Heterogeneous Peer Effects and Rank Concerns: Theory and Evidence

Michela M. Tincani[1]

January 24, 2017

[1]Tincani: University College London, 30 Gordon Street, London, WC1H0BE, UK, m.tincani@ucl.ac.uk. I am particularly grateful to Orazio Attanasio, Ed Hopkins, Aureo de Paula, Petra Todd and Ken Wolpin for their feedback and encouragement. I would also like to thank David Atkin, Richard Blundell, Denis Chetverikov, Martin Cripps, Mariacristina De Nardi, Steven Durlauf, Christian Dustmann, Jan Eeckhout, Eric Hanushek, Tatiana Kornienko, Dennis Kristensen, Stephen Machin, Costas Meghir, Lars Nesheim, Imran Rasul, Elie Tamer, Martin Weidner, Daniel Wilhelm and participants at various seminars and conferences for insightful comments and discussions. Julia Schmieder provided excellent research assistance. Research funding from the Centre for Microdata Methods and Practice and from the European Research Council's grant number IHKDC-249612 is gratefully acknowledged. I am grateful to the Chilean Ministry of Education and to the *Agencia de Calidad de la Educación* for access to some of the data used in this research. The views reported here are those of the author and do not necessarily reflect views at the Ministry or at the *Agencia*.

Abstract

Using a theoretical model where students care about achievement rank, I study effort choices in the classroom and show that rank concerns generate peer effects. The model's key empirical prediction is that the effect on own achievement of increasing the dispersion in peer cost of effort is heterogeneous, depending on a student's own cost of effort. To test this, I construct a longitudinal multi-cohort dataset of students, with data on the geographic propagation of building damages from the Chilean 2010 earthquake. I find that higher dispersion in home damages among one's classmates led, on average, to lower own Mathematics and Spanish test scores. To be able to test the theory, I develop a novel nonlinear difference-in-differences model that estimates effect heterogeneity and that relates observed damages to unobserved cost of effort. I find that some students at the tails of the predicted cost of effort distribution benefit from higher dispersion in peer cost of effort, as predicted by the theoretical model. This finding suggests that observed peer effects on test scores are, at least partly, governed by rank concerns.

1 Introduction

Peer effects have been widely studied by economists in many contexts, for example, career choices, health behaviours, crime and education. Typically, peer effect models describe an outcome of interest as a function of some feature of a peer group. The simplest model examines the importance of the mean of peer characteristics in shaping own behaviour. While this model seems to capture social influences in, for example, crime and drinking, it is rejected in other contexts.

In education, the linear-in-means model produces ambiguous and sometimes contradictory results: it is not always the case that own outcomes improve when peers are on average more able. In contrast, a growing body of empirical evidence points to the nonlinearity and heterogeneity of peer effects on test scores (see the survey by Sacerdote (2014)). For example, in one recent randomised experiment at the University of Amsterdam, low-ability students placed in tutorial groups with other low-ability students performed better than in groups with more able peers on average (Booij, Leuven, and Oosterbeek 2016).[1] In another randomised experiment among students at the U.S. Air Force Academy, low-ability students performed worse when placed in groups with other low-ability peers, but those groups included high-ability peers and no middle-ability peers (Carrell, Sacerdote, and West 2013). These experimental interventions altered the entire distribution of peer ability, and generated some results that may appear unrelated.

In this paper, I study the primitives of the student problem to provide a conceptual framework that can potentially explain disparate findings.[2] I propose a theoretical model of student effort choices in the classroom that has testable implications on the shape of peer effects. To test them, I exploit a natural experiment and a novel econometric approach that is closely guided by the theoretical model. An advantage of this method is that the treatment effect and other aspects of the econometric model have a direct structural interpretation. I show that all predictions are borne out in an administrative

[1] In a randomised experiment with Kenyan first-graders, Duflo, Dupas, and Kremer (2011) find positive effects of ability tracking at all ability levels.

[2] Other papers examine potential mechanisms behind peer effects. Blume, Brock, Durlauf, and Jayaraman (2014) and Fruehwirth (2013) provide microfoundations to the widely used linear-in-means peer effect specification, proving that it can be rationalized by a desire to conform. De Giorgi and Pellizzari (2013) develop and test behavioral models that can rationalize observed outcome clustering within classrooms at Bocconi University. Calvó-Armengol, Patacchini, and Zenou (2009) provide microfoundations to the Katz-Bonacich centrality measure in a network. Using a different approach, Lavy and Schlosser (2011) and Lavy, Paserman, and Schlosser (2012) use teacher and student surveys to understand why gender variation and proportion of low-ability students impact class outcomes. This paper differs from this stream of the literature in its intent to find a mechanism that is consistent with seemingly unrelated patterns of nonlinear and heterogeneous peer effects observed in the data.

dataset on Chilean students.

In the model, students care about achievement and achievement rank. Achievement is produced through costly study effort, and students are heterogeneous in terms of their cost of effort; that is, in terms of physical and/or psychological characteristics that affect their ability to study. In this context, how much study effort each student exerts depends, in equilibrium, on the cost of effort of her peers. For example, a student might "give up" in a group with considerably better peers, because improving rank is too costly in terms of effort. Therefore, rank concerns generate peer effects, independently from externalities in the test score production technology.[3]

Peer effects generated by rank concerns work through the entire distribution of peer characteristics. The comparative statics explore one feature of this distribution, its dispersion, because this is the empirically relevant margin in my data. The theory predicts that increasing the dispersion of peer cost of effort has heterogeneous effects on achievement. Students with a high cost improve their achievement, students with a medium cost decrease their achievement, and students with a low cost may increase or decrease their achievement, depending on the relative importance of rank concerns in their preferences. Intuitively, an important determinant of effort is how close to each other students are in the cost distribution, because this determines how easy or hard it is to improve ordinal rank in achievement. Changes in dispersion cause the proximity between students to change, and to do so differently on different portions of the cost distribution. This generates the heterogeneous patterns of response that lay the basis for the empirical tests.

To test the theoretical predictions, I combine a large administrative dataset on over $350,000$ Chilean students with information on the Chilean 2010 earthquake, which I use as a natural experiment that generated variation in some peer group attributes.[4] In

[3]The idea that rank concerns could generate peer effects dates back to at least Jencks and Mayer (1990). However, it has never been formalised before, even though there is evidence suggesting that students care about their achievement rank even in the absence of specific rank incentives (Tran and Zeckhauser 2012, Azmat and Iriberri 2010). Rank concerns have been studied also in various fields outside of education, for example, in the study of well-being at work and job satisfaction (Brown, Gardner, Oswald, and Qian 2008, Card, Mas, Moretti, and Saez 2012), of sport tournaments (Genakos and Pagliero 2012) and of personnel (Blanes i Vidal and Nossol 2011), among others.

[4]Natural disasters have been used before to identify peer effects in education, see, for example, Cipollone and Rosolia (2007), Imberman, Kugler, and Sacerdote (2012) and Sacerdote (2008). In contrast to some previous studies, this paper does not use forced relocations of students for identification. Rather, peer effects are identified by comparing peer groups with identical compositions, but that were subject to different earthquake shocks. This is in contrast also to the experimental and quasi-experimental literatures that use variation in assignment to peer groups, e.g. dorms (Sacerdote 2001, Zimmerman 2003, Stinebrickner and Stinebrickner 2006, Kremer and Levy 2008, Garlick 2016) or classrooms (Duflo, Dupas, and Kremer 2011, Whitmore 2005, Kang 2007).

terms of data construction, I use four waves of the SIMCE dataset (*Sistema de Medición de la Calidad de la Educación*, 2005, 2007, 2009, 2011), with information on students, teachers, classrooms and schools. The outcomes of interest are standardised test scores in Mathematics and Spanish in the 8^{th} grade. School and classroom identifiers allow me to match students to classmates, teachers and schools, making this dataset ideal to study peer effects at classroom level. I merge this educational data with a measure of damage to students' homes caused by the earthquake, which I obtained from the structural engineering literature. The measure is based on seismic intensity according to the Medvedev-Sponheuer-Karnik scale. The resulting dataset has three key features: it is longitudinal (test scores are observed twice for each student), it contains two cohorts (one affected by the earthquake in the second time period, one never affected), and it contains geographic variation in the intensity of exposure to the earthquake. Therefore, an appropriate tool of analysis is the difference-in-differences (DD) model. The preliminary and main data analyses use this model and its extensions, including a novel nonlinear one, for estimation and for tests of identification.

The preliminary data analysis indicates that being exposed to the earthquake reduced own test scores, on average, by up to 0.04 standard deviations. Students were affected differently depending on intensity of exposure, with a reduction of approximately 0.02 standard deviations for every additional USD 100 in home damages. Moreover, I find that *peer* intensity of exposure matters for *own* achievement. A linear regression DDD model with continuous treatment estimates that, keeping classroom composition and other school and teacher characteristics constant, the average level of damage among a student's peers had insignificant or negligible effects, while the dispersion in peer damages had a significant and sizeable negative effect on own test scores.[5]

To be useful for testing the theoretical prediction of heterogeneous impacts of cost dispersion, any empirical model must have two features: it must be able to detect effect heterogeneity across students (the linear model does not do this); and it must be able to relate the distribution of observed earthquake damages in the classroom to the distribution of unobserved students' cost of effort.[6] This allows the researcher to use variation

[5]Damages to the school are controlled for. This result holds irrespective of how dispersion is measured, e.g., standard deviation, coefficient of variation and various interquartile ranges. The identifying assumption for the damage dispersion effects, that is, that the correlation between damage dispersion and test score growth is constant across cohorts in the absence of treatment, is tested and not rejected using data on regions not affected by the earthquake.

[6]In $3,822$ out of the $5,574$ classrooms in earthquake regions there is within-classroom variation in earthquake intensity, because of a geographically dispersed student body. These are the classrooms that are used to identify the impact of damage dispersion.

across classrooms in damage dispersion as a shifter to the cost of effort dispersion, keeping constant usual confounders such as student composition and characteristics of the school, classroom and teacher. Guided by the theoretical model, I build a flexible econometric model that satisfies both conditions. It is a nonlinear matching estimator. In intuitive terms, rather than estimating a treatment effect parameter, it semiparametrically estimates a treatment effect function. This function traces the impact on achievement of an increase in the dispersion of cost of effort in the classroom, as a function of a student's predicted own cost of effort.[7] It is related to the non-linear DD models in, for example, Athey and Imbens (2006), Abadie (2005), Heckman, Ichimura, Smith, and Todd (1998), and Blundell, Costa Dias, Meghir, and Van Reenen (2004), but it addresses specific challenges of this context that those models are not well-suited to address.

Correlated effects may arise because classrooms with larger damage dispersion are also classrooms with a more geographically dispersed student body. Pre-earthquake data is used to net out any correlation between geographic dispersion and unobserved classroom characteristics that affect achievement. The identifying restriction is that the geographic dispersion of the student body does not determine how a classroom's unobserved inputs change across cohorts, for example, in response to the earthquake. This is a very weak restriction: unobserved inputs may change as a function of any other classroom variable, including those correlated with geographic dispersion like variance in household income. Moreover, the earthquake is allowed to have a direct and heterogeneous impact on students that is not due to damage to their own or their peers' homes. I develop an empirical test of the restriction at the school level, and show that identification is robust to school$\times$cohort fixed effects. Specifically, while changes to unobserved school characteristics may have occurred across cohorts and may have affected student outcomes, I show that they do not bias estimated damage dispersion effects. Moreover, I present additional empirical evidence suggesting that the latter are not due to changes in observed teacher productivity in the classroom. Finally, at the estimated parameter values I am able to rule out some alternative mechanisms, including self-selection into subgroups of friends within a classroom, a mechanisms proposed by Carrell, Sacerdote, and West (2013) to explain their experimental findings.

Findings indicate that estimates from the linear model mask considerable heterogeneity. In particular, not everyone is hurt by damage dispersion: some students at the tail of the cost of effort distribution benefit from an increase in damage dispersion. The es-

[7] Unobserved cost of effort is modelled as a single-index of observed student characteristics, including home damages. I validate this specification through various post-estimation tests, additional regression evidence, and survey evidence on elicited effort costs.

timated treatment effect functions have a direct structural interpretation as the effect on achievement of an increase in the dispersion of peer effort cost, and they are entirely consistent with the theoretical predictions, both in terms of point estimates and when used to statistically test those predictions.[8] This indicates that observed peer effects on test scores are, at least partly, governed by rank concerns.[9]

This paper is related to a growing literature on the effect of ability rank on achievement. Using data on primary schools in England and on high-schools in the U.S., Murphy and Weinhardt (2016), Elsner and Isphording (2016a) and Elsner and Isphording (2016b) find that a higher ability rank (keeping ability constant) improves academic outcomes and reduces risky behaviours later on, sometimes even years later. These results are consistent with the findings reported in Cicala, Fryer, and Spenkuch (2016): higher rank improved achievement in primary schools in Kenya, and it decreased misbehaviour in middle schools in the U.S.. Cicala, Fryer, and Spenkuch (2016) propose a theoretical framework and a field experiment to rationalise these results. This paper differs from this stream of the literature in a fundamental way. There is an important difference between rank in ability, a pre-determined student characteristic, and rank in achievement, an outcome. Concerns over achievement rank are not necessary nor sufficient for observing an impact on student outcomes of rank in a pre-determined characteristic.[10] Therefore, this paper does not specifically address the findings on the role of ability rank.

More generally, this paper speaks to the literature on nonlinear and heterogeneous peer effects. For example, one key idea is that proximity of students throughout the ability distribution matters, because it determines how easy it is to improve ordinal rank. Tincani (2014) shows that this idea can help understand, simultaneously, the positive effects of ability tracking in Booij, Leuven, and Oosterbeek (2016), and the (unexpected) findings in Carrell, Sacerdote, and West (2013), also qualitatively replicated in Booij, Leuven, and

[8]Moreover, rank concerns appear to be stronger in Mathematics than in Spanish, but this difference is not statistically significant.

[9]Finding direct evidence of rank concerns is difficult with available datasets, mainly because preferences are not observed. For this reason, this analysis cannot quantify what fraction of social influences is due to rank concerns, rather, it looks for symptoms of rank concerns in the shape of the peer effects, using an econometric framework that estimates treatment effects and other relations in the data that are fully interpretable with a structural model. This is the most that can be done with existing data.

[10]Moreover, depending on model specification, rank concerns may imply a positive or negative impact of ability rank on achievement. For example, in the randomised experiment in Carrell, Sacerdote, and West (2013), some low-ability students performed worse, even though the intervention improved their ability rank. Tincani (2014) shows that this result can be explained by rank concerns if students have a strong aversion to a low achievement rank, because they decrease their effort when they are moved away from the bottom of the ability distribution.

Oosterbeek (2016).[11] Proximity seems to be one important margin for policy intervention, especially when it is known that extrinsic rank incentives are in place. A novel policy implication, then, is that an educator can optimally exploit such rank concerns to promote study effort, much like an employer exploits social incentives to motivate workers.

The paper is organised as follows: section 2 introduces the theoretical model and predictions. Section 3 describes the data and presents preliminary results. Section 4 contains the empirical analysis of spill-overs generated by the earthquake, where the nonlinear model represents one of the main contributions of the paper. Findings from this model are presented in section 5. Section 6 addresses robustness of the empirical results, and it explains the link between the theoretical and the econometric models.[12]

2 A Theoretical Model of Social Interactions

The model is an application of the theory of conspicuous consumption in Hopkins and Kornienko (2004). Students choose how much costly effort e to exert, and effort increases achievement y. Students are heterogeneous in terms of how costly it is for them to exert effort.[13] They belong to a reference group (e.g. classroom, school). The empirical analysis finds support for and uses the classroom as the relevant reference group, but the theoretical results do not rely on a specific definition of a group. The main model assumptions are the following:

A.1 Students' utility is increasing in own achievement.

A.2 Students' utility is increasing in achievement rank in their reference group.

A.3 (optional). There are technological spill-overs in the production of achievement working through peer mean effort cost.

Assumption A.1 is standard.[14] Assumption A.2 is novel in the theoretical literature on educational peer effects. It generates peer effects even when spill-overs are not explicitly embedded in the achievement production function (A.3). That is, under A.2, own study

[11]Hoxby and Weingarth (2005) present direct evidence of the importance of proximity in determining student outcomes.

[12]There are four Appendices. Appendix A contains technical details on the preliminary data analysis and on the linear model of peer effects, including identification tests. Appendix B contains technical details of the nonlinear model, as well as details on robustness checks. Appendix C contains proofs and details relating to the theoretical model. Appendix D contains additional Tables and Figures.

[13]This corresponds to income heterogeneity in Hopkins and Kornienko (2004), where individuals choose how much of their income to spend on a consumption good and how much on a positional good. Here, achievement is at the same time a consumption good and a positional good, and it can be produced at a cost.

[14]For example, Blume, Brock, Durlauf, and Jayaraman (2014), Fruehwirth (2012), and De Giorgi and Pellizzari (2013) assume that students' utility is increasing in own achievement

effort depends on the distribution of peer characteristics, because it is determined by the equilibrium of a game of status. In the model, the relevant peer characteristic is how costly it is to produce achievement. For example, a student who finds it hard to exert study effort because she has low academic skills might "give up" in a classroom where all other classmates are highly skilled academically. On the other hand, the same student might instead engage in a healthy competition with her peers when they are similarly skilled academically. This intuition is explained in more detail in section 2.1. Assumption A.3 is standard and it gives rise to (exogenous) peer effects (Manski 1993).[15] It is optional in the sense that it is not necessary to prove the theoretical results. However, I allow for this additional kind of peer effect to emphasise that rank concerns generate a type of non-linear and heterogeneous peer effect that is easily distinguishable from the commonly assumed peer effects working through the mean of peer characteristics.

