, administratively independent from other elementary schools); the province where the school is located; the total number of students enrolled in 6, 7 and 8 grade, and the total number of classes for each grade; the total number of teachers hired in the school; the total number of support teachers for students with handicaps or language difficulties; the number of students with disabilities for each grade. Because of restrictions imposed by Privacy Law, I have the information of the municipality only in the case in which the school is located in a municipality with at least three junior high schools.

Catchment-area information. For each junior high school I define a 'catchment area' which identifies the area where the majority of school attendants live. A catchment area is composed by a number of census divisions linked to each school according to a given algorithm. The procedure for the association between school and census divisions assigns for each school the closest divisions (in terms of geographic distance) so that the 'relevant resident population' living in those divisions contains at least $k>1$ times the number of students enrolled in that particular school (Barbieri *et al.*, 2011, Appendix A). The 'relevant population' is defined according to the 10-14 years resident population in the census data, while the multiplicative factor k is set equal to ten and it allows the overlapping of census divisions among different (but geographically not distant) schools. As a result, the matching procedure links each school j with N_j census divisions constituting its 'catchment area'. For each school j the socio-economic background variables are obtained as average of the socio-economic variables of the school catchment area from 2001 Italian Population Census Survey.

Missing data correction. Missing values in school and catchment-area variables are due to the construction of the dataset. This fact would cause the number of schools in the regression estimates to shrink from 5611 in the estimation of eq. [8] in Table 5 to 4823. The variables containing missing values are two: the 'stock of non-native students' in the school (included in vector S_{st}) and the set of school specific catchment area variables (W_{st}). Preliminary analysis with probit regressions exclude any particular pattern in missing values due to geographical school location. The variable 'stock of non-native students' in the school is missing for 16% of schools due to school register data missing. I correct this variable replacing the missing values with the total stock of 7th grade non-natives students, one year lagged, from a different Invalsi data source. The correction replace all missing values. Catchment-area variables are missing for 6.3% of the schools. This is because the matching procedure between the school identifier and the census cells failed due to some non-perfect overlapping between the school identifier in the Invalsi data and the one in the Census data. I replaced the missing values of the socio-economic variables of the school catchment-area with the average value of the same variables taken from the schools which are located in the same municipality. This correction procedure shrinks missing data on catchment-area variables from 6.3% to 4.6% of schools. Table C.1 shows that implementing the correction procedures allows keeping all the observations but does not modify previous results, which, in turn, are not due to some selection pattern in the missing data.

TABLES CHAPTER 1

Table 1. Descriptive statistics. School level characteristics.

Panel A	2008			2009			2010		
	North	Centre	South	North	Centre	South	North	Centre	South
No. Students	201,650	89,870	204,339	208,575	91,639	200,643	205665	90993	197675
No. Schools	1762	832	2196	1837	857	2225	2080	906	2246
% Non-native students	0.10	0.09	0.02	0.13	0.11	0.02	0.11	0.10	0.02
Avg. No. Students per School	331.95	340.65	297.88	334.31	341.15	299.00	326.09	340.31	300.36
Avg. No. Students per Class	20.96	21.20	19.35	20.61	20.85	19.75	21.45	21.03	20.15
Panel B	Mean	Sd	P25	P50	P75	P95	Max	Min	N
% Non-native students	0.0683	0.0751	0.0053	0.0449	0.1075	0.2143	0.5	0	14941
Avg. No. Students per School	318.71	197.68	172	267	424	718	1340	11	14941
Avg. No. Students per Class	20.49	3.14	18.9	21	22.6	24.6	30	7	14941

Table 2. Descriptive statistics. Invalsi IC school mean test scores for native and non-native students.

	Language test							
	Native				Non-native			
	Mean	Sd	Max	Min	Mean	Sd	Max	Min
2008	68.5	6.22	93.87	16.57	59.23	11.57	98	8
2009	66.56	8.13	96.75	20	52.83	14.72	100	2.5
2010	64.97	6.97	89.3	0	55.6	10.92	100	0.76
	Math test							
	Native				Non-native			
	Mean	Sd	Max	Min	Mean	Sd	Max	Min
2008	53.92	8.73	92.73	8.77	47.31	12.1	100	9.09
2009	66	9.34	97.93	14.81	55.94	15.29	100	0
2010	55.56	8.04	88.09	24.72	49.65	10.85	95	0

Notes. Test scores range from 0 to 100 (percentage of right answers). The difference between test score means of native and non-native students is always statistically different from zero (p.val≤0.001); the ratio between test score sd of native and non-native students is always statistically different from one (p.val≤0.001).

Table 3. Gap in individual test scores between native and non-native students. Pooled OLS regressions on individual level Invalsi IC 2009-2010 data.

	Dep. variable: individual test score					
	Language			Math		
	(I)	(II)	(III)	(I)	(II)	(III)
Non-native	-11.6526***	-3.3199***	-3.4478***	-8.3664***	-2.6484***	-1.7929***
	(-0.121)	(-0.2072)	(-0.1892)	(-0.1378)	(-0.2665)	(-0.1952)
R sq.	0.064	0.1	0.199	0.171	0.19	0.3
Clusters	5514	5514	5514	5514	5514	5514
N	995190	995190	995190	995190	995190	995190
Cohort FE	yes	yes	yes	yes	yes	yes
Individual characteristics		yes	yes		yes	yes
School characteristics and school FE			yes			yes

Notes. Robust standard errors in parenthesis clustered at the school level. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are obtained from the dummy variable 'being non-native' through pooled OLS regressions performed at the individual level for Invalsi IC 2009 and 2010 waves. Individual control variables include dummies for gender, grade retention, having a father born in Italy, having a mother born in Italy, living in Italy since birth, living in Italy since elementary school, living in Italy since one year. School characteristics include total number of students per school and its square, average number of students per class and its square, pupil to teacher ratio.

Table 4. Control variables list: type (individual, school, catchment area) and description.

Name	Description
Individual level (X)	
female	Fraction of native females in school s (grade 8)
late	Fraction of native students retained in school s (grade 8)
father place of birth	Fraction of native students in school s and grade 8 with father born abroad
mother place of birth	Fraction of native students in school s and grade 8 with mother born abroad
elementary_Italy	Fraction of native students in school s grade 8 in Italy since elementary school
always_italy	Fraction of native students in school s grade 8 in Italy since birth
School level (S)	
Non-native school share	Fraction of non-native students in school s and grade 8.
nonnatives_stock	Total number of non-native students in the school (6, 7 and 8 grade)
Pt_ratio	Pupil to teacher ratio (8 grade)
school_size and school_size2	School size (total number of students in the school, 6, 7 and 8 grade) and its square.
avg_class and avg_class2	Average class size (average number of students in each 8 grade class) and its square.
Cheating_dummy_math and cheating_dummy_language	Dummy equal 1 if the school is in the 9th decile of the school cheating coefficient distribution (subject specific)
Catchment area level (W)	
lpop	Log of total resident population
illiterate	Fraction of illiterate pop.
university_edu	Fraction of pop. with university level education
m_occup_rate	Male occupation rate
f_occup_rate	Female occupation rate
foreign_citizens	No. of non-Italian residents
agri_oc	Fraction of workers occupied in agriculture
self_empl	Fraction workers self-employed
commuter	Fraction of residents commuting every day
avg_family_members	Average number of family members
house_poor	Fraction of houses without clean water
house_new	Fraction of houses built after 1980
avg_rooms	Average number of rooms per house

Table 5. Baseline model. Results from OLS regressions with school fixed-effects.

	Dep. Variable: School Mean Score for Native students					
	Language			Math		
	(I)	(II)	(III)	(I)	(II)	(III)
Non-native school share	-5.4294***	-4.7448***	-4.8530***	-4.7090**	-3.4601*	-3.5322*
	(1.5697)	(1.5370)	(1.5251)	(1.9606)	(1.8583)	(1.8465)
R sq.	0.206	0.316	0.325	0.512	0.631	0.633
Clusters	5611	5611	5611	5611	5611	5611
N	14941	14941	14941	14941	14941	14941
Individual characteristics, school FE and cohort FE	yes	yes	yes	yes	yes	yes
School characteristics		yes	yes		yes	yes
Catchment area characteristics			yes			yes

Notes. Robust standard errors in parenthesis clustered at the school level. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. See Table 4 for control variables list and description.

Table 6. Non-linear effects. Results from OLS regressions with school fixed-effects and spline linear functions with one structural break (T).

	Dep. Variable: School Mean Score for NATIVE students				
	Language				
	T=0.045 (P50)	T=0.068 (mean)	T=0.10 (≈P75)	T=0.15 (≈P90)	T=0.20 (≈P95)
Share < T	-3.0525	-1.9672	-2.0061	-3.3155*	-4.0949**
	(5.5378)	(3.8010)	(2.6718)	(1.9491)	(1.6728)
Share > T	-5.2763***	-5.9787***	-6.9270***	-7.6722***	-7.6323*
	(1.7303)	(1.8702)	(2.1805)	(2.9341)	(4.1333)
R sq.	0.325	0.325	0.325	0.325	0.325
Clusters	5611	5611	5611	5611	5611
N	14941	14941	14941	14941	14941
	Math				
	T=0.045 (P50)	T=0.068 (mean)	T=0.10 (≈P75)	T=0.15 (≈P90)	T=0.20 (≈P95)
Share < T	-0.2107	1.6601	1.6118	-0.0889	-1.4300
	(6.5980)	(4.6177)	(3.2537)	(2.3788)	(2.0450)
Share > T	-4.0741*	-5.2888**	-7.2790***	-9.2423**	-11.2379**
	(2.1109)	(2.3010)	(2.6745)	(3.5908)	(5.0951)
R sq.	0.633	0.633	0.633	0.633	0.633
Clusters	5611	5611	5611	5611	5611
N	14941	14941	14941	14941	14941
All Controls	yes	yes	yes	yes	yes

Notes. Robust standard errors in parenthesis clustered at the school level. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. See Table 4 for control variables list and description.

Table 7. Non-linear effects. Results from OLS regressions with school fixed-effects and spline linear functions with two break points (T1=0.10 and T2=0.15, 0.20, 0.30).

	Dep. Variable: School Mean Log Score for NATIVE students					
	Language			Math		
	T1=0.10 T2=0.15	T1=0.10 T2=0.20	T1=0.10 T2=0.30	T1=0.10 T2=0.15	T1=0.10 T2=0.20	T1=0.10 T2=0.30
Share < T1	-2.1148	-2.0087	-2.1650	0.9186	1.1445	1.4407
	(2.8096)	(2.7402)	(2.6924)	(3.3982)	(3.3171)	(3.2696)
T1<Share<T2	-6.4917	-7.1019**	-6.3709***	-2.7538	-4.7639	-6.1761**
	(4.4532)	(2.8148)	(2.2355)	(5.4557)	(3.4608)	(2.7377)
Share > T2	-7.1952**	-6.8600	-10.7928	-8.8420**	-10.1483**	-12.8443
	(3.0312)	(4.2071)	(8.8243)	(3.6920)	(5.1695)	(9.4230)
R sq.	0.325	0.325	0.325	0.633	0.633	0.633
Clusters	5611	5611	5611	5611	5611	5611
N	14941	14941	14941	14941	14941	14941
All Controls	yes	yes	yes	yes	yes	yes

Notes. Robust standard errors in parenthesis, clustered at the school level. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. See Table 4 for control variables list and description.

Table 8. Sensitivity analysis. Robustness to class size variations.

	Dep. Variable: School Mean Score for NATIVE students			
	Language		Math	
	(I)	(II)	(I)	(II)
Non-native school share	-4.8621***	-5.0857***	-3.5259*	-4.4172**
	(1.5262)	(1.7513)	(1.8490)	(2.1916)
Average class size	-0.0658	-0.0782	0.0456	-0.0041
	(0.3515)	(0.3569)	(0.3741)	(0.3764)
Average class size sq.	0.0027	0.0027	0.0019	0.0022
	(0.0072)	(0.0072)	(0.0077)	(0.0076)
Average class size * South dummy	0.0067	0.0173	0.2843	0.3267
	(0.3841)	(0.3879)	(0.4071)	(0.4078)
Average class size sq. * South dummy	-0.0013	-0.0013	-0.0056	-0.0059
	(0.0072)	(0.0072)	(0.0077)	(0.0076)
Non-native school share * big class dummy		0.4824		1.9226
		(1.4405)		(1.7893)
R sq.	0.325	0.325	0.633	0.633
Clusters	5611	5611	5611	5611
N	14941	14941	14941	14941
All Controls	yes	yes	yes	yes

Notes. Robust standard errors in parenthesis clustered at the school level. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. See Table 4 for control variables list and description.

Table 9. Sensitivity analysis. Robustness to retained non-native students.

	Dep. Variable: School Mean Score for NATIVE students			
	Language		Math	
	(I)	(II)	(I)	(II)
Non-native school share	-3.6392**		-3.3911	
	(1.8225)		(2.3977)	
'Retained' non-native students school share		0.2512		-1.3817
		(2.4162)		(3.1068)
'Non-retained' non-native students school share		-8.4408***		-5.8727
		(2.9118)		(3.8812)
R sq.	0.412	0.413	0.348	0.349
Clusters	5592	5592	5592	5592
N	10022	10022	10022	10022
All Controls	yes	yes	yes	yes
Only 2008 and 2010 cohorts	yes	yes	yes	yes

Notes. Robust standard errors in parenthesis clustered at the school level. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. See Table 4 for control variables list and description.

Table 10. Sensitivity analysis. Robustness to across-schools sorting.

	Dep. Variable: School Mean Score for NATIVE students		
	Language		
	Big municipalities	Small municipalities	Province by year FE
Non-native school share	-4.7285***	-6.8041**	-4.1518***
	(1.7199)	(3.4487)	(1.5099)
R sq.	0.331	0.331	0.362
Clusters	3903	1085	5611
N	11094	3012	14941
	Math		
	Big municipalities	Small municipalities	Province by year FE
Non-native school share	-3.1641	-7.0337	-3.8518**
	(1.9292)	(4.8337)	(1.8761)
R sq.	0.647	0.608	0.651
Clusters	3903	1085	5611
N	11094	3012	14941
All Controls	yes	yes	yes

Notes. Robust standard errors in parenthesis clustered at the school level. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. See Table 4 for control variables list and description.

Table 11. Sensitivity analysis. Non-linear effects adding higher order polynomials of non-native school share.

	Dep. Variable:School Mean Score for NATIVE students		
		Language	
Non-native school share (θ)	-1.9658	-3.5082	-0.7507
	(2.8974)	(4.9352)	(7.4583)
θ^2	-10.6773	2.2747	-38.3617
	(9.7352)	(33.5028)	(85.5224)
θ^3		-23.9852	152.1057
		(61.7825)	(351.1655)
θ^4			-221.6964
			(449.9035)
R sq.	0.325	0.325	0.325
Clusters	5611	5611	5611
N	14941	14941	14941
		Math	
Non-native school share (θ)	1.9595	2.1452	-3.8873
	(3.4891)	(5.7655)	(8.7567)
θ^2	-20.3081*	-21.8670	67.0276
	(11.5681)	(37.6217)	(97.7579)
θ^3		2.8866	-382.3169
		(66.0874)	(388.8570)
θ^4			484.9629
			(477.8939)
R sq.	0.633	0.633	0.634
Clusters	5611	5611	5611
N	14941	14941	14941
All Controls	yes	yes	yes

Notes. Robust standard errors in parenthesis clustered at the school level. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. See Table 4 for control variables list and description.

Table 12. Sensitivity analysis. Spline functions with intervals of five percentiles.

	Dep. Variable: School Mean Score for NATIVE students	
	Language	Math
pc1 (θ)	.	.
pc2 (θ)	-15.9934	-18.2352
	(11.6235)	(13.5156)
pc3 (θ)	5.1574	9.9346
	(7.0422)	(8.6081)
pc4(θ)	-3.9617	1.0133
	(3.8899)	(4.7890)
pc5(θ)	-6.9353***	-8.2250***
	(2.4676)	(2.9980)
R sq.	0.325	0.634
Clusters	5611	5611
N	14941	14941
All Controls	yes	yes

Notes. The first percentile is the omitted category. Robust standard errors in parenthesis clustered at the school level. Significance level: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. See Table 4 for control variables list and description.

Table A1. Non-native students in the Italian school system.

School Year	*All levels*		*Kindergarten*		*Primary school*		*Junior high*		*High- school*	
	Total No.	%	Total No.	%	Total No.	%	Total No.	%	Total No.	%
1996-97	59389	0.7	12809	0.8	26752	1.0	11991	0.6	7837	0.3
…	…	…	…	…	…	…	…	…	…	…
2001-02	181767	2.3	39445	2.5	84122	3.0	45253	2.5	27594	1.1
2002-03	232766	3.0	48072	3.0	100939	3.7	55907	3.1	34890	1.3
2003-04	282683	3.5	59500	3.6	123814	4.5	71447	4.0	52380	2.0
2004-05	361576	4.2	74348	4.5	147633	5.3	84989	4.7	63833	2.4
2005-06	424683	4.8	84058	5.0	165951	5.9	98150	5.6	83052	3.1
2006-07	501445	5.6	94712	5.7	190803	6.8	113076	6.5	102829	3.8
2007-08	574133	6.4	111044	6.7	217716	7.7	126396	7.3	118977	4.3
2008-09	629360	7.0	125092	7.6	234206	8.3	140050	8.0	130012	4.8
2009-10	673592	7.5	135632	8.1	244.457	8.7	150279	8.5	143224	5.3
2010-11	711046	7.9	144628	8.6	254.644	9.0	158261	8.8	153513	5.8

Notes. Elaboration from MIUR-ISMU Foundation (2011). Primary school (grades 1-5); Junior high school (grades 6-8); high-school (grades 9-13).Children enrolled in kindergartens are from 3 up to 5 years old and start primary school the month of September of the year they turn 6.

Table A.2. Descriptive statistics. The three main problems experienced by teachers in approaching non-native students.

Main problems faced by teachers	Average
North-West	
Difficulties in communication because of the language	2.7
Problematic family background	2.5
Slowing down teaching activities	2.3
North-East	
Problematic family background	2.9
Difficulties in communication because of the language	2.8
Adapting teaching activities to non-native students	2.6
Centre	
Difficulties in communication because of the language	3.1
Adapting teaching activities to non-native students	2.7
Slowing down teaching activities	2.7
South and Islands	
Difficulties in communication because of the language	2.7
Slowing down teaching activities	2.1
Problematic family background	2.1

Notes. Elaboration from CENSIS (2008, table 13), "Main problems faced by teachers in approaching non-native students, distribution by geographical macro-area". Average points: 1 means "no problems", 4 means "a lot of problems". I report the three answers with the highest average points.

Table A.3. Descriptive statistics. Evidence on the theoretical framework behavioural assumptions.

	Native	Non-native	High ability	Low ability	All
(a.1) "Studying Math is more difficult for me than for others"	29.76	37.54	23.26	27.17	30.54
(a.2) "Studying Language is more difficult for me than for others"	24.48	39.36	19.9	33.9	25.96
(b.1) "During Math lessons, we dedicate a lot of time to the same issue because class-mates do not understand"	58.89	58.11	59.88	57.54	58.81
(b.2) "During Language lessons, we dedicate a lot of time to the same issue because class-mates do not understand"	47.83	47.49	48.33	46.36	47.8
N	462,390	51,347	296,550	217,187	513,737

Notes. The data are taken from the Student Questionnaire of the Invalsi National Evaluation Program, s.y. 2009-10. The population refers to all 6^{th} grade students enrolled in Italian junior high schools. High ability and low ability students are classified according to whether the teachers' mark for each student at the end of the first semester (late January) is above or below median mark for all students.

Table C.1. Robustness to missing data correction.

	School Mean Score for NATIVE students	
	Language	
	Baseline	Without correction for missing data
Non-native school share	-4.8530***	-5.4214***
	(1.5251)	(-1.5641)
R sq.	0.325	0.328
Clusters	5611	4823
N	14941	13820
	Math	
	Baseline	Without correction for missing data
Non-native school share	-3.5322*	-4.4357**
	(1.8465)	(-1.8926)
R sq.	0.633	0.635
Clusters	5611	4823
N	14941	13820
All Controls	yes	yes

Notes. Robust standard errors in parenthesis clustered at the school level. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Correction for missing data is explained in Appendix C. See Table 4 for control variables list and description.

FIGURES CHAPTER 1

Figure 1. The 'disruption model'. The figure shows the concave relation between non-native school share (θ) and per student output (y) in the 'disruption model'.

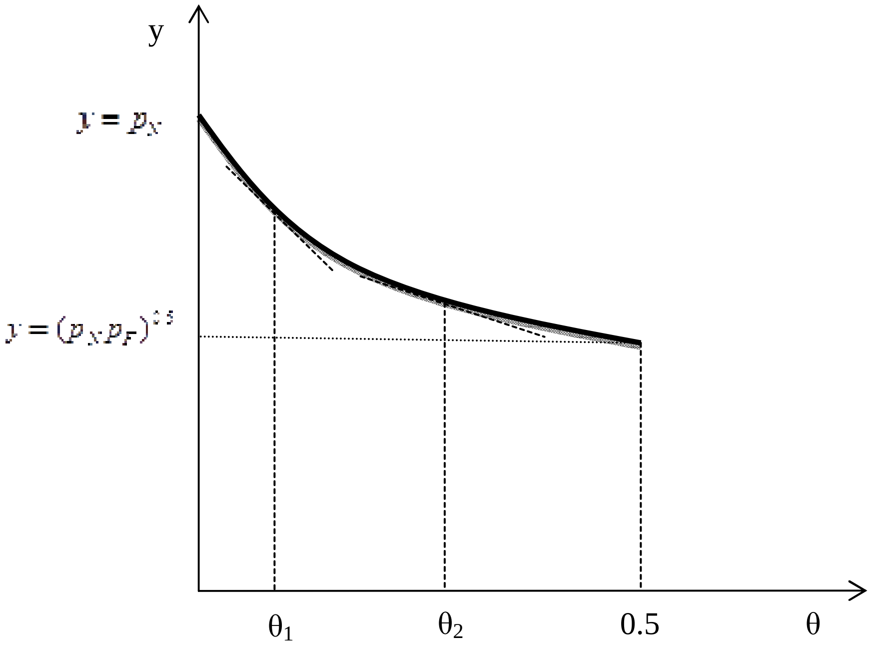


Figure 2. The ‘integration model’. The dotted and dashed lines in the figure show two possible shapes of the relation between non-native school share (θ) and per student output (y) consistent with the ‘integration model’. The dashed line is globally convex, the dotted line is convex for θ approaching zero.

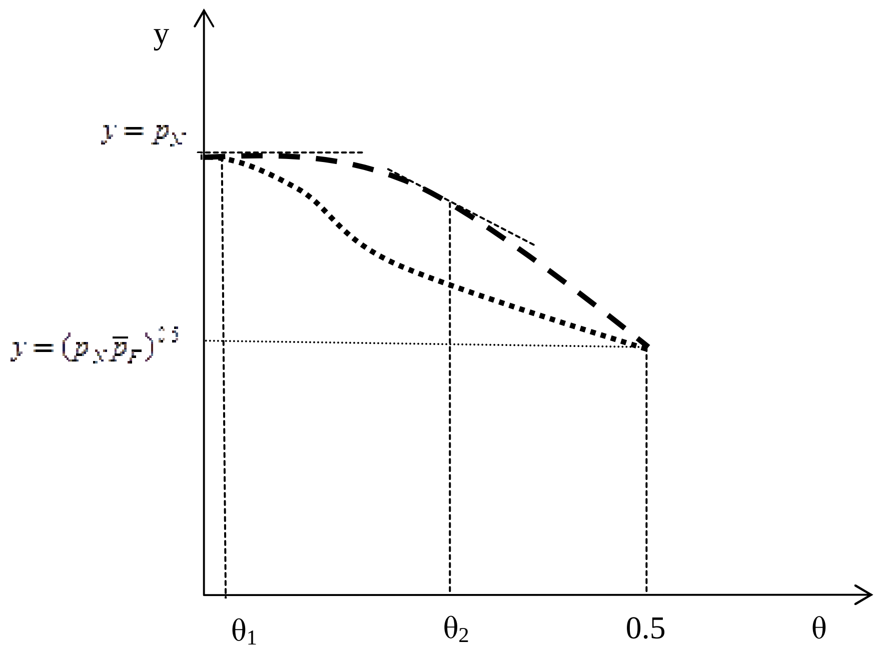
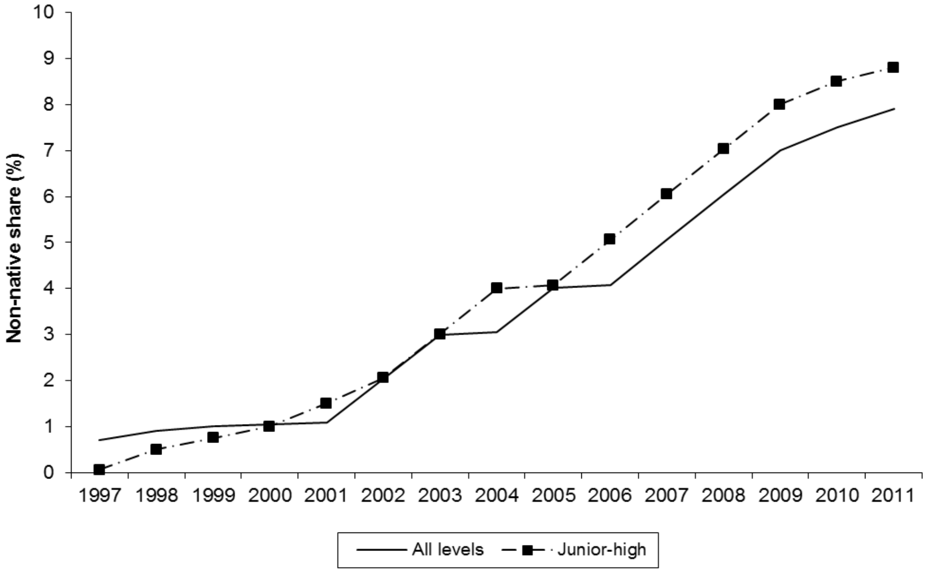
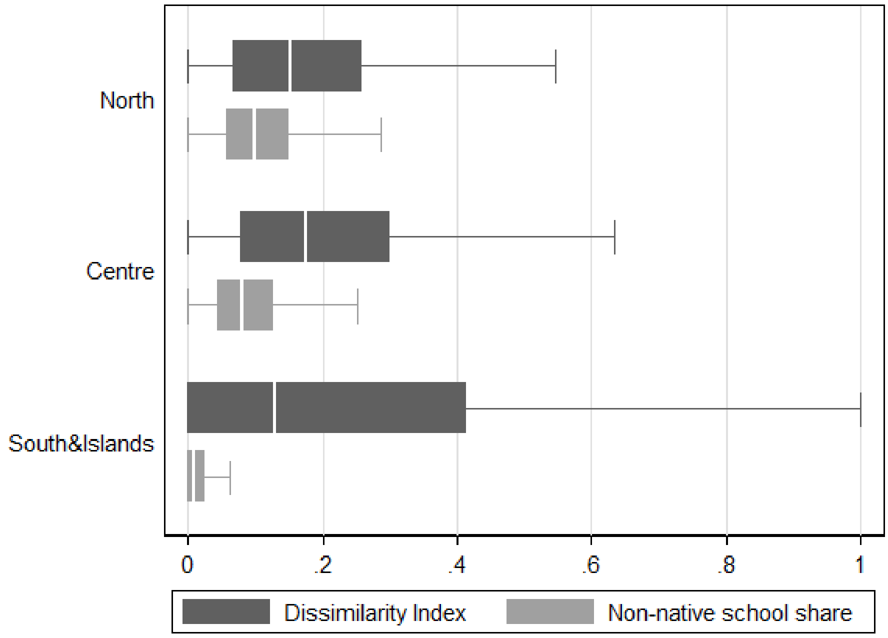


Figure 3. Non-native students the Italian school system. The graph shows the percentage of non-native students enrolled in Italian schools from s.y. 1996-07 to 2010-11for all education levels (primary, junior high and high schools) and junior high schools only.



Source: elaboration on data from MIUR (2011).

Figure A.1. Dissimilarity index and non-native school share. The figure shows a comparison between the dissimilarity index calculated at the school level and the non-native school share across three geographical macro-areas.



CHAPTER 2.

'ACTING-WHITE'?

SOCIAL INTERACTIONS AMONG NON-NATIVE STUDENTS

ABSTRACT

This paper focuses on social interactions within non-native students and analyses to which extent non-native students' cognitive outcomes depend on the exposure to non-native peers in the school. The sources of endogeneity due to non-random allocation of non-native students across classes (within schools) and across schools (within school-districts) are tackled with an instrumental variable approach. Exploiting a rich dataset on Italian junior high schools, I find negative within-group social interaction effects increasing with respect to the degree of school segregation and decreasing with respect to non-natives' assimilation. Increasing non-native school share has larger negative effects the more non-native students are unevenly distributed across classes in the same school and for the sub-group of first generation non-native students. These findings support the existence of 'oppositional culture' mechanisms (or 'acting-white behaviours') that exacerbate the negative social interactions effects within the non-native peer group.