Students differ in terms of a type c: those with a higher c incur a larger cost of effort. This is without loss of generality. For example, students could instead be heterogeneous in terms of how productive their effort is, and, under minor modifications to the assumptions on the utility function, the model would have conceptually the same implications. Type c captures all student characteristics, physical and/or psychological, that affect her ability to study, such as her cognitive or academic skills, her emotional well-being, access to a computer or books, availability of an appropriate space for studying, etc. Type c is distributed in the reference group according to c.d.f. $G(\cdot)$ on $[\underline{c}, \bar{c}]$. Each student's type c is private information, but the distribution of c in the reference group is common knowledge. There are no distributional assumptions on $G(\cdot)$.

The cost of effort is determined by an increasing and strictly quasi-convex function in effort: $q(e; c)$. Higher types c incur higher costs for every level of effort e, i.e. $\frac{\partial q(e;c)}{\partial c} > 0$ for all e. For this reason, type c is informally referred to as a student's cost of effort.

[15] Several papers model technological spill-overs as operating through mean peer characteristics, e.g. Arnott and Rowse (1987), Epple and Romano (1998), Epple and Romano (2008). In Manski's terminology, exogenous peer effects arise when peer characteristics affect own outcomes (e.g. achievement). Rank concerns generate what Manski would call endogenous peer effects, that is, a response of own outcome to peer outcomes, which occurs when individuals in a strategic setting best-respond to each other. This gives rise to a simultaneity problem known as the reflection problem. The goal of this paper is not to identify the shape of the best response function (i.e., to solve the reflection problem). Rather, it is, first, to find theoretically the shape of composite peer effects under rank concerns, and, second, to test whether this shape is borne out in the data. Composite peer effects describe how own outcome changes as a function of peer characteristics, and they combine exogenous and endogenous peer effects. Until recent advancements in network econometrics and data (Bramoullé, Djebbari, and Fortin 2009, De Giorgi, Pellizzari, and Redaelli 2010, Blume, Brock, Durlauf, and Jayaraman 2014), the goal of most empirical peer effect papers has been to identify composite peer effects (by solving the problem of correlated unobservables, an important focus of the empirical part of this paper).

Moreover, at higher types the marginal cost of effort is (weakly) higher: $\frac{\partial^2 q(e;c)}{\partial c \partial e} \geq 0$.

Effort increases achievement according to the production function $y(e) = a(\mu)e + u(\mu)$, with $a(\mu) > 0$, where μ is the mean of c among peers. Parameters $a(\mu)$ and $u(\mu)$ capture technological spill-overs working through the mean of peer ability (assumption A.3). For example, teacher productivity may depend on student's average skills, or more academically skilled students (lower μ) may ask relevant questions in class that facilitate their peers' learning. These technological spill-overs are allowed to affect both the level of achievement ($u(\mu)$) and the productivity of effort ($a(\mu)$).

The utility function can be decomposed into two elements: a utility that depends only on own test score y (in absolute terms) and effort cost q, $V(y, q)$, embedding assumption A.1; and a utility that depends on rank in terms of achievement, embedding assumption A.2. The utility from achievement in absolute terms net of effort cost is non-negative, increasing and linear in achievement, decreasing and linear in q, and it admits an interaction between utility from achievement and cost of effort such that at higher costs, the marginal utility from achievement is (weakly) lower ($V_{12} \leq 0$).[16] No specific functional form assumptions are made on $q(\cdot)$ and on the interaction between y and q, therefore, results from the model are valid under a broad class of preferences. For example, students with lower effort cost c may (or may not) have higher marginal utilities from achievement.

A student's classroom rank in terms of achievement is given by the c.d.f. of achievement computed at her own achievement level, $F_Y(y)$. This is the fraction of students with achievement lower than one's own, and it is a standard way to model rank in theoretical models of status seeking (Frank 1985). Because achievement is an increasing deterministic function of effort, rank in achievement is equal to rank in effort: $F_Y(y(e)) = F_E(e)$, where $F_E(\cdot)$ is the c.d.f. of effort. The utility from rank, $S(F_Y(y(e)))$, is given by $F_E(e) + \phi$, where ϕ is a positive constant. Overall utility $U(y, q; c)$ is the product of utility from achievement in absolute and in relative terms: $V(y, q; c)\,(F_E(e) + \phi)$.

Each student chooses effort to maximize overall utility. Focusing on symmetric Nash equilibria in pure strategies, and initially assuming that the equilibrium strategy $e(c)$ is strictly decreasing and differentiable with inverse function $c(e)$, rank in equilibrium can be rewritten as $1 - G(c(e_i))$, and i's utility as $V(y(e_i), q(e_i, c_i))(1 - G(c(e_i)))$.[17] The

[16] All results are valid under an alternative set of assumptions for the utility from achievement. These are: strictly quasi-concave utility of achievement, decreasing and linear utility from cost of effort ($V_2 < 0, V_{22} = 0$) with a linear cost function ($\frac{d^2 q}{d^2 e} = 0$) and additive separability between utility from achievement and cost of effort ($V_{12} = 0$).

[17] The probability that a student i of type c_i with effort choice $e_i = e(c_i)$ chooses a higher effort than another arbitrarily chosen individual j is $F(e_i) = Pr(e_i > e(c_j)) = Pr(e^{-1}(e_i) < c_j) = Pr(c(e_i) < c_j) = 1 - G(c(e_i))$, where $c(\cdot) = e^{-1}(\cdot)$. The function c maps e_i into the type c_i that chooses effort e_i under the

first-order condition then is:

$$\underbrace{V_1 \overbrace{a(\mu)}^{\text{Mg. increase in achiev.}}}_{\text{mg. ut. from increased achiev.}} + \underbrace{\frac{V(y,q)}{1-G(c(e_i))+\phi} \overbrace{g(c(e_i))(-c^{'}(e_i))}^{\text{Mg. increase in rank}}}_{\text{mg. ut. from increased rank}} = \underbrace{-V_2 \frac{\partial q}{\partial e}}_{\text{mg. cost}} \tag{1}$$

and it implies the first-order differential equation reported in equation 27 in Appendix C. The solution to this differential equation is a function $e(c)$ that is a symmetric equilibrium of the game. The assumptions on the utility function, on the cost of effort function and on the achievement production function guarantee that the results in Hopkins and Kornienko (2004) apply under appropriate proof adaptations.[18] In particular, while the differential equation does not have an explicit closed-form solution, existence and uniqueness of its solution and comparative statics results concerning the equilibrium strategies can be proved for any distribution function $G(c)$ twice continuously differentiable and with a strictly positive density on some interval $[\underline{c}, \bar{c}]$, with $\underline{c} \geq 0$. This means that it is possible to trace how the equilibrium distribution of achievement in the reference group changes as the distribution of peer characteristics changes, without the need to explicitly solve for the equilibrium effort function $e(c)$. That is, it is possible to derive the shape of the peer effects.

The first theoretical result is summarized in the following Proposition:

Proposition 2.1 (Adapted from Proposition 1 in Hopkins and Kornienko (2004)). *The unique solution to the differential equation* (27) *with the boundary condition* $e(\bar{c}) = e_{nr}(\bar{c})$, *where* e_{nr} *solves the first order condition in the absence of rank concerns* ($V_1 a(\mu)|_{e=e_{nr}} = -V_2 \frac{\partial q}{\partial e}|_{e=e_{nr}}$)*, is a unique symmetric Nash equilibrium of the game of status. Equilibrium effort* $e(c)$ *and equilibrium achievement* $y(c)$ *are both continuous and strictly decreasing in type* c.

Proof: see Appendix C.

The empirical analysis tests the monotonicity of the achievement function.[19]

equilibrium strategy. Strict monotonicity and differentiability of equilibrium $e(c)$ are initially assumed, and subsequently it is shown that equilibrium strategies must have these characteristics.

[18] One of the main differences with the model in Hopkins and Kornienko (2004) is that here equilibrium strategies $e(c)$ are decreasing in c, whereas there they are increasing. See the procurement auctions model in Hopkins and Kornienko (2007) for another example with decreasing strategies.

[19] Monotonicity rules out the case in which for large enough values of c students exert more effort, which would be akin to a backward-bending labor supply curve. For example, suppose that students have high disutility from very low values of achievement. Then, as the cost of effort increases, the "substitution" effect would induce individuals to exert less effort, but the "achievement effect" (like an income effect) would induce them to exert more effort to avoid very low values of achievement. The empirical test rejects such a scenario.

Now consider two reference groups, A and B, with two distributions of c, $G_A(c)$ and $G_B(c)$, that are such that they have the same mean, but G_B has larger dispersion than G_A in the Unimodal Likelihood Ratio sense ($G_A \succ_{ULR} G_B$), defined in Appendix C. This happens when, for example, G_B is a mean-preserving spread of G_A. In informal terms, one can show that the effect on achievement of moving from group A to group B is heterogenous across individuals, depending on a student's type c. For a formal statement of this comparative statics result see Proposition (C.1) in Appendix C. This result provides the main testable implication of the theoretical model, which concerns the shape of peer effects generated by rank concerns. It can informally be stated as follows:
Testable Comparative Statics: When the dispersion of c in the reference group increases (keeping the mean constant), middle-c students perform more poorly in terms of achievement and high-c students perform better, while low-c students may perform better or worse, depending on the relative strength of the preference for achievement rank in the utility function. These patterns are represented graphically in Figure 1.

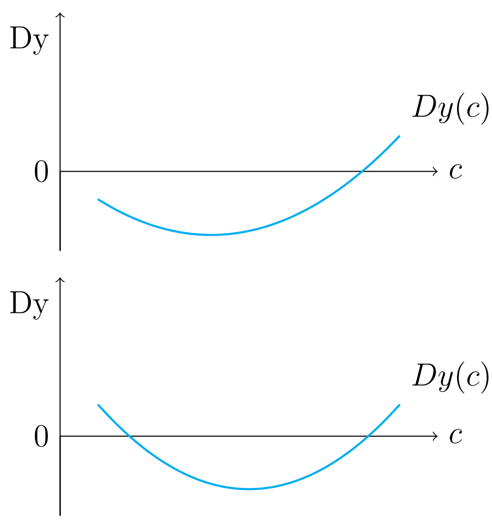

Figure 1: The function $Dy(c)$ traces the effect on achievement of increasing the dispersion of c, as a function of student type c. It can cross the x-axis once or twice. If it crosses it once (upper panel), the sequence of its signs, from low c to large c, is $-$, $+$. If it crosses it twice (lower panel), the sequence of its signs, from low c to large c, is $+$, $-$, $+$.

2.1 Model Intuition and Discussion of Assumptions

Rank concerns generate peer effects. When students have rank concerns, the entire distribution of effort cost in the reference group matters in determining the equilibrium distribution of achievement. For example, suppose that effort cost is entirely determined by academic skill. How much effort any student in the reference group exerts depends, among other things, on how much more skilled the person next up in the skill distribution is. If it is considerably more, she might "give up", knowing that it would be too costly to surpass this other student. If it is not, she might instead study hard in an effort to surpass him or her. This is true for all students in the reference group, therefore, the entire distribution of peer skills matters. The comparative statics focus on one feature of this distribution, its dispersion. First, it is convenient to focus on one feature of a distribution rather than working with an infinite dimensional object. Second, regression evidence indicates that dispersion is the empirically relevant feature (see section 4.1). Third, changes in dispersion of cost of effort affect different portions of the cost of effort distribution differently. As a result, they trigger complex nonlinear patterns of response in achievement that have empirical content, that is, that can be falsified by the data.

Intuition for the comparative statics results. Figure 2 shows two ability distributions $G(c)$ with different variances. As the variance increases moving from A to B, the density increases at the tails and decreases in the centre of the support. High-c students have an incentive to exert more effort in order to surpass the students with the same cost of effort as them, who are more numerous in group B. As can be seen from the first order condition in 1, the marginal utility from increasing one's own rank depends positively on the density at one's own type c, $g(c)$. In the discrete case, the key feature that corresponds to a change in density is a change in the distance between successive students.[20] High-c students in group B now also have a lower rank in c, because they face a larger fraction of peers with higher effort cost $(1 - G(c))$, and this decreases their incentive to exert effort: $1 - G(c)$ enters at the denominator of the marginal utility of effort in 1. This is a consequence of the multiplicative nature of preferences, discussed in the next section, which generates an aversion to a low achievement rank: the lower the rank in c is, the less desperate a student is to avoid a low achievement rank. The compar-

[20] For example, if a student is of cost type $c_1 = 10$ and the next more skilled student is of cost type $c_2 = 2$, the type 10 student might "give up" because surpassing a type 2 could be too costly in terms of effort. However, if the type 2 student becomes a type 9, the type 10 student has an increased incentive to exert effort, because surpassing the next student becomes less costly. The comparative statics results in this paper are proven with a continuum of students. Recent theoretical work shows that, in this class of models, the equilibrium strategy with a discrete number of players tends to the equilibrium strategy with a continuum of agents as the number of agents goes to infinity (Bhaskar and Hopkins 2013).

ative statics results demonstrate that, for high-c students, the incentive to improve their achievement prevails when moving from A to B. Middle-c students face two incentives: they face higher competition from the higher-c students, who are exerting more effort, but they also have an incentive to reduce their effort because of the lower density in their portion of the cost of effort distribution. The model predicts that the incentive to reduce effort prevails. Finally, also low-c students face two opposite incentives. First, the fatter tail at their end of the distribution gives them an incentive to increase their effort because of the larger density at their c level, and because they have a higher rank in c. Second, the weaker competition they face from middle-c students means that they can reduce their effort without reducing their rank. Which effect prevails depends on how strong the preference for rank is, relative to the utility from achievement in absolute terms net of the effort cost. This is determined by $V(\cdot)$ and $q(\cdot)$, on which the model does not make specific functional form assumptions. Intuitively, a stronger preference for rank leads to an increase in outcomes for low-c students, because the incentive to improve rank prevails over the incentive to pay a lower effort cost.[21]

Discussion of assumptions. The assumption that overall utility is multiplicative in utility in absolute and in relative terms may appear counterintuitive. However, it makes the problem's structure similar to that of a first-price sealed-bid auction, where expected payoff is the product of the value of winning (V) and the probability of winning (F). As noted in Hopkins and Kornienko (2004), "it is this formal resemblance to an auction that permits clear comparative statics results." While it would be desirable to analyse more general preference specifications, such a purely theoretical contribution would go beyond the scope of this paper. This paper is the first to apply theoretical tools at the frontier of the rank concerns theoretical literature to the field of peer effects in education, and to test them empirically in this context. Future extensions could specify different preferences and resort to numerical rather than analytical model solution tools.

Rank concerns can be modeled in many ways, and the comparative statics results depend on the assumed preference structure. For example, under the preference structure in this paper, both rank in and density of effort cost enter the first order condition of the student's problem. If, instead, utility were separable in rank and in achievement, and if rank utility were linear, then rank in effort cost would not be expected to affect behaviour.[22] Therefore, it is reassuring that the multiplicative preferences assumed in

[21]For example, it can be shown that if overall utility was $(y(e) - e)^{\alpha}\,(F(e) + \phi)$, then low-$c$ students would increase their effort when rank has a larger weight than achievement net of effort cost ($\alpha < 1$) and decrease it otherwise ($\alpha \geq 1$).

[22]If rank utility were convex in rank, those with higher ability rank would be more motivated to exert

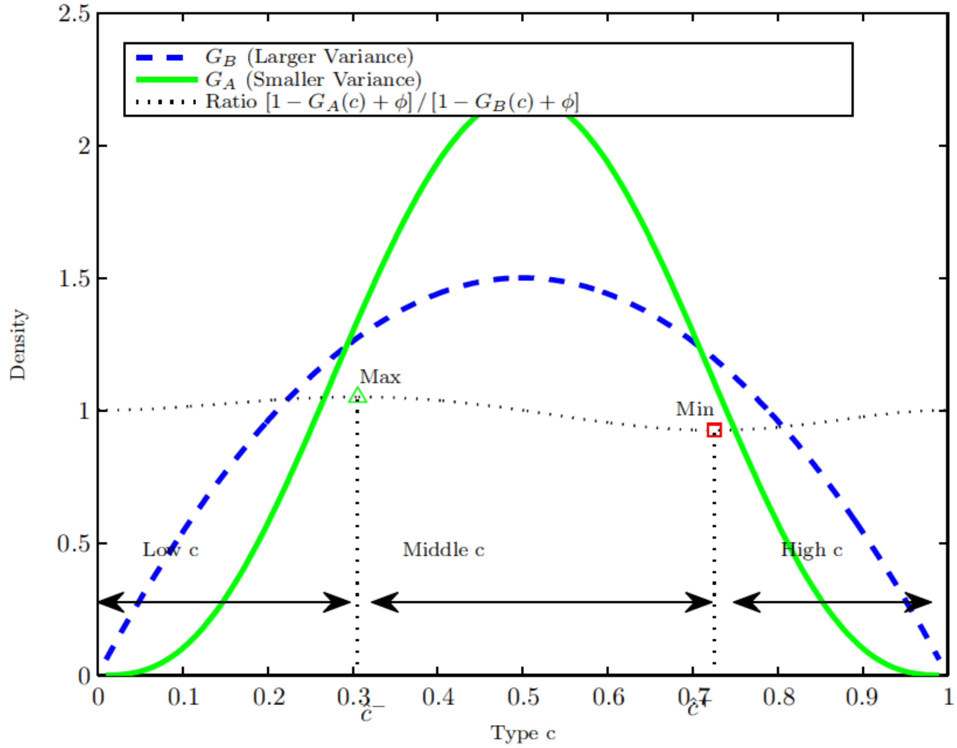

Figure 2: Type distributions in two reference groups, and cutoffs separating low-, middle- and high-c students.

this paper can explain important and distinct pieces of evidence. For example, in Tincani (2014) I show that a similar version of this model with a stronger aversion to a low rank (for laboratory evidence of this aversion see, for example, Kuziemko, Buell, Reich, and Norton (2014)) can explain all of the (unexpected) results of the randomised peer regrouping experiment in Carrell, Sacerdote, and West (2013), as well as the more recent experimental evidence in Booij, Leuven, and Oosterbeek (2016). Hence, the preference structure presented here does not seem inappropriate to describe rank concerns in various educational settings.

Finally, the model does not examine selection into a reference group. The empirical study of endogenous group formation is hard with standard available datasets. Moreover, understanding group formation is beyond the scope of this paper. Given allocations to groups, this paper's innovation is to study how outcomes are expected to vary when the

effort in order to reach the top of the achievement distribution. Conversely, if rank utility embedded a penalty for a very low achievement rank, those at the bottom of the ability distribution would exert more effort to improve their rank. Under nonlinear rank utility, both proximity between students and rank in terms of cost of effort would determine behaviour. These are the same features that matter under the nonseparable utility structure presented here.

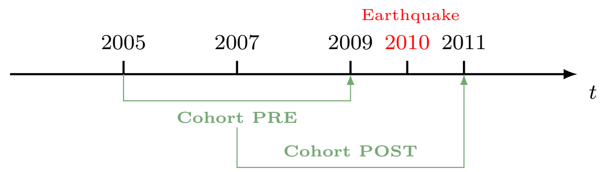

Figure 3: Data time-line.

features of the group vary exogenously, if there are rank concerns. The empirical part of the study uses a natural experiment to obtain the exogenous data variation needed to test the theory.

3 Data, Background and Preliminary Data Analysis

3.1 Data and Earthquake

Data. I use two cohorts of students from the SIMCE dataset (*Sistema de Medición de la Calidad de la Educación*), for a total sample size of $385,294$ students in $15,202$ classrooms. For both cohorts I observe administrative records on 8^{th} grade students' Math and Spanish standardized test scores, father's and mother's education, household income, gender, town of residence, and lagged (4^{th} grade) Math and Spanish test scores. Classroom level information includes class size and the experience, education, tenure at the school, gender, and type of contract (permanent or probationary) of both Spanish and Math teachers. School level information includes rurality and public or private status.

I refer to the two cohorts as pre- and post-earthquake cohorts. One cohort is observed in the 8^{th} grade in 2009, before the 2010 earthquake, while the other cohort is observed in the 8^{th} grade in 2011, after the earthquake, as shown in Figure 3. For both cohorts, identifiers are available at the student, teacher, classroom and school level, allowing me to identify each student's classmates, Math and Spanish teachers, and schools. This makes the dataset ideal to study spillovers down to the classroom level.

Earthquake. Just a few days before the start of the new school year, on February 27^{th} 2010, at 3.34 am local time, Chile was struck by a magnitude 8.8 earthquake, the fifth-largest ever instrumentally recorded and technically referred to as a mega-earthquake (Astroza, Ruiz, and Astroza 2012). Shaking was felt strongly throughout 500 km along the country, covering six regions that together make up about 80 percent of the country's population. While the death toll, as tragic as it was, was limited for such a strong earthquake (525 deaths), damage was widespread; $370,000$ housing units were damaged or destroyed. The Government implemented a national reconstruction plan to rebuild or repair $220,000$ units of low- and middle-income housing. Estimated total costs are around \$2.5 billion. The mega-earthquake had a continued impact on people's lives during the period covered by my sample. By the time the 2011 SIMCE sample was collected, i.e., 20-22 months after the earthquake struck, despite impressive efforts by the Government, only 24 percent of home reconstructions and repairs had been completed (Comerio 2013). This led to frustration in the population, as shown in Figure 10 in Appendix D.

Measure of earthquake intensity. I construct a measure of the intensity of shaking in each town in the sample using the Medvedev Sponheuer Karnik (MSK) scale. An advantage of this scale is that it can be mapped into a tangible measure of disruption: the expected level of damage to buildings of each earthquake resistance type in each town. Because reconstruction expenses were covered by the Government, this measure of damage reflects disruptions rather than shocks to household expenses.

For a given intensity of shaking, the level of damage depends on the construction type. For example, unreinforced masonries are less resistant than reinforced masonries, therefore, the same value of MSK-Intensity corresponds to larger damages in unreinforced masonries than in reinforced ones. The type of construction of students' homes is unobserved in my dataset, therefore, naive estimates of the impact of damage on achievement that use MSK-Intensity to measure damage without accounting for building type would be subject to measurement error. To account for unobserved construction type, the main empirical analysis restricts the sample to municipal (public) school students, who live in homes with similar earthquake resistance.[23] This sample restriction should eliminate any potential measurement error deriving from the unobservability of students' home types. Non-random location choices of parents are a remaining but separate concern that the empirical model is specifically designed to address.

To construct MSK-Intensity, I apply the intensity attenuation formula for the Chilean 2010 earthquake (Astroza, Ruiz, and Astroza 2012), which is a function of a town's distance from the earthquake's asperity. Using the geographical coordinates of each town and

[23] Astroza, Ruiz and Astroza (2012) report that around 60% of the poorest Chileans live in one of two house types with very similar earthquake resistance: old traditional adobe constructions (6.1%) and unreinforced masonry houses (51.9%). Given the striking school stratification in Chile, public school students belong to the poorest 50% of Chilean households. Therefore, it is reasonable to expect that all public school students live in one of these two building types. Notice that, as a measure of damages, MSK-Intensity may still contain residual noise because it averages damages within a town and, therefore, it may be vulnerable to classical measurement error inducing attenuation bias. This is inconsequential for the purpose of showing the existence of an effect of MSK-Intensity on achievement, because if an effect is found, one can conservatively conclude that an effect is present. Finally, MSK-Intensity is not informative on damages caused by the tsunami that afflicted some coastal areas after the earthquake.

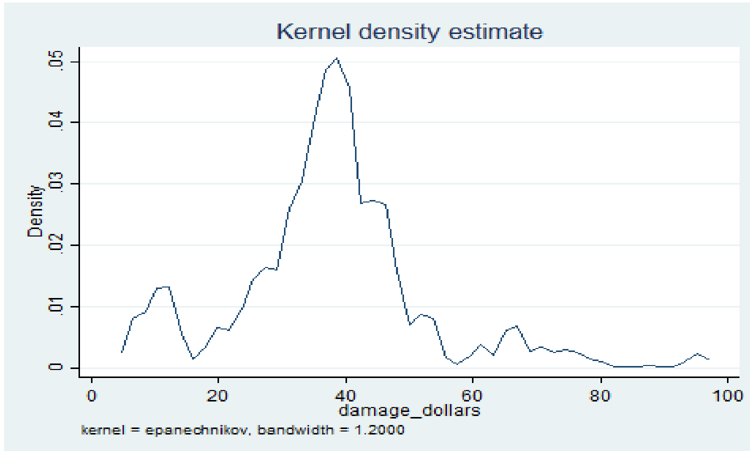

Figure 4: *Source*: SIMCE dataset and author's calculations. The right tail is truncated at USD 100 ($\sim 85^{th}$ percentile of the untruncated distribution) for ease of exposition.

of the asperity, I compute MSK-Intensity I according to: $I = 19.781 - 5.927\log_{10}(\Delta_A) + 0.00087\Delta_A$ ($R^2 = 0.9894$), where Δ_A is the distance from the main asperity. The formula is valid only at the town level and only for towns in the six regions affected by the earthquake (Astroza, Ruiz, and Astroza 2012). The Online Supplementary Material on the author's webpage contains additional technical details on this measure. There are two advantages to using this measure of shaking intensity as opposed to simple distance from the asperity of the earthquake. First, shaking intensity is a non-linear function of distance, therefore, using distance would introduce a non-classical measurement error. Second, the MSK-Intensity measure, coupled with other formulae borrowed from the structural engineering literature, allows me to express shaking in terms of the dollar amount of damage extent, which has intuitive meaning.

Distribution of earthquake damages. Figure 4 shows the distribution of earthquake intensities among the students in my sample, which, for illustrative purposes, are expressed in terms of reconstruction expenses in US dollars.[24] Intensities in the towns of the schools are also available and used in the analysis. On average, damages to homes are large, USD 170, or 24 percent of average household monthly income. The damage

[24] In the Online Supplementary Material I show the calculations and assumptions needed to map MSK-Intensity into dollar amounts. For example, one needs to make assumptions on the local reconstruction costs. The empirical models use the raw MSK-Intensity measure to limit the effect of these assumptions. However, I also express estimated effects in dollar amounts to give an idea of the magnitude of the effects. Notice that, unlike the main empirical analysis, Figure 4 includes also students from private schools, who live in different building types than public school students. Therefore, in this graph the translation of MSK-Intensity into dollar amounts is used only for illustrative purposes and must be interpreted with caution. The distribution of damages among Municipal school students alone is very similar.

distribution is right-skewed, with a median of USD 39, 6 percent of income, and a 90^{th} percentile of USD 303, 43 percent of income.

$3,822$ out of the $5,574$ classrooms in earthquake regions have a geographically dispersed student body. In those classrooms, not all classmates reside in the same town, and this generates variation in the measure of MSK-Intensity within the classroom. I use this variation to identify the effect of damage dispersion in the classroom on achievement. Figure 5 shows three examples of classrooms with students who do not all reside in the same town. Because of differences in soil type across towns, even classmates who live close to each other suffered very different damages. For example, the bottom panel of the Figure shows that students of the *La Florida* school who live 5.2 km apart from each other suffered a damage difference of USD 272, or 39 percent of average income. Large differences among neighbouring towns are not unusual, especially in areas closer to the asperity.[25] Among classrooms with a geographically dispersed student body, the within classroom standard deviation in damages is, on average, USD 79. Figure D in Appendix D shows the location of these classrooms on a map.

No earthquake-induced relocations. Finally, I obtained from the Ministry of Education the list of the schools that closed as a consequence of the earthquake, as well as the list of students at those schools.[26] I observe in what schools the displaced students enroll, and drop both the collapsed and receiving schools from the sample, for a total of 803 dropped schools, corresponding to 12 percent of the sample. This ensures that in my sample there are no earthquake-induced relocations of students across schools.[27] Such relocations could have both direct impacts on the relocated students, and spill-over impacts on the incumbent students in receiving schools. These effects would confound estimation of the effects of interest. In the preliminary data analysis, the effect of interest is the impact of own earthquake exposure on own achievement, controlling for classroom composition. In the main analysis, the effect of interest is the spill-over impact of the mean and dispersion of earthquake intensity among one's classmates on own achievement, controlling for classroom composition.

[25]Damage dispersion effects are calculated controlling for damage in the school town and average damages in the classroom, to account for correlation between these two variables and dispersion in damages. I also show results using the coefficient of variation, for the same purpose.

[26]They closed either because the buildings became unsafe, or because most of the students' homes were so badly damaged, that students had to relocate, reducing attendance below the minimum acceptable for a school to operate.

[27]Imberman, Kugler, and Sacerdote (2012) use the influx of Katrina evacuees in a school as an exogenous source of change to classroom composition. In Chile evacuees were spread across such a large number of schools that the influxes in each school are too small to detect any statistically significant impact.

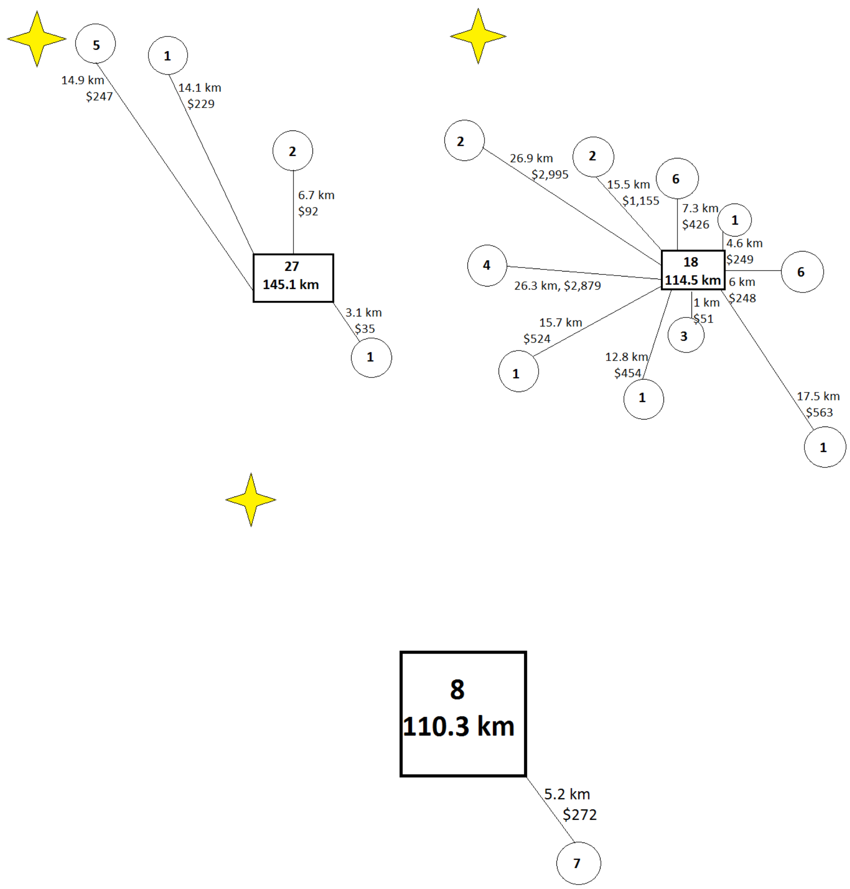

Figure 5: Examples of three schools where not all students are residents in the school town. At the top left is Colegio Santa Ines in the town of San Vicente, at the top right is Liceo Maria Auxiliadora in Santa Cruz, and at the bottom is Escuela La Florida in Talca. The squares represents the school town location relative to the earthquake asperity (star). In the square are the number of classmates residing in the town, and the distance from the asperity. The circles represent towns of residence of x classmates, where x is the number in the circle. The lines indicate the distance to school of each town, and the difference in damages suffered by an unreinforced masonry construction, the vastly most common type.

3.2 Balancing Tests and Preliminary Data Analysis

As a preliminary data analysis I estimate the impact of own exposure to the earthquake on own achievement, using both binary and continuous measures of exposure. To do so, I exploit: *i.* the longitudinal nature of the dataset (test scores are observed twice for each student), *ii.* the fact that different geographic locations were affected differently by the earthquake, and *iii.* the existence of two cohorts (one affected by the earthquake in the second time period, one never affected).

Table 1 presents sample descriptive statistics and balancing tests for four relevant subsamples in the data: the pre- and post- earthquake cohorts of students in regions affected

Table 1: Descriptive statistics and balancing tests

	Pre-earthquake Cohort		**Earthquake Regions**	
	(1) NER	**(2) Difference (ER-NER)**	**(3) Pre**	**(4) Difference (Post-Pre)**
Lagged Math Score	0.128	0.082***	0.210	-0.013***
	(0.922)		(0.948)	
Lagged Spanish Score	0.127	0.086***	0.213	-0.024***
	(0.912)		(0.941)	
Father's Education (years)	11.19	0.23***	11.43	-0.166***
	(3.76)		(3.890)	
Mother's Education (years)	10.90	0.34***	11.24	-0.105***
	(3.56)		(3.60)	
Monthly Household Income (USD)	637	48***	685	17***
	(713)		(834)	
Math Teachers				
% Female	0.582	0.02	0.590	-0.012
	(0.493)		(0.492)	
% Postgraduate Degree	0.433	0.03**	0.464	0.047***
	(0.496)		(0.499)	
Teaching Experience (years)	19.09	-0.36	18.72	-0.820**
	(12.89)		(12.57)	
Tenure at school (years)	9.92	0.510*	10.42	-0.253
	(9.73)		(10.16)	
Spanish Teachers				
% Female	0.803	0.03**	0.813	0.006
	(0.398)		(0.390)	
% Postgraduate Degree	0.414	0.06***	0.471	0.037***
	(0.492)		(0.499)	
Teaching Experience (years)	18.35	-0.461	18.08	-0.546**
	(12.75)		(12.66)	
Tenure at school (years)	9.53	0.85	10.46	-0.316
	(9.74)		(10.04)	
Sample Size	**Non-Earthquake Regions**		**Earthquake Regions**	
	Pre	**Post**	**Pre**	**Post**
Students	45,715	46,424	147,097	146,058
Classrooms	1,954	1,967	5,707	5,574

*** $p < 0.001$, ** $p < 0.05$, * $p < 0.10$ for a two-sided test of difference in means. Standard deviation in parenthesis. ER: Earthquake Region. NER: Non-Earthquake Region. Pre: Pre-earthquake cohort. Post: Post-Earthquake Cohort. Income measured in 8^{th} grade, therefore, post-earthquake income in ER includes earthquake subsidies.

and not affected by the earthquake. As is evident in column 1, students in earthquake regions (ER) have significantly higher lagged test scores and SES than students in regions not affected by the earthquake (NER, non-earthquake regions). The column shows pre-earthquake cohort data, but similar patterns across regions are present in the post-earthquake cohort. This unbalance implies that a simple DD model that compares test score growth of post-earthquake students in earthquake regions to that of post-earthquake students in non-earthquake regions may fail to identify the impact of being exposed to the earthquake if more highly skilled and higher SES students experience different trends in (unobserved) temporary shocks to test scores. To test the identifying assumption of such a DD model, I estimate the model on the pre-earthquake cohort sample of students. Because these students never experience treatment, a significant earthquake "effect" is an indication that the identifying assumption is violated. As shown in Appendix A.1.2, I find evidence of violation, with a positive and significant "effect". This could be due to non-random location choices of parents, that is, the choice of living in an earthquake or non-earthquake region is correlated with unobserved household characteristics affecting individual student trends in test scores.