JEL Classification: J15, I21, I28

Keywords: social interactions, acting-white, school segregation, non-native students

1. INTRODUCTION AND BACKGROUND

The recent developments in the study of social interactions between minority students and white peers underlines how a clear understanding of the internal dynamics of the 'minority peer group' is determinant to assess sources and mechanics of the minority students underachievement (Austen-Smith and Fryer, 2005; Fryer and Torelli, 2010; Fryer, 2010). Nevertheless, despite the great variety of studies on social interactions in educational settings, empirical evidence and theoretical models on peer effects between native and non-native students still present mixed findings and limited evidence on possible channels and mechanisms at work. Social interactions take place within the reference group (*within-group*) or between two separate groups of individuals (*across-groups*) and influence cognitive and non-cognitive outcomes of students' peers. What I name 'within-group' social interactions is generally referred to as 'within-race' social interactions in the U.S. literature and it refers to the specific aspects of peer effects *inside* the group of minority students and it has never been considered as an independent strand in the broad literature of 'racial peer effects'. Identification problems and policy interpretations are generally different with respect to the ones derived from estimations of 'between groups peer effects' and just a few and recent works look specifically at social interactions dynamics inside the 'minority students peer group' (Fryer and Torelli, 2010; Fairlie *et al.*, 2011).

In this work, I focus on *within-group interactions* and study to which extent non-native students' cognitive outcomes depend on the share of the same non-native peers in the school[28]. My primary aim is to disentangle the possible causal link between the size of the non-native peer group and its average test scores: does the size of the non-native group (i.e. non-native school share) influence the attainment of the same non-native students? If it is the case, in which way? The second aim of the analysis is to test the existence and relevance of two potential behavioural channels that might help to explain the underlying social interactions mechanisms: 'acting-white' and 'assimilation'. I label 'acting-white' the evidence that within-group *negative* social interactions are greater the greater is the segregation of minority students within each school. This is a sort of reinterpretation of the 'oppositional culture behaviours' sociological theory that asserts that minority students may underachieve and refuse assimilation to the majority behaviours in order to fit with their peers' (Fordham and Ogbu, 1986). The 'assimilation channel' is tested restricting the analysis on the sub-group of first-generation non-native students who plausibly experience more difficulties to

[28] I distinguish *native* and *non-native* students referring to a citizenship criterion. This is because data from Italian Ministry of Education (MIUR) only distinguish between *Italian* or *native* and *non-Italian* or *non-native* students. In the sensitivity analysis I test the robustness of the results using the 'immigrant' status definition as defined by OECD (OECD, 2010).

assimilate to the hosting country language and culture with respect to second generation peers.

The Italian context is an interesting case study under many aspects. From 1997 to 2007, Italy experienced one of the highest increase (+242.9%) in the percentage of foreign population among all OECD countries (OECD, 2011): foreign population more than doubled in less than one decade. Only Spain records an increase comparable to the Italian one. Of course, this huge phenomenon had direct consequences on students' population. Over the same period, the school system has recorded a growing number of non-native students enrolments: in 1996-97, only 0.7% of students in the Italian school system was non-native, while in 2009-10 the share has grown up to 7.9% (+106%, Figure 4) (MIUR 2009a, 2009b, 2010).

[Figure 4 here]

The pattern of immigration has also been changing in the last two decades. In the past, immigration flows mostly consisted in low-skilled, low-wage and often undocumented men seeking work. A lot of them were seasonal workers, and they normally arrived and stayed for brief periods without their families. Starting from the late Nineties immigrants show the intention to settle permanently: immigration flows consist more and more of complete families and the number of children in immigrant families has rapidly increased (Mencarini *et al.*, 2009). Consequently, in the last five years, second generation students have rapidly become part the Italian schooling population and constantly interact with first generation and native peers.

From the empirical point of view, in this specific setting the identification of 'social interactions effects' – defined as a combination of endogenous and exogenous peer effects (Manski, 1993) - has to solve three main threats: first, within school sorting given by non-random allocation of non-native students across classes in the same school; second, the separation of the effects of peers from other confounding influences in correlated effects; third, the endogeneity of non-native school share due to across schools sorting of non-native students generated by households' residential and working decisions. The identification strategy is based on school-level averages in order to sidestep the non-random allocation of non-native students within schools (across classes) while the endogeneity of the non-native school share due to across-schools sorting is tackled with an IV approach. The instruments exploit the existence of 'network effects' in the residential decisions of non-natives due to the evidence that early settlements of migrants tend to have an attractive power to successive migrants waves. I use as outcome measures standardized test scores from a unique and rich dataset combining the Italian national assessment program of educational attainment at the end of junior high schools (*INVALSI First Cycle Exams*[29]), with 2001 Census Survey data and administrative records on schools characteristics and socio-economic environment. The dataset overcomes problem of under-representation of non-native shares typical of survey data as it contains census information on all 8th grade students enrolled in junior high schools. I find robust evidence of negative within-group social interaction effects. Results also point to the existence of 'acting-white' behaviours among non-native students in Italian junior high schools.

The paper contribution is twofold. First, it is one of the few works that specifically looks at social interactions within the reference ethnic group. Despite the limited evidence in U.S. literature, this is one of the first times that 'within-group' peer effects are found in European school contexts (Aslund *et al.* 2011 is a notable exception). Second, I find evidence that two important mechanisms ('acting-white' and 'assimilation') are at work in the context under study and are likely to influence the social interactions dynamics within the non-native group.

The rest of the paper is organized as follows. In Section 2.2 I discuss the main results in the literature about the 'intra-race peer effects' and 'acting-white behaviours'. Section 2.3 describes the data and provides descriptive statistics. Section 2.4 discusses the identification strategy devoting particular care to the instrumental variable used. Section 2.5 contains the main results, while in Section 2.6 I speculate on the underlying social mechanisms. Section 2.7 contains several tests to corroborate the robustness of the findings. Section 2.8 concludes and provides some policy implications.

2.2. LITERATURE

A new strand of the social interaction literature tends to reinterpret the general result that 'intra-race' peer effects are stronger compared to 'extra-race' peer effects (Hoxby, 2000; Angrist and Lang, 2004; Hanushek *et al.*, 2009; Hanushek and Rivkin, 2009) under the light of the 'acting-white' theory. This is a reinterpretation of the 'oppositional culture behaviours' sociological theory asserting that minority students may underachieve and refuse assimilation to the majority behaviours in order to fit with their peers' (Fordham and Ogbu, 1986). Battu and Zenou (2010) exploit a similar intuition for outcomes of immigrant workers in the labor market. Fryer and Torelli (2010) provide the first empirical evidence using the National Longitudinal Study of Adolescent Health (*AddHealth*) and estimating the effects on achievement of an 'index of social status' based on the individuals' contacts with same-race

[29] 8th grade students are enrolled in their third year of the Italian middle grade comprehensive school (13-14 years old). After passing the final exam they gain the 'Junior High School Diploma' (ISCED level 2).

friends within the school. They show that this 'acting white' proxy variable varies a lot with respect to school characteristics and individual achievement and that the effect is concentrated in schools with more interracial contact. Their coefficient for the 'acting-white' indicator is twice as large in schools that are above the median in terms of segregation, whereas it is significantly lower where black students are more isolated. Fairlie *et al.* (2011) implement the same intuition of the 'acting white theory' to study the extent to which academic performance depends on students being of similar race or ethnicity to their instructors. They use detailed administrative data from one of the largest community colleges in the United States and address the concern of endogenous sorting using both student and classroom fixed effects. The authors find that the performance gap, in terms of class dropout and pass rates between white and minority students, falls by roughly a half when minority students are taught by a 'minority instructor', so that, for instance, African-American students perform particularly better when taught by African-American instructors. Friesen and Krauth (2011) use data on elementary school students in British Columbia (Canada) to assess the effects of the language spoken at home and attending 'enclave schools' on students' attainment. The authors broadly define 'enclave school' as schools with higher shares of same ethnic minority peers and identify non-natives with Aboriginal, Chinese and Punjabi ethnic minority groups. In contrast with the rest of the literature, the authors find within-group effects weaker compared to across-groups effects and that attending an 'enclave' school has differential effects with respect to the prevalent ethnic minority (slightly positive effects for Chinese, negative for Punjabi). According to Friesen and Krauth (2011) the evidence that effects on achievement of attending school with more same-language peers varies with the achievement level of one's own language group suggests that linguistic or ethno-cultural similarity to peers does not in itself play a significant role in immigrant success, but rather that human capital and cultural norms of peers is what matters.

In European contexts, Aslund *et al.* (2011) for Sweden, Maestri (2011) for the Netherlands, Jensen and Rasmussen (2011) for Denmark use three different identification strategies to answer a variety of research questions dealing, to some extent, to assess the impact of the presence of non-native students on natives' and non-natives' educational outcomes. Aslund *et al.* (2011) is one of the few work that specifically focuses on peer effects within the minority students' community and neighbourhood. They estimate to what extent the lower achievement of immigrant students is due to the characteristics of the neighbourhoods in which the immigrants grow up. The estimation strategy relies on a governmental placement policy that generated exogenous variation in the initial residential distribution[30]. They show that the size of the local ethnic community is positively related to compulsory school grades. Separating this effect into its components, the authors find that one standard deviation increase in the fraction of highly educated peers raises student performance by 0.9 percentile ranks and that one standard deviation increase in the size of the ethnic community has about the same effect, albeit less precisely estimated.

Maestri (2011) investigates how the heterogeneity of the ethnic minority composition within schools affects natives' and non-natives' attainment grounding on the idea that ethnic diversity can stimulate the creativity of students, push them to be proficient in the instructional language, and reduce the scope of ethnic identification with all its possible drawbacks as the 'acting white' effects. She exploits the within school cohort-to-cohort variation in the ethnic make-up of a rich dataset of primary schools in the Netherlands and finds that ethnic diversity has a positive impact on the test scores of minority students, in particular for language skills. She also finds evidence of a negative relationship between an ethnic diversity index, obtained as an inverted Hirschman-Herfindahl index, and an indirect measure of social interactions among pupils.

Finally, Jensen and Rasmussen (2011) analyse the effect of ethnic concentration in schools on the cognitive outcomes of children. They use a rich dataset for Danish 9th grade students, based on PISA test scores, administrative and census information on students, schools and neighbourhoods. In order to correct for the endogeneity in school ethnic concentration, the authors apply school fixed-effects and IV, using as instrumental variable the ethnic concentration in a larger geographical area where the school is located. Results show that there is a negative effect of ethnic concentration on students' outcomes but that these are statistically significant only for the native Danish children. In contrast to the majority of the results in the literature, they do not find statistically significant 'within-group' peer effects for immigrant children so that increasing non-native school share does not affect immigrant test scores. However, albeit using detailed individual level information, the authors apply an instrumental variable approach based on larger geographical area ethnic density to instrument for non-native presence in each school. This approach is likely to underestimate the effects - both in terms of social interactions within the non-native group and between natives and non-natives - as schools with different ethnic make-up within the same area are subject to the same value of the instrument.

[30] Between 1987–1991 Swedish authorities assigned refugees to their initial location, since individuals were not free to choose, Aslund *et al.* (2011) argue that the initial location was independent of (unobserved) individual characteristics.

2.3. DATA AND DESCRIPTIVE STATISTICS

I exploit a unique dataset that combines the Invalsi First Cycle Final Exam data[31], administrative records from Ministry of Education Statistical Office, and the Italian Population Census Survey 2001. Invalsi First Cycle Exam data (hereafter 'First Cycle' or 'IC') are the first experience of testing attainment levels of all students enrolled in Italian junior high schools. All 8th grade students sit the Invalsi First Cycle Exam in mid-June, at the end of the compulsory and comprehensive path of the Italian school system constituted by five years of primary education and three years of junior high school. The census dimension of Invalsi IC data allows us to overcome problems of underrepresentation of immigrant individuals and measurement errors typical of sample surveys (Aydemir and Borjas, 2010), while additional information about socio-economic family background are obtained as school-level averages of Census variables linked to each school using an original matching technique that identifies for each junior high school its 'catchment area' (Barbieri et al., 2011)[32].

In detail, Invalsi IC dataset contains Math and Language test scores, individual information and school level information for each 8th grade student enrolled in a public or private junior high school[33]. I exploit two waves corresponding to 2008-09 and 2009-10 school years final exams (about 500,000 students per wave). Individual information cover year of birth, gender, citizenship (Italian, non-Italian), place of birth; how long the student is in Italy if born abroad (from primary school, for 1-3 years, less than 1 year); mother's and father's place of birth (Italy, EU, European but non-EU, other non-European country), grade retention (if the student is 'regular' i.e. if he/she is 14 years old at the end of the school year; 'in advance' i.e. younger than 'regular' students, or 'retained' i.e. older than 'regular' students). Administrative records from Ministry of Education Statistical Office provide general information about school characteristics (i.e. type of school, public vs. private, number of students enrolled and number of teachers, average class size) and are matched to Invalsi First Cycle data through anonymous school identifiers. Each school is finally matched to a group of census divisions through an original matching technique designed to associate to each junior high school a group of census cells constituting its 'catchment area' (Barbieri et al., 2011)[34]. This procedure allows matching to each junior high school variables from 2001 Population Census Survey covering demographic and socio-economic information on resident population.

---

[31] INVALSI (*Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione*) is the independent public institute carring out the evaluation of Italian school system and test students' attainment levels.

[32] Notice also that this is the first time that a dataset with such a variety of information and covering the universe of 8th grade students is made available for the Italian school system.

[33] Test scores range from 0 to 100 and refer to the fraction of right answers for each of the two subjects.

[34] See the Appendix in Barbieri, Rossetti and Sestito (2011) for a detailed description of the matching technique used.

I identify native and non-native students according to a citizenship criterion. Sensitivity analysis on different categorizations never shows significant differences in the results (see Section 2.7). Although the empirical analysis will primarily focus on non-native students, I also distinguish between first and second generation students. 'First-generation' students are born abroad from parents born abroad, while 'second-generation' students are born inside the receiving country but from parents born abroad (OECD, 2010). The final population is constituted by all 8th grade non-native students enrolled in Italian junior high schools in 2008-09 and 2009-10 school years[35] (68,717 individuals).

[Table 13 here]

Panel A in Table 13 describes the distribution of these different categories across geographical macro-areas. For instance, referring to the IC 2009-10 wave, the overall share of non-native 8th grade students is 7.22%, but there are sharp differences across the country. The highest average school share of non-native students are in Northern and Centre regions (10.01% and 9.18%), while it dramatically falls in the South (1.97%). At the school level, Panel B in Table 13 describes school characteristics (share of public schools, pupil-teacher ratio, average school size, average class size) with respect to macro-area. On average, more than 76% of schools has at least one non-native student: this proportion is very high in the North and Centre (more than 90%) and sharply decreases in the South (58%).

[Table 14 here]

Table 14 contains general descriptive statistics with average test score results for first and second generation immigrants and native and non-native students, according to the definitions previously introduced. Second generation students perform better compared to first generation peers, and the difference is more pronounced in the Language skills (+5 points) than in Math (+2.2 points). Finally, it is worth noting that test scores gap between non-natives and natives does not change a lot along the test score distribution.

To provide evidence of the non-random allocation of non-native students across classes in the same school, I calculate a dissimilarity index (*D*) at the school level. The index was first proposed by Duncan and Duncan (1955) and then extensively used in school and residential segregation analysis (among the others, Clotfelter, 1999; Echenique and Fryer, 2007). It provides a measure of the evenness in the distribution of non-native students. Given that in each school there are N_j classes ($c=1...N_j$), the dissimilarity index at the school level (D^s) measures the percentage of non-native students that would have to change class for each class

[35] I exclude all individuals who did not sit either Maths or Italian Language test because absent the day of the exam (0.73% of the total students population).

to have the same percentage of non-native students as the one of the whole school (i.e. the non-native school share). In symbols:

$$D^j = \frac{1}{2}\sum_{c=1}^{N_j}\left|\frac{Natives_c}{Natives_j} - \frac{Non-natives_c}{Non-natives_j}\right| \qquad [1]$$

where Natives_c and Non-natives_c represent, respectively, the total number of native and non-native students in class c of school j, and Natives_j and Non-natives_j represent the total number of native and non-native students in school j. D^j ranges from 0 (perfectly even distribution, meaning 'no segregation') to 1 (perfectly uneven distribution, i.e. 'maximum segregation').

[Figure 5 here]

The graph box in Figure 5 portraits the results distinguishing among three geographical macro areas. The distribution of non-native students across classes cannot be considered even: median value of the dissimilarity index is 15% in the North, 17% in the Centre and 13% in the South, so that, for example, in the North, on average, 15% of non-native students have to be moved from one class to another to obtain an 'even' distribution within the school.

2.4. IDENTIFICATION STRATEGY

Starting from a standard linear-in-means reduced-form model with peer interactions (Manski, 1993), I assume that a student's outcome (y) depends on individual characteristics (X), the share of same-group peers experienced by each student i among the school and grade mates (P_s), contextual factors (μ_s) and an unobserved error term. Thus, for each non-native student i attending 8th grade, in school s, it yields[36]:

$$y_{ics}^{NN} = X_{ics}^{'NN}\gamma_x + \beta P_s^{NN} + \mu_s^{'}\tilde{\gamma}_\mu + \varsigma_{ics}^{NN} \qquad [2]$$

where:

$$P_s^{NN} = \frac{non-natives_s^{grade=8}}{(natives + non-natives)_s^{grade=8}} \qquad [3]$$

[36] Notice that in the reminder of the paper I simply refer to 'non-native school share' to easy the exposition. However, as expression [3] clarifies, P_s only refers to the share of non-native students attending grade 8 in the school s. P_s is a good proxy of peers' characteristics but it is also predetermined with respect to the outcome measure and thus not affected by common school-level shocks (i.e. the correlated effects, μ_s) (Ammermüller and Pischke, 2009; Angrist and Lang, 2004).

The identification of the 'social interactions effect' parameter β - which includes both endogenous and exogenous peer effects (Manski, 1993) - in equation [2] has to solve three main threats: first, within school sorting given by non-random allocation across classes of non-native students; second, the separation of the effects of peers from other confounding influences in correlated effects (i.e. omitting relevant components of the contextual effects (μ) that are correlated with P_s will bias the estimation of β); third, endogeneity of non-native school share (P_s) due to across schools sorting of non-native students generated by households' residential and working decisions. In the baseline model, I solve within-school sorting of non-native students moving from individual-level data to school-level averages: aggregation at the school level solves the problem of endogenous within-school sorting of non-native students across classes (Cutler and Glaeser, 1997; Card and Rothstein, 2007; Brunello and Rocco, 2011). To clarify this point, I specify the error term in three parts: a school-specific component (η_s) common to each student of the non-native group ($j=NN$) in school s, class specific component (u_{cs}) common to all non-native students in class c and school s, and a student-specific component (e_{ics}) [37]:

$$\varsigma_{ics} = \eta_s + u_{cs} + e_{ics} \qquad [4]$$

Any non-randomness in within school allocation of non-native students determines a correlation between the class specific component (u_{cs}) and the observable characteristics so that OLS estimates are biased. Under the assumption that the student-specific error and the class-specific error average to zero for each group j in each school s, taking school-level averages solves this problem[38]. Thus, the mean outcome for non-native student group in school s is given by:

$$\overline{y}_s = \overline{X}_s^{'}\alpha + \beta P_s + S_s^{'}\theta_S + W_s^{'}\theta_s + \eta_s \qquad [5]$$

[37] Henceforth, I suppress upper index ($j=NN$) to easy notation: $\overline{y}_s^{NN} = \overline{y}_s$.

[38] The assumption of zero mean for the class-specific error component might fail if some class level characteristics are not randomly allocated across classes (within each school). However, since the allocation of teachers across classes is a predetermined decision of the School Head and school resources are equally distributed in the school (across classes and grades) this issue is not particularly relevant in the Invalsi IC data.

where, $\bar{X}_s$ represents the mean characteristics of non-native students in school *s*, school (*S*) and catchment area characteristics (*W*) calculated as school-level mean characteristics (contained in the vector μ).

The averaging procedure sidesteps the problems due to correlation in within school allocation to classes of non-native students but leaves unsolved the endogeneity problems due to OVB in correlated effects and across schools sorting of non-native students. Concerning the possible omitted variable bias arising from correlated effects, I first point out that the possibility of correlation between *W* and the peers' variables in the equations is reduced in the estimates thanks to the original features of dataset used. In fact, catchment area variables are school-specific socio-economic indicators that are not directly obtained as an average of some peers' characteristics and that predate the outcome measure (they are obtained from the *Italian Population Census 2001*). These two characteristics reduce endogeneity problems in peer effects estimations limiting possible correlation with peers contemporaneous characteristics (*X* and *P*) (Ammermüller and Pischke, 2009)[39]. Then, the omitted variable bias is also reduced including school-district by year fixed-effects which capture all omitted and confounding factors that are common to all schools in the same school-district[40]. In other words, they capture unobserved heterogeneity mirrored by different socio-economic conditions of schools, underlying students' families populations and recent trends in immigrants' settlements across the territory. Thus, the baseline model to be estimated by OLS takes the following form:

$$\bar{y}_s = \bar{X}_s^{'}\alpha + \beta P_s + S_s^{'}\theta + W_s^{'}\gamma + \phi_t + \eta_s \qquad [6]$$

where ϕ_t represents the set of school-districts by year fixed-effects and year dummies (included to control for possible time trends in IC test score results in the two waves used).

2.4.1. IV model

I implement an IV approach to tackle the bias from sorting of non-native students across schools due to households' residential choices. For instance, in big cities immigrant families tend to settle in suburbs where location rents are lower. Within a given city, these areas generally reflect lower socio-economic status of both native and immigrant households living

---

[39] The different time pattern in the IC Invalsi data and catchment area level variables (*W*) is not a concern as socio-economic conditions across Italian territory did not change significantly in the period considered (Bank of Italy, 2008).

[40] School districts fixed-effects correspond to 110 dummies, one for each Italian province (NUTS 3 level, Eurostat Nomenclature of Territorial Units for Statistics).

there. Therefore, for schools located there the higher number of non-native students may be reasonably correlated to lower scores of both native and non-native peers. Nevertheless, this may be due not only because the exposure to higher numbers of non-native schoolmates *causes* negative externalities within the non-native peer group, but also because non-native students' test scores are lower *per se*, for instance, because of the underlying lower socio-economic status or because of negative externalities from disadvantaged native peers. With respect to the empirical framework proposed, any non-randomness in the sorting of students across schools or neighbourhoods produces a serious correlation between P_s and the school-error component and bias OLS estimates of β in the school-level equations.

To address this problem I instrument the non-native school share (P_s) with the number of non-native residents living in the school catchment area in 2001 (Z_s). This approach exploits the existence of 'network effects' in the residential decisions of immigrants due to the evidence that early settlements of migrants tend to have an attractive power to successive migrants waves, especially in urban areas (Borjas 1995; Card, 2001). This fact is confirmed also in the Italian context where important channels that could explain immigrants residential clustering have to do with the advantages of proximity to people in the same national, ethnic, linguistic, or socioeconomic group for information sharing purposes, reciprocal support and use of common local public goods (Barone and Mocetti, 2011; Boeri *et al.* 2011; Pellizzari, 2011).

Similar identification strategies are widely exploited in the migration and segregation literature. For instance, Boustan (2010) estimates the causal effect between 'white flight' from U.S. cities to suburbs and the arrival of blacks immigration waves. To solve the potential endogeneity in blacks settlements across cities she builds an instrumental variable making use of the fact that black migrants from given southern states clustered in particular northern cities. Saiz and Wachter (2011) take advantage of the immigrant clustering evidence to partially predict the patterns of new immigrant settlement in U.S. metropolitan areas and evaluate the causal impact of immigration on neighbourhood dynamics. To this purpose, they instrument for the actual number of new immigrants using the predictions of a geographic diffusion model that estimates the number of new immigrants in a neighbourhood using lagged densities of the foreign-born in surrounding neighbourhoods. Boeri *et al.* (2011) use houses characteristics form the 2001 Census data to instrument for immigrant segregation in eight Italian Northern cities. Their instrumental variable approach is very close to mine and hinges upon the same exogeneity condition, albeit applied to residential segregation and not to the school context. In educational settings, similar strategies have been used to determine the causal impact of immigrant concentration on students' outcomes. Dustmann and Preston (2001) and Jansen and Rasmussen (2011) use ethnic concentration in a larger geographical area to instrument for the school ethnic concentration. They ground on the assumption that larger geographic area immigrant concentration is a good predictor of school immigrant concentration while it does not directly affect children outcomes.

The IV strategy implemented combines insights from both streams of the literature, although being closer in spirit to arguments and the ideas typical of the migration one. The exclusion restriction claims that non-native residents in each school catchment-area in 2001 (Z_s) influence the test scores only through the effects on the actual share of non-native students in the school (P_s). The exogeneity of the instrument relies on the fact that it is antecedent to the outcome measures used and thus plausibly uncorrelated with test scores: the nine years temporal lag between the outcome variable and the instrument ensures the exogeneity condition to be met. The coefficient of the social interactions parameter ($\hat{\beta}$) is estimated from the reduced form equation [7] with 2SLS:

$$\bar{y}_{st} = \bar{X}_{st}^{'}\alpha + \hat{\beta}\hat{P}_{st} + S_{st}^{'}\theta + W_{s}^{'}\gamma + \phi_{t} + \eta_{st} \qquad [7]$$

2.4.2. Instrument relevance and validity: a discussion

In this paragraph I discuss possible concerns on the robustness of the instruments validity and relevance assumptions while in the sensitivity analysis I perform empirical falsification tests. The relevance of the instrument is based on the fact that the number of non-native individuals who lived in the school catchment-area in 2001 (Z_s) is a good predictor of the actual number of non-native residents in the school area, and, as a consequence, of the actual non-native composition of the school population.

[Figure 6 here]

Figure 6 compares the non-native students' concentration in junior high schools (average values for the two school years considered in the analysis, i.e. 2008-09 and 2009-10) with non-native resident population in 2001. The figures almost perfectly overlap corroborating the basic assumption on which the instrument relevance is grounded: non-natives tend to cluster only in particular areas of the country which can be predicted making adequate use of information on past immigration waves. First stage regressions confirm that Z_s is positively and strongly correlated with the endogenous variable.

Differently from Dustmann and Preston (2001) and Jansen and Rasmussen (2011), I exploit an instrument that is school-specific, thus it is more precise than instruments based on larger geographical areas ethnic concentration. In fact, larger geographical areas might contain more schools sharing the same value of the instrumental variable. This problem becomes particularly serious in urban areas, where different suburbs within the same city may show sharp differences in the school ethnic make-up which an instrumental variable approach based on ethnic concentration in larger geographical area would not capture.

Concerns upon the exogeneity of the instrument arise if Z_s is correlated with the outcome variable (non-native test scores) through some underling channels other than the presence of non-native students in the school. Indeed, the time lag between Z_s and y_s ensures the validity condition to be met. Using measures which are more distant in time with respect to the outcome variable would improve, in principle, the reliability of the exclusion restriction. This is not the case in the Italian context because the 1991 Census would not capture the 'network effects' in households' residential locations as the presence of non-native in Italy was totally negligible compared to the actual one (Billari and Della Zuanna, 2008; Mencarini *et al.*, 2009). In fact, in 1991 the non-native population in Italy was so small that such a hypothetical instrument would not have any predictive power on present non-native school shares. Boeri *et al.* (2011) do a similar exercise and conclude that 1991 Census data have not predictive power with respect to actual immigrant residential choices in Italy. They focus on eight cities in the North and Centre of Italy, but this result can be easily generalized to the whole country[41]. Possible concerns arise if I consider that some of the non-Italian resident population in the 2001 Census data may be constituted by the parents of the non-native students in Invalsi IC data who were 5 or 6 years old in 2001. To test for this, in Section 2.6.2 I repeat the main analysis only non-native students born abroad (i.e. first generation). First-generation non-native students are not born in Italy, thus it is likely that the majority of them and their parents either did not reside in Italy in 2001 or were undocumented and in 'illegal' status. In both cases they would not be recorded in the 2001 Census. Results are not qualitatively different from the main analysis and thus further support the reliability of the exogeneity condition[42].

To conclude, the instrumental variable chosen outperforms previous studies both in terms of relevance (stronger predictive power) and precision (Z_s is school specific), while the nine years lag supports the validity of the exclusion restriction.

2.5. RESULTS

This section contains the baseline OLS and IV results, while in Section 2.6 I provide evidence of two possible underlying mechanisms: 'acting-white' and 'assimilation'. In the empirical

[41] Notice also that Census 1991data are can not be matched to junior high schools and Invalsi test scores (Barbieri *et al*., 2011)

[42] Finally, it is worth to notice that it is not possible to build an instrumental variable based on 'supply-push' factors of immigration waves à la Card (2001) because of data availability constraints. Invalsi IC data only record whether the non-native student is from an EU country or not

analysis I exclude those schools where only one non-native student is enrolled, so that the final sample is constituted by almost 6,200 junior high schools. The outcome variable is expressed as the natural logarithm of the non-native school mean test in Math and Language.

2.5.1. OLS results

The complete list and detailed description of the control variables used can be found in Table 15.

[Table 15 here]

[Table 16 here]

Panel A in Table 16 contains OLS estimates of the social interaction parameter β from eq. [6]. I progressively add the complete set of explanatory variables: individual and family background mean characteristics (gender, school path regularity, place of birth, time spent in Italy since birth, parents' origins), school-district by year fixed effects; school characteristics (school type and size, ownership, average class size, pupil-teacher ratio, pupil-support teacher ratio, students with disabilities school share, average lesson hours per week, 'cheating' dummy[43]) and catchment-area socio-economic variables. OLS results show a negative and statistically significant impact of non-native school share (P_s) on non-natives' school mean test scores: increasing by 1% non-native school share is linked to a statistically significant decrease of -20.6% in Language and -16.7% in Math mean test scores. Results do not vary substantially once I control for individual characteristics, family background and add school-district*year fixed-effects, although school and catchment-area characteristics help to capture socio-economic features of the environment experienced by all students in each school (correlated effects). The general pattern of the OLS results induce to suppose that there are negative within-group peer effects. However, because of the endogeneity of the non-native school share causal links cannot be established.

The direction of the bias is *a priori* undetermined as it depends on many possible channels through which across school sorting and within-group mechanisms of social interactions might (or might not) be at work. Across schools sorting arises when non-native students tend to cluster in some schools which are plausibly located in urban areas characterized by lower socio-economic backgrounds. Given that non-natives have lower test scores than natives,

[43] Because of cheating evidence in IC data outlined by Invalsi (Invalsi 2009, 2010), I add a dummy variable that controls for all schools suspected to have 'cheated' on reporting test scores results. The dummy identifies the schools in the upper decile of the distribution of the school-specific cheating coefficient (ranging from 1, no cheating, to 0, full cheating) provided by the Invalsi Statistical Office. Indeed, cheating evidence is limited concerning the 2010 wave and sensitivity analysis on various specification of the 'cheating dummy' and on the use of 'cheating corrected' results do not find significant differences once I control for territorial dummies in the estimated specifications.

‘school segregation’ would determine negative spurious correlation and downward bias in the OLS estimates. This downward bias may be also exacerbated by negative spill-overs arising from disadvantaged native peers. On the contrary, the OLS estimates may be upward biased if both low-skilled and high-skilled non-native students tend to cluster in the same schools. This is a case that is well-suited for the Italian context where immigration is a relatively new phenomenon and immigrants tend to settle in the same areas irrespectively from their characteristics. Boeri *et al.* (2011) demonstrate that, in a general equilibrium model, even high-skilled immigrant (i.e. the ones with better education and, plausibly, better performances in the labour market) choose to settle in the same areas of the cities where low-skilled immigrants reside, even if their income would allow them to pay higher rents. Their work supports this ‘positive selection mechanism’ so that OLS estimates result to be upward biased compared to the IV case. A similar argument of ‘positive sorting’ can be translated in the school context so that high-skilled non-native students would plausibly attend the same schools of low-skilled non-natives and, for instance, do not enrol in schools with only native peers. This fact would generate an upward bias in OLS estimates.

2.5.2. IV results

2SLS estimations solve the endogeneity of the variable of interest (P_s) and the omitted variable bias induced if equation [6] fails to control for all relevant school and environment inputs, and obtain an *average causal response* measure to the increase of non-native school share on the same non-natives’ mean test scores. Panel B of Table 16 contains the 2SLS estimation (eq. [7]), Table 17 the first stage regressions.

[Table 17 here]

The instrumental variables used (Z_s) is the number of non-native residents in the school catchment-area in 2001, which is obtained matching the Census 2001 variables to each junior high school as described above. First stage estimations show that the coefficients of the instruments always have a positive and statistically significant impact on the endogenous variable (P_s), and first stage F-statistic strongly rejects the null of weak instrument (Yogo and Stock, 2005). Within-group social interactions are still negative, statistically significant and increased, in absolute terms, with respect to OLS estimations: a 1% increase in non-native school share causes a decrease of -81.6% in Math average test score of the non-native peer group, and a decrease of -73.3% in Language. The estimated effects of the increase of 1% in non-native school-share are quite huge: they correspond to a 2.37 times the standard deviation for Language and 2.56 for Math.

In Chapter 1 I estimate the social interactions effects caused by non-native students on natives' attainment so that I can compare those estimates with within-group effects in Table 16. The average causal response for an increase of 1% in the non-native school share is different for the two types of peer interactions. *Within* the non-native peer group, a one percentage point increase in the non-native school share lowers their own mean test scores by 70-80%. On the contrary, the same increase causes small negative effects or no effects at all on natives' attainment. Although not directly comparable, these results are in line with U.S. literature on '*racial peer effects'* which finds evidence of negative and sizable 'intra-race peer effects' (Hoxby 2000; Hanushek *et al.* 2009) but are new in the European schools context. In fact, Aslund *et al.* (2011) document the strong social interactions effects taking place intra-ethnic groups in Sweden and find that positive externalities on immigrant children education may arise if the neighbourhood ethnic reference group contains a higher fraction of educated adults. Jensen and Rasmussen (2011) find no significant within-group effects. However, these studies focus on Northern Europe countries (Sweden and Denmark) which experience a different kind of migration with respect to countries such as Italy, U.K., Portugal and Spain where prevalently low-skilled (and often undocumented) immigrants had great impact on the labour market and school systems only in the last two decades. Indeed, these results are in line with Boeri *et al.* (2011) who study the effects of residential segregation on immigrants' labour market outcomes in eight cities in Northern Italy. Their instrumental variable approach uncovers a positive sorting process between segregation and immigrants' employment which results in IV negative estimates greater, in absolute terms, than the baseline OLS.

2.6. MECHANISMS

The evidence of strong and negative effects in within-group social interactions in the non-native peer group can be explained through many possible underlying channels. In this Section I explore two main mechanisms: 'assimilation' and 'acting-white'. In fact, within-group negative social interactions might be exacerbated by within-school segregation of the non-native group with respect to native peers up to generate 'acting-white' behaviours and, on the contrary, the same negative effects might be attenuated the more non-native children and their families are assimilated in the hosting country society. The 'acting-white behaviour' comes from the social interactions literature (Austen-Smith and Fryer, 2005; Fryer and Torelli, 2011) while the 'assimilation' mechanism prevalently follows the migration literature (Dustmann and Glitz, 2011). I focus on these mechanisms for two main reasons. First, they can be easily linked to direct policy implications as the assimilation and integration of immigrants are nowadays considered key elements of a good immigration policy (Dustman *et al.*, 2012). Second, they are the two main elements considered in the literature as potential determinants of non-native students' underachievement. Given that it is not possible to quantify the single contribution of each channel, I indirectly test whether there is evidence to support or reject the hypothesis that these two mechanisms are at work in the setting under study. To this purpose, I first seek for non-linearity in the effects with respect to a school segregation index and then separately estimate the model for the subgroup of 'first-generation' non-native students.

2.6.1. Oppositional cultural behaviours: 'acting white'

The linear specifications estimated so far do not take into account that, conditional on non-native school share, the average test scores of non-natives may also vary with the degree of segregation experienced by non-native students in each school (Brunello and Rocco, 2011). It is particularly interesting to verify this hypothesis in the light of the 'oppositional culture behaviours' that may arise in cases in which strong within-school segregation of non-natives lead them to cluster, do not interact with native peers, and even refuse to be integrated. The general refusal of assimilation to native peers' could lead non-natives to under-achieve in order not to fit-in with 'native stereotypes of good students'. This mechanism, known in the sociological literature as 'oppositional identity' (Fordham and Ogbu, 1986; Portes, 1987), has been used to explain the evidence of 'acting-white' behaviours in U.S. schools (Fryer and Torelli, 2011).

In this context, I name 'acting-white' the particular form of 'oppositional culture behaviours' that might arise when negative within-group effects increase with the degree of school segregation. To test for this hypothesis I add non-linearity in the within-group social interactions effects interacting the non-native school share with the school dissimilarity index (eq. [1]) and estimating the following model:

$$\bar{y}_{st} = \bar{X}_{st}^{'}\alpha + \hat{\beta}\hat{I}_{st} + S_{st}^{'}\theta + W_{s}^{'}\gamma + \phi_t + \eta_{st} \qquad [8]$$

$$\text{where } I_{st} = P_{st} \cdot D_{st}$$

The new interaction variable (I_{st}) represents a weighted version of the simple non-native school share where the within-group social interactions effects are weighted by the degree of segregation experienced by non-native students in each school. Given that D_{st} (as well as the interaction variable I_{st}) is a refinement of the simple non-native school share variable (P_{st})

used in previous equations, I exploit the same instrumental variable (Z_s) of the main IV analysis. Results (Table 18) confirm the existence of non-linear effects increasing in within-school segregation. The estimated coefficients for the interaction term between the non-native school share and the dissimilarity index are negative and statistically significant both for Language and Math, while first stage F-statistic show that the IV estimates are strongly identified. Hence, increasing non-native school share has larger negative effects the more the school is segregated. That is, the more non-native students are allocated together in the same class the greater are the negative within-group social interactions effects. This finding supports the existence of 'oppositional culture' mechanisms that exacerbates the negative within-group social interactions effects. While the 'acting-white' theory has been recently debated in the American literature (Torelli and Fryer, 2010, for the U.S., and Friesen and Krauth, 2011, for Canada), it is new to the European context and these findings are the first that support the possible existence of 'acting-white' mechanisms also in the school systems of European countries that experienced relatively recent and massive migration waves.

Indeed, results are in line with the limited European evidence on peer effects within the non-natives at school. In particular, these findings are in line with Maestri (2011) who provides evidence to support the benefits from ethnic diversity (which is opposed to ethnic segregation) on students attainments in Dutch primary schools. Using cross-country data based on PISA test scores, also Brunello and Rocco (2011) establish that school segregation exacerbates the (small) negative effects on natives' attainments due to the presence of immigrant peers.

6.2. First generation students and assimilation effects

Non-natives' children assimilation in the hosting country is a complex process that involves, at least, three main factors: (*i*) parental background and parental decisions on children education; (*ii*) school system characteristics; (*iii*) the social context and 'ethnic capital' in which children grow up (Dustmann and Glitz, 2011; Schneeweis, 2011; Schnepf, 2007). In general terms, better assimilated households help the assimilation of non-native children in schools and, more broadly, in any aspect of the social life in the hosting country so that within-group negative social interactions should be decreasing with respect to degree of assimilation of the immigrants' households. Notice that I refer to 'assimilation' in an extremely broad sense, ranging from the acquisition of skills in the use non-mother tongue language to the degree of integration the non-native family has reached in the host country.

To test this hypothesis, I rerun the analysis focusing on the sub-group first-generation students (*1G*, non-native students born abroad) making the assumption that, after controlling for a proxy of arrival time, first-generation non-native students usually face more difficulties in assimilation processes compared to second generation peers[44] (Dustmann, Machin and Schonberg, 2010; Dustmann, Frattini and Lazzara, 2012). In fact, international surveys on students' attainment generally find that second generation usually perform better than first (especially in language skills) and seem to benefit from their longer stay in the hosting country (Schnepf, 2007; Dustman and Glitz, 2011).

To test this assumption in the Invalsi IC data, I perform OLS regressions on the Invalsi IC 2009-2010 individual test scores (for about 870,000 8th grade students, both native and non-native) using school-fixed effects to capture unobserved heterogeneity at the school level, year dummies and individual characteristics as controls (dummies for gender, retention, immigrant status, first and second generation immigrants).

[Table 18 here]

In the first regression (Table 18, column a) I simply use the dummy for being a non-native student, while in the second regression (Table 18, column b) I distinguish between first and second generation immigrants. *Coeteris paribus*, being a non-native student implies a Language test score 12.41% lower than native peers and 6.16% lower in Math. First generation students score 13.74% lower in Language and 6.05% lower in Math than native peers. Second generation gaps are reduced for Language (-8.41%), while there is not statistically significant difference in Math test score between first and second generation students. Thus, a descriptive pattern emerges concerning differences in achievement gaps between subjects and immigrants' generations. Although non-native students show a sizable gap, it is more pronounced in Language than in Math. Moreover, the difference between first and second generation students' achievement gaps with respect to native peers is relevant in Language skills, but not for Math.

This evidence suggests that the greater difficulties in achievement could be potentially linked to language difficulties and difficulties to interact with teachers and native peers and that second generation students benefit from their longer stay in the hosting country showing a greater assimilation to the hosting country language. Further descriptive evidence can be drawn from a recent survey on non-native adolescent integration in the society which confirms that first-generation non-native have more difficulties at school mainly driven by shyness, language, difficulties in interactions with classmates and teachers (CNEL, 2011). Thus, I focus on first generation non-natives and their contribution to the within-group social effects estimating the following model:

[44] Ideally, I would also perform the analysis on second-generation non-native students (the ones born in Italy) but given their limited presence, results cannot be considered robust.

$$\bar{y}_{st} = \bar{X}_{st}^{'1G}\alpha + \hat{\beta}_{1G}\hat{P}_{st}^{1G} + S_{st}^{'1G}\theta + W_s^{'}\gamma + \phi_t + \eta_{st} \quad [9]$$

If the 'assimilation' channel plays a role in the within-group social interaction effects I would expect different estimates of the social parameter with respect to the baseline model. In detail, if lower assimilation of first generation non-natives is associated with greater within-group negative externalities I would expect the social interaction parameter to be negative and greater in absolute terms with respect to the baseline IV estimates of the effects of the whole non-native group (i.e. $\left|\hat{\beta}_{1G}\right| > \left|\hat{\beta}\right|$). On the contrary, if assimilation does not play a role, I would not find differences (i.e. $\left|\hat{\beta}_{1G}\right| \cong \left|\hat{\beta}\right|$).

[Table 19 here]

Table 19 shows the results both for the OLS and IV estimates. Focusing on IV estimates, I find that social interaction effects are negative and greater in absolute terms for first-generation peers with respect to the estimates for the whole group of non-native students ($\left|\hat{\beta}_{1G}\right| > \left|\hat{\beta}\right|$). The hypothesis that assimilation plays a substantive role in shaping within-group peer interactions cannot be rejected. In fact, the 'less assimilated' sub-group of first-generation students generate greater negative within-group social effects with respect to the whole group of the non-native students. Thus, 'difficulties in assimilation' can be considered a channel that exacerbates the negative within-group externalities. In this sense, integration through education is a powerful tool to favour assimilation of immigrants, at least by reducing the negative within-group social interactions effects.

2.7. SENSITIVITY ANALYSIS

In this section I test the robustness of the results paying particular attention to the instruments used. I repeat the analysis to test the robustness of the results according to: (*i*) a different specification of the endogenous variable; (*ii*) a different definition of the 'non-native group'; (*iii*) alternative indices to measure segregation at the school level. The sensitivity analysis shows that, once the assumptions on the instrumental variable Z_s are accepted, results are stable and robust.

2.7.1. Different specification for the endogenous variable and for the 'non-native group'

I test the robustness of the results using a different specification of the endogenous variable. Instead of the non-native school share in 8^{th} grade students I simply use the number of 8^{th} grade non-native students in the school.

[Table 20 here]

Table 20 contains the estimation of the baseline model (eq. [7]) substituting the endogenous variable P_s with the simple number of non-native students in the school. First-stage F-statistics show that the coefficients are always strongly identified. There are not significant differences in the results, apart from the interpretation of the coefficients. Including the full set of explanatory variables, 2SLS results show that there is a negative impact of the number of non-native students on the test scores of the same non-native students: one additional non-native student determines a decrease in the average test score by -0.8% in Language and -0.9% Math.

So far, I have focused the analysis on the effects within the peer group composed by the non-native students, where for the identification of the 'non-native status' I rooted upon a simple citizenship criterion: non-native students are students without Italian citizenship which means that both student's parents do not have the Italian citizenship (so called *ius sanguinis* rule). This is because the citizenship is an administrative record and does not show severe missing-values problems in the Invalsi IC data. However, I also rerun the analysis exploiting a slightly different specification and identifying the 'immigrant' group following the OECD-PISA definition already introduced (OECD, 2010). This definition partially overlaps with the 'non-native status' definition, but it is stricter and include less students. Notice also that Invalsi IC data offers for the first time the chance to identify separately first and second generation immigrant students in Italy although this classification could be less precise. In fact, 'student citizenship' is an administrative compulsory information that parents are obliged to give to schools staff at the moment of the enrolment, while the information about parents' place of birth used to identify the 'immigrant status' is given on voluntary basis. I rerun the analysis using the 'immigrant school share' as endogenous variable. Results (Table 20) always point to a negative effect between immigrant share (or number) and the average test score results of the same immigrant peer group; first stage F-statistics always reject the null of weak instrument. A 1% increase in immigrant school share determines a decrease of -27% in Language and -34% in Math. Even if the general pattern of the results confirm the ones obtained in the main analysis, the magnitude of the effects is smaller compared to the results obtained with the 'non-native status' definition. This can be due to the fact that the 'immigrant' definition is 'stricter' as it encompasses only students born from *both* foreign-born parents and to measurement error due to missing data.

2.7.2. Alternative segregation measures

Finally, I also implement two alternative measures of segregation at the school level: the isolation index and the inverse of the exposure index (Clothfelter, 1999). Both indexes range from 0 (no segregation) to 1 (maximum segregation). The isolation index measures the extent to which non-natives are exposed only to one other, rather than to natives: at the school level, the index is computed as the non-native-weighted average of each class non-native population. In symbols, using the same terminology as the one of the dissimilarity index:

$$I^{j} = \sum_{c=1}^{N_j} \left(\frac{Non-natives_c}{Non-natives_j} \cdot \frac{Non-natives_c}{(Non-natives_j + Natives_j)} \right) \qquad [10]$$

The inverse of the exposure index is a measure of isolation similar to the isolation index, and it is computed as the inverse of the standard exposure index (Echenique and Freyer, 2006). I choose these two indexes because they offer a measure of school segregation which is based on the ‘extra-groups contacts’ between natives and non-natives rather than on the unevenness of the distribution of non-natives, as the dissimilarity index.

[Table 21 here]

The estimated coefficients (Table 21) are all negative and statistically significant, confirming the robustness of the results. In general, increasing segregation at the school level is associated with greater within-group negative social interaction effects.

2.8. CONCLUSIVE REMARKS

Although U.S. scholars have long focused on the social interaction effects of desegregation policies on minority students attainments (Angrist and Lang, 2004; Hanushek et al., 2009), European literature has rarely focused on peer effects as a potential channel which contributes to explain the well-known gap in achievement between natives and non-natives. Recent studies investigating the causes of non-native students’ underachievement in European schools have primarily focused on immigrant families socio-economic background, without investigating the contribution that school segregation and social interactions have on explaining the gap. In this light, a clearer understanding of the internal dynamics of the minority students groups of peers is fundamental in order to understand whether and following which channels social interactions within non-native students widen the existing attainment gap and under which mechanisms (Fryer, 2010).

From the methodological point of view, I solve the endogeneity of non-native school share exploiting the original features of the dataset and building an instrumental variable which is school-specific, shows a strong predictive power and that is plausibly exogenous to the outcome measures. Once the assumptions on the instrumental variable are accepted, results are stable and robust under a variety of falsification tests and alternative specifications. The population for the analysis is constituted by all non-native students enrolled in the 8^{th} grade of Italian junior high schools in the 2008-09 and 2009-10 school years and the Italian background is similar to those of European countries (among the others, Spain and U.K.) which experienced massive migration waves of unskilled and often undocumented individuals in the last two decades.

I find strong negative within-group social interactions effects: increasing by 1% the non-native school share determines a decrease of -73% in Language and -82% in Math mean test scores of the same non-native students. The analysis of the mechanisms at work supports the evidence of ‘acting-native’ behaviours in Italian junior-high schools. That is, I find that the negative within-group effects are stronger the greater is the degree of within-school segregation (i.e. the less uniformly non-native students are allocated across the classes of the same school) and the lower is the degree of assimilation (focusing on first generation students results are negative and greater, in absolute terms, compared to the entire group of non-natives). Findings are robust to different specifications of the segregation index and to different definitions of the endogenous variable and of the ‘non-native’ group. These results are new in the European context while in line with the negative ‘within-race effects’ found in the U.S. and close to Maestri (2011) who finds that ‘ethnic diversity’ within classes has positive externalities both on attainments levels and behaviours of native and non-native students.

The evidence on the existence of acting-white behaviours and negative within-group effects supports the general idea that a successful immigration policy has to be concerned with the assimilation and integration of the immigrants (Dustman and Glitz, 2011) and that successful integration policies for immigrant children should start as early as possible in the school path. Focusing on school policy implications, this work suggests avoiding any sort of segregation of non-native students across schools and across classes within the same schools. Schneeweis (2011) points out that school segregation of non-native students in European contexts is primarily due to two factors: immigrants’ families residential segregation and (explicit or hidden) selectivity criteria in the school system. However, the formation of ‘enclave classes’ by school heads or the (more or less explicit) constitution of ‘segregated schools’ has often found policy justification under the idea that ‘special schools’ or ‘special classes’ might favour the work of teachers which can concentrate their efforts on a group of non-native students with homogeneous educational needs and difficulties. In some sense, if schools are only concerned in maximizing their profit function, or equivalently, in maximising the average school performance in test scores and national assessment programs, it can be shown that the non-native students’ segregation in ‘enclave’ classes may lead to the result (Lazear, 2001). However, this view totally disregards the social implications of such programs. I show that increasing school segregation is detrimental to the educational outcomes of non-native students and to their integration with native peers. Similar policies heavily neglect the importance played by social interactions: ‘oppositional culture’ behaviours and the ‘lack of assimilation’ with native peers could dramatically exacerbate the existing attainment gaps (Fryer, 2010). On the contrary, adequate mixing rules in schools and class composition criteria could easily mitigate these negative effects.

Moreover, from a school-path perspective, within-school segregation must be avoided especially in the lowest levels of the education path, where pupils could be more easily integrated with native peers. This aspect is particularly relevant in educational systems with explicit tacking. For instance, Ludemann and Schwerdt (2012) show that, conditional on students’ attainments, the early tracking system in German schools generates greater negative effects for second generation immigrant students which largely explain the wage gap differential between native and second generation immigrants. It is also relevant in educational systems, like the Italian one, where an ‘implicit tracking’ for non-native students takes place at the end of the 8^{th} grade, in the passage from the comprehensive compulsory education to the upper secondary education. In fact, non-native students tend to cluster in vocational schools or even drop out after junior high school (MIUR, 2010). As an indirect result, the sorting effects leading to enclave vocational schools in secondary education potentially prevent non-natives from an effective assimilation if this has not started as early as possible during the school path, from primary and junior high schools.

TABLES CHAPTER 2

Table 13. Individual and school level descriptive statistics.

Panel A: Individual Level	IC 2008-09				IC 2009-10			
	North	*Centre*	*South*	*Tot.*	*North*	*Centre*	*South*	*Tot.*
No. Students	211,567	93,440	205,856	510,863	206,530	91,629	199,405	497,564
% Non-natives	11.20	8.97	1.83	7.04	11.24	9.28	1.84	7.12
% Immigrants	11.42	9.02	1.98	7.21	11.24	9.47	1.99	7.22
% First Gen. Imm.	8.56	6.62	1.28	5.25	8.41	6.82	1.24	5.21
% Second Gen. Imm.	2.22	1.63	0.45	1.37	1.01	1.17	0.40	0.78
Panel B: School Level	IC 2008-09				IC 2009-10			
	North	*Centre*	*South*	*Tot.*	*North*	*Centre*	*South*	*Tot.*
No. Schools	2359	1017	2427	5803	2368	1009	2356	5733
No. Schools with non-native students	2103	932	1425	4460	2096	911	1356	4363
% Schools with non-native students	89.15	91.64	58.71	76.86	88.51	90.29	57.55	76.10
% Public Schools	83.59	86.52	95.09	88.91	83.78	86.72	95.33	89.04
% K-8 schools	66.21	63.32	95.10	63.26	69.04	67.19	66.04	67.49
Avg. No. Students per School	341.94	342.03	298.58	322.21	306.61	315.50	291.01	301.72
Avg. No. Students per Class	21.20	20.83	19.75	20.48	21.30	21.00	20.08	20.74
Pupil-teacher Ratio	11.60	11.64	10.30	11.02	21.15	15.74	13.12	16.77
% Schools linked to Catchment Area Info.	94.82	94.00	93.08	93.95	93.12	92.86	91.85	92.55

Table 14. Descriptive statistics at the individual level on IC 2010: students' origins and test scores.

	Language test score						
	% Students	*Mean*	*Median*	*2nd Q.*	*3rd Q.*	*Variance*	*Δ Mean [(a)-(b)]*
Native (a)	92.878	60.904	63.110	51.125	73.440	304.074	7.418*
Non Native (b)	7.122	53.486	54.251	42.418	65.250	274.005	
1st Gen. Imm. (a)	6.021	53.398	54.165	42.351	65.067	271.577	-5.046*
2nd Gen. Imm. (b)	0.931	58.444	59.983	48.699	69.612	256.839	
	Math test score						
	% Students	*Mean*	*Median*	*2nd Q.*	*3rd Q.*	*Variance*	*Δ Mean [(a)-(b)]*
Native (a)	92.878	52.262	52.195	41.865	64.049	272.017	4.602*
Non Native (b)	7.122	47.659	47.126	37.226	57.292	229.224	
1st Gen. Imm. (a)	6.021	47.663	47.134	37.232	57.263	226.876	-2.253*
2nd Gen. Imm. (b)	0.931	49.916	49.731	39.671	59.884	240.780	

Notes. Test scores range from 0 to 100 (percentage of right answers) and are cheating-corrected. The last column contains standard t-test (with different variances) results on the difference between means of each (a) – (b) category; star indicates whether the mean difference is statistically significant (p.val≤0.05).

Table 15. Variables description.

Type	*Name*	*Description*	*Source*
Individual (X)	*female*	Fraction of non-native females in school s	Invalsi
	late	Fraction of non-native retained students in school s	
	father place of birth	Fraction of non-native students in school s with father born abroad	
	mother place of birth	Fraction of non-native students in school s with mother born abroad	
	always_italy	Fraction of non-native students in school s in Italy since birth	
School level (S)	*istituto*	Dummy equal 1 if "K-8 school"	Invalsi
	statale	Dummy equal 1 if State school	
	tot_alunni *tot_alunni2*	School size, given by the total number of students in the school and its square	MIUR / Invalsi
	avg_class *avg_class2*	Average class size in each school and its square	
	handicap_percent	Percentage of students with disabilities in the school	
	pt_ratio	Pupil-to-teacher ratio	
	it_ratio	Non-native students-to-support Teacher ratio	
	tl_class_iii	Fraction of 40-hours classes in 8^{th} gr	
	High_cheating_dummy (subject specific)	Dummy equal 1 if the school is in the 9^{th} decile of the school cheating coefficient distribution	
Province by year Fixed Effects	*provyearFE_**	Interaction dummies for provinces (103 dummies) and years (2 dummies)	
Catchment Area (W)	*lpop*	Log of total resident population	Census 2001
	illiterate	Fraction of illiterate pop.	
	university_edu	Fraction of pop. with university level education	
	m_occup_rate	Male occupation rate	
	f_occup_rate	Female occupation rate	
	agri_oc	Fraction of workers occupied in agriculture	
	self_empl	Fraction workers self-employed	
	commuter	Fraction of resident commuting every day for school or working reasons	
	avg_family_members	Average number of family members	
	house_poor	Fraction of houses without clean water	
	house_new	Fraction of houses built after 1980	
	avg_rooms	Average number of rooms per house	

Table 16. Baseline estimates OLS and IV: effect of non-native school share on school mean test of non-native peers.

	Dep. Var.: log School Mean Score for NON-NATIVE students					
	Panel A: OLS estimates					
	Language			Math		
Non-Native School Share	-0.2891***	-0.1976***	-0.2059***	-0.2316***	-0.1594***	-0.1659***
	(0.0399)	(0.0548)	(0.0566)	(0.0405)	(0.0559)	(0.0579)
R sq.	0.187	0.219	0.222	0.202	0.237	0.240
Adj.R sq.	0.159	0.190	0.193	0.174	0.210	0.211
N	6201	6201	6201	6201	6201	6201
	Panel B: IV estimates					
	Language			Math		
Non-Native School Share	-0.4656***	-0.6856***	-0.7328***	-0.5130***	-0.7652***	-0.8156***
	(0.1320)	(0.2108)	(0.2561)	(0.1330)	(0.2113)	(0.2524)
R sq.	0.184	0.210	0.212	0.196	0.223	0.224
Adj.R sq.	0.156	0.181	0.182	0.168	0.195	0.195
N	6201	6201	6201	6201	6201	6201
1st stage F-statistic	294.56	231.64	203.22	294.56	231.11	202.75
Individual Charact. (X) and Province*Year FE	yes	yes	yes	yes	yes	yes
School Charact. (S)		yes	yes		yes	yes
Catchment Area (W)			yes			yes

Notes. Sig. level: * p<0.1, ** p<0.05, *** p<0.01.

Table 17. First stage regressions.

	Endogenous Dep. Var.: Non-native students school share		
Non-Italian residents in the school catchment area in 2001 (Population CENSUS 2001)	0.00871***	0.00568***	0.00521***
	(0.00051)	(0.00037)	(0.00036)
First stage F-statistics	294.56	231.64	203.22
R sq.	0.340	0.629	0.640
Adj.R sq.	0.317	0.616	0.627
N	6201	6201	6201
Individual Charact. (X) and Province*Year FE	yes	yes	yes
School Charact. (S)		yes	yes
Catchment Area (W)			yes

Notes. Sig. level: * p<0.1, ** p<0.05, *** p<0.01.

Table 18. Individual level OLS estimates.