DDD model to estimate effect of earthquake exposure (binary treatment). To overcome this, I estimate the effect of being exposed to the earthquake using a difference-in-difference-in-differences (DDD) approach that allows unobserved trends in temporary shocks to differ across regions. This is in the spirit of the differential-trend-adjusted estimator proposed by Bell, Blundell, and Van Reenen (1999) (see also Blundell and Costa Dias (2000) for a precise description of this estimator). Details of this model are in Appendix A.1.2. Intuitively, the "effect" of future treatment on test score growth in the pre-earthquake cohort is used to net out biases from the post-earthquake cohort effect. The identifying assumption is that the difference in (unobserved) trends across region types (ER and NER) must be constant across cohorts. The model allows also for cohort effects, which can either directly affect test scores, or affect parameters of the test score production function. Estimates from the DDD model are reported in Table 2 and they indicate that being exposed to the earthquake (binary variable) had a negative and statistically significant impact on test scores ranging from -0.02 to -0.04 standard deviations (see the coefficient on $T_r \times E_i \cdot d_g$). The first row of the Table shows the correlation between the earthquake region dummy and test score growth due to non-random location choices, while the second row shows cohort effects on test scores.[28]

[28] Estimates for public (municipal) schools are reported in Table 13 in Appendix D.

Table 2: Effect of being exposed to the earthquake (binary variable) on Math and Spanish test scores (TS). DDD model.

	(1) Math TS	(2) Math TS	(3) Spanish TS	(4) Spanish TS
$E_i \cdot d_g$	0.153***	0.067***	0.101***	0.022***
	(0.005)	(0.005)	(0.005)	(0.005)
T_i	0.024***	0.007	0.038***	0.025***
	(0.007)	(0.006)	(0.007)	(0.006)
$T_i \times E_i \cdot d_g$	-0.024**	-0.027***	-0.040***	-0.027***
	(0.007)	(0.007)	(0.008)	(0.007)
Lagged TS		0.689***		0.676***
		(0.002)		(0.002)
Controls	No	Yes	No	Yes
Observations	365,328	206,666	365,239	202,365

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Controls: whether the student lives in the same town where the school is, gender, mother's education, father's education, household income, class size, whether the Math or Spanish teacher is female, has a postgraduate degree, has a permanent contract, her tenure at the school, her teaching experience, and whether the school is public. A constant is always included.

DD model to estimate effect of earthquake exposure (continuous treatment). A limitation of this DDD model is that the identifying assumption of a constant difference in region effects across cohorts cannot be tested. To overcome this limitation I exploit an additional margin of variation: variation in earthquake intensity. Specifically, I estimate the impact of earthquake intensity using a DD model whose identifying assumption is testable. First, I restrict attention to earthquake regions, where variation in earthquake intensity is observed (Astroza, Ruiz, and Astroza (2012) caution that the MSK-Intensity measure is valid only in these regions). Second, I estimate a test score value added DD model with continuous treatment, similar to the model in Card (1992), on post-earthquake cohort data. Third, I use pre-earthquake cohort data to perform a placebo test of the identifying assumption. Results from the test are reported in the first row, columns 3 and 4 of Table 3. Conditional on student, teacher and school characteristics, the measure of intensity of the (future) earthquake is not correlated with individual unobserved trends in test scores in the absence of treatment, satisfying the identifying assumption of this DD model with continuous treatment. Formal details of the model and of the placebo test can be found in Appendix A.1.3.

The first row in columns 1 and 2 in Table 3 show that the point estimate of the impact of earthquake intensity on test scores is around -0.040 in both Mathematics and Spanish, corresponding to a decrease in 0.02 standard deviations for every additional USD 100 in damages. The estimate is significant only for Mathematics.[29] Estimation is performed only on municipal schools to minimise measurement error due to house type unobservability, as explained in section 3.1.

Table 3: Effect of earthquake intensity on Math and Spanish test scores (TS) in Municipal Schools.

	Post-earthquake data		**Pre-earthquake data (placebo test)**	
	(1)	(2)	(3)	(4)
	Math TS	Spanish TS	Math TS	Spanish TS
$I_i \cdot d_g$	-0.049*	-0.037	-0.083	-0.021
	(0.028)	(0.030)	(0.051)	(0.054)
Lagged TS	0.658***	0.671***	0.672***	0.699***
	(0.004)	(0.004)	(0.004)	(0.005)
Controls	Yes	Yes	Yes	Yes
Observations	32,519	31,072	26,146	24,619

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Controls: whether the student lives in the same town where the school is, gender, mother's education, father's education, household income, class size, whether the Math or Spanish teacher is female, has a postgraduate degree, has a permanent contract, her tenure at the school, her teaching experience, average MSK-Intensity among classmates, MSK-Intensity in the school town, and whether the school is rural.

Direct effect of own earthquake exposure on own achievement. The DDD model with binary treatment and the DD model with continuous treatment control for teacher and school characteristics, suggesting that the estimated effects are, at least in part, due to a direct impact of the earthquake on a student's individual determinants of achievement rather than being entirely mediated by teachers and schools. While evidence of a direct impact is difficult to obtain with available data (because of potentially unobserved teacher and classroom characteristics), there are a number of reasons why it may occur. First, there is established evidence in the medical literature that earthquake exposure

[29] In fact, estimation of a DDD model that nets out pre-earthquake cohort trends from post-earthquake effects yields a statistically significant and negative earthquake intensity effect in Spanish (-0.120). Full regression Tables for the DDD model are available upon request.

affects brain function and that it can cause Post Traumatic Stress Disorder (PTSD), and this may last for several months after the earthquake.[30] Moreover, the severity of PTSD has been found to increase with seismic intensity (Groome and Soureti 2004), and not just with a binary exposure measure. Second, the disruption to the home environment may increase the opportunity cost of time, because students may need to spend their free time helping their parents with home repairs.[31] Additionally, students may not have access anymore to the areas of the home that they used for doing their homework. The findings in this paper are independent of the channel through which earthquake exposure directly affects a student's ability to produce achievement, but rely on the existence of such an effect. Importantly, this effect is not imposed in the main nonlinear analysis, rather, it is identified. Moreover, there is no restriction that this be the only channel through which the earthquake affects achievement. Sections 6.1 and 6.2 discuss robustness of the main findings to school and teacher effects, while section 6.3 presents further evidence in favour of a direct effect on a student's individual ability to produce achievement.

4 Main Empirical Analysis: Earthquake Spill-overs

The preliminary analysis suggests that own earthquake exposure is bad for own achievement. The main empirical analysis examines if there are spill-overs: keeping classroom composition constant, does *peer* earthquake exposure affect *own* achievement? Because students in the same classroom are either all affected if the classroom is in an earthquake region, or all not affected otherwise, the necessary data variation must come from variation in classmates' intensity of exposure in earthquake regions ("peers" and "classmates" are used interchangeably). I study the effect of two peer variables: mean and (various measures of) dispersion of intensity of exposure in the classroom. First, I use a linear model to examine average spill-over effects. Second, guided by the theoretical model, I build a nonlinear empirical model that allows me to detect effect heterogeneity across students and to test if the theoretical predictions are borne out in the data. The nonlinear analysis is one of the main contributions of the paper.

[30] See, for example, Altindag, Ozen, et al. (2005), Lui, Huang, Chen, Tang, Zhang, Li, Li, Kuang, Chan, Mechelli, et al. (2009), Giannopoulou, Strouthos, Smith, Dikaiakou, Galanopoulou, and Yule (2006).

[31] This is particularly likely to have occurred among the low-income Chilean families that my sample focuses on, because most of the government subsidies were in the form of vouchers for purchasing the materials needed for the repairs, and families were expected to perform the repairs themselves (Comerio 2013). In the main semi-parametric empirical model, I allow for heterogeneous impacts of exposure depending on household characteristics.

4.1 Estimating Average Spill-over Effects with a Linear Model

The mean and the dispersion in peer intensity of exposure are continuous treatment variables. Therefore, a model in the spirit of the regression DD with continuous treatment in Card (1992) could be adopted again.[32] However, a placebo test reveals that the identifying assumption for the DD model of spill-overs is not satisfied in my data. Specifically, using location information, I build the classroom mean and dispersion in (future) MSK-Intensity in the pre-earthquake cohort of students in earthquake regions. Because the earthquake never affects this cohort of students, any "effect" of these variables would capture a correlation between the treatment variables and unobserved trends in test scores between 4^{th} and 8^{th} grade. As shown in Table 7 in Appendix A.2, the coefficient on mean MSK-Intensity is small and insignificant, but the coefficient on its dispersion is positive and statistically significant. Therefore, there are confounding effects associated with the dispersion in MSK-Intensity treatment variable.

This is not surprising: classrooms with higher dispersion in MSK-Intensity, mechanically, have a more geographically dispersed student body. In my data, geographic dispersion is positively correlated with unobserved classroom characteristics that increase test scores. For example, classrooms that attract students from further away could do so because they are of better unobserved quality, or the student body in those classroom could itself be of better unobserved quality. Regardless of its cause, this correlation implies that the treatment variable "dispersion in earthquake exposure among one's peers" violates the assumptions of a standard continuous treatment DD model.

To address this, I extend the model in Card (1992) to allow for correlation between the treatment variables and test score trends between grades. Appendix A.2 formally introduces the model and a test of its identifying assumption. Here, I explain its intuition. I set up a continuous treatment regression DDD model that uses the pre-earthquake cohort of students to net out any potential biases due to a correlation between treatment and unobserved test score trends. The main difference with a discrete DDD model is that the second difference is a derivative due to the continuous nature of treatment. The model allows also for cohort effects, accounting for any nation-wide policies which may have affected the two cohorts differently, such as the 2008 voucher reform (Neilson 2013).

The identifying assumption is that any correlation, in the absence of treatment, between the treatment variables (MSK-Intensity mean and dispersion) and trends in unobserved

[32] This would amount to estimating equation (17) in Appendix A.2 on the post-earthquake cohort data, comparing test score growth over time across classrooms with different mean and variance in earthquake intensity.

shocks between 4^{th} and 8^{th} grade is constant across cohorts. In practice, the placebo test indicates that the only treatment variable correlated with unobserved trends is dispersion. Therefore, to test the identifying assumption, I verify that the correlation between dispersion and unobserved trends is constant across cohorts in the absence of treatment. To do so, I use classroom geographic dispersion as the treatment variable, and I estimate a DDD model on the sample of non-earthquake regions, which are not used in estimation and are never subject to treatment. Any "effect" of geographic dispersion would indicate that the triple difference technique does not effectively net out bias. Reassuringly, no "effects" are found, therefore, the test does not reject the identifying assumption.

Table 4 reports estimation results from this DDD model for the (preferred) sample of Municipal schools, using four different measures of dispersion.[33] All specifications control for school and individual MSK-Intensity, and the even numbered columns include additional regressors. There are two main results. First, the first line in each panel shows that mean damages in the classroom almost never have a statistically significant impact on test scores. When the estimate of the coefficient on the triple interaction $(\bar{I}_r \cdot d_g)T_r$ is significant (mostly in specifications without the additional controls), it is very small. Second, the dispersion of peer damages has a significant and sizeable negative effect on test scores, regardless of what measure of dispersion is used, as can be seen from the third row of each panel. Unsurprisingly, the magnitude of this effect depends on the definition of dispersion, but it is always sizeable and negative. Notice that the second row of each panel shows the pre-existing positive correlation between geographic dispersion and unobserved test score trends, which would bias effect estimates in a simpler DD model. Measuring damage dispersion through the standard deviation of MSK-Intensity yields the most conservative estimates, with effect estimates ranging from -0.17 to -0.26. Assigning a dollar amount to damage dispersion in each classroom, effects range from -1.2 percent to -7.6 percent of a standard deviation of test scores when moving up one standard deviation in this distribution.

These findings are robust to controlling for various student composition variables that could be correlated with mean and dispersion of MSK-Intensity in the classroom like, for example, mean and dispersion in income. Moreover, classroom level damage dispersion matters even after controlling for school level damage dispersion, indicating that the

[33] Results for the sample of all schools are available upon request. In general, the coefficient signs are the same, while the magnitudes are lower in absolute values in the sample of all schools. The *caveat* in the sample of all schools is unobserved variation in the earthquake resistance of students' homes, introducing measurement error on the variables constructed from MSK-Intensity. For this reason, the sample of Municipal school students is preferred.

classroom is a relevant unit of analysis. Focusing on the classroom has the additional advantage that it allows me to examine the model's robustness to the inclusion of school fixed effects. See Appendix A.2.1 for these robustness checks.

4.2 A Novel Nonlinear Model to Estimate Spill-over Heterogeneity

The linear model of spill-overs has several limitations that make it unfit to test the theoretical predictions. To overcome them, I develop an econometric approach that combines semi-parametric difference-in-differences with matching. The model has three key features. First, it can detect heterogeneity of the dispersion effect across students. To obtain interpretable results, it aggregates each student's vector of characteristics - including earthquake intensity at her home - into a single scalar capturing all the student-level determinants of achievement. It then expresses effect heterogeneity with respect to this scalar. Second, this scalar provides an empirical counterpart to the theoretical model's cost of effort, which is not directly observed in the data (see section 6.3). This allows me to use earthquake shocks as shifters to the predicted cost of effort distribution, a variation needed for testing. Third, it can accommodate the unbalance in observed student and school characteristics between the untreated pre-earthquake cohort and the treated post-earthquake cohort, as shown in column 4 of Table 1.[34] The econometric model is related to the nonlinear difference-in-differences models in Athey and Imbens (2006), Abadie (2005), Heckman, Ichimura, Smith, and Todd (1998), and Blundell, Costa Dias, Meghir, and Van Reenen (2004). However, it is designed to address specific challenges of this context that those models are not well-suited to address, as discussed in more detail in section 6.4.

4.2.1 The Model

Damage dispersion as measured by variance in MSK-Intensity is a continuous variable. Existing non-linear difference-in-differences models cannot be used in this context because they do not accommodate continuous treatment.[35] Two notables exceptions are the mod-

[34]While the Table shows unbalance for students of all schools, similar unbalance across cohorts is found when restricting the sample to Municipal schools, which is the sample used in the non-linear analysis. This could reflect movements of students across public and private schools occurring in between the two cohorts, for example, as an effect of the voucher reform analysed in Neilson (2013). The model fully accounts for this unbalance using high-dimensional matching.

[35]For example, the changes-in-changes and quantile-difference-in-differences models in Athey and Imbens (2006) compare outcome distributions across multiple groups and time periods, however, treat-

Table 4: Linear model of spill-overs, sample of Municipal schools in earthquake regions.

	(1) Math TS	(2) Math TS	(3) Spanish TS	(4) Spanish TS
Dispersion measured by standard deviation				
$(\bar{I}_r \cdot d_g) \times T_i$	0.012^{+}	-0.004	0.010	-0.011
	(0.007)	(0.007)	(0.007)	(0.007)
$(DI_r \cdot d_g)$	0.366^{***}	0.180^{***}	0.468^{***}	0.211^{***}
	(0.037)	(0.042)	(0.039)	(0.045)
$(DI_r \cdot d_g) \times T_i$	-0.259^{***}	-0.138^{**}	-0.319^{***}	-0.174^{**}
	(0.047)	(0.049)	(0.050)	(0.053)
Dispersion measured by coefficient of variation				
$(\bar{I}_r \cdot d_g) \times T_i$	0.010	-0.005	0.008	-0.013^{+}
	(0.007)	(0.007)	(0.007)	(0.007)
$(DI_r \cdot d_g)$	1.776^{***}	0.187^{***}	2.354^{***}	1.339^{***}
	(0.207)	(0.255)	(0.219)	(0.272)
$(DI_r \cdot d_g) \times T_i$	-1.249^{***}	-0.906^{**}	-1.692^{***}	-1.119^{***}
	(0.259)	(0.296)	(0.273)	(0.321)
Dispersion measured by interquartile range				
$(\bar{I}_r \cdot d_g) \times T_i$	0.013^{+}	-0.006	0.011	-0.012^{+}
	(0.007)	(0.007)	(0.007)	(0.007)
$(DI_r \cdot d_g)$	1.132^{***}	0.388^{***}	1.120^{***}	0.373^{***}
	(0.069)	(0.072)	(0.073)	(0.077)
$(DI_r \cdot d_g) \times T_i$	-0.898^{***}	-0.383^{***}	-0.854^{***}	-0.102
	(0.098)	(0.097)	(0.104)	(0.104)
Dispersion measured by range between 90^{th} and 10^{th} percentiles				
$(\bar{I}_r \cdot d_g) \times T_i$	0.023^{***}	0.003	0.021^{***}	-0.003
	(0.006)	(0.006)	(0.006)	(0.006)
$(DI_r \cdot d_g)$	0.959^{***}	0.347^{***}	0.890^{***}	0.280^{***}
	(0.028)	(0.032)	(0.029)	(0.033)
$(DI_r \cdot d_g) \times T_i$	-0.874^{***}	-0.312^{***}	-0.820^{***}	-0.287^{***}
	(0.035)	(0.038)	(0.037)	(0.039)
Controls	No	Yes	No	Yes
Observations	110075	58661	110500	56687

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Controls: whether the student lives in the same town where the school is, mother's education, father's education, household income, intensity of earthquake in hometown and in school town, gender, lagged test score in Math or Spanish, class size, the teaching experience of the Math or Spanish teacher, whether he/she is female, has a postgraduate degree, has a permanent contract, her tenure at the school, average MSK-intensities among classmates, cohort dummy. A constant is always included.

els in Abadie (2005) and D'Haultfoeuille, Hoderlein, and Sasaki (2015) that, however, are not well-suited to accommodate other features of this context, as detailed in section 6.4. To deal with a continuous treatment variable in a non-linear DD model, my analysis starts from the observation that, to test the theoretical model, it is sufficient to estimate the effect of an increase of any amount in the treatment variable (damage dispersion). The continuous dispersion measure provides rich data variation for this purpose: it allows me to make all possible pair-wise comparisons between classrooms, and consider the classroom with the larger variance within the pair as the treated classroom. This approach avoids making arbitrary assumptions to categorise classrooms into "high-" or "low-variance" classrooms, it uses the data more efficiently than under such a categorisation, and it is fully coherent with the theoretical model, where the comparative statics results are obtained from the pair-wise comparison of a classroom with a relatively higher variance in cost of effort to one with a relatively lower variance. The large number of classrooms in the data afford me the opportunity to perform a very large number of such comparisons, and average the results across classroom pairs. This simple idea is embedded into a more sophisticated weighting scheme that accounts for potentially unbalanced confounding factors, as explained in more detail below.