	Dep. Var.: log (individual test score)			
	Language		Math	
	(a)	(b)	(a)	(b)
Female	-0.0042***	-0.0042***	0.0085***	0.0081***
	(0.0006)	(0.0006)	(0.0007)	(0.0007)
Retained	-0.1888***	-0.1860***	-0.1619***	-0.1635***
	(0.0016)	(0.0016)	(0.0015)	(0.0016)
Advance	0.0633***	0.0634***	0.0600***	0.0598***
	(0.0018)	(0.0018)	(0.0021)	(0.0021)
Non-native	-0.1241***		-0.0616***	
	(0.0017)		(0.0017)	
First Gen. Imm		-0.1374***		-0.0605***
		(0.0021)		(0.0020)
Second Gen. Imm		-0.0841***		-0.0690***
		(0.0034)		(0.0035)
R sq.	0.173	0.173	0.259	0.259
Adj.R sq.	0.167	0.167	0.254	0.254
Clusters	47887	47872	47890	47875
N	874157	868203	874185	868233
Year fixed effects	X	X	X	X
School Fixed Effects	X	X	X	X

Notes. Robust std. errors clustered at class level. Sig. Lev. * p<0.1, ** p<0.05, *** p<0.01. Female: dummy equal 1 if female; Retained: dummy equal 1 if retained; Advance: dummy equal 1 if younger than normal age for 8^{th} grade (i.e. enrolled one year in advance); Non-native: dummy equal 1 if non-native; First Gen. Imm.: dummy equal 1 if first generation immigrant; Second Gen. Imm.: dummy equal 1 if second generation immigrant

Table 19. Mechanisms: non-linear effects with the interaction of the Dissimilarity Index and assimilation effects in the subgroup of first generation non-native students.

	Dep. Var.: log School Mean Score for Non-native students			
	Language		Math	
	OLS	IV	OLS	IV
Non-native SS*D_Index	0.0052*	-0.0897**	0.0080***	-0.1002***
	(0.0027)	(0.0361)	(0.0027)	(0.0369)
R sq.	0.221	0.081	0.240	0.049
Adj.R sq.	0.192	0.047	0.211	0.013
N	6200	6200	6200	6200
1st stage F-statistic		26.54		26.49
	Language		Math	
	OLS	IV	OLS	IV
First Gen. school share	0.2604***	-0.9829***	0.3241***	-1.0981***
	(0.0590)	(0.3588)	(0.0592)	(0.3549)
R sq.	0.223	0.174	0.242	0.175
Adj.R sq.	0.194	0.143	0.214	0.144
N	6200	6200	6200	6200
1st stage F-statistic		122.00		121.81
All Controls	yes	yes	yes	yes

Notes. Sig. level: * p<0.1, ** p<0.05, *** p<0.01.

Table 20. Sensitivity analysis: different specification of the endogenous variable (Panel A) and of the immigrant group (Panel B).

	Dep. Var.: log School Mean Score for Non-native students			
Panel A	Language		Math	
	OLS	IV	OLS	IV
Non-native No.	-0.0022***	-0.0084***	-0.0015***	-0.0094***
	(0.0004)	(0.0030)	(0.0004)	(0.0029)
R sq.	0.223	0.202	0.240	0.205
Adj.R sq.	0.194	0.172	0.211	0.175
N	6201	6201	6201	6201
1st stage F-statistic		134.23		133.53
	Language		Math	
Panel B	OLS	IV	OLS	IV
Immigrant school share	-0.0395	-0.2787	-0.0683***	-0.3400
	(0.0283)	(0.2356)	(0.0259)	(0.2247)
R sq.	0.174	0.162	0.219	0.203
Adj.R sq.	0.142	0.129	0.189	0.172
N	6021	6021	6021	6021
1st stage F-statistic		60.11		59.88
All Controls	yes	yes	yes	yes

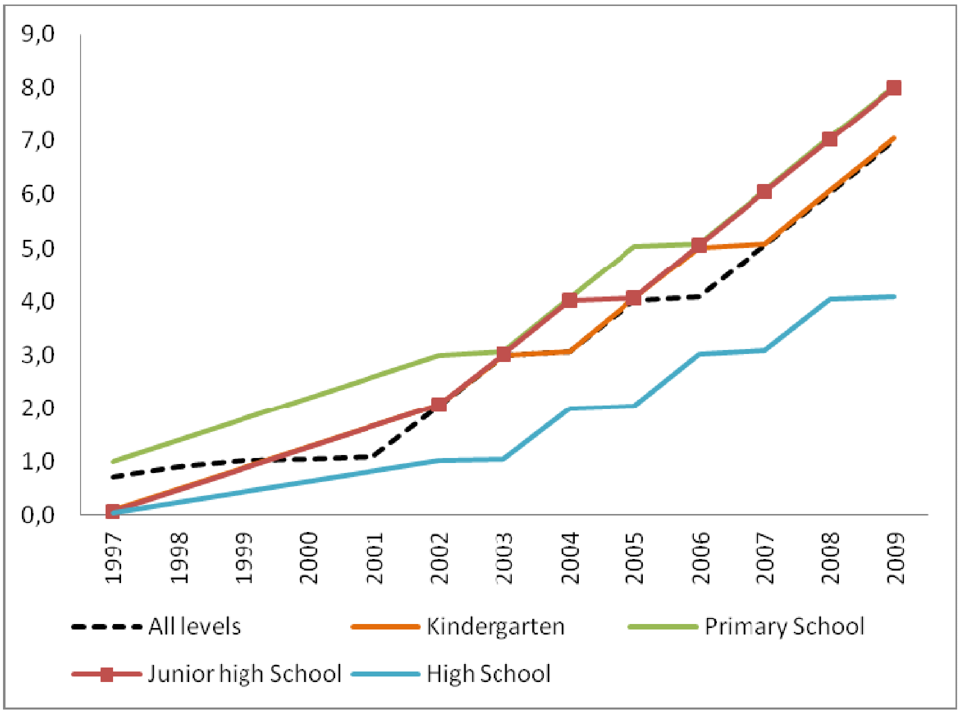
Notes. Sig. level: * p<0.1, ** p<0.05, *** p<0.01.

Table 21. Sensitivity analysis: different segregation measures.

	Dep. Var.: log School Mean Score for Non-Native students					
	Language			Math		
	D	I	IE	D	I	IE
Non-native SS*Index	-0.0719***	-0.0929**	-0.0014***	-0.0880***	-0.1138***	-0.0017***
	(0.0276)	(0.0380)	(0.0004)	(0.0277)	(0.0386)	(0.0004)
R sq.	0.152	0.041	0.369	0.111	-0.031	0.354
Adj.R sq.	0.120	0.004	0.345	0.077	-0.070	0.329
First Stage F-stat Zs	47.33	35.13	32.8	47.05	34.89	32.5
N	6201	6201	6201	6201	6201	6201
All Controls	yes	yes	yes	yes	yes	yes

Notes. Sig. level: * p<0.1, ** p<0.05, *** p<0.01.

Figure 4. Non-native students percentage in the Italian school system, from s.y. 1996-07 to 2008-09.



Source: elaboration on Ministry of Education Statistical Office data (2009).

Figure 5. Dissimilarity Index and non-native school share.

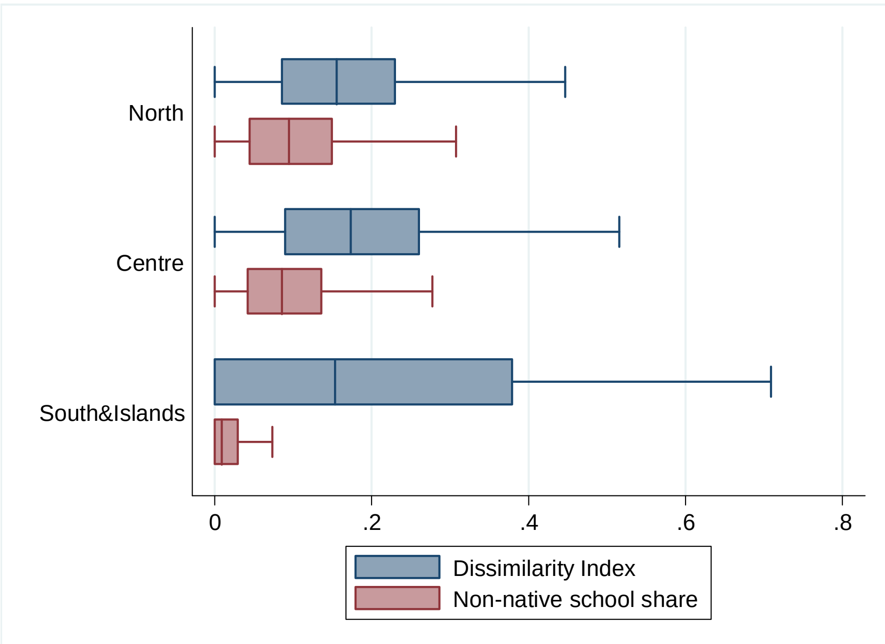
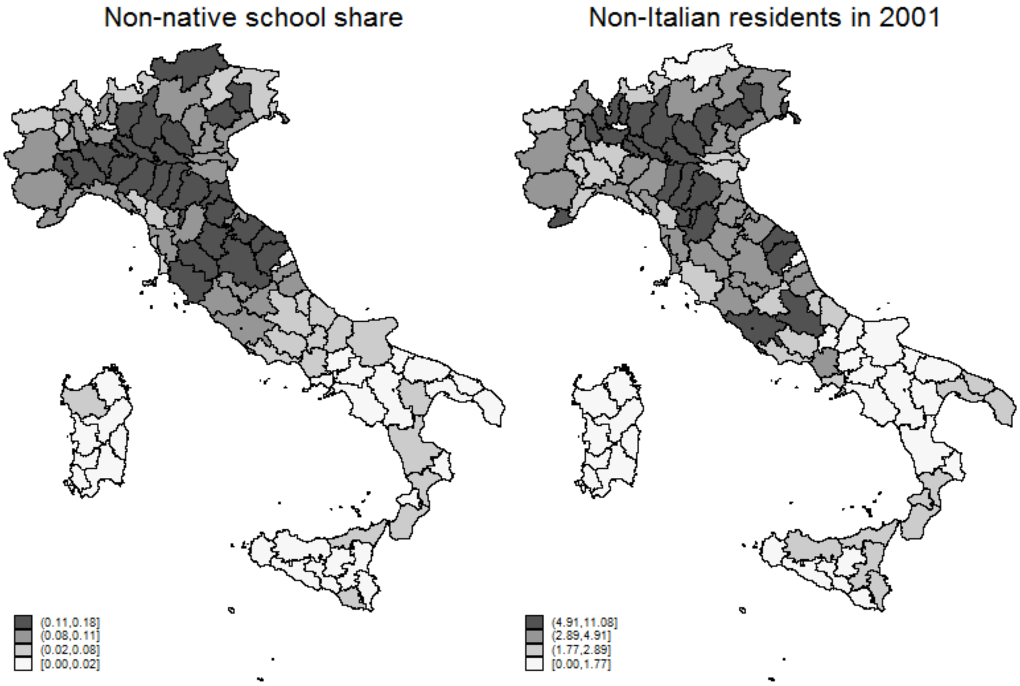


Figure 6. Comparison between non-native school share (endogenous variable), on the left, and non-Italian residents in 2001 (instrument), on the right. Geographical distribution by province.



CHAPTER 3.

STUDENTS' CHEATING AS A SOCIAL INTERACTION: EVIDENCE FROM A RANDOMIZED EXPERIMENT IN A NATIONAL EVALUATION PROGRAM[*]

ABSTRACT

We analyze students' cheating behavior during a national evaluation test. We model the mechanisms that trigger cheating interactions between students and show that, when monitoring is not sufficiently accurate, a social multiplier may magnify the effects on students' achievements. We exploit a randomized experiment, which envisaged the presence of an external inspector in the administration and marking of the tests, to estimate a structural (endogenous) social multiplier in students' cheating. The empirical strategy exploits the Excess-Variance approach (Graham, 2008). We find a strong amplifying role played by social interactions within classrooms: students' cheating behaviors more than double the class average test scores results. The effects are found to be larger when students are more homogeneous in terms of parental background characteristics and social ties.

JEL Classification: C31, D62, I21.

Keywords: social multiplier, students' cheating, randomized experiment

---

[*] This chapter is a joint work with Claudio Lucifora (Università Cattolica and IZA).

"It's seen as helping your friend out. If you ask people, they'd say it's not cheating. I have your back, you have mine." senior student at Stuyvesant High School in Manhattan.

"We want to be famous and successful, we think our colleagues are cutting corners, we'll be damned if we'll lose out to them, and some day, when we've made it, we'll be role models. But until then, give us a pass." student at Harvard Graduate School of Education.

The New York Times, September 25th, 2012

3.1. INTRODUCTION

In many social and economic contexts individuals often face the choice to adopt different types of opportunistic or even illicit behavior to increase their welfare taking advantage of others for personal interests. Leaving aside major crimes, there is abundant evidence indicating that cheating on taxes, free riding on public goods, claiming benefits without entitlement, bribing and corrupting public officials, abusing of drug and drinking, smoking when not permitted, as well as other types of dishonest behaviors are widely diffused phenomena in most countries (Kleven et al. 2011; Fortin et al. 2007; Powell et al. 2010; Gaviria and Raphael 2001; Clark and Loheac, 2007).

In this paper, we focus attention on a specific type of such fraudulent behavior, that is students' cheating when taking an exam. Several surveys document that students' cheating has grown, over the last decades, hand in hand with the more extensive use of testing programs (Davies et al., 2009; McCabe 2005; Rimer 2003)[45], yet there is little evidence on the effects of cheating behavior for educational outcomes, as well as on the measures taken to contrast its diffusion[46]. Students' cheating behavior can have important consequences in the process of human capital accumulation and for the functioning of the labor market. For example, cheating can interfere with the evaluators ability to assess students' performance and

[45] Large-scale cheating has been uncovered over the last year at some of the US most competitive schools, like *Stuyvesant High School* in Manhattan, the *Air Force Academy* and, most recently, *Harvard University* (The New York Times, September 7, 2012). A survey conducted as part of the Academic Integrity Assessment Project by the Center for Academic Integrity (Duke University) and covering 80,000 students and 12,000 faculties in the U.S. and Canada, between 2002 and 2005, reported that 21% of undergraduates admitted to have cheated on exams at least once a year (McCabe, 2005). Another survey run - in 2010, on a national sample of U.S. public and private high schools students - by the Josephson Institute of Ethics – '*Report on honesty and integrity*' (2011) - found that 59.3% of the U.S. students interviewed cheated at least once during a test, while more than 80% of them copied from others' homework at least once.

[46] In many countries, policy interventions make extensive use of test scores to determine the allocation of resources across schools and to evaluate teachers' work but little has been done to develop objective measures of students' cheating.

can decrease the external validity of grades (Anderman and Murdock, 2007). 'Cheating bias' may contaminate the information used in many educational decisions, such as: promoting students from one grade to the next, or awarding a diploma without the required knowledge. In one case, cheating detracts from the signaling validity of education titles on the labor market; in the other case, it determines negative externalities on the learning processes, for example, slowing down the teaching activity[47].

Moreover, students' cheating raises a number of concerns not just for the unfairness with respect to students who do not cheat, but more generally for the externalities that are created on others (McCabe, Treviño and Butterfield, 1999; Carrel et al. 2008; Dee and Jacob, 2012). In particular, when a student breaks an ethical code of behavior exchanging information, cooperating with other students or using any prohibited materials during an exam (Cizek, 2003), many others – who might otherwise have behaved honestly - end up being influenced thus reacting to such behavior. Many students may feel that they cannot afford to be disadvantaged by those who cheat without being reported or punished by school authorities[48]. In this context, even an isolate cheating behavior may propagate and become larger through social interactions. Hence, as widely discussed in the social interaction literature, the aggregate outcome is likely to depend on a direct effect (a reaction *via* private incentives to cheat) and an indirect effect on behavior (a reaction to the cheating behavior of others): the ratio between the equilibrium aggregate response and the sum of the reactions of individuals to cheating is the so-called social multiplier (Glaeser et al., 2003). The cheating outcome is amplified by the multiplier generating large differences in variance across different groups (i.e. school, classroom, etc.) with otherwise similar characteristics. While unobserved heterogeneity and sorting of individuals across groups may account for part of the differences in cheating behavior, social interactions within group of students linked by different types of contextual ties are often necessary to explain the excess variation that is observed in the data (Manski, 1993; Brock and Durlauf, 2001).

Note that in many circumstances the driving force for dishonest or illicit student's behavior during an exam may be found in some personal traits, such as: greed, envy, competitive pressure, etc.; however, social norms, low trust, a widespread acceptance of illicit behavior

[47] The consequences of cheating can be even more severe in educational settings in which the school system is based on a strict tracking system (e.g. Germany).

[48] Note that reporting the offenders, as contemplated in many schools' ethical codes, is required to halt the diffusion of cheating behaviors, nevertheless it should be noted that small transgressions and dishonest behavior are very often overlooked or tolerated within many schools, either because students do not like to be directly involved in the accusation, or because schools themselves do not want to be associated to the judiciary procedures required to support the allegations of student's dishonesty.

and other background characteristics may also increase the likelihood of dishonesty within students.

In other words, cheating behavior can be seen a genuine free-riding problem, where students, for any given level of effort, try to maximize their performance (i.e. pass-rate probability, exam grades, test scores, etc.) and exploit the possibility of opportunistic behavior – i.e. exchanging information or cooperating - anytime the monitoring system tolerates it or is not efficient in reporting the offenders[49]. The interdependencies between students' decisions to cheat are at the basis of the positive covariance in individual behavior that triggers the (social) multiplier effect. In terms of the framework introduced by Manski (1993), and extensively discussed in the literature on social interactions, the above cheating behavior represents the endogenous part of social effects (Bramoullé et al. 2009; De Giorgi et al. 2010)[50].

The literature on social interactions in education has largely focused on peer effects in students achievements in classrooms and schools, or on social outcomes within fraternity (or sorority) membership (Sacerdote 2001; Zimmerman 2003; Stinebrickner and Stinebrickner 2006; Foster 2006; Graham 2008; Hanushek, et al. 2003; Lyle, 2007; Lefgren 2004; Carrel et al. 2009; Lavy et al. 2012). Conversely, the effect of students' cheating interactions has not received much attention and even less is known about the potential mechanisms that may drive cheating behavior.

An extensive literature in educational psychology has documented cheating behavior in schools[51], while only few papers have addressed the issue of social interactions in cheating behavior using a credible identification strategy. Most papers in the literature use statistical techniques that cannot reliably separate the endogenous and exogenous effects – i.e. the effect

[49] Monitoring activities are introduced to validate testing procedures in national evaluation programs. However, contrary to international programs of students' assessments (e.g. PISA, TIMSS, PIRLS) - which are usually conducted on a survey basis and sampled students sit the test under the supervision of inspectors -, national assessments programs are conducted on a census basis and the same school teachers supervise students while tacking the exam (U.S. Department of Education, 2009; Eurydice, 2009).

[50] Manski (1993) identifies three main factors that are likely to influence social interactions: exogenous (or contextual) effects (i.e. when the propensity of an individual to behave in some way varies with the exogenous characteristics of the group), correlated effects (i.e. common shared group-level factors) and endogenous social interactions (i.e. when the propensity of an individual to behave in some way varies with the behavior of the group). Only the latter effect can determine the social multiplier.

[51] Stephens and Gehlbach (2007) count more than an hundred empirical studies on this issue over the last decade. Research in this area documents that cheating occurs among students from all grades, from elementary schools to colleges, and even in graduate schools. From a developmental perspective, Miller et al. (2007) find that cheating tend to occur less in younger children than in adolescents. These developmental differences are due to changes both in students' cognitive abilities and in the social structure of the educational contexts in which children and adolescents interact (Murdock et al., 2001). From a motivational perspective, Anderman and Murdock (2006) document different reasons for engaging in academic cheating: some students cheat because they are highly focused on extrinsic outcomes such as grades; others cheat because they are concerned with maintaining a certain image to themselves or to their peers or because they lack the requisite self-efficacy to engage in complex tasks.

of the group upon an individual from the effect of an individual upon the group due to the well-known reflection problem (Carrel et al. 2008). Starting from an early study by Stanard and Bowers (1970), where it was shown that cheating tended to be higher among members of a fraternity or sorority, the psychological literature has focused attention on how social norms, peer pressure, environmental pressure and self-perception of cheating behavior affect individual cheating decision. McCabe and Trevino (1997), for example, found peer-related contextual factors to be the strongest predictors of cheating in their multi-campus investigation of individual and contextual influences related to academic dishonesty. Students who perceived that their peers disapproved academic dishonesty were less likely to cheat, while those who perceived higher levels of cheating among their peers were more likely to report cheating. Grimes and Rezek (2005) estimate a probit regression model to determine the factors that contribute to the probability of cheating. Their results indicate that the most important determinants are personal beliefs about the ethics and social acceptability of cheating and various attributes of the classroom environment[52]. Carrel et al. (2008) are the first to analyze cheating behavior as a social interaction using separate estimation procedures to identify an exogenous (contextual or pre-treatment) peer effect and an endogenous (during treatment) peer effect. Their model assumes that peer effects are completely driven either through experiences of cheating behavior at high school or completely through peers' behavior while at college. Their results for the endogenous peer effects indicate that one additional college cheater 'creates' approximately 0.61–0.75 additional college cheaters.

There is also a parallel literature that has focused on other forms of cheating, for example cheating on taxes is one of the most interesting cases. Kleven et al. (2011) analyze a tax enforcement field experiment in Denmark confronting different types of tax reporting methods (i.e. third-party reporting vs. self-reported income), as well as different auditing methods faced by tax filers. The authors show that tax cheating is close to zero for income subject to third-party reporting, but substantial for self-reported income and that prior audits and threat-of-audit letters have significant effects in reducing cheating on self-reported income. Galbiati and Zanella (2012) estimate a social multiplier effect in tax cheating generated by the congestion of the auditing resources. They use a rich dataset from Italian Local Tax Authorities and find that an exogenous shock altering concealed income independently across individuals produces an equilibrium variation that is up to three times the initial response.

[52] Kerkvliet and Sigmund (1999) explore the determinants of source-specific cheating behavior including student characteristics and deterrent measures. They conclude that large alcohol consumption and low grade point average increase the probability of cheating. Jordan (2001) finds a significant correlation between college students' perceived social norms and their self-reported cheating.

We develop a simple theoretical model to highlight the mechanisms that may drive social interactions in cheating behavior and to derive testable predictions. We show that students may optimally decide whether to engage in cooperative effort exchanging information (e.g. conform to other student cheating behavior) and do so taking into account other students' best response. The equilibrium solution takes the form of a linear-in-means model with (endogenous) social interactions *à-la* Manski, so that we can attach a structural interpretation to our estimate of the multiplier (Cooley 2010a,b). In particular, our model posits a specific social spillover: by observing or expecting that student achievements depend also on cheating interactions, students adjust their behavior in response to the cheating behavior in the classroom.

We use a unique data set drawn from the 'National Survey of Students' Attainments' (henceforth SNV) (in both Mathematics and Language), which is compulsory for all schools and students attending different grades of primary and junior-high school in Italy, and exploit a randomized experiment which envisaged the presence of an external inspector in the administration and marking of the tests.

In particular, we contrast the behavior of students in classrooms where the test is administered only by the school teachers, with the behavior of students in classroom where an external inspector invigilates over students' behavior during the exam, to identify students' social externalities in cheating behavior. In the non-monitored classrooms (i.e. our control group), we may expect monitoring to be more 'benevolent' *vis-à-vis* student interactions during the exam, while no interactions are expected to occur in the monitored classrooms (i.e. our treated group).

In this context, we interpret the presence of a positive covariance in students' behavior, when exchanging information or engaging in any sort of collaborative behavior during the test in the non- monitored classroom, as a form a behavioral externality which may produce a social multiplier[53]. Students' cheating behavior during the test has few relevant implications. First, it generates excess variance in individual behavior with respect to individual and group characteristics in the monitored classrooms. Second, it introduces a difference among the between-group and the within-group variance of individual behavior. These two features are the foundations of the empirical strategy proposed by Graham (2008), which exploits the Excess-Variance (henceforth, E-V) approach to separate the part of variability due to

[53] Note that students' cheating during an exam is an interesting case study of social interactions in the classroom, since it is likely to capture the same network of friendships and cooperative behaviors that take place during the school year. Students are more likely to collaborate with closer friends, with classmates they share out-of-the school activities (like sport practice), as well as with classmates sitting closer.

individual and group level heterogeneity from the excess variability genuinely originating from social interactions.

We contribute to different strands of literature. First, to the literature on the identification of grade inflation due to various types of cheating behaviors (Dee and Jacob, 2012, for plagiarism; Carrel et al., 2008, for students' cheating; Jacob and Levitt, 2003, Jacob 2005 for teachers' cheating)[54]. Focusing on students' cheating behavior, our approach departs from Carrel *et al.* (2008) since we do not identify the effects of a given 'share of cheaters' on individual test score, rather we provide a measure of endogenous interactions due to students' cheating behavior. In this sense, we contribute to the part of the literature on social interactions which tries to overcome the 'reflection problem' and directly estimate the effects of the endogenous social multiplier (among others: De Giorgi et al. 2010; Calvò-Armengol et al., 2009; Bramoullé et al. 2009)[55]. Second, we use data on test scores and other individual characteristics drawn from the whole student population at different grades in a national evaluation test, which is a significant improvement from studies which rely on representative samples. We also match our data with other administrative archives, at the school level, and with a follow-up survey to get additional information on parental background characteristics as well as motivational questions concerning the test. Third, we implement a rather innovative estimation method based on the Excess-Variance approach to estimate (endogenous) social interactions by exploiting an exclusion restriction provided by a randomized experiment and illustrate the presence of heterogeneity in the estimated social effects[56].

We find a strong amplifying role played by cheating social interaction within students in the classroom: in the baseline estimates we identify a social multiplier ranging between 2.26 and 2.43 for Math, and between 2.05 and 2.18 for Language. This implies that students' cheating behavior more than double the class average test scores results and the effects are found to be

[54] Evidence of cheating behavior mostly refers to academia (Mc Cabe and Trevino, 1999; Mc Cabe, 2005; Carrel et al. 2008), less from other type of schools. In the Italian context, Ferrer-Esteban (2012) and Bertoni et al. (2012) use SNV dataset to study the effects of supervision on students' performance.

[55] Grounding on Manski's seminal works (1993), empirical literature on peer effects has focused on the estimation of reduced form equations which collapse the endogenous and the exogenous effects into one parameter of interest, that is identifiable and defined to as 'social effect' parameter (Ammermüller and Pischke, 2009; Lavy et al. 2012). Recent works in the field of social interactions in education addressed the reflection problem in the estimation of the classical linear-in-means model à la Manski (1993) using data where social groups are endogenously defined (i.e. networks, Calvo-Armengol et al. 2009; Bramoullé et al. 2009), introducing appropriate exclusion restrictions (e.g. partially overlapping groups in De Giorgi et al. 2010; group-size variations in Davezies et al. 2009), or even just plugging into the reduced form equation a lagged value of peers' achievement as proxy of the contemporaneous one (Hanushek et al. 2003, 2009). In these cases, random assignment is usually ensured in the specific characteristics of the data used (e.g. random assignment to classes and courses at the first year of college), or controlled for using multiple levels fixed-effects.

[56] Galbiati and Zanella (2012) also implement the Excess-Variance approach to tax cheating behavior using a more standard exclusion restriction given by group-size variations.

larger when students are more homogeneous in terms of parental background characteristics and 'social ties'.

Our findings show that tolerating cheating behavior, as it is often done, can amplify the negative effects on students' performance, alter the signaling role of education in the labor market, and raise collective indulgence with respect to various forms of dishonest practices. Also, given that increasing competition in school achievement and in the job market are likely to exerts considerable pressure on students to perform well in exams, more resources should be devoted to monitoring activities in order to avoid cheating interactions to become widespread.

The paper is organized as follows. In Section 3.2 we build a theoretical framework to define the social multiplier parameter. Section 3.3 describes the institutional setting, the data and the randomized experiment, and provides some descriptive statistics. Section 3.4 discusses the identification strategy while Section 3.5 and 3.6 present the main results and some robustness checks. Section 3.7 concludes and provides some policy implications.

3.2. Theoretical framework

We develop a simple model to investigate the mechanisms that may drive social interactions in cheating behavior. We consider the (endogenous) decision students face, when taking an exam, as to whether work individually or, alternatively when the monitoring technology is loose, engage in any kind of prohibited cooperation exchanging information with other students. We assume that students derive utility from achievement, which depends on own (costly) effort and on the effort of classmates[57]. Since own effort and peers' effort in the classroom are complementary inputs in the achievement function, students may decide to cheat choosing the optimal level of cooperative effort to be shared with their peers (Anderman et al., 2007). In this context, cheating originates a behavioral externality among individuals, who simultaneously choose their utility-maximizing level of effort taking into account peers' best response to each level of effort chosen (Brock and Durlauf, 2001). Note that the type of social externality that emerges from student's cheating behavior is different from the traditional peers' achievement externality (i.e. based on predetermined characteristics of the students, such as unobserved ability or 'quality') considered in the literature, since here individual decisions play an important role in shaping students' behavior which, in turn, originates the endogenous effects needed to determine the social multiplier (Sacerdote, 2001;

[57] Notice that we assume no cost of cheating. This is consistent with the institutional setting (and the empirical application, i.e. SNV surveys) as in practice disciplinary measures or sanctions have never been applied to students and teachers who behave dishonestly.

Cooley, 2010a,b; De Giorgi and Pellizzari, 2011; Imberman et al. 2012; Lavy et al. 2012)[58]. In particular, in the present context, peers' achievement *per se* may not affect a student's achievement when cheating or any other form of behavioral interactions are absent[59].

We model students' achievement (y_i) as dependent on the following elements: x_i and x_j are, respectively, individual and peers' predetermined characteristics (i.e. gender, parental background, non-native status, etc), μ represents shared class-level factors (i.e. school and class environment, teacher's experience), while e_i and e_j are, respectively, unobservable individual and peers' endogenous behaviors where *j* indicates any other student different from *i*:

$$y_i = x_i^{'}\pi_x + x_j^{'}\tilde{\pi}_x + e_i + \tilde{\pi}_e e_j + \mu \qquad [1]$$

The above specification describes the Achievement Production Function (APF) suggesting that achievement is increasing in both the student and peers' unobservable behavior, such that individual achievement may improve when cheating externalities are present (i.e. $\tilde{\pi}_e > 0$). The parameter of interest here is $\tilde{\pi}_e$ which identifies the endogenous social interactions characterizing students' behavior in the presence of cheating. The other parameters π_x, and $\tilde{\pi}_x$ describe exogenous (or contextual) effects (Cooley, 2010b)[60]. To get further insights on

[58] In terms of behavioral interactions, the literature on drug use, smoke habit or alcohol addiction provides a better illustration of a social mechanism through which group's behavior directly affects individual decision. In these cases, we have endogenous peer effects whenever the 'average behavior' of the reference group directly influences the individual behavior or choice. It is the group's decision to drink, smoke or use drugs that influences the individual decision to take some action, and both group and individual behaviors are directly captured by some quantifiable measures (alcoholic drinks per day, binary decision to smoke/not smoke, or cigarettes per day etc.) (Cooley, 2010b; Gaviria and Raphael, 2001; Sacerdote, 2001).

[59] The empirical literature on peer effects, traditionally, does not distinguish between the effect on test scores deriving from unobservable pre-determined characteristics of the students and their unobservable behavioral choices (Sacerdote, 2001; Imberman et al. 2012; Lavy et al. 2012). However, as noted by Cooley (2010b, p. 7) "*[…]Annual standardized exams are often the outcome of interest, and, in the absence of cheating, are not a group effort. Thus, peer achievement per se may not affect a student's achievement. In contrast, the decision of a teenager to smoke or drink alcohol might be readily affected by having peers that engage in these behaviors*". Examples of endogenous peer behavior on achievement are discussed in Lazear (2001) where peer disruptive behavior imposes negative externalities on other students in the classroom. Similarly, Figlio (2007), Lavy and Schlosser (2011) and Kinsler (2006) present empirical evidence that disruptive peers may negatively affect achievement. In the robustness section we also test whether achievement peer effects play a role in our data.

[60] Note that human capital externalities still operate in the APF (eq. [1]) but, in some sense, they can be thought as being part of the individual and peers' predetermined characteristics and contribute to individual outcome as 'endowment effects' (i.e. exogenous effects incorporated in $\tilde{\pi}_x$). Given that we only want to estimate the endogenous component of the social interactions process due to cheating behavior during the test, and that students' quality is likely to be same in the two sub-populations used for the empirical estimation, we assume that x_i also includes unobserved predetermined individual characteristics.

how student's behave, we specify individual's utility (U_i) as a quadratic function that depends positively on achievement ($\beta_y \geq 0$) and it is concave in own effort costs ($\beta_e \geq 0$):

$$U_i\left(y_i, e_i, e_j, c\right) = \beta_y y_i - \frac{1}{2}\beta_e e_i^2 + \tilde{\beta}_e e_i e_j - \bar{c}(\mu) \tag{2}$$

The component $\bar{c}(\mu)$ represents an exogenous cost due to teacher's monitoring activity during the test. All individuals have to bear this cost which is likely to depend on class-level characteristics (strictness in teacher monitoring, class physical dimension, desks allocations, etc.). Notice that peers' behavior matters as long as there are social interactions during the exam (i.e. loose or benign monitoring allows cheating) and students are willing to share their effort cooperating with other students (i.e. conforming to other students' cheating behavior): hence individual utility increases with peers' effort ($\tilde{\beta}_e > 0$). Students maximize utility choosing the level effort as best response to peers' (simultaneous) effort choices and subject to the achievement function (i.e. given by the structural APF):

$$\begin{aligned} &\underset{e_i}{Max}\ U_i = \beta_y y_i - \frac{1}{2}\beta_e e_i^2 + \tilde{\beta}_e e_i e_j - \bar{c}(\mu) \\ &s.t. \qquad y_i = x_i^{'}\pi_x + x_j^{'}\tilde{\pi}_x + e_i + \tilde{\pi}_e e_j + \mu \end{aligned} \tag{3}$$

Solving for e_i the first order condition yields the effort best response function:

$$e_i^{BR} = \frac{\beta_y}{\beta_e} + \frac{\tilde{\beta}_e}{\beta_e} e_j \tag{4}$$

The effort best response is a function of the marginal utility of effort relative to the cost and is increasing in the average effort of peers when cheating interactions occurs (i.e. cooperative peers' effort, $\tilde{\beta}_e > 0$). Given the assumption that achievement is monotonically increasing in cooperative effort behavior (e_i), the effort best response can be mapped into an achievement best response which is observable to the researcher[61]:

$$y_i^{BR} = \delta_0 + x_i^{'}\delta_x + x_j^{'}\tilde{\delta}_x + J y_j + \mu^{'}\tilde{\delta}_\mu \tag{5}$$

[61] See Appendix A for the detailed derivation.

Given the linear-in-parameters form of the achievement best response, it can be shown that a unique Nash equilibrium exists ($y_i^*; y_j^*$) so that equation [9] can be rewritten as:

$$y_i^* = \delta_0 + x_i^{'}\delta_x + x_j^{'}\tilde{\delta}_x + Jy_j^* + \mu^{'}\tilde{\delta}_\mu \qquad [6]$$

Under the assumption that the achievement observed during any exam or test (i.e. grade, tests scores, etc.) originates from the described utility-maximising behavior - when cheating occurs - we can use peer achievement to proxy for peer cooperative behavior (effort) such that equation [6] expresses individual achievement as function of individual and peers' characteristics as well as peers' achievement. The parameter *J* corresponds to the 'unobserved endogenous social effects' and it is a measure of the endogenously determined effect of individual behavior on the reference group average behavior:

$$J = \frac{\tilde{\beta}_e + \beta_e\tilde{\pi}_e}{\beta_e + \tilde{\beta}_e\tilde{\pi}_e} \qquad [7]$$

It is composed by three structural parameters: the marginal (dis)utility from own effort exerted in cheating activities (β_e), the marginal utility derived from peers' effort in cooperative cheating behavior ($\tilde{\beta}_e$), and the marginal effect of peers' effort exerted in cheating on individual achievement ($\tilde{\pi}_e$).

The linear-in-means model in equation [6] requires $J < 1$ (i.e. a stability condition to ensure that a small change in cheating behavior will not determine a diverging response in aggregate), and this is true if two restrictions are imposed to the structural parameters: that is $\tilde{\beta}_e < \beta_e$ and $\tilde{\pi}_e < 1$ [62]. The first condition states that the utility from cooperative cheating behavior (i.e. peers' effort) must be smaller than the disutility from own effort; the second condition requires the marginal contribution of peers' effort on individual achievement to be smaller than own contribution (i.e. normalized to 1 in the APF, see equation [1]). Both conditions are rather intuitive and realistically met in our framework. Notice also that, when monitoring allows cheating to occur, the assumption of cooperative peer effort (i.e. $\tilde{\beta}_e > 0$), implies that *J* is always positive ($J \geq 0$)[63]. In other words, as we show in the descriptive evidence, when the monitoring technology prevents students to interact or cooperate during

[62] $\pi_x > 0$ is without loss of generality, assuming that covariates are constructed accordingly.

[63] It is easy to show that, since we have assumed cooperative peer effects ($\tilde{\beta}_e > 0$), and given that $\beta_e > \tilde{\beta}_e$, this necessarily implies that $\beta_e > 0$ and also $\tilde{\pi}_e > 0$ thus ensuring that $J \geq 0$.

the test, their achievements (or test-scores) tend to be more dissimilar and exhibit a larger within-class variance as compared to the achievements of the non-monitored students where behavioral interactions are present.

2.1. The social multiplier

The simple model described above implies a social multiplier, such that any shock to individual behavior - *via* social interactions - determines relatively larger aggregate responses. To frame the model in a way suitable for empirical estimation, we need to retrieve an expression for the social multiplier. First, without loss of generality, we can rearrange equation [6] substituting average peers' characteristics and average peers' achievement:

$$y_{ic}^* = \delta_0 + x_i^{'}\delta_x + \bar{x}_c^{'}\tilde{\delta}_x + J\,\bar{y}_c^* + \mu^{'}\tilde{\delta}_\mu \qquad [8]$$

Averaging within the reference group (i.e. the classroom) and solving for $\bar{y}_c$ yields [64]:

$$\bar{y}_c = \gamma\delta_0 + \bar{x}_c^{'}\gamma(\delta_x + \tilde{\delta}_x) + \mu^{'}\gamma\tilde{\delta}_\mu \qquad [9]$$

Where $\gamma = (1-J)^{-1}$ represents the social multiplier in students' cooperative efforts, when during the exam cheating can occur (Glaeser *et al.*, 2003). Substituting equation [9] into [8] we obtain the following reduced form model:

$$y_i^* = \gamma\delta_0 + x_i^{'}\delta_x + \gamma\tilde{\delta}_x\bar{x}_c^{'} + (\gamma-1)\delta_x\bar{x}_c^{'} + \gamma\delta_\mu\mu^{'} \qquad [10]$$

The achievement best response takes the form of the classical linear-in-means model of social interaction *à la* Manski (1993). While this has been obtained at the cost of introducing some *ad hoc* linear functional forms, it has some clear advantages[65]: first, it highlights the mechanism through which cheating behavior may generate students' social interactions; second, it provides a specification that allows direct estimation of the social multiplier parameter (γ) using the Excess-Variance approach (Graham, 2008).

[64] In the social interactions literature the group whose (average) behavior influences the behavior of each individual is considered a "reference group", in our setting the classroom is the natural reference group to be considered in the empirical analysis.
[65] See Cooley (2010a) for an illustration of the general case.

Note that, in some sense, the interpretation of the social interaction parameter as students' cooperative effort is specific to our model, since cheating is the only social externality we are modeling. However, while we think that cheating externalities are the main driving force in the estimation of our structural social multiplier, we cannot exclude that other social mechanisms may also play a role. We briefly discuss some alternative interpretations hereafter.

One hypothesis, also discussed in the literature (Jacob and Levitt, 2003a,b; Jacob, 2005; Lavy, 2009), is that our social multiplier parameter may originate also from explicit teacher cheating rather than students' cooperative efforts in exchanging information when monitoring is more benevolent or looser. Teacher cheating may take the form of suggesting the right answers to all students, or even altering students' answers sheets during the marking phase. Indeed, besides the ethical implications of such behavior, there are several reasons why teachers may want to alter students' outcomes: for example, they may wish to improve their students' results in the exams, alternatively teachers may dislike sharp differences in results across classes within the same school, or feel pressure because of monetary incentives linked to student performance, or because the allocation of resources to schools depends on students outcomes (Jacob, 2005; Lavy 2009). A second hypothesis is that students in classroom with an external inspector feel intimidated and are negatively affected in their performance during the test (Bertoni et al. 2012). Finally social effects may also derive from the presence of some ethical norms of behavior whose strength decreases with the extent of cheating itself (Algan et al. 2011; Myles and Naylor, 1996).

In a later section we provide evidence to prove the robustness of our results to these alternative effects and their interpretation. Moreover, while we cannot exclude that some of the above effects is at work, it should be stressed that their presence does not invalidate our estimation procedure to provide a structural estimation of the social multiplier, while the randomized experiment in the data allows us to identify precisely behavioral interactions (i.e. cheating) during an exam.

3. INSTITUTIONAL CONTEXT, DATA AND DESCRIPTIVE STATISTICS

The Protocol for the SNV survey entails the use of external inspectors for the administration of the tests, in a representative and random sample of classrooms. We define a 'sampled school' as a school where there are one or more 'monitored classrooms', and a 'monitored classrooms' (in a sampled school) as a classroom where an inspector is present during the test. Moreover, a 'non-monitored classroom in a sampled school' is a classroom in a sampled school where the inspector was not present. The natural experiment in SNV surveys administration determines a random variation in the type of classrooms subject to and not-subject to the external monitoring (monitored *versus* non-monitored classrooms) which is exploited to identify social spillovers due to students' cheating behaviors.

3.1. The National Survey of Students'Attainments

Starting from 2009-10 school-year the 'National Institute for the Evaluation of the Education System' (Invalsi, from now onwards), carries out a yearly evaluation of students' attainment and schools quality administering the SNV survey based on questionnaires and test scores evaluations[66]. SNV takes the form of an annual census, since it is compulsory for all schools and students attending the second and fifth grade, in primary schools, and the sixth and eighth grade, in junior high schools (about 500,000 students in each grade) [67]. Each student takes a test in Mathematics and Language in two different days in late May. Test administration and marking is carried out by school teachers, while Invalsi enforces a detailed Protocol (i.e. Invalsi, SNV Report 2010) for the administration and marking of the tests to reduce the possibility of teachers' cheating. For example, as often done in National Evaluation programs (Eurydice, 2009), the test is not administrated by the class teacher but by teachers of other classes and specialized in a different subject with respect to the one that is tested. All school teachers are simultaneously involved in the marking process, so that they cross-check each other during the marking, and the School-head - who is responsible for the correct implementation of the Protocol - supervises the whole process. Finally, an external specialized institution is charged to compute the test scores using an automatic procedure. However, what cannot be excluded *a priori* is that teachers adopt forms of soft monitoring. Teachers might simply adopt some form benevolent supervision because they allow students to exchange information or use prohibited material, or even because they are not able to implement a strict monitoring simply because of classrooms dimensions. Another kind of teachers' benevolent behavior which is not possible to control ex ante concerns the so called 'teaching to the test activity' (Lazear, 2006; Jacob, 2005; Kohn, 2007). For instance, since the

---

[66] Since 2005-06 school year a similar survey was carried out on a representative sample of schools, while all the other schools not in the survey sample were invited to participate on a voluntary basis.

[67] The choice of these grades corresponds to the requirement to test students' abilities at the beginning and at the end of the education path in primary and junior-high school levels. Formally, 8^{th} grade test is part of the final exam at the end of the junior high school and follows different procedures and protocols. Pupils with disabilities are recognized by a team of specialists since the beginning of their schooling path, sit special formats of the tests and their results are not included in the official reports. In any case, it is not possible to change their 'disability' status during the school year.

beginning of the SNV surveys in 2008, it has become a common practice in many schools as teachers want to prepare students to test and quiz like the ones that they have to solve the day of the exam[68].

3.2. The randomized experiment in SNV data

External inspectors are sent to administrate and mark the SNV tests in a representative and random sample of classrooms both to validate the general results of the survey and give each school a 'certified' benchmark. In particular, inspectors are required to perform a number of tasks in the selected classrooms: (i) invigilate students during the tests, (ii) provide specific information on the test administration, (iii) compute the test scores and send results and documentation to Invalsi within a couple of days (Invalsi, 2010).

The allocation of inspectors to a random sample of classroom in the SNV data provides the ideal framework for our empirical strategy, for it introduces a random treatment with respect to the possibility of students to interact exchanging information or cooperating during the test – i.e. 'monitored classrooms' constitute the treated group of students, while 'non-monitored classrooms' are the control group. While, the possibility of any interactions among the students (cheating behavior) in the monitored classrooms is totally excluded and rigorously tested by Invalsi (Invalsi, 2010)69, there is evidence that students in the non-monitored classroom received a more 'benevolent' supervision allowing the possibility of exchange of information and cooperative interactions. The latter is also confirmed by a number of studies which have used Invalsi data to investigate the extent of 'cheating bias' in test scores (Invalsi, 2010; Ferrer-Esteban, 2012; Bertoni et al. 2012; Castellano et al. 2009)[70]. Given that the choice of the monitored classrooms was random and done after classrooms formation, there is no sorting or matching between the treatment and school or classroom characteristics. The only exclusion criterion from the sample is constituted by classrooms with less than 10 students[71]: this feature will require a careful analysis in the empirical estimations (see Section 3.5). On average, monitored students

[68] A confirmation can be easily found looking at how text books have changed with the introduction of the SNV Program and started to include tests and quiz similar to the SNV exams structure.

[69] To test this Invalsi implemented sophisticated statistical techniques based on fuzzy-logic algorithms – i.e. see, Castellano et al. (2009) - and reported no evidence of cheating in the monitored classes.

[70] Bertoni et al. (2012) find that the presence of the external inspector reduces the average score (i.e. in terms of percent of correct answers) in the classroom by 5.5 to 8.5 per cent as compared to classrooms in schools with no external monitoring. They also find evidence of indirect effects on non-monitored classrooms in sampled schools, although the magnitude of the effect in this case is much smaller.

[71] In case in which a class with less than 10 students was selected, it was discarded and replaced with another class.

correspond to 7-8% of the total student population in each grade, while sampled classrooms corresponds to 6-7% of the total number of classrooms in each grade.

3.3. Data and descriptive statistics

In the empirical analysis we use the 2009-10 SNV data for sixth grades[72], for each student SNV data provides the test score for Math and Language and micro-data containing individual level information which are discussed in detail in Appendix B. Test scores are obtained as percentage of right answers for each subject and are standardized with zero mean and unitary standard deviation for the empirical analysis[73]. Individual characteristics cover information on gender, year and place of birth, Italian citizenship, grade retention, kindergarten attendance and school and class (anonymous) identifier. Table 22 sums up the major characteristics of the dataset: number of schools, classes and students by each grade tested, average number of students per school and class while Table 23 shows that the two groups are not different in terms of observable characteristics.

[Table 22 here]

[Table 23 here]

The only systematic difference is found in the presence of immigrant students who are oversampled. This feature suggests particular care when estimating the social multiplier (see Section 3.5). The two groups mainly differ because of the cheating behaviors of non-monitored students. Invalsi excludes the possibility of any interactions among students in monitored classes (SNV 2010 Report, Appendix 10, p.330) and provides statistical evidence of cheating behavior occurring in non-monitored classes by computing an index of ‘cheating’ (i.e. a class-level and subject-specific indicator ranging from 1, cheating is high, to 0, no cheating)[74]. The statistical method implemented by Invalsi highlights a high probability of cheating behaviors in non-monitored classrooms: the average cheating coefficients are .97 for Math and .92 for Language tests (Table 22). On the contrary, Invalsi Report shows that cheating coefficients for monitored classrooms are statistically not different from 0.

[Table 24 here]

Finally, Table 24 provides statistical evidence on the differences in test score results between monitored and non-monitored students. The mean and the median test score of non-monitored students is generally higher compared to monitored students, while the total variance is lower.

[72] We also repeat the analysis using SNV 5th and 2nd grade data in the robustness section.

[73] Students with special education needs take appropriate versions of the tests compatible with their physical or mental disability. Their results are not available due to privacy regulation restrictions.

[74] Invalsi uses these techniques to detect cheating behaviors also in other surveys and official national examinations. For further details about the “fuzzy c-means clustering” technique which is at the base of the indicator, see Castellano et al. (2009), Dunn (1973), Bezdek (1981).

The effect of the inspector's supervision becomes more clear when we decompose the total variance in its within- and between-class components[75]: within-class variance is greater in monitored classes while the between class variance is lower.

4. EMPIRICAL STRATEGY

To identify the endogenous social multiplier effect originating from students' cheating behavior, we implement the Excess-Variance approach developed by Graham (2008). This approach, by relying only on the cross-group variation that originates from endogenous social effects, allows a direct estimation of the (structural) social multiplier - i.e. parameter γ in equations [9] and [10] as derived in the theoretical section. One advantage of this empirical strategy is that it bypasses most of the identification problems that characterize the classical reduced-form linear-in-means model[76]. For example, most studies in the social interactions literature (i.e. Gleaser *et al.* 1996, 2003; Entorf and Lauk, 2006; Entorf and Tatsi, 2009) have not been able to reliably separate the different sources of variability of individual and group level heterogeneity from the 'excess variability' genuinely originating from social interactions (Sacerdote, 2010)[77]. Moreover, the E-V approach has other notable advantages: first, it is robust to individual and group-level heterogeneity; second, the data requirements necessary to overcome the bias originating from standard omitted bias variable – i.e. which is a rather fundamental problem in social interactions setting due to the various sources of correlated effects - are very limited[78].

In practice, we observe N classrooms, each composed of N_c students. For each student we observe y_i, the outcome variable (test score), Z_c and Ψ_c, vectors containing group-level information, while individual-level (ε_i) and classroom-level heterogeneity (μ_c) are unobserved latent variables. Following Galbiati and Zanella (2012), we can rewrite the reduced form model from equations [10] and [9] in variance-components: let the classroom-

[75] The formula is corrected with appropriate weights to take into account the different size of the subgroups (i.e. classes) (see Ammermüller and Pischke, 2009)

[76] For example using proxy for peers' education level (Hanushek et al. 2003, 2009) or having to rely on specific exclusion restrictions (De Giorgi et al. 2011; Bramoullé et al. 2009).

[77] Some recent papers in the social interaction literature refer to the concept of social multiplier as the 'multiplicative effect due to social interactions' and derive the estimation of the multiplier indirectly (e.g. Maurin and Moschion, 2009; for female labour market participation decisions; Drago and Galbiati, 2012, for crime and recidivism).

[78] Durlauf and Tanaka (2008) discuss the advantages of the E-V approach compared to the regression approach and conclude that the former requires stronger assumptions on the variance covariance matrix which are not needed in the classical estimation of peer effects parameters from linear-in-means models. However, the authors suggest that E-V can be better justified whenever the sort of exclusion restriction needed on the variance covariance matrix of the outcomes can be substituted by appropriate prior information on the variance matrix structure. Our implementation of the EVA follows exactly this direction: we implement EVA exploiting the exclusion restriction which directly arises from the natural experiment in Invalsi SNV data.

level heterogeneity be, $\mu_c = \tilde{\delta}_x \bar{x}_c + \delta_\mu \mu + \delta_0$; the individual-level heterogeneity, $\varepsilon_i = \delta_x x_i$; and the classroom-level average of individual heterogeneity, $\varepsilon_c = \delta_x \bar{x}_c$. This transformation yields the following behavioral equations:

$$y_{ic} = \varepsilon_i + (\gamma - 1)\varepsilon_c + \gamma \mu_c \quad [11]$$

$$\bar{y}_c = \gamma \varepsilon_c + \gamma \mu_c \quad [12]$$

The social multiplier parameter to be estimated is, γ (with $\gamma \geq 1$), which captures the equilibrium social effect on individual achievement (i.e. test score) due students' cheating cooperative behavior during the exam. Equation [12] shows that the social multiplier is related to both the average of classroom-level (individual) heterogeneity, ε_c, as well as to the classroom-level heterogeneity, μ_c, such that - as implied by the theoretical model – exogenous shocks to contextual factors can also contribute (feeding-back through individual behaviors) to amplify the effects social externalities[79].

3.4.1. The Excence-Variance approach

Following Galbiati and Zanella (2012), a simplified notation for the conditional variances and covariance of individual and group-level heterogeneity is given hereafter: let $\sigma_\varepsilon^2(Z_c, \Psi_c) = \sigma_\varepsilon^2$ be the conditional variance (i.e. on Z_c and Ψ_c) of individual-level heterogeneity; $\sigma_{\varepsilon\varepsilon}(Z_c, \Psi_c) = \sigma_{\varepsilon\varepsilon}$ the conditional covariance of across individuals heterogeneity; $\sigma_\mu^2(Z_c, \Psi_c) = \sigma_\mu^2$ the conditional variance of group-level heterogeneity; $\sigma_{\mu\varepsilon}(Z_c, \Psi_c) = \sigma_{\mu\varepsilon}$ the conditional covariance of group-level heterogeneity with individual heterogeneity; while $V_c^w(Z_c, \Psi_c) = V_c^w$ and $V_c^b(Z_c, \Psi_c) = V_c^b$ are, respectively, the within-group and the between-groups conditional variance. Notice that: $\sigma_{\varepsilon\varepsilon}$ can be considered a measure of the degree of student sorting across classrooms; while σ_μ^2 represents the variance of unobserved teachers' characteristics, such as experience, strictness, ability and effectiveness, as well as the variance of all other unobserved characteristics that are common to all students

[79] Note that Graham (2008) defines the social multiplier parameter as a combination of both endogenous and exogenous peer effects - as group level heterogeneity there is obtained through group level averages of both observable individual characteristics and unobservable behaviors –, while in our setting it incorporates only the endogenous part of the cheating interactions.

in a classroom; $\sigma_{\mu\varepsilon}$ is a measure of 'matching' between these characteristics and the students. The latter is non-zero any time teachers (or classroom characteristics) and students are not randomly allocated – i.e. student can choose the school or, within each school, the classroom in which enrol. Then assuming that: $\sigma_{\varepsilon\varepsilon}(Z_c,\Psi_c)=\sigma_{\varepsilon\varepsilon}$, $\sigma_{\mu}^2(Z_c,\Psi_c)=\sigma_{\mu}^2$, $\sigma_{\mu\varepsilon}(Z_c,\Psi_c)=\sigma_{\mu\varepsilon}$ are independent of Z_c; and that the portion of the between-group variance independent from the within-group variance can be approximated by a linear function, such as:

$$\gamma^2\left[\sigma_{\mu}^2(\Psi_c)+2\sigma_{\mu\varepsilon}(\Psi_c)+\sigma_{\varepsilon\varepsilon}(\Psi_c)\right]=\pi\Psi_c \qquad [13]$$

Graham (2008) shows that V_c^w and V_c^b can be rewritten as follows:

$$V_c^w=\mathrm{E}\left[\frac{\sigma_{\varepsilon}^2(Z_c,\Psi_c)-\sigma_{\varepsilon\varepsilon}(\Psi_c)}{N_c}\,|\,Z_c,\Psi_c\right] \qquad [14]$$

$$V_c^b=\gamma^2\left[\sigma_{\mu}^2(\Psi_c)+2\sigma_{\mu\varepsilon}(\Psi_c)+\sigma_{\varepsilon\varepsilon}(\Psi_c)+V_c^w\right] \qquad [15]$$

where, substituting expression [13] into [14], it yields:

$$V_c^b=\pi\Psi_c+\gamma^2V_c^w \qquad [16]$$

It is easy to show that the within-group variance of students' achievements in classroom *c* (denoted V_c^w in equation [14] above) is independent of social interactions and classroom-level heterogeneity. Note that, within-classroom differences in individual cheating behavior, when teachers are not sufficiently scrupulous in supervising students during exams such that cheating occurs, cannot be attributed to social externalities – since in our model, by definition, are the same for all students - but only to differences in individual characteristics and the covariances arising from students' sorting. Conversely, the between-group variance (denoted V_c^b in equation [15] above) depends on classroom heterogeneity and, when students' cheating behavior occurs, is magnified by social externalities. In this case, part of the variability in students' achievement between two different classrooms, one in which teachers do not strictly supervise students and another where strict monitoring is efficiently enforced, must necessarily depend on supervision. Then, since students' achievement in a classroom is also driven by cheating interactions, the cross-classroom variation will be affected. Cheating interactions introduce a wedge between the variance of students' achievements (measured by test scores) at different levels of aggregation, which is what we exploit to identify the social multiplier.

Expressing the conditional variances, as in [14] and [15] above, as conditional expectations of the relative within- (G_c^w) and between-classroom (G_c^b) statistics, namely: $V_c^w = \mathrm{E}\left[G_c^w \mid Z_c, \Psi_c\right]$ and $V_c^b = \mathrm{E}\left[G_c^b \mid Z_c, \Psi_c\right]$, we can rewrite equation [16] as[80]:

$$\mathrm{E}\left(G_c^b \mid Z_c, \Psi_c\right) = \pi\Psi_c + \gamma^2\left[\mathrm{E}\left(G_c^w \mid Z_c, \Psi_c\right)\right] \qquad [17]$$

which implies the following conditional and unconditional moment restrictions, respectively:

$$\mathrm{E}\left[G_c^b - \pi\Psi_c - G_g^w \mid Z_c, \Psi_c\right] = 0 \qquad [18]$$

$$\mathrm{E}\left[\begin{pmatrix} Z_c \\ \Psi_c \end{pmatrix}\left(G_c^b - \pi\Psi_c - \gamma^2 G_c^w\right)\right] = 0 \qquad [19]$$

Equation [19] delivers the appropriate specification to estimate (i.e. by GMM) the social multiplier, γ^2, using Z_c as instrumental variable.

3.4.2. The identifying assumption

The randomized experiment in Invalsi SNV data provides the ideal setting for identification. We observe two classrooms with, otherwise identical, students interacting in different ways: in one classroom achievement can also be attained by student cooperative behavior (i.e. control group); in another classroom external monitoring limits students' possibilities to interact, such that achievement is only based on individual effort (i.e. treatment group)[81]. Given the perfect randomization in treatment assignment, both individual and group level heterogeneity are likely to be the same across the two classrooms, such that the only difference in achievement between the two is the one originating from social externalities in students' cheating behavior: which are present only in the control group. Notice, that the

[80] For each class *c* we observe the outcome for a (random) sample of students ($n_c \le N_c$) given by all students who sit both Language and Math test scores. For this reason we rewrite expressions [14] and [15] using the appropriate statistics containing correction terms to take into account the difference between the sample and the population means. See Galbiati and Zanella (2012) web supplement for a formal derivation of conditional expectations.