A semi-parametric DD model. Define treatment as an increase in damage dispersion in the classroom. Consider a pair of classrooms, $\{r, r'\}$, with variances in damages equal to σ_H^2 and σ_L^2, with $\sigma_H^2 - \sigma_L^2 = \delta > 0$. Treatment is the additional damage dispersion δ. The observed outcome (achievement) of student i in classroom r is Y_{ir}, and let Y_{ir}^1 denote the potential outcome for student i if that student receives treatment δ, and Y_{ir}^0 if she does not receive the treatment. Damage dispersion is measured by the variance in MSK-Intensity in the post-earthquake cohort. In the pre-earthquake cohort, variance in MSK-Intensity reflects only the geographic dispersion of the student body, because the earthquake has not occurred yet. Let the indicator $G_{r,r'}$ be equal to 1 if classroom r has the higher MSK-Intensity variance within the pair (r, r'), i.e., $\sigma_r^2 = \sigma_H^2$, and 0 otherwise, and let $T_r = 1$ indicate the post-earthquake cohort (the treatment group) and $T_r = 0$ the pre-earthquake cohort (the control group). I assume that potential outcomes satisfy:

ment status is a binary variable. Similarly, Heckman, Ichimura, Smith, and Todd (1998) and Blundell, Costa Dias, Meghir, and Van Reenen (2004) develop semi-parametric difference-in-differences models based on propensity score matching, where the propensity score is based on a treatment dummy.

$$
\begin{aligned}
Y^0_{ir} &= h^0(c^{T_r}(X_i), W_r, T_r, \sigma^2_r) + \epsilon^0_{ir} \\
&= \phi(c_i, W_r) + \lambda^{GD}(c_i, W_r, \sigma^2_L) + T_r \cdot \left[\lambda^E(c_i, W_r) + \lambda^{DD}(c_i, W_r, \sigma^2_L)\right] + \\
&\quad + G_{r,r'} \cdot \lambda^{GD}(c_i, W_r, \delta) + \epsilon^0_{ir} \qquad (2)
\end{aligned}
$$

$$
\begin{aligned}
Y^1_{ir} &= h^1(c^{T_r}(X_i), W_r, T_r, \sigma^2_r) + \epsilon^1_{ir} \\
&= \phi(c_i, W_r) + \lambda^{GD}(c_i, W_r, \sigma^2_L) + T_r \cdot \left[\lambda^E(c_i, W_r) + \lambda^{DD}(c_i, W_r, \sigma^2_L)\right] + \\
&\quad + G_{r,r'} \cdot \lambda^{GD}(c_i, W_r, \delta) + \lambda^{DD}(c_i, W_r, \delta) + \epsilon^1_{ir} \qquad (3)
\end{aligned}
$$

where $\phi(\cdot), \lambda^{GD}(\cdot), \lambda^{DD}$ and $\lambda^E(\cdot)$ are semi-parametric functions. $\phi(c_i, W_r)$ captures the amount of achievement that is produced independently of the variance in MSK-Intensity, and net of any cohort effects. It transforms student type $c_i = c^{T_r}(X_i)$ (described in detail below) and classroom, school and teacher characteristics W_r into achievement Y_i.[36] Vector W_{ir} contains class-size, teacher characteristics including experience, location as measured by MSK-Intensity in the school town, and classroom student composition described by various moments of the distribution of student characteristics within the classroom.[37] This function allows for any kind of interaction between classroom, school, teacher, peer and student characteristics in the production of achievement.

$\lambda^E(c_i, W_r)$ are cohort effects, capturing any policy change across cohorts as well as the direct effect of the earthquake on achievement. Being modeled semi-parametrically, they are very flexible. For example, they allow the direct impact of the earthquake to be different for higher and lower SES students, and/or for more or less affluent schools. They also allow for policies like the 2008 voucher reform to have a different impact across different classrooms and students.

$\lambda^{GD}(c_i, W_r, \sigma^2_r)$ are geographic dispersion effects, and they are assumed to be linear in σ^2_r.[38] Estimates from the linear model indicate that classrooms with higher geographic dispersion have on average higher test scores. The function $\lambda^{GD}(\cdot)$ captures these effects

[36] With an abuse of notation, I do not add the r index to c_i. Type c_i depends on r only through cohort T_r.

[37] W includes: mean, variance, skewness and kurtosis of the classroom distribution of the elements of X (including mean, skewness and kurtosis - but not variance - of MSK-Intensity), MSK-Intensity in the school's town (to account for school damage post-earthquake and school location pre-earthquake), class size, and all observed teacher characteristics. X includes measures of student ability, gender and SES.

[38] This assumption allows me to substitute the sum $\lambda^{GD}(c_i, W_r, \sigma^2_L) + G_{r,r'}\lambda^{GD}(c_i, W_r, \delta)$ for $\lambda^{GD}(c_i, W_r, \sigma^2_r)$ in 2 and 3.

and allows them to vary arbitrarily with student and school type, generalising the linear model. The identifying assumption is similar to the linear model's one: conditional on classroom, teacher and school characteristics W_r and on student's type c_i, geographic dispersion effects do not depend on cohort T_r.[39]

$\lambda^{DD}(c_i, W_r, \sigma_r^2)$ are damage dispersion effects, also assumed to be linear in σ_r^2. In the post-earthquake cohort, this variance reflects not only geographic dispersion, but also the variance in damages. Function $\lambda^{DD}(c_i, W_r, \delta)$ traces how a δ increase in damage dispersion affects student achievement as a function of student and classroom characteristics.

The assumption of linearity in σ_r^2 of λ^{GD} and λ^{DD} does not restrict their arguments to interact arbitrarily. Any kind of heterogeneity of these dispersion effects along student and classroom characteristics is allowed. For example, damage dispersion effects may be stronger in classrooms with less experienced teachers or, as implied by the theoretical model, they may vary with student ability. Estimating the latter type of heterogeneity is the central focus of this non-linear analysis.

Let $\tau_{irr'}$ be an indicator for treatment. The realised (observed) outcome for individual i in classroom r is:

$$Y_{ir} = \tau_{rr'} Y_{ir}^1 + (1 - \tau_{rr'}) Y_{ir}^0.$$

We have that $\tau_{rr'} = T_r \times G_{r,r'}$, i.e., given a pair of classrooms r and r', a student is treated if she is in the classroom with the higher MSK-Intensity dispersion in the post-earthquake cohort, untreated otherwise. The realised outcome can be expressed as:

$$\begin{aligned} Y_{ir} &= h(c^{T_r}(X_i), W_r, T_r, \sigma_r^2) + \epsilon_{ir} \\ &= \phi(c_i, W_r) + \lambda^{GD}(c_i, W_r, \sigma_L^2) + T_r \cdot \left[\lambda^E(c_i, W_r) + \lambda^{DD}(c_i, W_r, \sigma_L^2)\right] + \\ &\quad + G_{r,r'} \cdot \lambda^{GD}(c_i, W_r, \delta) + T_r \cdot G_{rr'} \cdot \lambda^{DD}(c_i, W_r, \delta) + \epsilon_{ir} \qquad (4) \end{aligned}$$

where the error term is assumed to be mean-independent, i.e. $E[\epsilon_{ir} | c^{T_r}(X_i), W_r, T_r, G_{rr'}, \sigma_r^2] = 0$.[40] Notice the similarity of this equation with a standard linear difference-in-differences model.[41] The main differences are that the regression coefficients are replaced by semi-

[39] Notice, however, that the identifying assumption does not require geographic dispersion effects to be independent of the cohort T_r conditional on the vector of student characteristics X_i, rather, independence is required only conditional on c_i, a weaker assumption.

[40] This is a weaker assumption than the full-independence assumptions imposed in the nonlinear DD models in Athey and Imbens (2006).

[41] The cohort here represents a group (treatment or control), while the low or high variance status plays the role that is typically played by time, with geographic dispersion effects corresponding to time trends

parametric functions of student type and classroom characteristics, and that the treatment indicator has a double index rr', because treatment status is determined within each pair of classrooms. Like in standard difference-in-differences, the estimand of interest can be obtained through double differences, however, in this non-linear model the differences are taken between functions, point-by-point. Before describing what is identifiable, I discuss student types, which are an important feature of the model.

Student types. Scalar c_i is an unobserved student type, and I assume that it is a cohort-specific index function $c^{T_r}(X_i) = \theta^{T_r} X_i$ of the vector of observed student characteristics X_i. Aggregating multiple student characteristics into a single scalar accomplishes four goals. First, it reduces the curse of dimensionality when the difference-in-differences estimator is computed conditional on the scalar c_i rather than the vector X_i, improving precision of the estimator (Abadie 2005, Horowitz 2010). Second, it allows me to obtain interpretable results, because heterogeneity of the treatment effect can be graphed simply with respect to a single scalar. Third, it provides a natural empirical counterpart to the theoretical cost of effort. In the theoretical model, students are allowed only one dimension of heterogeneity for tractability (game of status models like the one proposed here have never been analytically solved with more than one level of heterogeneity). Section 6.3 discusses and provides evidence for the interpretation of c_i as the theoretical cost of effort. Fourth, and perhaps most importantly, letting the index function be cohort specific addresses a key feature of the data, that is, post-earthquake (treated) students have all been affected by the earthquake, which may have increased their unobserved cost of effort, while pre-earthquake (control) students have not. Therefore, a post-earthquake student may have a different (unobserved) effort cost from a pre-earthquake student with the same observed characteristics X_i. Modelling student type as an unobserved scalar which is a cohort specific function of student characteristics allows me to correctly match control and treated students on cost of effort, rather than on X_i. This would be precluded in, for example, a quantile DD approach (see Appendix B.5).

Formally, the index function varies with cohort T_r to let earthquake damage at a student's home, measured by MSK-Intensity I_i at her home, affect c_i in the post-earthquake

in the typical DD application.

cohort, but not in the pre-earthquake cohort:

$$c_i = c^{T_r}(X_i;\theta) = \begin{cases} \theta_1 y_{i,t-1} + \theta_2 parental\_educ_i + \theta_3 income_i + \theta_4 gender_i, & \text{if } T_r = 0 \\ c^{T_r=0}(X_i) + \theta_5 I_i + \theta_6 I_i X_i^-, & \text{if } T_r = 1 \end{cases}$$

where $y_{i,t-1}$ is lagged achievement, capturing academic skills, X_i^- is a vector grouping together student's skills, parental education (i.e., the average between mother's and father's), household income and gender, with $X_i = [X_i^-, I_i]$, and where $I_i X_i^-$ is an interaction term capturing individual heterogeneity in how damage affects a student's unobserved type c_i.[42] Parameters θ_5 and θ_6 capture the direct effect of earthquake exposure on a student's type.

Identifiable classroom specific functions. Consider a classroom r. Conditional on W_r, T_r, and σ_r^2, under regularity conditions set out in Ichimura (1993), the mapping from c_i to achievement ($h(c_i, \cdot)$ as a function of c_i in equation 4) and the parameters of the index $c^{T_r}(X_i) = \theta^{T_r} X_i$ are identifiable:[43]

$$Y_{ir} = h(c_i; W_r, T_r, \sigma_r^2) + \epsilon_{ir} \quad c_i = c^{T_r}(X_i) = \theta^{T_r} X_i \quad E[\epsilon_{ir} | c_i, W_r, T_r, \sigma_r^2] = 0. \tag{5}$$

Conditional double difference of functions. Just like in a standard linear DD model, within each classroom r the different components of $h(\cdot)$ as described in equation 4 are not separately identified. However, the treatment effect of interest, which is the function premultiplied by $T_r G_{rr'}$ in equation 4, is identified through a double difference. In contrast to the standard linear model, the double difference is now performed between four classroom specific *functions*, point-by-point. Consider two pairs of classrooms: one pair in the post-earthquake cohort, r and r', and one pair in the pre-earthquake cohort, s and s'. Assume that $G_{rr'} = 1$ and $G_{ss'} = 1$, that is, r and s are the classrooms with the relatively higher variance in MSK-Intensity within the pair. Moreover, assume that $\delta_{rr'} = \delta_{ss'} = \delta$, that is, the treatment intensity within pair is the same across pairs. Finally, assume that, except

[42] For example, wealthier parents may try to attenuate the impact of the earthquake by providing more resources to an affected child, or the psychological impact of I_i may vary by gender. In fact, I find that the positive impact of I_i on c_i is significantly stronger for female students, and this is compatible with findings in the medical literature. For example, on a sample of young adults who survived the L'Aquila 2009 earthquake, females were significantly more likely to suffer from PTSD (Dell'Osso, Carmassi, Massimetti, Daneluzzo, Di Tommaso, and Rossi 2011).

[43] The θ parameters are identified up to a normalisation (here, I set the coefficient on $y_{i,t-1}$ to -1, assuming that higher academic skills reduce cost of effort). The regularity conditions include assuming that X_i has at least one continuously distributed component whose θ coefficient is nonzero, and h is differentiable and nonconstant in c_i (Ichimura 1993).

for the variance in MSK-Intensity, these four classrooms share all other characteristics, that is $W_r = W_{r'} = W_s = W_{s'} = W$. W is high-dimensional, this means not only that the four classrooms share teacher and school characteristics, but also that the student composition in terms of the *distribution* of X within the classroom is the same across classrooms. Conditional on W and δ, the damage dispersion effect can be obtained as a function of c_i through the following double difference (which is visualised in Figures 6 and 7):

$$\begin{aligned}\lambda^{DD}(c_i; W, \delta) &= \Big(h_r(c_i) - h_{r'}(c_i)\Big) - \Big(h_s(c_i) - h_{s'}(c_i)\Big) \\ &= \Big(\lambda^{GD}(c_i; W, \delta) + \lambda^{DD}(c_i; W, \delta)\Big) - \lambda^{GD}(c_i; W, \delta). \end{aligned} \tag{6}$$

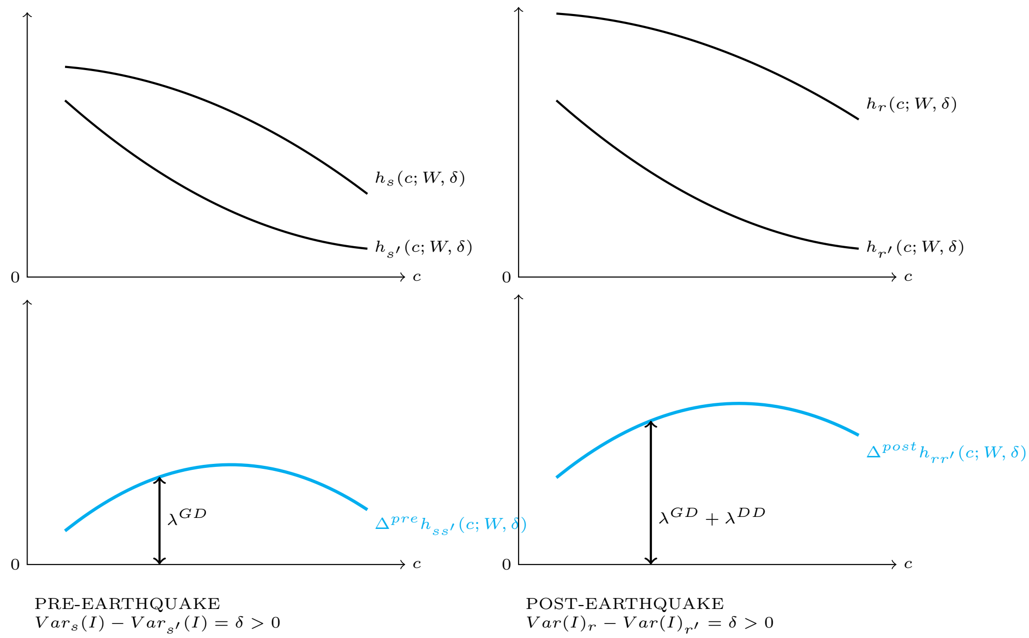

Figure 6: Classrooms r, r', s, and s' have identical W and within-pair δ. λ^{GD} is the geographic dispersion effect function, λ^{DD} is the effect on achievement of increasing damage dispersion by δ, as a function of student type c, conditional on W.

The differences are taken point-by-point, that is, for all values of c_i. The fact that the index function $c^{T_r}(X_i; \theta)$ is cohort specific is what permits correct comparisons between pre- and post-earthquake students in this point-by-point differencing.[44]

[44] For example, suppose that we estimate that pre-earthquake $c_i = 1 \cdot y_{i,t-1}$ and post-earthquake $c_i = 1 \cdot y_{i,t-1} - 2 \cdot I_i$. Then a pre-earthquake student of academic skill $y_{i,t-1} = 10$ is the appropriate control for a post-earthquake (treated) student of higher academic skill $y_{i,t-1} = 12$, but who is hit by an earthquake shock of 1 ($12 - 2 \cdot 1 = 10$), rather than for a post-earthquake student of the same academic skill $y_{i,t-1} = 10$ who is hit by the same shock ($10 - 2 \cdot 1 = 8 \neq 10$). The fact that I estimate the parameters

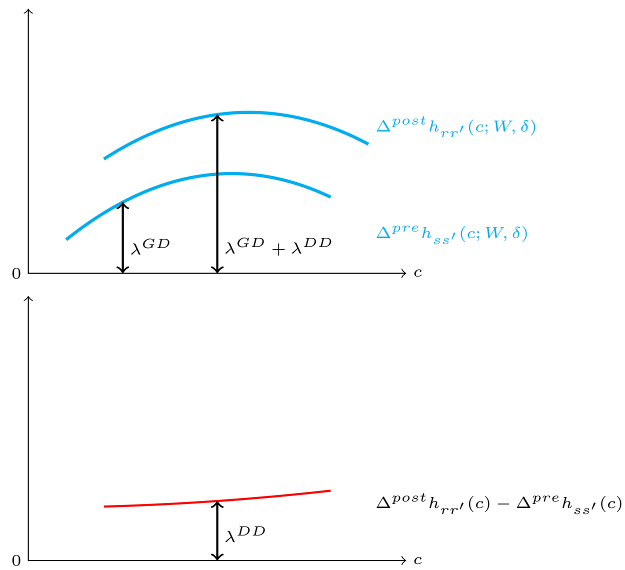

Figure 7: Netting out the geographic dispersion effects. Notice that the difference between the Δh functions can be taken only over the overlapping portion of the two domains. The domain of $\Delta^{post} h_{rr'}$ is shifted to the right with respect to the domain of $\Delta^{pre} h_{ss'}$ because all students in the post-earthquake cohort have been affected by the earthquake and their cost of effort c is expected (and estimated) to be larger than for pre-earthquake students. This plays the same role as, for example, the support restriction in the nonlinear difference-in-differences models in Athey and Imbens (2006).

Matching and integration to obtain unconditional treatment effects under unbalanced covariates. For each pair of pairs of classrooms in the data that are matched on W and δ, the effect of damage dispersion on achievement as a function of student's type c_i is identified through the conditional double difference in 6. Like with other matching DD estimators, matching on W addresses the unbalance in the distribution of covariates W between the control (pre-earthquake) and the treatment (post-earthquake) groups (Smith and Todd 2005), shown in Table 1.

In principle, because in the data there are multiple quadruplets of classrooms that share different values of W and δ, it is possible to identify non-parametrically how this function varies with W and δ.[45] However, to test the theoretical model it is sufficient to trace how the damage dispersion effect varies with student type c_i, regardless of the value of treatment intensity δ (as long as $\delta > 0$) and characteristics W. Therefore, classroom char-

of the $c^{T_r}(X_i, \theta)$ function allows me to make the correct comparisons between pre-earthquake control and post-earthquake treated students when taking the point-by-point differences.

[45] While there is non-parametric identification, the standard error of an estimator that traced heterogeneity of the treatment effect by W and δ would be large, with poor power properties.

acteristics and treatment intensities are integrated out using their empirical distribution. First, I match classroom quadruplets in terms of W and δ and perform the conditional double difference in 6. Second, let $f(W, \delta)$ indicate the empirical distribution of quadruplets of classrooms with the same W and δ. The effect of increasing damage dispersion, unconditional on W and δ, is obtained by averaging the treatment effect over W and δ, with $\delta > 0$:

$$TE(c_i) = E[\lambda^{DD}(c_i; W, \delta)] = \int \lambda^{DD}(c_i; W, \delta) I[\delta > 0] f(W, \delta) dW d\delta \tag{7}$$

where $TE(c_i)$ is the treatment effect of interest, and $I[\cdot]$ is an indicator function equal to 1 if its argument is true. In practice, matching quadruplets of classrooms with respect to W and δ is performed by kernel weighting, in the spirit of Ahn and Powell (1993). Because of the high-dimensionality of W, it would be difficult to find a quadruplet of classrooms that are exactly identical in all elements of W; nearest neighbour matching is preferable.[46] Weights are built with multivariate standard normal kernel functions. Details of the weighting procedure can be found in Appendix B.1.2. Appendix B.1.1 presents the algorithm for the estimation of the semiparametric single-index model.

Implementation. Given estimated $h_r(\cdot)$ functions for all classrooms in the sample, and given kernel weights $\omega_{rr'ss'}$ for each quadruplet of classrooms in the sample (two from each cohort), at each candidate value of c the estimate of the treatment effect $TE(c)$ is obtained through the following sample mean:

$$\hat{TE}(c) = \frac{\sum_{s=1}^{N^{pre}-1} \sum_{s'=s+1}^{N^{pre}} \sum_{r=1}^{N^{post}-1} \sum_{r'=r+1}^{N^{post}} \omega_{rr'ss'} (\Delta^{post} \hat{h}_{rr'}(c; W, \delta) - \Delta^{pre} \hat{h}_{ss'}(c; W, \delta))}{\sum_{s=1}^{N^{pre}-1} \sum_{s'=s+1}^{N^{pre}} \sum_{r=1}^{N^{post}-1} \sum_{r'=r+1}^{N^{post}} \omega_{rr'ss'}} \tag{8}$$

where N^{post} and N^{pre} are the sample number of classrooms in the post- and pre-earthquake cohorts, and the double difference at point c is at the numerator. Estimating $\hat{TE}(c)$ over a grid of values for c allows me to trace the treatment effect $TE(c)$ as a function of student types. Computing $\hat{TE}(c)$ at each grid point requires a number of calculations of the order of 10^{12}, therefore, parallel processing is required. Using $\sim 2,000$ nodes on the UCL Legion cluster, estimation is completed in around 70 hours.

[46] One additional regularity condition is required to apply kernel matching: the functions $\phi(c^{T_r}(X_i), W_r)$ and $\lambda^E(c^{T_r}(X_i), W_r)$ must be continuous in W_r, and the functions $\lambda^{GD}(c_r^T(X_i), W_r, \sigma_r^2)$ and $\lambda^{DD}(c_r^T(X_i), W_r, \sigma_r^2)$ must be continuous in W_r and σ_r^2. In Ahn and Powell (1993), this assumptions corresponds to the continuity of the selection function (see page 9 of their paper). This guarantees that there are no jumps when we compare pairs of classrooms that are similar but not identical in W and δ.

5 Findings from the Nonlinear Empirical Model

Table 5 presents the parameter estimates. As expected, earthquake intensity is estimated to increase student type and there are interactions with income and gender. The model fit is very good, as can be seen in Table 6.

Table 5: Parameter Estimates (bootstrapped standard errors in parentheses)

Parameter	Coefficient on	Math	Spanish
θ_2	Parental Education	-0.0116^{***}	-0.0212^{***}
		(0.0052)	(0.0045)
θ_3	High Income Dummy	-0.0560^{***}	-0.0356^{**}
		(0.0162)	(0.0175)
θ_4	Female	0.1290^{***}	-0.2303^{***}
		(0.0195)	(0.0350)
θ_5	MSK-Intensity	0.0326	0.0946
		(0.0596)	(0.1438)
θ_{61}	MSK-Intensity*High Income	-0.0004^{***}	-0.0004
		(0.0000)	(0.0027)
θ_{62}	MSK-Intensity*Female	-0.0031	0.0550^{*}
		(0.0288)	(0.0334)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The estimates of the treatment effect functions $TE(c_i)$ for Spanish and Mathematics are reported in Figure 8. The effect of increasing damage dispersion on student test scores is heterogeneous depending on a student's type c_i. It is worth noting that some students benefit from an increase in damage dispersion, indicating that results from the linear model mask considerable heterogeneity. Going from low to high c, the function $\widehat{TE(c)}$ is negative and then positive for Spanish test scores, while it is positive, then negative and then positive for Math test scores. This means that increasing the variance of damage dispersion has a negative impact on the test scores of middle-c students, and a positive impact on the test scores of high-c students, while it has a negative impact on low-c Spanish students, and a positive impact on low-c Mathematics students. These patterns are identical to those predicted by the comparative statics results reported in Figure 1, with Mathematics following the lower panel pattern and Spanish the upper panel pattern. They are consistent with stronger rank concerns in Mathematics than in Spanish.

Figure 8 reports also bounds useful for testing the sign of $TE(c)$ over its domain. The statistical tests support the consistency of the empirical patterns with the theoretical predictions: $TE(c)$ is statistically negative for middle-c students and statistically positive

Table 6: Model Fit, Test Scores

	Mathematics		Spanish	
	Actual	Model	Actual	Model
Pre-Earthquake Cohort				
Overall	-.185	-.189	-.121	-.123
Female	-.304	-.283	-.050	-.063
Male	-.058	-.089	-.196	-.186
Female				
Urban	-.300	-.279	-.052	-.064
Rural	-.322	-.302	-.043	-.056
Male				
Urban	-.035	-.066	-.180	-.172
Rural	-.159	-.188	-.262	-.249
Female				
Lower Income	-.414	-.387	-.148	-.155
Higher Income	-.130	-.120	.104	.083
Male				
Lower Income	-.222	-.246	-.348	-.328
Higher Income	.155	.116	.001	-.003
Post-Earthquake Cohort				
Overall	-.222	-.228	-.153	-.156
Female	-.307	-.292	-.058	-.078
Male	-.132	-.159	-.254	-.239
Female				
Urban	-.302	-.287	-.071	-.086
Rural	-.329	-.315	.001	-.039
Male				
Urban	-.120	-.148	-.257	-.246
Rural	-.180	-.205	-.242	-.209
Female				
Lower Income	-.414	-.388	-.146	-.160
Higher Income	-.151	-.151	.071	.042
Male				
Lower Income	-.237	-.262	-.351	-.322
Higher Income	-.0004	-.0304	-.133	-.136

for high-c students, while for low-c students it is statistically negative in Spanish, and ambiguous in Mathematics. This is consistent with the model, which allows $TE(c)$ to take on any sign for low-c students. However, the statistical tests do not support a difference in signs between Mathematics and Spanish for lower-c students.

In sum, the statistical test results support the model predictions, and indicate that Spanish follows the patterns that the theory predicts under weaker rank preferences (top panel of Figure 1). Moreover, the point estimates indicate a stronger rank preference in Mathematics than in Spanish, however, identical patterns between Spanish and Mathematics cannot be ruled out statistically. [47]

6 Discussion and Robustness

6.1 Identification and Robustness to Fixed Effects

An earthquake is a complex disruption affecting not only students and households, but also teachers and schools. Moreover, not all teacher and school characteristics are observed. Therefore, the effect of damage dispersion may be entirely due to an endogenous response of unobserved teacher and/or school inputs. If this was the case, the interpretation of the empirical findings through the game of status model would be problematic. Here I show, theoretically and empirically, that the results are not driven by the reaction to the earthquake of unobserved school-level characteristics. Specifically, I exploit the fact that some of the schools in the data have multiple classrooms for the same grade. I consider school×cohort fixed effects that account for any change across cohorts of unobserved school inputs, including changes due to an endogenous response to the earthquake (or to any policy introduced in between cohorts).

Identification requires robustness to classroom level unobservables. Proposition 6.1 formalises robustness in terms of a restriction on the fixed effects. However, this restriction cannot be tested directly on classroom fixed effects because the treatment effect of interest is at classroom level. For this reason, I formalize and test the identifying restriction at school level. Additionally, in section 6.2 I provide evidence that classroom level characteristics that were not used in estimation (the productivity of the Spanish and

[47] The magnitudes of the effects cannot be directly compared to the magnitudes of the effects from the linear model. The nonlinear model's effects are averages over the empirical distribution of treatment intensities δ, while the coefficient in the linear model represents the effect of a marginal increase in δ. Because the nonlinear model imposes fewer restrictions, it is more data-demanding, therefore, conditioning on treatment intensity δ would generate power issues. Moreover, conditioning on δ is not needed to test the comparative statics results.

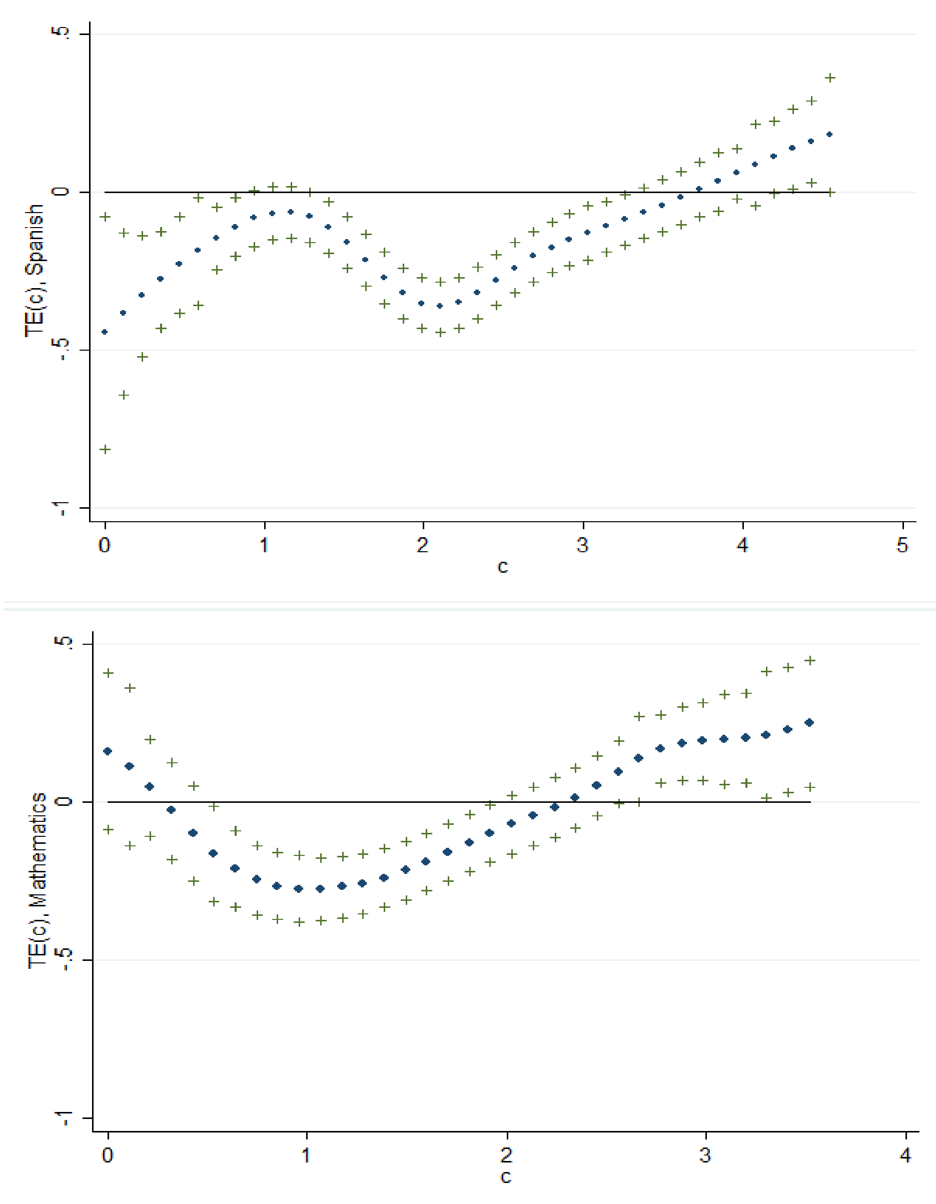

Figure 8: Estimated $\widehat{TE(c_i)}$ for Spanish (top) and Mathematics (bottom) test scores. Bounds for one-sided significance tests at the 10 percent significance level are reported. When the lower bound is above 0, we accept the hypothesis that $TE(c) > 0$, when the upper bound is below 0, we accept $TE(c) < 0$.

Mathematics teachers) satisfy this restriction.

First, I show theoretically that the nonlinear model is robust to the presence of school×cohort fixed effects in the Data Generating Process (DGP), subject to a (weak) restriction on the distribution of the fixed effects.

Proposition 6.1 Robustness of the non-linear model to fixed effects. *If the DGP presents school×cohort fixed effects, the treatment effect $TE(c)$ in 7 is identified in the nonlinear model defined in 4 and 5 if*

$$\begin{aligned} &\Big[E[\alpha_{MT}|T_M=1,G_{MM'}=1,c,W,\delta]-E[\alpha_{MT}|T_M=1,G_{MM'}=0,c,W,\delta]\Big]= \\ =\;&\Big[E[\alpha_{MT}|T_M=0,G_{MM'}=1,c,W,\delta]-E[\alpha_{MT}|T_M=0,G_{MM'}=0,c,W,\delta]\Big] \quad \forall\delta,c,W \end{aligned} \tag{9}$$

where α_{MT} is the fixed effect.
Under this condition (9), in the presence of school×cohort fixed effects in the DGP, the point estimator of the treatment effect $\hat{TE}(c)$ (equation 8) is unbiased at all points c. The estimate converges in probability to the true treatment effect as the number of schools in the sample goes to infinity.

Proof: see Appendix B.2.