[81] We may also expect that supervision is more efficient in treated-group classroom simply because of the joint presence of the inspector and a school teacher rather than just one teacher as in the control-group classroom. Note that in this case, the test score incorporates both the 'endowment type' peer effects (i.e. ability) and the 'behavioral peer effects' due to students' cheating interactions (see Section 3.6.3).

presence of an inspector, by virtue of randomization, has no effect on the allocation of students and teachers to classroom, nor any effect on matching and sorting process of students' characteristics. According to our main identifying assumption (i.e. see equations [14] and [15]), Z_c generates an exogenous variation that affects the between-classroom variance in students' achievement only via the effect that cheating interactions have on the within-classroom variance. That is, by comparing the conditional variance of individual behavior within and between classrooms that we can identify the contribution due to endogenous social interactions only. In practice, we define a dummy variable identifying classrooms with external monitoring, ($Z_c=1$), and classrooms without external monitoring ($Z_c=0$)[82]. The standard rank condition for Z_c to be a valid instrument can be easily assessed empirically: $\mathrm{E}\left(G_c^w \mid Z_c = 1, \Psi_c\right) \neq \mathrm{E}\left(G_g^w \mid Z_c = 0, \Psi_c\right)$.

Since the model is just-identified, we can simply estimate it by two-stage least squares and given that the instrument, Z_c, is a dummy variable, the estimator of the social multiplier takes the form of a Wald estimator:

$$\gamma^2 = \frac{\mathrm{E}\left(G_c^b \mid Z_c = 1\right) - \mathrm{E}\left(G_c^b \mid Z_c = 0\right)}{\mathrm{E}\left(G_c^w \mid Z_c = 1\right) - \mathrm{E}\left(G_c^w \mid Z_c = 0\right)} \qquad [20]$$

The numerator is a contrast of observed (or actual) between-classroom variance in student achievement across treatment states (i.e. $Z_c=1$ *versus* $Z_c=0$). As discussed above, under perfect randomization, this contrast is purged of the influence of teacher heterogeneity, matching, and sorting; thus it solely reflects differences in the variance of achievements across the above treatment states as amplified by the cheating interactions. The denominator also equals the difference in the variance of achievements across the treatment states, but unaffected by social interactions (Graham, 2008; Sacerdote, 2010).

Finally, the feasible estimator requires an estimate of the conditional expectation of students' achievement $\mathrm{E}\left(y_{ic} \mid Z_c, \Psi_c\right)$ which we obtain from a regression of y_{ic} on Z_c and Ψ_c. We then use the residuals to replace G_c^b with $\hat{G}_c^b = (\hat{y}_c - Z_c^{'}\hat{\pi}_1 - \Psi_c^{'}\hat{\pi}_2)^2$, where $\hat{\pi}_1$ and $\hat{\pi}_2$ are least squares estimates.

Randomization also implies that (in principle) we do not need to include any variable in the vector Ψ_c to control for sorting or matching of students with respect to assignment to treatment, Z_c, and class characteristics. Descriptive evidence provided in Section 3.3 shows

[82] Graham posits that identification relies on: "[…] two subpopulations of social groups where assignment to groups is as if random" (Graham, 2008, p. 658). In his paper, Graham identifies a social multiplier arising from differences in peer quality across groups, in our setting however peer quality is homogeneous across groups the only source of excess variation being cheating behavior.

that the two subgroups constitute a representative and random sample of the students population for the sixth grade (see also Appendix B). There are, however, a couple of matters for concern: first, we may need to control for the share of immigrant students as they appeared to be slightly oversampled in treated classroom (see Table 23); second, there may also be spill-over effects of external monitoring in treated classroom on non-monitored classrooms of sampled schools, which we need to control for (Bertoni et al. 2012). For these reasons, we include two additional controls: a dummy variable indicating whether a classroom is a 'non-monitored class in a sampled school', and a dummy variable indicating whether there is a 'high share' of immigrant students in the classroom (i.e. takes value 1 if the immigrant share is greater than the 75^{th} or the 90^{th} percentile of the immigrant class share distribution). We discuss further extensions in the following section.

3.5. RESULTS

The estimates of the social multiplier are obtained through two-stages least squares where we regress the feasible estimator for the between-groups variance ($\hat{G}_c^b$) on the additional controls (Ψ_c), and on the within-groups variance, $\hat{G}_c^w$, instrumented by the class type indicator (Z_c). We first report our estimates of equation [20], without including any control variable (i.e. baseline social multiplier), then we progressively add other control variables to the vector Ψ_c to account for selected features of randomization, or test the existence of spill-over effects. Social externalities exist if the social multiplier is different from one (eq. [11] and [12]), thus we test the null that $\gamma^2=1$ and report the correspondent p-value in each table. To allow for the comparability of the results across subjects, we focus on all students who sit both Language and Math test scores. In fact, given that the tests were in two different, although subsequent, days there are students who sit just one of the two tests and students who do not sit none of them because they are absent in both days. The percentage of absent students in 6^{th} grade is about 0.6%. As previously discussed, the only criterion Invalsi used in the randomized experiment to drop, *a priori*, some classrooms from receiving the treatment (i.e. external monitoring) was classroom size – i.e. less than 10 students (723 classes for corresponding to 2.7% of the total number). For this reason we conduct the analysis dropping classes with less than 10 students, while robustness checks to the inclusion of these classes are tested in the next section.

Note, that the E-V approach leaves the sign of the social multiplier (γ^2), in principle, undetermined (since we estimate its square). Hence, the sign has to be inferred from the underlying theoretical model which, in our case, posits a positive effect of social multiplier (γ and $J>0$) due to the assumption of students' cooperative effort, such that cheating interactions among students during the exam are likely to increase each student's achievement and the class average performance. Next, we explore the heterogeneous effect of social interactions comparing sub-populations with a different degree of heterogeneity according to a set of selected (exogenous) characteristics.

3.5.1. Baseline estimates

First stage F-statistics reported in Table 25 show that instruments are not weak and the standard rank condition is always satisfied (the coefficient of the excluded instrument is always positive and statistically different from zero at 1% significance level).

[Table 25 here]

First-stage results, not reported in the tables, indicate that in monitored classrooms the variance of the tests scores is higher compared to non-monitored classrooms. This reflects the larger dispersion of individual heterogeneity in test scores when behavioral interactions are not at work and students cannot exchange information or engage in any cooperative effort.

From our baseline specification we obtain an estimate γ^2 of 5.13 for Math and 4.18 for Language. Progressively adding the control variables described above does not alter the results: estimates for Math range between 5.13 and 5.89, while estimates for Language range between 4.18 and 4.77. This confirms that the two subgroups are (almost) identical in terms of observable characteristics, and that adding control variables (included in the Ψ_c vector) only has a negligible effect on the estimated social multiplier. All estimates are significantly different from 1 at 1% confidence level: this means that we can strongly reject the null of 'no social interactions' (i.e. that $\gamma=1$, Graham, 2008).

Our results imply a strong amplifying role played by social interactions within students in the classroom. The above estimates correspond to values for γ ranging between 2.26 and 2.43 for Math, and between 2.05 and 2.18 for Language, and values for J ranging between 0.56 and 0.59 for Math, and are slightly lower for Language (0.51 - 0.54)[83]. In terms of our structural parameters, a cheating social multiplier close to two (i.e. $\gamma \in [2.05;2.43]$) means that cooperative behaviors, when external monitoring is loose or benevolent, may generate a change in the equilibrium of students' achievements that is twice as big as the class average

[83] Standard errors for the model parameters (γ, J) are obtained using the delta method. The delta method expands a function of a random variable (i.e. the estimated parameters) about its mean with a one-step Taylor approximation. Then, it computes the variance to obtain an estimate of the standard errors (see Davidson and MacKinnon 2004, chap. 5.6).

achievement without behavioral interactions (equation [9]). In terms of individual test score, the estimates for J (i.e. $J \in [.51;.59]$) imply that the marginal contribution due to cheating increases individual test score by almost a half of the standard deviation (equation [10]), which corresponds to almost 10 points in Math and 8 points in Language tests[84].

For what concerns the general pattern of the results with respect to the two subjects, the magnitude of the estimated social multiplier is slightly larger in Math with respect to Language. This small difference can be explained considering that cheating behavior may be easier for mathematics, which are based on closed answers and quiz, rather than in language since text comprehension exercises require more effort and longer time to get through the text, to interpret it and derive the answers. This result is also in line with educational psychology literature which finds that cheating occurs more frequently in the hard sciences compared to the arts and social sciences (Miller *et al.*, 2007).

Our results, although not directly comparable, confirm in general the evidence available from other studies in the social interactions literature (see Carrel et al. 2008; Glaeser et al., 1996; Drago and Galbiati, 2012; Maurin and Moschion, 2009) which find social multipliers between 2 and 3 in order of magnitude. A more direct comparison can be done with those studies that use the E-V approach to recover an estimate of the social multiplier. In his analysis of students' peer effects in class learning activities, using Project STAR data, Graham (2008) reports an estimate for the social multiplier of approximately 1.9 for Math, and 2.29 for Reading. Galbiati and Zanella (2012) estimate a social multiplier arising from congestion externalities in tax cheating between 3.1 and 3.2. In other words, in all the above settings an exogenous shock altering the variable subject to social interactions (respectively, school achievement and concealed income) produces an equilibrium variation that is between two and three times the initial response. Note, however, that when comparing the results reported in Graham (2008) and Galbiati and Zanella (2012) to our own, some important differences should be born in mind. First, while we exploit the identifying restriction given by the natural experiment in Invalsi SNV data, both Graham (2008) and Galbiati and Zanella (2012) identify the social multiplier through exogenous variations in the size of the reference group. As standard in this literature (Sacerdote, 2001; Imberman et al. 2012), Graham's social multiplier due to peer interactions in achievement embeds both exogenous and endogenous effects[85]. Galbiati and Zanella (2012) provide a structural interpretation of the social multiplier generated by externalities in concealed income due to tax congestion within Local Tax

[84] According to the corresponding values of *J*, the cheating marginal contribution for Math ranges between 10.1 – 10.7 points. For Language it is sensibly smaller (between 7.8 – 8.2 points).
[85] Graham (2008) points out that the estimated structural parameter for the social multiplier should be referred to an explicit structural model to highlight the underlying social mechanisms which originate the peer effects.

Authorities so that their social multiplier represents an upper bound of the long run effects of the endogenous effects of tax cheating.

3.5.2. Heterogeneous effects

We exploit the richness of individual-level information in the SNV data to explore different dimensions of students' characteristics which may give rise to heterogeneous effects in cheating behavior. In practice, we test whether the social multiplier differs across selected subpopulations of classrooms characterized by large amounts of heterogeneity in some observed students' attributes, with respect to a subpopulation of classrooms with low heterogeneity (Graham, 2008). Since cheating requires some cooperative effort between students within each classroom, one may expect that classrooms in which students are more homogeneous with respect to some exogenous attributes[86] exhibit stronger social interactions as compared to classrooms in which students are more heterogeneous. This corresponds to test whether there is complementarity or substitutability between the intensity of cheating behaviour, due to looser external monitoring, and the strength of classroom social ties. With complementarity, moving a group of students with more homogeneous characteristics and stronger social ties (i.e. low heterogeneity subpopulations) to a non-monitored classroom should, in addition to increase average test scores, reduce its variance more than for a comparable group of students with less homogeneous characteristics (i.e. high heterogeneity subpopulations). Thus, if external monitoring and classroom heterogeneity are complementary, the social multiplier estimated on the low heterogeneity subpopulations should be greater compared to the one calculated on the high heterogeneity subpopulations. If they are substitutes, the opposite will occur.

In particular, we select the following attributes for the subpopulations: number of books at home, sport practice (outside school), participation to outside school activities (other than sport, e.g. music, arts and foreign languages courses) and time spent playing with friends (outside school)[87]. In all the above cases, we split the sample of classrooms into two groups characterized by high and low degrees of heterogeneity. We refer to the number of books that students have at home as a proxy for heterogeneity of parental background in terms of education and socio-economic status (Ammermüller and Pischke, 2009). In this case, the high (low) heterogeneity group is defined as the subpopulation of classrooms having a standard deviation higher or equal (lower) to the median standard deviation observed in the entire

[86] Note that all the attributes are considered exogenously pre-determined with respect to students' achievement during the exam.
[87] See Appendix B for further details on the definition of the variables.

classroom population. The ‘sport’, the ‘outside school activities’ and the ‘time spent playing with friends’ variables are themselves a proxy of the strength of the social links within each classroom, measured as the amount of time classmates meet and spend time together outside the school. Classrooms in which social ties, proxied by the above variables, are below the median level of the whole population belong to the high heterogeneity group. In other words, classrooms above the median level of these variables encompass situations in which a lot of students interact more outside school (sport, music, arts, playing with friends) thus showing stronger social ties. An opposite reasoning is true for classrooms below the median levels.

[Table 26 here]

Table 26 shows the main results: for each selected attribute we report γ^2 - the square of the social multiplier - for the group of classrooms with high and low heterogeneity, respectively, and test the null of no differences (p-values reported)[88]. We exclude from the analysis students with missing values in any of the four variables used and drop classrooms with less than ten students because of the above discussions[89]. First stage F-statistics show that the effect is always strongly identified. We find that the social multiplier is larger in the subpopulation of classroom with low heterogeneity with respect to parental background characteristics and students’ outside school activities both in Language and Math. For Language, the same result holds also for the sport variable. No statistically significant difference is detected with respect to time spent playing with friends. This suggests that higher strength of social ties and more homogeneous classrooms in terms of family socio-economic background favour social interactions in cheating behavior[90].

In general, we find support for the hypothesis that cooperative efforts in cheating interactions require a more homogeneous pool of classmates and deliver a greater social multiplier. In particular, the results for the sport practice and the outside school activities variables, seem to suggest that practicing sport with classmates outside school and doing other leisure activities such as arts and music courses are to be considered complementary to the social links that are useful to support cheating.

[88] We test the null, H_0: $\gamma^2_H = \gamma^2_L$, using the Sargan-Hansen test of over-identification associated with the estimates of the combined sample where the binary instrument (monitored/non-monitored classrooms) and its interaction with the high heterogeneity dummy serve as excluded instruments (Graham, 2008).

[89] The same pattern of results holds keeping classrooms with less than 10 students. Dropping these classrooms slightly improves p-values for the ‘books at home’ variable.

[90] We also calculate heterogeneous effects with respect to classrooms showing high and low heterogeneity in teachers’ marks given to students at the end of the first semester, in late January. We find that cheating social multiplier is higher the more the class is homogeneous in terms of ‘perceived’ ability level (as proxied by teachers marks). We do not include these results as teachers’ marks cannot be considered plausibly exogenous to students’ test scores results.

3.6. ROBUSTNESS

We test the robustness of the empirical results taking into account different forms of social mechanisms that could affect our estimates of the cheating social multiplier. For example, we investigate whether teachers' cheating in non-monitored classes, or stress induced by the presence of an external inspector in monitored classes may explain (part of) the gap in performance between monitored and non-monitored students, as opposed to students' cheating. Next, we replicate on our data Graham's empirical analysis of achievement peer effects in the Tennessee Schools STAR Project. All the robustness checks support our identification strategy and show that estimated values of the cheating multiplier are not affected by alternative mechanisms that could bias the results. There alternatives are discussed hereafter.

6.1. Teachers' cheating

Several forms of teachers' cheating are discussed in the literature. There could be totally illicit activities, so called 'explicit cheating', such as changing student responses on answer sheets, providing correct answers to students, or obtaining copies of an exam illegitimately prior to the test date and teaching students using knowledge of the precise exam questions. There is 'hidden cheating' in which educators attempt to raise a school overall performance profile by retaining low-scoring students in grade, classifying more students as 'special needs' in order to exclude their scores from school averages, or lavishing attention on students who are close to passing, and ignoring those who are sure to do well and those likely to fail (Kohn, 2007). Additionally, there could also be 'soft' forms of teacher cheating such as 'teaching to the test'. One reason why teachers' cheating should not play a significant role in the Italian schools is due to the fact that the career of teachers follows a simple experience-age rule and is not linked in any way to students' performance. In fact, teachers' cheating has been found to be a substantial problem when high-stakes testing programs are introduced in the school system (Jacob, 2005; Jacob and Levitt, 2003a,b). Moreover, Invalsi controls that the SNV Protocol is strictly followed by school teachers and School-heads are responsible for any illicit behavior of the school staff. However, teachers may be induced in illicit behavior because, for example, they simply dislike sharp differences in results across classes within the same school (Bertoni et al. 2012). Anytime teachers help students in suggesting the right answers or changing their answers while marking the test, the estimates for the social multiplier will also include this component and be upward biased.

Ferrer-Esteban (2012) and Bertoni et al. (2012) analyse the effects of monitoring on students test scores using SNV data and show that external monitoring has a negative effects on students' test scores. Bertoni et al. (2012) use Math tests of elementary school students (5^{th} grade) and argue that the better performance of classes without the external inspector is due to the manipulation of tests by students and/or teachers. The authors do not distinguish between students and teachers' cheating so that they interpret the performance gap between monitored and non-monitored classrooms as a measure of the average intensity of (generalized) cheating taking place in non-monitored classrooms. They also show that spill-overs effects are present in non-monitored classrooms of sampled schools. This fact also justifies the inclusion of the 'non-monitored classroom in sampled school' indicator variable in the vector of controls. Ferrer-Esteban (2012) uses data both from elementary schools (2^{nd} to 5^{th} grades) and junior high schools (6^{th} to 8^{th} grades) in the 2009-10 SNV to build an individual level cheating indicator. Similarly to Jacob and Levitt (2003 a,b), a student is suspected of cheating if the entire path of the answers of the test - item by item, independently of whether answers are right or wrong - is equal to the one of a class-mate. He shows that the distribution of 'suspected cheaters' conditional on the result in the tests is sharply different across grades. In the elementary schools 'suspected cheaters' are all distributed in the upper tail of the test score performance distribution while in the junior high schools 'suspected cheaters' are normally distributed along the test score performance range of results. The author interprets this evidence as teachers' cheating playing a substantive role especially in the elementary schools, as 'suspected cheaters' always give right answers as if they are suggested by teachers and not by each other copying or cheating. Taken together, Bertoni et al. (2012) and Ferrer-Esteban (2012) studies suggest that teachers' cheating - if any - is particularly concentrated in elementary schools and less in the junior high schools.

As robustness check, we replicate the analysis on elementary school students in 5^{th} and 2^{nd} grades who sit the 2009-10 SNV test (see Appendix B for details). Grounding on the aforementioned studies, we expect cheating social multiplier to be higher in magnitude than the 6^{th} grade as it potentially includes bias given by teachers' cheating which is likely to increase class average test scores.

[Table 27 here]

Table 27 shows descriptive evidence on test score means and variances across grades. It is easy to notice that the gap between mean test scores of monitored and non-monitored students is much higher in the elementary grades compared to 6^{th} grade (Language test score gap between sixth grade monitored and non-monitored students is not even statistically different).

The same is true for the total variances, while the variance within classes is always higher in monitored classes.

[Table 28 here]

The estimates of the cheating social multiplier for 5th and 2nd grade students do not show significant differences in terms of strength of identification and statistical significance, but they are always higher in magnitude (Table 28)[91]. This confirms that teachers' monitoring is looser in elementary schools as compared to junior high schools. Restricting our main analysis to 6th grade students, thus minimizes possible bias due to teachers' cheating behavior.

3.6.2. Stress induced by external monitoring

The presence of an external inspector in the classroom during the test (under the external monitoring regime) may exert psychological pressure or induce stress among student, which might alter their performance and lower the average test score in monitored classroom. In this case, the observed gap in test scores between monitored and non-monitored classrooms might incorporate a component that is due to psychological stress. We use the SNV 'Student Questionnaire' (see Appendix B), which contains a set of motivational questions that students have to answer immediately after taking the test, to ascertain the emotional feelings and psychological pressures that students experience while taking the test or preparing for it[92].

[Figure 7 here]

We find no difference in the answers to the motivational questions between monitored and non-monitored students (Figure 7), which leads us to exclude that our estimates might be biased (upward) due to the stress induced by external monitoring. Exploiting the same variables for elementary schools (5th grade), Bertoni et al. (2012) discuss in detail the possibility that young students under-perform as a consequence of the distraction induced by the presence of a stranger in the class and find no evidence that being in a classroom with an external inspector increases anxiety or nervousness.

3.6.3. Achievement peer-effects and class-size

[91] Because of the differences in the test structure, 2nd grade Language results are not directly comparable across grades. School and family background information are not provided for 2nd graders as students do not have to fill in the 'Student Questionnaire'. The estimates obtained without dropping classes with less than 10 students (not included in the text) do not change the overall pattern of the results and confirm their robustness.

[92] Students are asked whether they totally agree / partially agree / partially disagree / totally disagree with the following statements: 'I already was worried before taking the tests'; 'I was so nervous I could not find the answers'; 'While taking the test I was calm'.

Since randomization ensures that students' quality across monitored and non-monitored classrooms is the same, social interactions can only arise from students' cheating behavior. In this section, we test this proposition and investigate whether a more conventional 'peer effects in achievement' may also influence the social multiplier we estimate. Peer effects may work either *via* peers' characteristics (contextual effects such as aptitude to learn, readiness, ability to focus), or *via* alternative endogenous social interactions (such as information gathering, endogenous preference formation, congestion externalities) (Sacerdote, 2001). We replicate the empirical strategy proposed by Graham (2008), which relies on classroom-size variation as instrument, estimating our model separately on for monitored and non-monitored classrooms (Lazear, 2001; Graham, 2008; Carrel et al. 2009; Cooley, 2010a,b)[93]. The key assumption, in this case, is absence of sorting and unobserved heterogeneity across small and large classrooms. Since, general rules for class size formation in junior high schools are considerably influenced at the school-district level by the availably of tenured versus non-tenured teachers and the allocation of resources across schools in the same district, we include in our baseline specification the usual classrooms level controls (Ψ_c), as well as school-district fixed effects (i.e. 110 dummies corresponding to Italian provinces, NUTS 5 level). Specifically, we run the analysis separately for monitored and non-monitored classrooms and calculate an achievement (squared) social multiplier, that we label γ_a^2 to keep it distinguished from the usual cheating social multiplier[94]. We expect the value of the social multiplier estimated for the group of non-monitored classrooms ($\gamma_a^2|Z_c=0$) to be larger than the social multiplier estimated for the group of the monitored classrooms ($\gamma_a^2|Z_c=1$), since the former is likely to be inflated by cheating interactions while the latter is not.

[Table 29 here]

The instrument we use is a dummy for 'small class size' that takes value 1 if class size is below the median class size. Table 29 contains the different estimates for the (squared) social multiplier. The standard rank condition is satisfied, as the coefficient of the excluded instrument – not reported in the Table - is always positive and significantly different from zero at 1 per cent confidence level, and the first-stage F-statistics show that the effect is always strongly identified. The positive sign in the first stage regressions confirms that small class size tends to increase individual-level heterogeneity. Interestingly, estimates of the (squared) social multiplier are found to be not statistically different from 1 in the subgroup of monitored classes - where only interactions in achievement may have taken place -, while the

[93] Group-size is a good instrument for the E-V approach because, provided that group-level heterogeneity is the same across the two subpopulations (small vs. large classes), the dispersion of individual heterogeneity typically is not the same.

[94] We run the analysis on the whole population as well as excluding classes with less than 10 students. Results do not change.

estimates show up statistically different from 1, ranging between 2.08 and 3.21 (close to Graham's estimations), in the subgroup of non-monitored classes[95]. In other words, since we cannot reject the null of 'no-achievement social interactions' in the monitored classrooms (both for Language and Math), while we find sizable social interactions in the non-monitored classes, it seems reasonable to expect any effect of 'achievement social interactions' to be negligible as compared to the effect of 'cheating social interactions'.

[Table 30 here]

A final concern with respect to class-size might arise with respect to classrooms with less than 10 students which were dropped from the main analysis to meet the only ex-ante selectivity criteria implemented by Invalsi in the random selection of the monitored classrooms. To assess whether this threshold introduced some selectivity in the sample of treated versus control classrooms, we repeated the analysis also including all the classrooms with less of 10 students (723 classes for grade 6 corresponding to 2.7% of the total number)[96]. Results reported in Table 30 show no significant differences with respect to the baseline estimates.

3.7. DISCUSSION AND CONLUSION

There is abundant evidence showing that students' cheating has worsened over the last few decades, becoming a widespread practice in schools, college and high-ranked universities (Dee and Jacob, 2012). Experts say that cheating has grown hand in hand with high-stakes testing systems, such as the No-Child-Left-Behind-Act (2001) in the U.S. (Jacob, 2005), and it has become easier and more widely tolerated, as both schools and parents fail to give students clear messages about what is allowed and what is prohibited (*The New York Times*, September 7, 2012). In this paper we provide evidence on the social interactions which are generated when students' cheat - either exchanging information and cooperating with other students, or using any prohibited materials - while taking an exam. We develop a simple theoretical model describing the mechanisms that drive social interactions in cheating behavior, and show that students optimally decide whether or not to cheat taking into account other students' best response. We estimate the social multiplier generated by cheating behaviors using data from a randomized experiment in a national evaluation tests. Our findings suggest a strong amplifying role played by cheating social interactions in the classroom, which increases in the strength of social ties. The value of the social multiplier

[95] Graham (2008) finds a (squared) social multiplier of 2.33 for Math and 2.11 for the Reading test scores in the complete specification. However, while we exploit junior high school students, Graham (2008) focuses on kindergarten students.

[96] Due to students absence on the day of the test, we do find classrooms with less than 10 students also in the treated group. This, of course, was not known ex-ante, and absent students re-sit in September.

implied by students' cheating behaviors is estimated to be between 2 and 3 in all the specifications, suggesting that cooperative behaviors, when a strict external monitoring is missing, may generate a change in the equilibrium of students' achievements that is twice as big as the class average achievement. In terms of individual test score, the marginal contribution of cheating interactions increases individual test score by almost half of the standard deviation (i.e. between 7 and 10 points). Heterogeneous effects show that the strength of social ties in the classroom is a complementary input to cheating behaviors such that the effect is larger the more the classroom is homogeneous. Several sensitivity checks confirm the overall robustness of our results.

Our findings have a number of relevant policy implications. First, we show that tolerating cheating behavior, as it is often done in schools, is a very dangerous practice, since the social multiplier magnifies the negative effects on both students' performance and on the signaling role of education in the labor market. McCabe (2005) documents that a large share of college students considers cheating and other forms of illicit collaboration with classmates as a minor offence or no offence at all. He also finds that most high school teachers and college professors fail to report and pursue most of the violations that are detected. Moreover, commitment to academic integrity and sanctions to violations are still not adequately considered: few schools place any meaningful emphasis on academic integrity, and colleges are even more indifferent than high schools[97]. Our estimates also show that tolerating such behaviors is particularly relevant as cheating is likely to feedback onto social norms thus raising collective indulgence with respect to various forms of dishonest practices. In other words, ethical or honor codes of behavior in schools should be strictly enforced and students' cheating behavior reported and sanctioned. Second, given that increasing competition in the job market and high-stakes testing systems are likely to exert considerable pressure on students to perform well in exams, it should be recognized that where (and when) the pressure is higher, more resources should be devoted to monitoring activities in order to avoid cheating interactions to become widespread. In this sense, the social multiplier mechanism would also magnify the effects of policies directed to stricter monitoring and sanctioning of cheaters. From the policymaker perspective a commitment to rigorous monitoring and sanctioning - by changing the individual's private incentives to cheat, would deliver significantly larger social effects (Durlauf and Cohen-Cole, 2004). Our results also show that strong social links among classmates are likely to facilitate social interactions and cheating behaviors. In this context, a rather inexpensive way to reduce students' illicit behaviors would consist in a random reshuffling of students and teachers across classrooms, within any given school, so to reduce

[97] Michael Josephson, president of the Institute for Academic Integrity, *The New York Times*, September 7, 2012.

students' tendency to conform to other students' behavior. Finally, the presence of spill-over effects of monitoring in non-monitored classrooms of sampled schools, suggests that another rather inexpensive intervention to contrast cheating would be to spread the inspectors on more schools as non-monitored classrooms in sampled schools show a significantly lower degree of cheating interactions.

APPENDIX A. DERIVATION OF THE ACHIEVEMENT BEST RESPONSE

Under the assumption that achievement is monotonically increasing in cooperative cheating effort behavior (henceforth simply referred to as *effort*), we can solve from the APF (eq. [1]) for the unobservable effort. Thus, for individual *i* and *j* - where *j* represents any *i*'s classmate peer - we have:

$$e_i = y_i - x_i\pi_x - x_j\tilde{\pi}_x - e_j\tilde{\pi}_e - \mu \qquad \text{[A.1]}$$

$$e_j = y_j - x_j\pi_x - x_i\tilde{\pi}_x - e_i\tilde{\pi}_e - \mu \qquad \text{[A.2]}$$

Plugging the expression for e_j from equation [A.2] into [A.1] and solving for e_i, we obtain:

$$e_i = \left(\frac{1}{1-\tilde{\pi}_e}\right)\left[y_i - y_j\tilde{\pi}_e - x_i(\pi_x - \tilde{\pi}_x\tilde{\pi}_e) - x_j(\tilde{\pi}_x - \pi_x\tilde{\pi}_e) + \mu(\tilde{\pi}_e - 1)\right] \qquad \text{[A.3]}$$

Similarly, for individual *j*:

$$e_j = \left(\frac{1}{1-\tilde{\pi}_e}\right)\left[y_j - y_i\tilde{\pi}_e - x_j(\pi_x - \tilde{\pi}_x\tilde{\pi}_e) - x_i(\tilde{\pi}_x - \pi_x\tilde{\pi}_e) + \mu(\tilde{\pi}_e - 1)\right] \qquad \text{[A.4]}$$

Substituting equation [A.4] into the effort best response of individual *i* from equation [4] yields:

$$e_i^{BR} = \frac{\beta_y}{\beta_e} + \frac{\tilde{\beta}_e}{\beta_e}\left(\frac{1}{1-\tilde{\pi}_e}\right)\left[y_j - y_i\tilde{\pi}_e - x_j(\pi_x - \tilde{\pi}_x\tilde{\pi}_e) - x_i(\tilde{\pi}_x - \pi_x\tilde{\pi}_e) + \mu(\tilde{\pi}_e - 1)\right] \qquad \text{[A.5]}$$

Finally, substituting the LHS of equation [A.3] with the equation of the best response effort function from equation [A.5] and rearranging we obtain the expression of the achievement best response function (Cooley, 2010b):

$$\begin{aligned}&\frac{\beta_y}{\beta_e} + \frac{\tilde{\beta}_e}{\beta_e}\left(\frac{1}{1-\tilde{\pi}_e}\right)\left[y_j - y_i\tilde{\pi}_e - x_j(\pi_x - \tilde{\pi}_x\tilde{\pi}_e) - x_i(\tilde{\pi}_x - \pi_x\tilde{\pi}_e) + \mu(\tilde{\pi}_e - 1)\right] = \\ &= \left(\frac{1}{1-\tilde{\pi}_e}\right)\left[y_i - y_j\tilde{\pi}_e - x_i(\pi_x - \tilde{\pi}_x\tilde{\pi}_e) - x_j(\tilde{\pi}_x - \pi_x\tilde{\pi}_e) + \mu(\tilde{\pi}_e - 1)\right]\end{aligned} \qquad \text{[A.6]}$$

$$
\begin{aligned}
y_i &= \frac{\beta_y(1-\tilde{\pi}_e^2)}{\beta_e+\tilde{\beta}_e\tilde{\pi}_e} + x_i\left[\pi_x - \frac{\tilde{\beta}_e+\beta_e\tilde{\pi}_e}{\beta_e+\tilde{\beta}_e\tilde{\pi}_e}\tilde{\pi}_e\right] + x_j\left[\tilde{\pi}_x - \frac{\tilde{\beta}_e+\beta_e\tilde{\pi}_e}{\beta_e+\tilde{\beta}_e\tilde{\pi}_e}\pi_e\right] + \\
&\quad + y_j\left(\frac{\tilde{\beta}_e+\beta_e\tilde{\pi}_e}{\beta_e+\tilde{\beta}_e\tilde{\pi}_e}\right) + \mu\left(1-\frac{\tilde{\beta}_e+\beta_e\tilde{\pi}_e}{\beta_e+\tilde{\beta}_e\tilde{\pi}_e}\right); \\
&= \frac{\beta_y(1-\tilde{\pi}_e^2)}{\beta_e+\tilde{\beta}_e\tilde{\pi}_e} + x_i\left[\pi_x - J\tilde{\pi}_e\right] + x_j\left[\tilde{\pi}_x - J\pi_e\right] + y_j J + \mu(1-J); \\
&= \delta_0 + x_i\delta_x + x_j\tilde{\delta}_x + y_j J + \mu\delta_\mu
\end{aligned} \tag{A.7}
$$

APPENDIX B. INVALSI SNV DATA DESCRIPTION

School system in Italy starts with five years of primary school (grades 1 to 5, corresponding to ISCED level 1) and three years of junior high school (grades 6 to 8, ISCED level 2). These two form the ‘first cycle’ of the educational system which is compulsory and identical for all students, while secondary education lasts three to five years depending on the path chosen (vocational, technical, academic). Children enrol in the first grade of the primary school the year they turn six, and start the junior high school when they turn eleven. Primary and junior high schools are quite different in terms of organization and types of teaching activities. In primary schools pupils spend almost all school time with two teachers, one teaches Language, History, Geography and the other teaches Math and Science. The two ‘reference teachers’ usually follow the pupils from the first to the fifth grade establishing a strong personal link. Junior high school is more similar to high school. Students experience a kind of more rigorous teaching, with several professors, one for each subject, and acquire a wide range of core skills necessary to succeed in high schools.

Invalsi SNV data contain test scores results and individual level information. Individual level information are gathered in the dataset from three different sources: (i) students’ general information from school administrative records compiled directly from school administrative staff on each student’s answer sheet; (ii) family background information collected through a ‘Family Questionnaire’ sent to each family some days before the test; (iii) additional individual information on family, school and environmental characteristics collected through a ‘Student Questionnaire’ taken by each 5th and 6th grade students the same day of one of the test (after finishing the exam). They are collected by the school administrative staff on the same answer sheets of the students’ test and are taken from the administrative register data which are given by the families at the moment of the child’s enrolment (at the beginning of each school year in September). Other parental background information are available and cover mother's and father's place of birth (Italy, EU, European but non-EU, other non-European country), occupation and education level.

Some variables (e.g. kindergarten and pre-kindergarten attendance; parental occupation and education) which are not administrative data records kept by the school but rather provided by the families filling in the 'Family Questionnaire' suffer from a relevant problem of missing information (from 9% to 30% depending on the grade and variable). This problem may be considerably mitigated for 5th and 6th grade students exploiting additional information about school characteristics and family background contained the 'Student Questionnaire' which is filled by each student in the class the first day of the test and does not entail problems of missing data. Second grade students did not have to fill such additional information. The 'Student Questionnaire' is different for 6 and 5 graders, but the more relevant variables are common to both. From these sources we obtain variables that are commonly used as proxy for socio-economic background and family information in international programs of students' attainments testing (e.g. PISA, TIMSS) and applied research. For instance, students have to answer questions such as "How many books have you at home?", "Which language do you usually speak at home?"; "Do you currently speak dialect at home?".

In the heterogeneous effects analysis (Section 3.5.1) we exploit some variables taken from the 'Student Questionnaire'. The 'number of books at home' is a categorical variable with 5 levels (0-10 books; 11-25; 26-100; 101-200; more than 200). The 'sport' variable asks students how many times per week he/she practices sport activities outside school (never, 1 or 2, 3 or 4, more than 4). Similarly, the 'outside school activities' variable asks students how many times per week he/she takes part to leisure activities outside the school time (e.g. music, arts, theatre or foreign language courses)(never, 1 or 2, 3 or 4, more than 4). Finally, the variable indicating the time spent each day playing with friends outside school takes the following values: never, 1 hour or less, 1 or 2 hours, more than 2 hours.

B.1 The sampling procedures: randomness and representation

Invalsi exploits a simple random computer routine that ensures the representation of the sampled group of students, classes and schools. First, for each of the 20 Italian regions they randomly choose a representative sample of primary schools for grade 2 and 5, and a sample of junior high schools for grade 6. Then, within each school they randomly picked up one or two classrooms for each grade. The sampling procedure starts at the regional level, so that the final sampled group is representative of the whole student population at the national and regional level. However, also the province dimension (NUT5) was implicitly taken into account so that the final sample can be considered also representative at the province level (Invalsi, 2010). The number of units to be sampled within each region to ensure the sample representation was calculated on the basis of past SNV surveys using the Neyman procedure which is able to generate a sample size in such a way that both the dimension and the variability of the phenomenon under study are correctly mirrored in the sampled units (Invalsi, 2010). Sampled schools could not refuse to receive the inspectors and were informed just a couple of weeks before the test was taken.

To test the effective goodness of these subsamples, we repeat the same analysis as in Table 23 using two other subsamples which are defined according to whether a school is a monitored school or not. Thus, the group of students in monitored schools contains the subgroup of the treated, but is larger because it also contains students in non-monitored classrooms of a sampled school.

[Table B.1 here]

Results are shown in Table B.1. Although now the group of the students in monitored schools is much larger (more than 20% of the population) the t-test for the comparison of the means are statistically significant for almost all the observable characteristics we observe in our dataset. We take this piece of evidence as a further confirmation about the goodness of representation of the two subsamples given by ‘monitored’ and ‘non-monitored’ classrooms. We can conclude that the treatment randomly splits the students population of each grade into two equally representative subgroups. Finally, notice that the same analysis performed on elementary schools data used in the robustness checks (2^{nd} and 5^{th} grade students) confirm the same results.

TABLES CHAPTER 3

Table 22. Descriptive statistics.

% Sampled Schools	22.48
% Monitored classes	7.78
% Monitored students (*)	8.01
% Non-monitored class in sampled school	13.07
% Absent students	0.71
Average school size	131.75
Average class size	20.58
Average cheating coefficient in non-monitored classrooms: Math	0.97
Average cheating coefficient in non-monitored classrooms: Language	0.91
Total no. schools	5,824
Total no. classrooms	26,707
Total no. students	522,655

Notes. (*) the percentage of 'monitored students' is calculated over the total number of students excluding absents. A student is considered 'absent' if he/she does not sit either Math or Language test, or both. Average class and school size refer to the average number of students in the class or school; the total no. of classrooms includes 25 classrooms with missing values in test scores results which are excluded from the empirical analysis. **Source**: SNV Invalsi 2009-10, 6th grade.

Table 23. Mean comparison of observable individual characteristics.

	Monitored students	Non-monitored students	Δ	Missing (% over total)
Female	48.3	48.34	0.04	1.3
Retained	7.31	7.09	-0.22	1.47
Immigrant	10.26	9.94	-0.32**	1.68
First gen. immigrants	6.59	6.54	-0.05	3.47
Second gen. immigrants	4	3.68	-0.32**	6.32
Kindergarten attendance	96.83	96.82	-0.01	22.6
Speak dialect at home	16.93	17.08	0.15	5.13
N (% over total)	41,550 (8.01)	477,395 (91.99)		

Notes. Absent students are excluded: a student is considered 'absent' if he/she does not sit either Math or Language test, or both. Δ indicates the difference between mean characteristics in the two groups; asterisks indicate whether the difference is statistically significant at 0.01 (***), 0.05 (**), 0.1 (*) confidence levels. **Source**: SNV Invalsi 2009-10, 6th grade.

Table 24. Descriptive statistics: test scores mean, median and variance decomposition.

	Language			Math		
	All Pop.	Monitored	Non-monitored	All Pop.	Monitored	Non-monitored
Mean	61.44	61.39	61.45	51.95	51.42	51.99
Median	63.79	63.80	63.79	50.00	50.00	50.00
Total Var.	232.21	235.75	231.91	329.03	329.78	328.94
Var. Between Classrooms	48.79	42.88	49.30	76.80	69.44	77.41
Var. Within Classrooms	183.43	192.86	182.60	252.24	260.34	251.53

Notes. The formula is corrected with appropriate weights to take into account the different size of the subgroups (i.e. classrooms) (Ammermüller and Pischke, 2009, p.323)**. Source:** SNV Invalsi 2009-10, 6th grade.

Table 25. Baseline estimates of the social multiplier.

	MATH					
γ^2	5.135	5.136	5.889	5.390	5.926	5.401
	(0.211)	(0.211)	(0.301)	(0.244)	(0.291)	(0.240)
P-value (H_0: γ^2=1)	0.00	0.00	0.00	0.00	0.00	0.00
Model Parameters						
γ	2.266	2.266	2.427	2.322	2.434	2.324
	(0.047)	(0.047)	(0.062)	(0.053)	(0.060)	(0.052)
J	0.559	0.559	0.588	0.569	0.589	0.570
	(0.009)	(0.009)	(0.011)	(0.010)	(0.010)	(0.010)
First Stage F-Statistic	10772.44	10772.02	3231.63	6161.20	3820.05	6838.82
	LANGUAGE					
γ^2	4.189	4.182	4.713	4.370	4.774	4.383
	(0.169)	(0.169)	(0.241)	(0.198)	(0.234)	(0.195)
P-value (H_0: γ^2=1)	0.00	0.00	0.00	0.00	0.00	0.00
Model Parameters						
γ	2.047	2.045	2.171	2.090	2.185	2.094
	(0.041)	(0.041)	(0.056)	(0.047)	(0.053)	(0.047)
J	0.511	0.511	0.539	0.522	0.542	0.522
	(0.010)	(0.010)	(0.012)	(0.011)	(0.011)	(0.011)
First Stage F-Stat	8290.73	8290.42	2868.01	5172.91	3315.70	5641.34
No. Classrooms	25959	25959	25959	25959	25959	25959
Additional controls (Ψ_c)						
Non-monitored class in sampled school		yes			yes	yes
High immigrant share (>P75)			yes		yes	
High immigrant share (>P90)				yes		yes

Notes. Classes with less than 10 students are dropped from the sample. Additional controls (Ψ_c) include the dummy for non-monitored classrooms in sampled school and the dummy for high share of immigrants, respectively, for immigrant class shares greater than P75 or P90. **Source:** SNV Invalsi 2009-10, 6^{th} grade.

Table 26. Heterogeneous effects in the cheating social multiplier.

PANEL A	MATH			
	High heterogeneity	Low heterogeneity	High heterogeneity	Low heterogeneity
	Books at home		*Outside-school activities*	
γ^2	4.9199	5.8522	4.7388	6.1082
	(0.3203)	(0.3781)	(0.2932)	(0.4092)
First Stage F-Statistic	2876.05	3635.15	3270.83	3187.27
P-value H0: $\gamma^2_H=\gamma^2_L$	0.06		0.01	
No. Classrooms	12031	12107	12040	12098
	Play with friends		*Sport practice*	
γ^2	5.4594	5.2770	5.5002	5.0890
	(0.4135)	(0.3044)	(0.3907)	(0.3151)
First Stage F-Statistic	2633.23	3811.77	2517.54	4066.95
P-value H0: γ $\gamma^2_H=\gamma^2_L$	0.72		0.41	
No. Classrooms	11897	12241	11903	12235
PANEL B	**LANGUAGE**			
	High heterogeneity	Low heterogeneity	High heterogeneity	Low heterogeneity
	Books at home		*Outside-school activities*	
γ^2	4.0339	4.7029	3.8383	4.9494
	(0.2606)	(0.2817)	(0.2410)	(0.3025)
First Stage F-Statistic	2679.83	2875.71	2881.15	2682.26
P-value H0: $\gamma^2_H=\gamma^2_L$	0.08		0.00	
No. Classrooms	12031	12107	12040	12098
	Play with friends		*Sport practice*	
γ^2	4.4492	4.2456	3.8128	4.6742
	(0.3320)	(0.2303)	(0.2577)	(0.2737)
First Stage F-Statistic	2133.98	3435.55	2244.61	3547.14
P-value H0: $\gamma^2_H=\gamma^2_L$	0.61		0.02	
No. Classrooms	11897	12241	11903	12235
Additional controls (Ψ_c)	yes	yes	yes	yes

Notes. Classes with less than 10 students and with missing values in the relevant variables are dropped from the sample. Additional controls (Ψ_c) include the dummy for non-monitored classrooms in sampled school and the dummy for high share of immigrants (immigrant class share greater than P90). **Source:** SNV Invalsi 2009-10, 6th grade.

Table 27. Robustness checks. Descriptive statistics elementary school students.

	LANGUAGE			MATH		
	All Pop.	Monitored	Non monitored	All Pop.	Monitored	Non monitored
5th grade						
Mean	70.23	67.54	70.44	64.76	61.89	65.38
Median	73.91	71.01	73.91	65.91	61.36	65.91
Var. Tot.	144.27	146.62	143.81	65.39	62.5	65.44
Var. Between Classes	45.90	35.03	46.46	27.05	19.7	27.45
Var. Within Classes	98.38	111.59	97.35	38.34	42.8	38
N (students)	475,343	34,554	440,789	475,343	34,554	440,789
2nd grade						
Mean	65.94	62.05	66.24	62.52	57.17	62.94
Median	69.23	65.38	69.23	60.71	57.14	64.28
Var. Tot.	34.85	35.5	34.71	30.81	27.2	30.89
Var. Between Classes	10.39	7.51	10.52	14.52	8.37	14.8
Var. Within Classes	24.46	27.99	24.18	16.29	18.83	16.09
N (students)	466,536	34,201	432,335	466,536	34,201	432,335

Source: SNV Invalsi 2009-10, 5^{th} and 2^{nd} grade.

Table 28. Robustness checks. Social multiplier estimates for elementary school students.

Panel A: 5th grade						
	MATH					
γ^2	7.482	7.471	7.812	7.562	8.027	7.590
	(0.365)	(0.364)	(0.513)	(0.408)	(0.484)	(0.399)
P-value H_0: γ^2=1	0.00	0.00	0.00	0.00	0.00	0.00
First Stage F-Statistic	7113.20	7112.93	2508.88	4840.38	3158.73	5393.57
	LANGUAGE					
γ^2	5.245	5.227	5.402	5.267	5.524	5.280
	(0.323)	(0.326)	(0.457)	(0.368)	(0.438)	(0.361)
P-value H_0: γ^2=1	0.00	0.00	0.00	0.00	0.00	0.00
First Stage F-Statistic	6504.80	6504.56	2429.27	4440.85	3004.18	4911.43
No. Classrooms	26942	26942	26942	26942	26942	26942
Panel B: 2nd grade						
	MATH					
γ^2	7.419	7.364	6.816	7.228	7.112	7.297
	(0.379)	(0.375)	(0.502)	(0.418)	(0.478)	(0.411)
P-value H_0: γ^2=1	0.00	0.00	0.00	0.00	0.00	0.00
First Stage F-Statistic	7867.09	7866.80	2987.77	5115.35	3557.85	5648.61
	LANGUAGE					
γ^2	4.437	4.385	4.246	4.273	4.348	4.296
	(0.201)	(0.198)	(0.269)	(0.219)	(0.257)	(0.215)
P-value H_0: γ^2=1	0.00	0.00	0.00	0.00	0.00	0.00
First Stage F-Statistic	8809.94	8809.61	3236.33	5608.41	3865.24	6218.09
No. Classrooms	26850	26850	26850	26850	26850	26850
Additional controls (Ψ_c)						
Non-monitored class in sampled school		yes			yes	yes
High immigrant share (>P75)			yes		yes	
High immigrant share (>P90)				yes		yes

Notes. Classes with less than 10 students are dropped from the sample. Additional controls (Ψ_c) include the dummy for non-monitored classrooms in sampled school and the dummy for high share of immigrants, respectively, for immigrant class shares greater than P75 or P90. **Source:** SNV Invalsi 2009-10, 5th and 2nd grade.

Table 29. Robustness checks. Achievement peer effects using an alternative instrument (class size).

PANEL A	Non-monitored classrooms			
	MATH		LANGUAGE	
γ_a^2	5.9350	3.2160	5.0901	2.0785
	(0.1059)	(0.4398)	(0.0976)	(0.3375)
P-value H_0: $\gamma_a^2=1$	0.00	0.00	0.00	0.00
First Stage F-Statistic	49602.35	3333.62	48037.98	3914.55
No. Classrooms	23901	23901	23901	23901
PANEL B	**Monitored classrooms**			
	MATH		LANGUAGE	
γ_a^2	5.3363	1.7149	4.2599	1.5793
	(0.2956)	(1.2544)	(0.2333)	(0.6559)
P-value H_0: $\gamma_a^2=1$	0.01	0.57	0.00	0.38
First Stage F-Statistic	5457.98	257.31	4767.00	309.41
No. Classrooms	2058	2058	2058	2058
Additional controls				
Class level variables (Ψ_c)	yes	yes	yes	yes
School-district fixed effects		yes		yes

Notes. Class level variables (Ψ_c) include the dummy for non-monitored classrooms in sampled school and the dummy for high share of immigrants (immigrant class share greater than P90). School-districts fixed effects correspond to 110 dummies. Classes with less than 10 students are dropped from the sample. **Source:** SNV Invalsi 2009-10, 6th grade.

Table 30. Robustness checks. Baseline results including classrooms with less than 10 students.

	MATH					
γ^2	5.172	5.173	5.965	5.447	5.994	5.454
	(0.210)	(0.210)	(0.305)	(0.245)	(0.293)	(0.241)
P-value (H$_0$: γ^2=1)	0.00	0.00	0.00	0.00	0.00	0.00
Model Parameters						
γ	2.274	2.274	2.442	2.334	2.448	2.335
	(0.046)	(0.046)	(0.062)	(0.053)	(0.060)	(0.052)
J	0.560	0.560	0.591	0.572	0.592	0.572
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.009)
First Stage F-Statistic	8267.11	8266.80	2538.80	4739.77	2960.45	5200.40
	LANGUAGE					
γ^2	4.272	4.265	4.819	4.456	4.879	4.468
	(0.175)	(0.176)	(0.255)	(0.208)	(0.247)	(0.205)
P-value (H$_0$: γ^2=1)	0.00	0.00	0.00	0.00	0.00	0.00
Model Parameters						
γ	2.067	2.065	2.195	2.111	2.209	2.114
	(0.042)	(0.043)	(0.058)	(0.049)	(0.056)	(0.049)
J	0.516	0.516	0.544	0.526	0.547	0.527
	(0.010)	(0.010)	(0.012)	(0.011)	(0.011)	(0.011)
First Stage F-Stat	7353.47	7353.20	2519.57	4558.58	2901.45	4956.61
No. Classrooms	26682	26682	26682	26628	26628	26628
Additional controls (Ψ_c)						
Non-monitored class in sampled school		yes			yes	yes
High immigrant share (>P75)			yes		yes	
High immigrant share (>P90)				yes		yes

Notes. Additional controls (Ψ_c) include the dummy for non-monitored classrooms in sampled school and the dummy for high share of immigrants, respectively, for immigrant class shares greater than P75 or P90. **Source:** SNV Invalsi 2009-10, 6th grade.

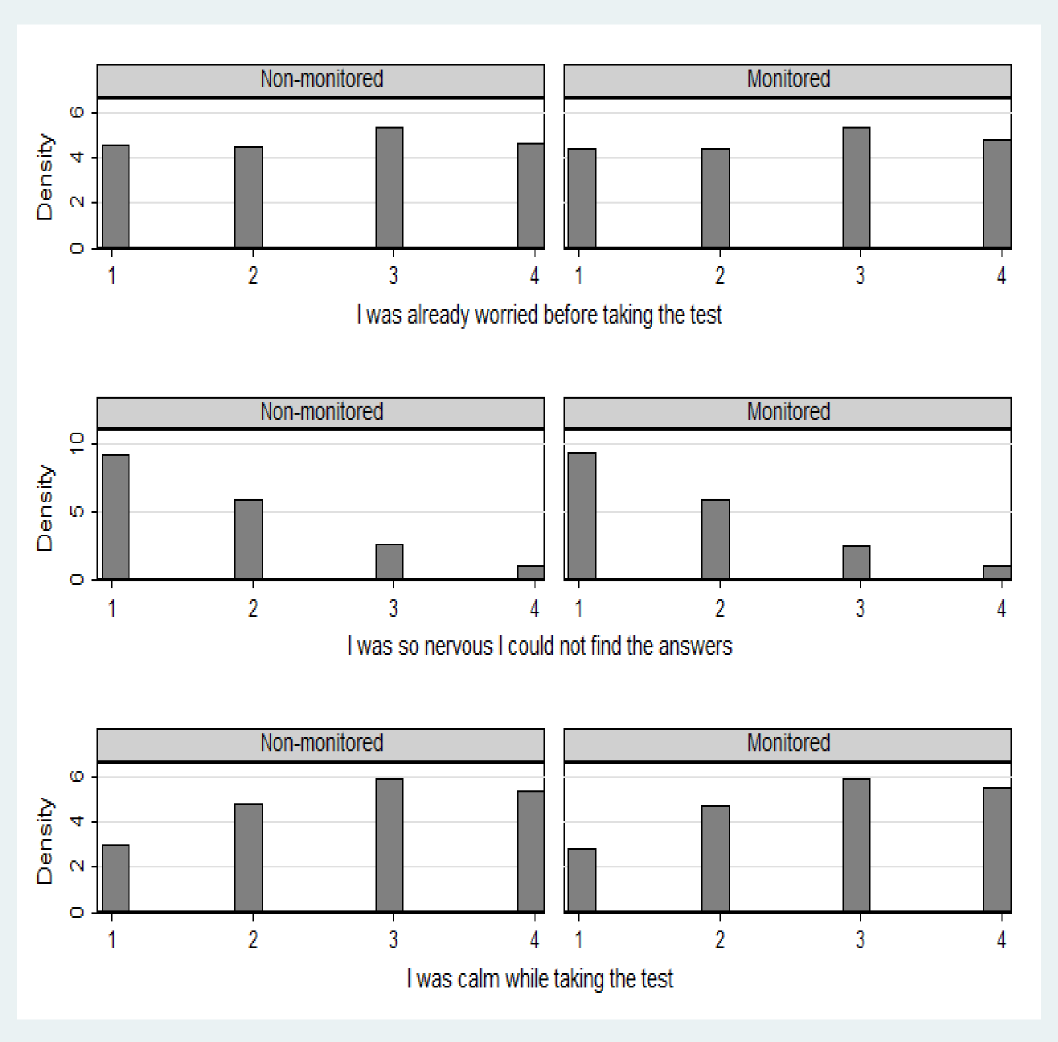
Table B.1. Mean comparison between students in sampled and non-sampled schools.

	Students in sampled schools	Students in non-sampled schools	Δ
Female	48.35	48.34	-0.01
Retained	7.27	7.07	-0.2**
Immigrant	10.36	9.85	-0.51***
First gen. immigrants	6.89	6.46	-0.43***
Second gen immigrants	3.83	3.67	-0.16***
Kindergarten attendance	96.61	96.87	0.26***
Speak dialect at home	16.06	17.35	1.29***
N (% over total)	111,497 (21.48)	407,448 (78.52)	

Notes. Absent students are excluded: a student is considered 'absent' if he/she does not sit either Math or Language test, or both. Δ indicates the difference between mean characteristics in the two groups; asterisks indicate whether the difference is statistically significant at 0.01 (***), 0.05 (**), 0.1 (*) confidence levels. **Source**: SNV Invalsi 2009-10, 6th grade.

Figures Chapter 3

Figure 7. Students stress while taking the test. Comparison between monitored and non-monitored students' answers to motivational questions (1=totally disagree; 2=partially disagree; 3=partially agree; 4=totally agree).



Notes. Students are asked whether they totally agree/partially agree/partially disagree/totally disagree with the following statements: 'I already was worried before taking the tests' (top histogram); 'I was so nervous I could not find the answers' (central histogram); 'While taking the test I was calm' (bottom histogram).

REFERENCES

Akerlof, G. A., Kranton, R. E. (2002), "*Identity and Schooling: Some Lessons for the Economics of Education*", Journal of Economic Literature, 40 (4), 1167-1201.

Algan, Y., Cahuc, P., Shleifer, A. (2011), "*Teaching Practices and Social Capital*", IZA DP. No. 6052.

Ammermüller, A., Pischke, J. S. (2009), "*Peer Effects in European Primary Schools: Evidence from PIRLS*", Journal of Labour Economics, 27(3), 315-348.

Anderman, E. M., Murdock, T.B. (2007), "*The psychology of academic cheating*", Academic Press Inc.

Angrist, J. D., Lang, K. (2004), "*Does School Integration Generate Peer Effects? Evidence from Boston's Metco program*", American Economic Review, 94(5), 1613–1634.

Angrist, J. D., PischkE, J.S. (2008), "*Mostly Harmless Econometrics: An Empiricist's Companion*", Princeton University Press.

Aslund. O., Edin Per-Anders, Fredriksson, P., Gronqvist, H. (2011), "*Peers, Neighborhoods, and Immigrant Student Achievement: Evidence from a Placement Policy*", American Economic Journal: Applied Economics, vol. 3(2): 67-95.

Austen-Smith, D., Fryer, R. (2005), "*An economic analysis of 'acting white*", Quarterly Journal of Economics 551–583.

Aydemir, A., Borjas, G. J. (2010). A comparative analysis of the labor market impact of international migration: Canada, Mexico and the United States. *Journal of the European Economic Association*, 5(4), 663-708.

Bank of Italy (2008), "*Household Wealth in Italy*", Bank of Italy Occasional Paper, 2008.

Barbieri, G., Rossetti, C., Sestito, P. (2011), "*The determinants of teacher mobility: evidence using Italian teachers' transfer applications*", Economics of Education Review, 30(6), 1430-1444.

Barbieri, G., Rossetti, C., Sestito, P. (2011), "*The determinants of teachers mobility. Evidence using Italian teachers transfers applications*", Economics of Education Review, 30(6): 1430-1444.

Barone, G., Mocetti, S. (2011), "*With a little help from abroad: the effect of low-skilled immigration on female labor supply*", Labour Economics, Elsevier, vol. 18(5): 664-675.

Battu, H., Zenou, Y. (2010), "*Oppositional identities and employment for ethnic minorities: evidence from England*", The Economic Journal, 120 (February), F52–F71.

Becker, G. (1968), "*Crime and Punishment: An economic approach*", Journal of Political Economy, 76(2): 168-217.