First, an advantage of this result is that the number of classrooms per school, which is typically finite in the population and small in my sample, does not need to go to infinity for consistency. Second, the fairly weak restriction in Proposition 6.1 allows for a flexible distribution of school×cohort fixed effects and, as shown in the proof, the fixed effects may even have heterogeneous impacts across students ($\alpha_{MT}(c)$). Specifically, the expectation of the fixed effect conditional on observed student, classroom, teacher and school characteristics (c, W) can be an arbitrary function, and it is allowed to change across cohorts to capture a response to the earthquake of unobserved school inputs. Schools with different observables can change their unobserved inputs differently in response to the earthquake. Moreover, this change may have heterogeneous impacts on students. For example, schools with better resources or schools subject to a stronger earthquake intensity may increase the instructional hours by more than other schools, and they may increase their focus on the most vulnerable students. What is not allowed is for the geographic dispersion of the student body to determine how a school reacts to the earthquake in terms of unobserved school inputs.[48] This is not especially restrictive, because schools are allowed to respond

[48] To see this, notice that $G_{MM'}=1$ means that $\sigma^2_M=\sigma^2_H$ and $\sigma^2_{M'}=\sigma^2_L$, with $\sigma^2_H-\sigma^2_L=\delta$. Using this result and rearranging, condition 9 becomes:

$$\begin{aligned} &\big[E[\alpha_{MT}|T_M=1,\sigma^2_H,c,W]-E[\alpha_{MT}|T_M=0,\sigma^2_H,c,W]\big]= \\ =\;&\big[E[\alpha_{MT}|T_M=1,\sigma^2_L,c,W]-E[\alpha_{MT}|T_M=0,\sigma^2_L,c,W]\big] \quad \forall\sigma^2_H,\sigma^2_L,c,W. \end{aligned}$$

The change in the mean of the fixed effect across cohorts must be the same for the high (σ^2_H) and for the low (σ^2_L) geographic dispersion classroom, and this must be true $\forall\sigma^2_H,\sigma^2_L$ in the data.

differently depending on their students' composition, including variables that are correlated with geographic dispersion like income variance.

Second, I propose an empirical test for this restriction.

Proposition 6.2 Identification test of the non-linear model in the presence of fixed effects. *If the conditional expectation of the school×cohort fixed effect is a linear function of σ_m^2 and T_M, with W and c entering in an additively separable way, then a sufficient condition for identification of the treatment effect $TE(c)$ in the nonlinear model defined in 4 and 5 is that $\beta_3 = 0$ in the following equation:*

$$\alpha_{MT} = \beta_0 + \beta_1\sigma_m^2 + \beta_2 T_M + \beta_3\sigma_m^2 T_M + g(W_m, c_i; \beta_4) + \epsilon_{mM} \quad \forall m \in M. \tag{10}$$

Proof: see Appendix B.2.

To verify if this condition is satisfied in the data, I estimate the linear model of spillovers in 18 in Appendix A.2 with the addition of school fixed effects, and compute predicted fixed effects $\hat{\alpha}_{MT}$ using the estimated parameters. While the fixed effects are identified also in the semiparametric model, those predicted from the linear model have lower variance.[49] I then verify if the condition for identification under linearity of the fixed effects in 10 is rejected in the data by estimating the following linear model:

$$\hat{\alpha}_{RT} = \beta_0 + \beta_1\sigma_r^2 + \beta_2 T_R + \beta_3\sigma_r^2 T_R + \beta_4 W_R + \beta_5\tilde{c}_R + \epsilon_R \tag{11}$$

where, for simplicity, I have replaced $g(W, c; \beta_4)$ with a linear function, and where $\tilde{c}_R$ are student characteristics aggregated at the school level like, for example, average income and average parental education.[50] A t-test on the significance of the $\hat{\beta}_3$ estimated coefficient

[49] However, in the linear model only the average fixed effect across students in the same school can be estimated. That is, the heterogeneity of the school effect across students cannot be captured in the linear model. Therefore, the empirical application of the identification test that I present here tests robustness of $\hat{TE}(c)$ to a standard kind of school fixed effects (that is, a constant school effect for all students in the same school). If the heterogeneity of the school effects could be estimated precisely, robustness of $\hat{TE}(c)$ to these more general effects could be tested. In the data only 37 percent of schools have multiple classrooms, making precise estimation of the heterogeneous school effects difficult.

To keep a close correspondence between the semiparametric and the parametric model, the linear regression used to compute the fixed effects contains the same set of W and X characteristics as the non-linear model.

[50] For each school there are as many equations as there are classrooms, because σ_r^2 is classroom specific. Therefore, the most appropriate model is a seemingly unrelated regression (SUR). However, given the high correlation in σ_r^2 across classrooms within the same school, for simplicity, I compute overall school level σ_R^2, and estimate the regression in 11 using σ_R^2 in place of σ_r^2. Because the SUR estimator is expected to have a higher variance, ceteris paribus, it would reject $\beta_3 = 0$ less often than the simpler model that I

cannot reject $\beta_3 = 0$, for both Mathematics and Spanish (p-values: 0.171 and 0.682 respectively). Estimation results are reported in Table 11 in Appendix B.2.2. Therefore, the identifying assumption under fixed effects is not rejected.

6.2 Teachers

One important alternative channel that could generate the treatment effect is a response to the earthquake of unobserved teacher inputs. The damage dispersion effect patterns $\hat{TE}(c)$ have been shown to be identified and estimated unbiasedly even in the presence of school×cohort fixed effects. Therefore, all changes to unobserved teacher inputs at the *school* level are accounted for.[51]

However, robustness to school×cohort fixed effects cannot account for changes in unobserved teacher inputs at the *classroom* level. To rule out this channel, one approach would be to include classroom×cohort fixed effects, which would capture any change across cohorts in unobserved classroom inputs. However, the treatment variable is at the classroom level, therefore, the treatment effect would not be identified. Instead, to verify empirically whether teacher inputs could explain the treatment effect patterns, I use a measure of teacher productivity in the classroom that is available both before and after the earthquake: the fraction of the national curriculum covered in class during the year by the Spanish and Mathematics teachers in the sample. This is a unique feature of the dataset because measures of teachers' effort/productivity in the classroom are typically unavailable in large administrative datasets like the SIMCE. However, there are three *caveats* to the use of this variable. First, it is subject to considerable non-response (35 percent for Math and 30 percent for Spanish teachers) and this non-response is non-random (for example, the mean of Math test scores when the variable is non-missing is 0.035 and it is -0.066 when it is missing). Second, it is self-reported, and there may be legitimate concerns of mis-reporting. Third, the survey question in the questionnaire for Math teachers changed slightly in between cohorts. In spite of these *caveats*, it would still cause concern if this (imperfect) productivity measure did not pass the empirical tests presented here.

First, I use this variable to run a similar identification test to that presented in Proposition 6.2. Intuitively, for teacher productivity to explain any part of the damage dispersion effect, its response to the earthquake must vary with the geographic dispersion of the stu-

estimate, which has higher power. Because it is desirable to detect a wrong null with high probability if $\beta_3 \neq 0$, the single regression model that I estimate (11) is preferable to SUR.

[51] For example, the treatment effect patterns are identified in the nonlinear model even if, in some or all schools, all teachers were required to teach more hours because of the earthquake.

dents in class.[52] Finding that its response (if any) is independent of student geographic dispersion is evidence against a contribution of teacher productivity to the damage dispersion effect, for the same logic presented in Proposition 6.1. Formally, I test for $\beta_3 = 0$ in a regression like 11, where the dependent variable is the productivity of the teacher in classroom. Regression results are reported in the first two columns of Table 14. $\beta_3 = 0$ cannot be rejected at any conventional significance level for both Mathematics and Spanish, suggesting that the estimated treatment effects $\hat{TE}(c)$ are not due to a change in teacher productivity.

Second, I use teacher productivity as an additional regressor in the linear model of spillovers in equation (18), to verify if it explains away average damage dispersion effects. Results in the third row, columns 3 and 4 of Table 14 in Appendix D show that not only the average damage dispersion effects are not explained away, but also that their point estimates are very similar to those from the model that does not include teacher productivity as regressors (compare with third row, columns 2 and 3 of Table 4). Therefore, consistently with the previous result, this suggests that also the estimated average treatment effects are not due to a change in teacher productivity.

While it is not possible to draw definitive conclusions from these results, they are consistent with teacher productivity at the classroom level not entirely explaining the damage dispersion effects and their patterns across students.

6.3 Linking the Empirical and the Theoretical Models

Discussion. In the theoretical model, students differ in terms of c, that summarises all individual level determinants of achievement like, for example, academic skills, home environment, and any other factor that affects the ability to study. c increases cost of study effort, and is informally referred to as cost of effort. To test the theoretical predictions, an exogenous shock to the classroom variance of this cost is needed. For this to happen, individuals in the same classroom must be subject to different shocks. Such shocks are rare in observational data, and they are difficult to generate within a randomised controlled trial for ethical reasons. However, they were generated by the earthquake, which offers a unique opportunity for testing. Any empirical model wanting to exploit this within-classroom variation must posses a specific feature: it must capture how earthquake intensity at a student's home affects her unobserved cost of effort. Only in this scenario can variation in earthquake exposure across classmates generate the variation to the classroom variance

[52] An equivalent way to phrase this condition is that the correlation between geographic dispersion of the students and teacher productivity must change after the earthquake.

of cost of effort that is needed to test the theory.

The empirical model estimates how own intensity of earthquake exposure affects a student's unobserved type c_i through the function $c^{T_r}(X_i;\theta)$. This section provides evidence in support of the interpretation of type c_i as the theoretical cost of effort. Notice that the empirical model does not restrict the impact of own earthquake exposure on achievement to work only through c_i: $\lambda^E(\cdot)$ in 4 captures cohort effects, including the effect of being subject to the earthquake (dummy exposure measure), and it lets them depend arbitrarily on student, classroom, school and teacher characteristics. Therefore, the empirical model allows for multiple channels through which own earthquake exposure impacts own achievement. Any impact on the student type c_i is only one of them.

Post-estimation: parameter signs. Under the (necessary) normalisation that past achievement decreases unobserved type c_i ($\theta_1 = -1$), the other estimated θ parameters have signs that are consistent with an interpretation of c_i as cost of effort. For example, higher household income is estimated to decrease c_i, and earthquake intensity to increase it, and to increase it by more for poorer families. Moreover, in Spanish there is a significant positive interaction between MSK-Intensity and gender, which is consistent with the medical finding of higher incidence of PTSD among female earthquake survivors (Dell'Osso, Carmassi, Massimetti, Daneluzzo, Di Tommaso, and Rossi 2011). Additionally, at the estimated parameter values the classroom variance of predicted c_i ($\hat{c}_i = c^{T_r}(X_i;\hat{\theta})$) is increasing in the classroom variance of MSK-Intensity. Therefore, under the interpretation of type c_i as cost of effort, variation across classrooms in damage dispersion generates the necessary variation in cost of effort dispersion.

Post-estimation: monotonicity test and why c_i does not depend on an unobservable. One of the theoretical model's implications is that, under the interpretation of c_i as cost of effort, achievement is monotonically decreasing in c_i (see Proposition 2.1). To verify this, in estimation I do not impose monotonicity of $h_r(c_i)$ $\forall r$, the empirical counterpart of the theoretical model's achievement function $y(c)$. Rather, monotonicity is formally tested for. If the function that defines c_i, $c^{T_r}(X_i;\hat{\theta})$, included an unobservable argument, monotonicity of $h_r(c_i)$ would have to be imposed for identification. By not letting c_i depend on an unobservable and, therefore, not imposing monotonicity, I can test for it. Figure 12 in Appendix D shows an example of two estimated classroom-specific functions $\hat{h}_r(\hat{c}_i)$. As can be seen, the higher a student's type $\hat{c}_i$ is, the lower achievement is. I formally test monotonicity of $h_r(c_i)$ in c_i using the method developed in Chetverikov (2013). For all classrooms, the null hypothesis that the h function is monotonically decreasing in c_i is not rejected at the $\alpha = 0.10$ significance level (see Appendix B.3). This supports the interpretation of c_i as cost of effort.

Survey evidence. Survey evidence suggests that seismic intensity at a student's home affected a student's self-reported cost of exerting study effort.[53] Conditional on student academic skills and on parental education and income, post-earthquake students affected by higher earthquake intensity report that it is more costly for them to study, as shown in Table 15 in Appendix D. This is consistent with the interpretation of c_i as cost of effort.

Additional regression evidence. Type c_i is increasing in earthquake intensity and decreasing in academic skills. Therefore, if it is indeed cost of effort, we would expect to see that it matters for achievement if it is the high- or the low-skill students in the classroom who are more or less affected by the earthquake. Moreover, these effects should be consistent with the theory.

Using a linear DDD model similar to 18, I find that, in Mathematics, it is worse for achievement when it is the high skill rather than the low skill students who are affected more harshly by the earthquake. In the main sample of Municipal schools, this effect is driven entirely by a reduction in test scores of the most academically skilled students, as can be seen in the second to fifth columns of Table (12) in Appendix B.4, where model details can also be found. That is, *conditional* on own damage and on the skill composition in the class, a high-skill student performs worse in a classroom where high-skill students are hit more harshly than in one where low-skill students are hit more harshly. This is consistent with the theoretical model. The empirically relevant prediction for Mathematics is that low-cost students do worse when there are fewer other low-cost students (lower panel of Figure 1). Therefore, conditional on own damage, high-skill (low-cost) students are expected to perform worse in Mathematics when the proportion of other low-cost students in the classroom has been decreased by the earthquake because the other high-skill students have been hit more harshly. This is precisely what this additional regression analysis finds, corroborating the specification for c_i and its interpretation as cost of effort.

6.4 Relationship to Other Nonlinear DD Models

Like the semiparametric model presented here, other nonlinear difference-in-differences models estimate heterogeneity of the effects. The closest models to the one presented here

[53]Students were asked to rate how much they agree with sentences such as "It costs me to concentrate and pay attention in class" and "Studying Mathematics costs me more than it costs my classmates". I combine the answers to these questions into a single factor.

are the quantile Difference-in-Differences (QDID) and Changes-in-Changes (CIC) models in Athey and Imbens (2006), and the nonlinear difference in differences models in Blundell and Costa Dias (2000), Bell, Blundell, and Van Reenen (1999), Blundell, Costa Dias, Meghir, and Van Reenen (2004), Smith and Todd (2005), Heckman, Ichimura, Smith, and Todd (1998), Abadie (2005) and D'Haultfoeuille, Hoderlein, and Sasaki (2015). However, three features set the model in this paper aside: nonseparability in X, continuous treatment, and collapsing of the covariates in X into a single scalar. These three features are necessary for testing the theoretical predictions, however, no existing method possesses all three, as discussed in detail in Appendix B.5.

6.5 Self-selection into Subgroups of Friends

I have presented arguments and evidence against unobservables at the school level (fixed effects) and teaching in the classroom driving the results. Here, I discuss another potential alternative mechanism: self-selection into peer subgroups formed mainly of peers with a similar ability (cost of effort), a mechanism proposed by Carrell, Sacerdote, and West (2013) to explain their experimental findings.[54] If only mean peer ability matters and if students choose more often to become friends with similarly able peers when their availability increases, then we should observe a worsening of the outcomes of high-cost (low-ability) students whenever there are more high-cost (low-ability) classmates. This mechanism cannot explain my data, because high-cost (low-ability) students in both Mathematics and Spanish classes increase their test scores when the proportion of high-cost (low-ability) classmates increases following a variance increase.

7 Conclusions

Drawing on the theoretical literature on rank preferences, I propose a theory of peer effects based on rank concerns. When it is not just an interest in learning, but also rank concerns that motivate students to study, their effort choices are an equilibrium outcome. They depend on the distribution of peer characteristics because, intuitively, how easy or hard it is to obtain a good achievement rank depends on how able the other students are. An important implication is that peer effects are heterogeneous and nonlinear.

My empirical findings show that the model predictions on the shape of peer effects are borne out in a large dataset on Chilean 8th graders. I use an empirical strategy that

[54] Appendix B.6 discusses additional alternative mechanisms.

combines a novel nonlinear DD model with data variation from a natural experiment used for identification. The Chilean 2010 earthquake did not always affect students in the same classroom equally, generating useful variation in the distribution of peer group attributes. The empirical model is specifically designed to exploit this variation, and it is carefully informed by the theoretical model. As a result, the treatment effects have a direct structural interpretation.

Together with the fact that the theory of rank concerns seems to be useful to understand other pieces of evidence in the literature, this finding indicates that rank concerns are one possible channel generating peer effects. This has important implications for the estimation of peer effects and for policy. First, rank concerns do not necessarily generate outcome clustering around the mean of a reference group. Therefore, techniques that look for evidence of clustering may yield false negatives on the existence of peer effects.[55] Second, a normative implication of the paper is that rank concerns could be exploited by policy-makers to increase study effort. To do so, it may be possible to optimally design classroom allocations while simultaneously intervening on rank incentives. While rank incentives in education have typically been the domain of affirmative action studies, and classroom allocation rules of peer effect studies, this paper demonstrates that there could be combined benefits from these two types of policies - and literatures - that have not been explored yet.[56]

References

Abadie, A. (2005): "Semiparametric difference-in-differences estimators," *The Review of Economic Studies*, 72(1), 1–19.

Ahn, H., and J. L. Powell (1993): "Semiparametric estimation of censored selection models with a nonparametric selection mechanism," *Journal of Econometrics*, 58(1), 3–29.

Altindag, A., S. Ozen, et al. (2005): "One-year follow-up study of posttraumatic stress disorder among earthquake survivors in Turkey," *Comprehensive psychiatry*, 46(5), 328–333.

[55] See, for example, conditional variance restriction methods (Glaeser, Sacerdote, and Scheinkman 1996, Graham 2008).