Becker, Gary S (1974), “A Theory of Social Interactions”, Journal of Political Economy, Vol. 82, No. 6, pp. 1063-1093

Bertoni, M., Brunello, G., Rocco, L. (2012), “*When the Cat Is Near, the Mice Won't Play: The Effect of External Examiners in Italian Schools*”, IZA DP. 6629, Institute for the Study of Labor (IZA).

Betts, J.R., Fairlie, R.W. (2003), “*Does immigration induce ‘native flight’ from public schools into private schools?*”, Journal of Public Economics, 87(5-6), 987–1012.

Billari, F. C., Dalla Zuanna, G. (2008), “*La rivoluzione nella culla: il declino che non c’è*”, Milan: Egea.

Blume, Lawrence E. and Steven N. Durlauf (2005), “Identifying social interactions: a review”, unpublished working paper.

Boeri, T., Philippis M., Patacchini E., Pellizzari M. (2011), “*Moving to segregation: evidence from 8 Italian cities*”, IGER Working Paper No. 390, Bocconi University (Milan).

Borjas, G. J. (1995), “*Ethnicity, Neighborhoods, and Human-Capital Externalities*”, The American Economic Review, Vol. 85, No. 3, pp. 365-390.

Boustan, L. P. (2010), “*Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration*”, The Quarterly Journal of Economics, MIT Press, vol. 125(1), pages 417-443.

Bramoullé, Y., Djebbari, H., Fortin, B. (2009), “*Identification of peer effects through social networks*”, Journal of Econometrics 150 (2009) 41:55.

Bratti M., Checchi, D., Filippin, A. (2011), “*Should you compete or cooperate with your schoolmates?*”, Education Economics, 19(3), 275-289.

Brock, W. Durlauf, S. (2001b), *“Interactions-Based Models”*, in Handbook of Econometrics 5, Heckman, J., Leamereds, E.., Elsevier, 3297-3380.

Brock,W., Durlauf, S. (2001), “*Discrete Choice with Social Interaction*”, Review of Economic Studies, 68, 235-260.

Brock,W., Durlauf, S. (2001a), “*Discrete Choice with Social Interaction*”, Review of Economic Studies, 68, 235-260.

Brunello, G., De Paola, M., Scoppa, V. (2010), “*Peer effects in higher education: does the field of study matter?*”, Economic Inquiry, 48(3), 621–634.

Brunello, G., Rocco, L. (2011), “*The Effect of Immigration on the School Performance of Natives: Cross Country Evidence Using PISA Test Scores*”, IZA DP No. 5479.

Brunello, G., Rocco, L. (2011), “*The Effect of Immigration on the School Performance of Natives: Cross Country Evidence Using PISA Test Scores*”, IZA DP No. 5479.

Brunello, Giorgio, Maria De Paola and Vincenzo Scoppa (2010), “Peer effects in higher education: does the field of study matter?”, *Economic Inquiry*, Vol. 48, No. 3, July 2010, 621–634.

Calvo-Armengol, A., Patacchini, E., Zenou, Y. (2009), “*Peer Effects and Social Networks in Education*”, Review of Economic Studies 76:4, 1239-1267.

Card, D. (2001), *“Immigrant inflows, native outflows, and the local labour market impacts of higher immigration”*, Journal of Labour Economics, 19:22-64.

Card, D., Rothstein, J. (2007), “*Racial Segregation and the Black-white Test Score Gap*”, Journal of Public Economics, 91(11–12), 2158–2184.

Card, Da., Rothstein, J. (2007) “*Racial Segregation and the Black-white Test Score Gap*” Journal of Public Economics, 91: 2158–2184.

Carrel, S. E., Fullerton, R.,West J. E. (2009), “*Does your cohort matter? Measuring peer effects in college achievement*”, Journal of Labour Economics, 27(3): 439-464.

Carrel, S. E., Malstrom, F. V., West J. E. (2008), “*Peer effects in academic cheating*”, Journal of Human Resources, XLIII(1): 173-206.

Castellano R., Longobardi S., Quintano C. (2009), “*A fuzzy clustering approach to improve the accuracy of Italian student data. An experimental procedure to correct the impact of outliers on assessment test scores*”, StatisticaApplicazioni, Vol. VII, July-December 2009.

Caucutt EM. (2002), “Educational policy when there are peer group effects-size matters”, *International Economic Review* 43(1): 195–222.

CENSIS (2008), “*Vissuti ed esiti della scolarizzazione dei minori di origine immigrata*”, Censis Survey 2007.

Cipollone, P., Rosolia, A. (2007), “*Social Interactions in High School: Lessons From an Earthquake*”, American Economic Review, 97(3).

Cipollone, Piero and Alfonso Rosolia (2007), “Social Interactions in High School: Lessons From an Earthquake”, *American Economic Review*, Vol. 97, No. 3.

Cizek, G. J. (2003), “*Detecting and preventing classroom cheating: Promoting integrity in assessment*”, Thousand Oaks, CA: Corwin Press.

Clark A.E., Loheac, Y. (2007), “*It wasn’t me, it was them!. Social influence in risky behavior by adolescents*”, Journal of Health Economics, 26 (2007), pp. 763–784.

Clotfelter, C. T. (1999), “*Public school segregation in metropolitan areas*”, Land Economics, 75(4): 487-504.

CNEL (2011), “*Le seconde generazioni e il problema dell’identità culturale*”, Fondazione Silvano Andolfi, Italian National Council for Labour and Economics (CNEL).

Coleman, J. S.. (1966), “*Equality of Educational Opportunity*”, Washington D.C. U.S., Department of Health, Education and Welfare, Office of Education, OE-38001.

Coleman, James S. (1966), “Equality of Educational Opportunity,” Washington D.C., U.S., Department of Health, Education and Welfare, Office of Education, OE-38001.

Contini, D. (2011), “*Immigrant background peer effects in Italian schools*”, IRVAPP D.P., No. 2011-1.

Cook, T. D., (1984), “*What have black children gained academically from school desegregation? A review of the meta-analytic evidence*”, in National Institute of Education, School desegregation, Washington, DC.

Cooley, J- (2010a), “*Can Achievement Peer Effect Estimates Inform Policy? A View from Inside the Black Box*”, University of Wisconsin-Madison, Economics Department, working paper.

Cooley, J. (2010), “*Desegregation and the Achievement Gap: Do Diverse Peers Help”*, Duke University Working Paper.

Corno Lucia (2009), “Peer effects on criminal behavior. Evidence from the homeless in Italy”, Bocconi University, working paper.

Crain, R. L., Mahard, R. E. (1978), “*Desegregation and black achievement: A review of the research*”. Law and Contemporary Problems, 42(3), 17-53.

Cutler, D. M., Glaeser, E. L. (1997), “*Are Ghettos Good or Bad?*”, Quarterly Journal of Economics, 112, 827-72.

Cutler, D. M., Glaeser, E. L. (2010), “*Social Interactions and Smoking*”, NBER Chapters, in: Research Findings in the Economics of Aging, pages 123-141 National Bureau of Economic Research, Inc.

Cutler, D. M., Glaeser, E. L., Vigdor, J. L. (2008), “*When are ghettos bad? Lessons from immigrant segregation in the United States*”, Journal of Urban Economics 63, 759–774.

Davezies, L., D’Haultfoeuille, X., Fougèere, D. (2009), “*Identification of peer effects using group size variation*”, Econometrics Journal (2009), volume 12, pp. 397–413.

Davidson, R., MacKinnon, J. (2004), “*Econometric theory and methods*”, Oxford University press, New York.

Davis, S. F.,. Drinan, P. F, Gallant, T. B. (2009), “*Cheating in school*”, Wiley-Blackwell, U.K..

De Giorgi, G., Pellizzari M. (2011), “*Understanding Social Interactions: Evidence from the Classroom*”, IZA DP No. 5624.

De Giorgi, G., Pellizzari, M., Redaelli, S. (2010), “*Identification of Social Interactions through Partially Overlapping Peer Groups*”, American Economic Journal: Applied Economics, 2(2): 241-75.

De Giorgi, Giacomo, Michele Pellizzari and Silvia Redaelli (2010), "Identification of Social Interactions through Partially Overlapping Peer Groups", *American Economic Journal: Applied Economics*, 2 (April 2010): 241–275.

De Paola, Maria (2010), "Absenteeism and peer interaction effects: Evidence from an Italian Public Institute", *The Journal of Socio-Economics,* 39 (2010): 420–428.

De Paola, Maria and Vincenzo Scoppa (2010), "Peer group effects on the academic performance of Italian students", *Applied Economics*, 42: 17, 2203-2215.

Dee, T. S., Brian A. J. (2012), "*Rational Ignorance in Education: A Field Experiment in Student Plagiarism*", Journal of Human Resources, University of Wisconsin Press, vol. 47(2), pp. 397-434.

Drago, F., Galbiati, R. (2012), "*Indirect Effects of a Policy Altering Criminal Behaviour: Evidence from the Italian Prison Experiment*", American Economic Journal: Applied Economics, 4(2): 199-218.

Duflo, E., Dupas, P., Kremer, M. (2011) "*Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya*", American Economic Review, 101(5): 1739–74.

Duncan, O. D., Duncan, B. (1955), "*A Methodological Analysis of Segregation Indexes*", American Sociological Review, 20: 210-217.

Durlauf, S. N., Cohen-Cole, E. (2004), "*Social interactions models*", in "*The Encyclopedia of Social Measurement*", K. Lempf-Leonard eds., Academic Press.

Durlauf, S. N., Yannis M. Ioannides (2009), "*Social Interactions*", Tufts University, Economic Department Working Paper, Forthcoming in Annual Review of Economics (2010).

Durlauf, S., Tanaka, H. (2008), "*Regression Versus Variance Tests for Social Interactions.*", Economic Inquiry, 46(1), 25-28.

Durlauf, Steven N. (2004), "Neighbourhood effects", in the *Handbook of Regional and Urban Economics*, Volume 4, Economics, J. V. Henderson and j. F. Thisse Eds.

Durlauf, Steven N. and Yannis M. Ioannides (2009), "Social Interactions", Tufts University, Economic Department Working Paper. Forthcoming in *Annual Review of Economics*.

Dustmann, C., Machin, S., Schonberg, U. (2011), "*Ethnicity and educational achievement in compulsory schooling"*, Economic Journal, 120(546): F272-F297.

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

Dustmann, C., Glitz, A. (2011), "Migration and education", in Hanushek, E. A., Machin, S., Woessmann, L. (eds), "*Handbook of Economics of Education"*, Vol. 4, Ch. 4, Elsevier.

Dustmann, C., Glitz, A. (2011), "*Migration and education*", in: Handbook of Economics of Education, Vol. 4, Ch. 4, Elsevier.

Dustmann, C., Machin, S., Schonberg, U. (2011), "*Ethnicity and educational achievement in compulsory schooling*", Economic Journal, 120(546): F272-F297.

Dustmann, C., Preston, I. (2001), "*Attitudes to ethnic minorities, ethnic context and location decisions*", Economic Journal, 111(470), 353-373.

Dustmann, Christian and Albrecht Glitz (2011), "Migration and education", *Handbook of Economics of Education,* Vol. 4, Ch. 4, Elsevier.

Echenique, F., Fryer, R. G. Jr. (2007), *"A measure of segregation based on social interactions",* The Quarterly Journal of Economics, Vol. CXXII, Issue 2: 441-485.

Echenique, Federico and Fryer, Roland G. Jr. and Kaufman, Alex (2006), "Is school segregation good or bad?", *The American Economic Review*, Vol. 96(2): 265-269.

Entorf, H., Eirini, T. (2009), "*Migrants at School: Educational Inequality and Social Interaction in the UK and German*y", IZA DP No. 4175.

Entorf, H., Lauk, M. (2006), *"Peer Effects, Social Multipliers and Migrants at School: An International Comparison"*, IZA DP No. 2182.

Epple D., Romano, R.D. (2011), "*Peer effects in education: a survey of the theory and evidence"*, Handbook of Social Economics, Vol. 1B, Ch. 20.

Eurydice (2009), "*National testing of pupils in Europe: objectives, organization and use of the results*", Education, Audiovisual and Culture Executive Agency (European Commission).

Eurydice (2012), Key data on education in Europe, *Indire Reports*.

Evans, W. N., Oates, W. E., Schwab, R. M. (1992), "*Measuring Peer Group Effects: A Study of Teenage Behavior*", Journal of Political Economy, 100(5), 966-91.

Fairlie, R., Hoffmann, F., Oreopoulos, P. (2011), "*A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom*", NBER Working Paper No. 17381

Falk, Armin and Andrea Ichino (2006), "Clean Evidence on Peer Effects", *Journal of Labor Economics*, 2006, vol. 24, no. 1.

Ferrer-Esteban, G. (2012), "*Cheating to the test in the Italian standardized assessment system: rationale and incentives*", FGA Working Paper, (in press), Giovanni Agnelli Foundation (Turin).

Figlio, D. (2007), "*Boys Named Sue: Disruptive Children and Their Peers*", Education Finance and Policy Vol. 2, No. 4, Pages 376-394.

Fordham, S., Ogbu, J. (1986), "*Black students' school successes: coping with the burden of 'acting white'*", The Urban Review, 18, 176–206.

Fortin B., Lacroix G., Villeval M. C. (2007), “*Tax evasion and social interactions*”, Journal of Public Economics, 91: 2089–2112

Foster, G. (2006), “*It’s not your peers and it’s not your friends: Some progress towards understanding the educational peer effect mechanism”*, Journal of Public Economics 90, 1455 – 1475.

Foster, G., Frijters, P. (2009), “*Are there complementarities in educational peer effects?*”, *mimeo*, University of South Australia and Queensland University of Technology.

Friesen, J., Krauth, B. (2011), “*Ethnic enclaves in the classroom*”, Labour Economics, 18: 656-663.

Fryer, R. G. Jr. (2010), “*The Importance of Segregation, Discrimination, Peer Dynamics, and Identity in Explaining Trends in the Racial Achievement Gap”*, Handbook of Social Economics, Elsevier.

Fryer, R. G. Jr. (2011), “*The importance of segregation, discrimination, peer dynamics, and identity in explaining trends in the racial gap*”, in Handbook of Social Economics, Vol. 1B, Ch. 21.

Fryer, R. G. Jr., Torelli, P. (2010), “*An empirical analysis of ‘acting white’*”, Journal of Public Economics, 94 (2010), 380–396

Galbiati, R., Zanella, G. (2012), “*The tax evasion social multiplier. Evidence from Italy*”, Journal of Public Economics, vol. 96(5): 485-494.

Gaviria, A., Raphael S. (2001), “*School-Based Peer Effects and Juvenile Behavior*”, The Review of Economics and Statistics, Vol. 83, No. 2, Pages 257-268.

Geay, C., Mc Nally, S., Telhaj, S. (2012). “*Non-native speakers of English in the classroom: what are the effects on pupil performance?*”, IZA DP. No. 6451.

Gibbons, Stephen and Telhaj Shqiponje (2006), “Peer effects and pupil attainment: evidence from secondary school transition”, *Centre for the Economics of Education Working Paper* (LSE), May 2006.

Glaeser, E., Sacerdote, B., Scheinkman, J. (1996): “*Crime and Social Interactions*”, The Quarterly Journal of Economics, 111, 507–548.

Glaeser, E., Sacerdote, B., Scheinkman, J. (2003), “*The Social Multiplier*”, Journal of the European Economic Association, 1, 2-3, 345-353.

Glaeser, E.L., B.I. Sacerdore, J.A. Scheinkman (2003) ‘The Social Multiplier’ *Journal of the European Economic Association,* 1 (2-3): 345-353.

Gould, E. D., Lavy, V., Paserman, M. D.(2009), “*Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence*”, The Economic Journal 119(540), 1243-1269.

Graham, B. (2008), "*Identifying social interactions through conditional variance restrictions*", Econometrica, 76(3), 643-660

Graham, B. (2008). "*Identifying Social Interactions through Conditional Variance Restrictions*", Econometrica, 76(3), 643-660

Grimes, P. W., Rezek, J.P. (2005), "*The Determinants of Cheating by High School Economics Students: A Comparative Study of Academic Dishonesty in the Transitional Economies"*, International Review of Economics Education, volume 4, issue 2, pp. 23-45.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2009). New Evidence About Brown V. Board of Education: The Complex Effects of School Racial Composition on Achievement. *Journal of Labor Economics,* 27(3).

Hanushek, E. A., Kain, J. F., Martin, J. M., Rivkin, S. G. (2003), "*Does peer ability affect student achievement?*", Journal of Applied Econometrics, 18, 527-544.

Hanushek, E. A., Rivkin, S. G (2009), "*Harming the Best: How Schools Affect the Black–White Achievement Gap*", Journal of Policy Analysis and Management, Vol. 28, No. 3, 366–393.

Hanushek, E. A., Rivkin, S. G. (2009), "*Harming the Best: How Schools Affect the Black–White Achievement Gap"*, Journal of Policy Analysis and Management, 28(3), 366–393.

Hanushek, Eric A., Kain J. F., Martin, J. M., Rivkin, S. G. (2003), "*Does peer ability affect student achievement?*", Journal of Applied Econometrics, 2003, 18: 527-544.

Hanushek, Eric A., Kain, J. F., Rivkin, S. G. (2009), "*New Evidence About Brown V. Board of Education: The Complex Effects of School Racial Composition on Achievement*", Journal of Labor Economics, 2009, vol. 27, no. 3.

Heckman, J. J. (2011), "*The American Family in Black and White: A Post-Racial Strategy for Improving Skills to Promote Equality*", IZA DP. No. 5495.

Hoxby, C. M. (2000), "*Peer Effects in the Classroom: Learning from Gender and Race Variation*", National Bureau of Economic Research: Working Paper No. 7867.

Hoxby, C. M. (2000), "*Peer Effects in the Classroom: Learning from Gender and Race Variation*", National Bureau of Economic Research: Working Paper No. 7867.

Hoxby, C. M., Weingarth, G. (2006), "*Taking race out of the equation: school reassignment and the structure of peer effects*", Unpublished working paper.

Hoxby, Caroline M. (2000), "Peer Effects in the Classroom: Learning from Gender and Race Variation", *National Bureau of Economic Research, Working Paper No. 7867*.

Imberman, S. A., Kugler, A., Sacerdote B. I. (2012), "*Katrina's children: evidence on the structure of peer effcts from hurricane evaquees*", American Economic Review, 102(5), 2048-2082.

INVALSI (2008, 2009, 2010), "*Esame di Stato di I Ciclo, Prova Nazionale: Rapporti Tecnici e Analisi*", INVALSI online publication.

INVALSI (2010b), "*Prova Nazionale di Valutazione degli apprendimenti degli studenti italiani*. *Rapporto Tecnico e Analisi*", INVALSI Reports.

Jacob B., Levitt, S. D. (2003b), "*Catching Cheating Teachers: The Results of an Unusual Experiment in Implementing Theory*", NBER Working Paper No. 9414.

Jacob, B. A. (2005), "*Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools*", Journal of Public Economics, 89:761-796.

Jacob, B. A., Levitt, S. D., (2003a), "*Rotten Apples: An Investigation Of The Prevalence And Predictors Of Teacher Cheating*", The Quarterly Journal of Economics, MIT Press, vol. 118(3), pages 843-877, August.

Jansen, P., Rasmussen, A. W. (2011), "*The Effect of Ethnic Concentration Schools on native and immigrant children's reading and math skills*", Economics of Education Review, 30(6): 1503-1515.

Kerkvliet, J., Sigmund, C. L. (1999), *"Can we control cheating in the classroom?"*, Journal of Economic Education 30 (Fall): 33 -143.

Kinsler, J. (2010), "*School Discipline: A Source or Salve for the Racial Achievement Gap?"*, November 12, 2010, Duke University, Economics Department Working Paper.

Kleven, H. J., Knudsen, M. B., Thustrup, Kreiner, Claus, Pedersen, Søren,Saez, Emmanuel, (2011), "*Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark*", Econometrica, Vol. 79, No. 3, 651–692.

Lavy, V. (2009), "*Performance Pay and Teachers' Effort, Productivity, and Grading Ethics*", American Economic Review, Vol. 99, No. 5: Pages 1979-2011.

Lavy, V., Olmo, S., Weinhardt, F. (2012), *"The good, the bad and the average: evidence on the scale and the nature of ability peer effects at school"*, Journal of Labor Economics, Vol. 30, No. 2, pp. 367-414.

Lavy, V., Schlosser, A. (2011), "*Mechanisms and Impacts of Gender Peer Effects at School*", American Economic Journal: Applied Economics", vol. 3(2), pages 1-33, April.

Lazear, E. P. (1999), "*Culture and Language*", Journal of Political Economy, 107(S6), S95–S126.

Lazear, E. P. (2001), "*Educational production*", The Quarterly Journal of Economics, 116(3): 551-583.

Lazear, E. P. (2006), "*Speeding, terrorism, and teaching to the test*", The Quarterly Journal of Economics, MIT Press, vol. 121(3), pages 1029-1061, August.

Ludemann, E., Schwerdt G. (2012), "*Migration background and educational tracking. Is there a double disadvantage for second –generation students?"*, Journal of Population Economics, DOI: 10.1007/s00148-012-0414-z.

Lyle, D. S. (2007), "*Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point*", Review of Economics and Statistics 89(2), 289 – 299.

Maestri, V. (2011), "*A deeper insight into the ethnic make-up of school cohorts. Diversity and school achievement*", AIAS Working Papers wp108, Amsterdam Institute for Advanced Labour Studies.

Manski Charles F. (1995), "Identification Problems in the Social Sciences", *Harward University Press*.

Manski Charles F., (2000), "Economic Analysis of Social Interactions", *Journal of Economic Perspectives*, vol. 14(3), pages 115-136.

Manski, C. F., (1993), "*Identification of Endogenous Social Effects: The Reflection Problem*", Review of Economic Studies, vol. 60(3), pages 531-42, July.

Mas Alexandre and Enrico Moretti (2009), "Peers at Work", American Economic Review, American Economic Association, vol. 99(1), pages 112-45, March

Maurin, E., Moschion, J. (2009), "*The Social Multiplier and Labor Market Participation of Mothers*", American Economic Journal: Applied Economics 2009, 1:1, 251–272.

McCabe, D. L. (2005), "*Cheating among college and university students: A North American perspective*", International Journal for Educational Integrity 1 (1).

McCabe, D. L., Klebe Trevino, L. (1993), ''*Academic Dishonesty: Honour Codes and Other Contextual Influences*'', Journal of Higher Education 64(5):522–38.

McCabe, D. L., Klebe Trevino, L. (1997), "*Individual and contextual influences on academic dishonesty: a multicampus investigation*", Research in Higher Education, 38(3): 379-376.

Mckee, G. J., Rivkin, S. G., Sims, K. (2010), "*Disruption, achievement and the heterogeneous benefits of smaller classes*", NBER Working Paper No.15812.

Mencarini, L., Baldoni E., Dalla Zuanna, G. (2009), "*The situation of children in immigrant families in italy: changes and challenges"*, Innocenti Working Paper: Special Series on Children in Immigrant Families in Affluent Societies, IWP-2009-15.

Miller, A., Murdock, T., Anderman, E. M., Poindexter, A. L. (2007), "*Who are all these cheaters? Characteristics of academically dishonest students*", In E. M. Anderman and T. B. Murdock (Eds.), "*Psychology of Academic Cheating*", Boston: Elsevier.

MIUR (2009a), "Alunni con cittadinanza non italiana Scuole statali e non statali. Anno scolastico 2007/2008", Aprile 2009, MIUR: Servizio Statistico.

MIUR (2009b), "*Gli alunni stranieri nel sistema scolastico italiano*", Anno scolastico 2008/2009", Dicembre 2009, MIUR: Servizio Statistico.

MIUR (2010), “*Focus in breve sulla scuola: la presenza degli alunni stranieri nelle scuole statali*”, Marzo 2010, MIUR: Servizio Statistico.

MIUR (2011). Gli alunni stranieri nel sistema scolastico italiano. Anno scolastico 2010/2011. Italian Ministry of Education Statistical Office and ISMU Foundation.

Moretti Enrico, (2011), “Social Learning and Peer Effects in Consumption: Evidence from Movie Sales”, Review of Economic Studies, Oxford University Press, vol. 78(1), pages 356-393

Moscone F., Tosetti E. and Vittadini G. (2011), “Social Interaction in Patients’ Hospital Choice: Evidences from Italy”, Journal of the Royal Statistical Society: Series A, 175(2), 453–472.

Mueller, S. (2011), “*Teacher experience and the class size effect – experimental evidence*”, IWQW discussion paper series No. 07/2011.

Murdock, T. B., Anderman, E. M. (2006), “*Motivational perspectives on student cheating: Towards an integrated model of academic dishonesty*”, Educational Psychologist, 41(3), 129–145.

Murdock, T. B., Hale, N. M., Weber, M. J. (2001), “*Predictors of cheating among early adolescents: Academic and social motivations*”, Contemporary Educational Psychology, 26(1), 96-115.

Neidell, Matthew and Jane Waldfogel (2010), “Cognitive and noncognitive peer effects in early education”, *The Review of Economics and Statistics*, August 2010, 92(3): 562–576.

OECD (2010), “*OECD Reviews of Migrant Education”*, online publications, Paris: OECD.

OECD (2011), “*International Migration Outlook”*, SOPEMI 2011, Paris: OECD.

Ohinata, A., Van Ours, J. C. (2011), “*How immigrant children affect the academic achievement of native Dutch children”*, IZA DP. No. 6212.

Patacchini, E., Zenou, Y. (2012), “*Juvenile delinquency and conformism*”, Journal of the Law, Economics, and Organization, 28, 1-13.

Patacchini, E., Zenou, Y.(2006), “*Racial Identity and Education*”, IZA DP. No. 2046.

Pellizzari, M. (2011), “*The use of welfare by migrants in Italy*”, IZA DP. No. 5613.

Portes, A. (1987), “*The social origins of the Cuban enclave economy of Miami*”, Sociological Perspectives 30, 340–372.

Powell B., Manish, G.P., Nair M. (2010), *“Corruption, crime and economic growth”*, in “*Handbook on the Economics of Crime*”, Benson B. L. and Zimmerman P. R. eds., Edward Elgar Publishing, Northampton.

Sacerdote Bruce (2010), *"Peer effects in education: how might they work, how big are they, and how much do we know so far?"*, in *"Handbook of the Economics of Education"*, Volume 3, Ch. 4, Eric A. Hanushek, Stephen Machin, Ludger Woessmann Eds.

Sacerdote, B. (2001), "*Peer Effects with Random Assignment: Results for Dartmouth Roommates"*, Quarterly Journal of Economics, 681-704.

Sacerdote, B. (2001), "Peer Effects with Random Assignment: Results for Dartmouth Roommates", *The Quarterly Journal of Economics*, vol. 116(2), pp. 681-704.

Saiz, A., Wachter, S. (2011), "*Immigration and the Neighborhood*", American Economic Journal: Economic Policy, American Economic Association, vol. 3(2), pages 169-88, May

Scheinkman, José A. (2008), "Social Interactions", Published in *The New Palgrave Dictionary of Economics*, 2nd edition, S. Durlauf and L. Blume (eds.), Palgrave Macmillan, 2008.

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

Schnepf, S. V.(2007), "*Immigrants' educational disadvantage: an examination across ten countries and three surveys"*, Journal of Population Economics, 20, 527-545.

Schnepf, V. (2007),"*Immigrants' educational disadvantage: an examination across three surveys and ten countries*", Journal of Population Economics, 20(3): 527-545.

Shepard E. M, Blackely, P. R. (2010), "*Economics of crime and drugs: prohibition and public policies for illicit drug control*", in "*Handbook on the Economics of Crime*", Benson B. L., Zimmerman P. R. eds., Edward Elgar Publishing, Northampton.

Stanard, C. I., Bowers, W. J. (1970), ''*The College Fraternity as an Opportunity Structure for Meeting Academic Demands*'', Social Problems 17(3):371–90.

Stanat, P., Christensen, G. *(2006)*, "*Where immigrant students succeed: A comparative review of performance and engagement in PISA 2003*", Paris: Organisation for Economic Co-operation and Development.

Steele, S., Aronson, J. (1998), "*Stereotype threat and the test performance of academically successful African Americans*", in Jencks, C., & Phillips, M. (Eds.), "*The Black–White Test Score Gap"*, Brookings Institution Press, Washington, DC, pp. 401–430.

Stinebrickner, T. R., Stinebrickner, R. (2005) "*What Can Be Learned About Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds*", CIBC Working Paper Series, 2005-4.

Stock, J., Motohiro, Y. (2005), "*Testing for Weak Instruments in Linear IV Regression*", in "*Identification and Inference for Econometric Models Essays in Honor of Thomas Rothenberg***",** Edited by Donald W. K. Andrews and James H. Stock, Cambridge University Press.

U.S. Department of Education (2009), "*Standards and Assessments Peer Review Guidance: Information and Examples for Meeting Requirements of the No Child Left Behind Act of 2001*", online publication.

Zanella, Giulio (2004), "Social Interactions and economic behaviour", *Quaderni del Dipartimento di Economia Politica*, Università di Siena, No. 441, November 2004.

Zimmerman, David J. (2003), "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment", *Review of Economics and Statistics*, vol. 85(1), pp. 9-23.

“Lui conobbe lei e se stesso, perché in verità non s'era mai saputo.

E lei conobbe lui e se stessa, perché pur essendosi saputa sempre,

mai s'era potuta riconoscere così.”

Italo Calvino, Il barone rampante.