[56] For studies that frame affirmative action policies in terms of relative incentives see, for example, Cotton, Hickman, and Price (2015) and Hickman (2016). In ongoing work, I am collecting data on student rank concerns within a large scale randomised controlled trial on affirmative action.

Arnott, R., and J. Rowse (1987): "Peer group effects and educational attainment," *Journal of Public Economics*, 32(3), 287–305.

Astroza, M., S. Ruiz, and R. Astroza (2012): "Damage Assessment and Seismic Intensity Analysis of the 2010 (Mw 8.8) Maule Earthquake," *Earthquake Spectra*, 28(S1), S145–S164.

Athey, S., and G. W. Imbens (2006): "Identification and inference in nonlinear difference-in-differences models," *Econometrica*, 74(2), 431–497.

Azmat, G., and N. Iriberri (2010): "The importance of relative performance feedback information: Evidence from a natural experiment using high school students," *Journal of Public Economics*, 94(7), 435–452.

Bandura, A. (1986): *Social foundations of thought and action.* Englewood Cliffs, NJ Prentice Hall.

Bell, B., R. Blundell, and J. Van Reenen (1999): "Getting the unemployed back to work: the role of targeted wage subsidies," *International tax and public finance*, 6(3), 339–360.

Bhaskar, V., and E. Hopkins (2013): "Marriage as a rat race: Noisy pre-marital investments with assortative matching," .

Blanes i Vidal, J., and M. Nossol (2011): "Tournaments without prizes: Evidence from personnel records," *Management Science*, 57(10), 1721–1736.

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman (2014): "Linear social interactions models," *forthcoming, Journal of Political Economy.*

Blundell, R., and M. Costa Dias (2000): "Evaluation methods for non-experimental data," *Fiscal studies*, 21(4), 427–468.

Blundell, R., M. Costa Dias, C. Meghir, and J. Van Reenen (2004): "Evaluating the employment impact of a mandatory job search program," *Journal of the European economic association*, 2(4), 569–606.

Booij, A., E. Leuven, and H. Oosterbeek (2016): "Ability Peer Effects in University: Evidence from a Randomized Experiment," *Review of Economic Studies, forthcoming.*

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of econometrics*, 150(1), 41–55.

Brown, G. D., J. Gardner, A. J. Oswald, and J. Qian (2008): "Does Wage Rank Affect Employees' Well-being?," *Industrial Relations: A Journal of Economy and Society*, 47(3), 355–389.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009): "Peer effects and social networks in education," *The Review of Economic Studies*, 76(4), 1239–1267.

Card, D. (1992): "Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wage," *Industrial and Labor Relations Review*, 46(1), 22–37.

Card, D., A. Mas, E. Moretti, and E. Saez (2012): "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction," *American Economic Review*, 102(6), 2981–3003.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013): "From natural variation to optimal policy? The importance of endogenous peer group formation," *Econometrica*, 81(3), 855–882.

Chetverikov, D. (2013): "Testing regression monotonicity in econometric models," *mimeo, UCLA*.

Cicala, S., R. G. Fryer, and J. L. Spenkuch (2016): "Self-Selecion and Comparative Advantage in Social Interactions," *Available at SSRN 1780652*.

Cipollone, P., and A. Rosolia (2007): "Social interactions in high school: Lessons from an earthquake," *The American economic review*, pp. 948–965.

Comerio, M. C. (2013): *Housing recovery in Chile: A qualitative mid-program review*. Pacific Earthquake Engineering Research Center Headquarters at the University of California.

Cotton, C., B. R. Hickman, and J. P. Price (2015): "Affirmative Action and Human Capital Investment: Theory and Evidence from a Randomized Field Experiment," Discussion paper.

De Giorgi, G., and M. Pellizzari (2013): "Understanding social interactions: Evidence from the classroom," *The Economic Journal*.

De Giorgi, G., M. Pellizzari, and S. Redaelli (2010): "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics*, 2(2), 241–275.

Dell'Osso, L., C. Carmassi, G. Massimetti, E. Daneluzzo, S. Di Tommaso, and A. Rossi (2011): "Full and partial PTSD among young adult survivors 10 months after the L'Aquila 2009 earthquake: Gender differences," *Journal of affective disorders*, 131(1), 79–83.

D'Haultfoeuille, X., S. Hoderlein, and Y. Sasaki (2015): "Nonlinear Difference-in-Differences in Repeated Cross Sections with Continuous Treatments," .

Duflo, E., P. Dupas, and M. Kremer (2011): "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *The American Economic Review*, 101, 1739–1774.

Elsner, B., and I. Isphording (2016a): "A big fish in a small pond: Ability rank and human capital investment," *forthcoming, Journal of Labor Economics.*

Elsner, B., and I. E. Isphording (2016b): "Rank, Sex, Drugs, and Crime," .

Epple, D., and R. Romano (2008): "Educational Vouchers and Cream Skimming," *International Economic Review*, 49(4), 1395–1435.

Epple, D., and R. E. Romano (1998): "Competition between private and public schools, vouchers, and peer-group effects," *American Economic Review*, pp. 33–62.

Frank, R. H. (1985): *Choosing the right pond: Human behavior and the quest for status.* Oxford University Press.

Fruehwirth, J. C. (2012): "Can achievement peer effect estimates inform policy? A view from inside the black box," *Review of Economics and Statistics*, (00).

——— (2013): "Identifying peer achievement spillovers: Implications for desegregation and the achievement gap," *Quantitative Economics*, 4(1), 85–124.

Garlick, R. (2016): "Academic Peer Effects with Different Group Assignment Rules: Residential Tracking versus Random Assignment," Discussion paper, Working paper, World Bank Development Research Group, Washington, DC.

Genakos, C., and M. Pagliero (2012): "Interim Rank, Risk Taking, and Performance in Dynamic Tournaments," *Journal of Political Economy*, 120(4), 782–813.

Ghosal, S., A. Sen, and A. W. Van Der Vaart (2000): "Testing monotonicity of regression," *Annals of statistics*, pp. 1054–1082.

Giannopoulou, I., M. Strouthos, P. Smith, A. Dikaiakou, V. Galanopoulou, and W. Yule (2006): "Post-traumatic stress reactions of children and adolescents exposed to the Athens 1999 earthquake," *European Psychiatry*, 21(3), 160–166.

Glaeser, E. L., B. Sacerdote, and J. A. Scheinkman (1996): "Crime and Social Interactions," *Quarterly Journal of Economics*, 111(2).

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

Groome, D., and A. Soureti (2004): "Post-traumatic stress disorder and anxiety symptoms in children exposed to the 1999 Greek earthquake," *British Journal of Psychology*, 95(3), 387–397.

Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66(5), 1017–1098.

Hickman, B. R. (2016): "Pre-college human capital investment and affirmative action: a structural policy analysis of US college admissions," *University of Chicago. Unpublished.*

Hopkins, E., and T. Kornienko (2004): "Running to keep in the same place: consumer choice as a game of status," *American Economic Review*, pp. 1085–1107.

——— (2007): "Cross and double cross: comparative statics in first price and all pay auctions," *The BE Journal of Theoretical Economics*, 7(1).

Horowitz, J. L. (2010): *Semiparametric and nonparametric methods in econometrics*, vol. 692. Springer.

Hoxby, C. M., and G. Weingarth (2005): "Taking race out of the equation: School reassignment and the structure of peer effects," Discussion paper, Working paper.

Ichimura, H. (1993): "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models," *Journal of Econometrics*, 58(1), 71–120.

Imberman, S. A., A. D. Kugler, and B. I. Sacerdote (2012): "Katrina's children: Evidence on the structure of peer effects from hurricane evacuees," *The American Economic Review*, 102(5), 2048–2082.

Jencks, C., and S. E. Mayer (1990): "The social consequences of growing up in a poor neighborhood," *Inner-city poverty in the United States*, 111, 186.

Kang, C. (2007): "Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea," *Journal of Urban Economics*, 61(3), 458–495.

Kremer, M., and D. Levy (2008): "Peer effects and alcohol use among college students," *The Journal of Economic Perspectives*, 22(3), 189–189.

Kuziemko, I., R. Buell, T. Reich, and M. Norton (2014): "Last-place Aversion: Evidence and Redistributive Implications," *Quarterly Journal of Economics*, 129(1), 105–149.

Lavy, V., M. D. Paserman, and A. Schlosser (2012): "Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom*," *The Economic Journal*, 122(559), 208–237.

Lavy, V., and A. Schlosser (2011): "Mechanisms and impacts of gender peer effects at school," *American Economic Journal: Applied Economics*, pp. 1–33.

Lui, S., X. Huang, L. Chen, H. Tang, T. Zhang, X. Li, D. Li, W. Kuang, R. C. Chan, A. Mechelli, et al. (2009): "High-field MRI reveals an acute impact on brain function in survivors of the magnitude 8.0 earthquake in China," *Proceedings of the National Academy of Sciences*, 106(36), 15412–15417.

Manski, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 60(3), 531–542.

Murphy, R., and F. Weinhardt (2016): "Top of the Class: The Importance of Ordinal Rank," Discussion paper, mimeo.

Neilson, C. (2013): "Targeted vouchers, competition among schools, and the academic achievement of poor students," *Job Market Paper*.

Pagan, A., and A. Ullah (1999): *Nonparametric econometrics*. Cambridge university press.

SACERDOTE, B. (2008): "When the saints come marching in: Effects of Hurricanes Katrina and Rita on student evacuees," Discussion paper, National Bureau of Economic Research.

——— (2014): "Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward?," *Annual Review of Economics*, (0).

SACERDOTE, B. I. (2001): "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *Quarterly Journal of Economics*, 116(2), 681–704.

SCHUNK, D. H. (1996): "Learning theories," *Printice Hall Inc., New Jersey*.

SMITH, J. A., AND P. E. TODD (2005): "Does matching overcome LaLonde's critique of nonexperimental estimators?," *Journal of econometrics*, 125(1), 305–353.

STINEBRICKNER, R., AND T. R. STINEBRICKNER (2006): "What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds," *Journal of public Economics*, 90(8), 1435–1454.

TINCANI, M. (2014): "On the Nature of Social Interactions in Education: An Explanation for Recent Evidence," Discussion paper, UCL.

TODD, P. E., AND K. I. WOLPIN (2003): "On the specification and estimation of the production function for cognitive achievement," *The Economic Journal*, 113(485), F3–F33.

TRAN, A., AND R. ZECKHAUSER (2012): "Rank as an inherent incentive: Evidence from a field experiment," *Journal of Public Economics*, 96(9), 645–650.

WHITMORE, D. (2005): "Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment," *American Economic Review, P&P*, pp. 199–203.

ZIMMERMAN, D. J. (2003): "Peer effects in academic outcomes: Evidence from a natural experiment," *Review of Economics and Statistics*, 85(1), 9–23.

A Appendix: Preliminary Analysis and Linear Model of Peer Effects. FOR ONLINE PUBLICATION ONLY

A.1 Effect of Own Earthquake Exposure

A.1.1 Test score value added model as a panel data model in weighted first differences

The test score value added model is commonly used to account for unobserved student characteristics when longitudinal data on student test scores are available, but only contemporary inputs (or proxies of inputs) are available. Following Todd and Wolpin (2003), I show under what assumptions this model yields consistent estimates of the test score production function.

In each cohort, each student's test scores are observed twice: in 4^{th} grade and in 8^{th} grade. I use index g for the 8^{th} grade and $g-1$ for the 4^{th} grade. The test score value added model nets out individual unobserved heterogeneity in a similar fashion to the fixed effect model in first differences. The main difference is that in order for the value added specification to yield consistent estimates, it must be that the individual's permanent unobserved heterogeneity impacts test scores at a geometrically declining rate over time. Formally, assume that the production function of the test score of student i in classroom r and grade g is the following (ignoring, for the moment, the earthquake treatment):

$$y_{irg} = \alpha_1 X_{ig} + \alpha_2 Z_{rg} + \tilde{\alpha}_1 X_{i,g-1} + \tilde{\alpha}_2 Z_{r,g-1} + \alpha_{3g}\phi_i + \lambda_g + \epsilon_{irg} \tag{12}$$

where ϕ_i captures individual's unobserved heterogeneity and λ_g captures grade effects affecting all students in the grade equally. Vectors X and Z contain student and classroom characteristics, respectively. The latter include average peer characteristics (e.g. average income, average parental education). The following transformation takes first differences where past test scores are weighed by γ:

$$\begin{aligned} y_{irg} - \gamma y_{ir,g-1} \quad = \quad & \alpha_1 X_{ig} + \alpha_2 Z_{rg} + (\tilde{\alpha}_1 - \gamma\alpha_1) X_{i,g-1} + (\tilde{\alpha}_2 - \gamma\alpha_2) Z_{r,g-1} + \\ & (\alpha_{3g} - \gamma\alpha_{3,g-1})\phi_i + (\lambda_g - \gamma\lambda_{g-1}) + \epsilon_{irg} - \gamma\epsilon_{ir,g-1}. \end{aligned}$$

When the effect of student and classroom characteristics and of permanent unobserved heterogeneity decline geometrically over time, the coefficients on $X_{i,g-1}$, $Z_{r,g-1}$, and ϕ_i are zero, and this first difference model is equivalent to a test score value added model:

$$y_{irg} = \gamma y_{ir,g-1} + \alpha_1 X_{ig} + \alpha_2 Z_{rg} + \Delta\lambda + \Delta\epsilon_{ir} \quad E[\Delta\epsilon_{ir}|y_{ir,g-1}, X_{ig}, Z_{rg}] = 0 \tag{13}$$

where Δ is an operator performing the weighted first difference ($\Delta\omega_i = \omega_{ig} - \gamma\omega_{i,g-1}$). $\Delta\lambda$ is now the constant, capturing the value added by 8^{th} grade with respect to 4^{th} grade. Under the identifying assumptions that observed inputs and unobserved student characteristics have geometrically declining effects over time, and that trends in temporary shocks ($\Delta\epsilon_{ir}$) are uncorrelated with lagged test scores and observed contemporary inputs, estimation of 13 by OLS yields consistent estimates of the test score production function parameters.

A.1.2 Effect of earthquake dummy: a DDD model

Consider the post-earthquake cohort of students. Test scores are observed at two points in time, and treatment occurs only in the second period. Moreover, only students in earthquake regions are affected ($E_i = 1$), those in non-earthquake regions are never affected ($E_i = 0$). In principle one could use a panel data difference-in-differences model (DD) in first differences to net out individual specific fixed effects and compare test score growth in $E_i = 1$ to test score growth in $E_i = 0$. Rather than a model in first differences, I use a test-score value added model because it is more in line with the literature on test score production functions. Section A.1.1 shows that this can be interpreted as a panel data model in *weighted* first differences. The test score value added model with earthquake treatment becomes:

$$y_{irg} = \gamma y_{ir,g-1} + \alpha_1 X_{ig} + \alpha_2 Z_{rg} + \delta(E_i \cdot d_g) + \Delta\lambda + \Delta\epsilon_{ir} \quad E[\Delta\epsilon_{ir}|y_{ir,g-1}, X_{ig}, Z_{rg}, E_i \cdot d_g] = 0 \tag{14}$$

where $d_g = 1$ for students in the second period (8^{th} grade), 0 otherwise. The vector of classroom and school characteristics, Z_{rg}, contains also MSK-Intensity in the school town. For δ to identify the causal impact of being in an earthquake region on test scores, being in an earthquake region is allowed to be correlated with permanent unobserved student characteristics (ϕ_i in 12), but it must be uncorrelated with unobserved individual trends in test scores ($\Delta\epsilon_{ir}$). In particular, there cannot be any differential trends in earthquake and non-earthquake regions.

Estimation of 14 on the pre-earthquake cohort yields a statistically significant and positive coefficient on the earthquake dummy: 0.064 for Math and 0.020 for Spanish.[57] Because the earthquake never happens in the pre-earthquake cohort, this is an indication that the earthquake region dummy is correlated with unobserved trends in test scores, violating the identifying assumption for the panel data DD model. Formally, $\delta^{pre} = E[\Delta\epsilon_{ir}^{pre}|y_{ir,g-1}, X_{ig}, Z_{rg}, E_i \cdot d_g = 1] - E[\Delta\epsilon_{ir}^{pre}|y_{ir,g-1}, X_{ig}, Z_{rg}, E_i \cdot d_g = 0]$, that is, δ in the pre-earthquake cohort captures the difference across regions in expected shock trends. This could be due to non-random locations of households, that is, the choice of living in an earthquake region or in a non-earthquake region is correlated with unobserved household characteristics affecting individual student trends in test scores.

To net out this potential bias, I use a panel data difference-in-differences-in-differences model (DDD), where the additional difference is across cohorts. Under the assumption that the difference in trends between regions is constant across cohorts, any earthquake region effect δ in the pre-earthquake cohort captures the bias that must be netted out from the earthquake region effects estimated on the post-earthquake cohort. This is in the spirit of the differentially adjusted estimator proposed by Bell, Blundell, and Van Reenen (1999). Formally, letting $T_i = 1$ if a student is in the post-earthquake cohort and $T_i = 0$ if she is in the pre-earthquake cohort, and allowing the parameters of the test score production function in 14 to change across cohorts, I estimate the following regression DDD model:

$$y_{irg} = \gamma y_{ir,g-1} + \theta_0 + \theta_1 T_i + \theta_2 X_{ig} + \theta_3 X_{ig} T_i + \theta_4 Z_{rg} + \theta_5 Z_{rg} T_i + \theta_6 (E_i \cdot d_g) + \theta_7 (E_i \cdot d_g \cdot T_i) + \zeta_{ir}, \tag{15}$$

where $\theta_0 = \Delta\lambda^{pre}$ is the effect of 8^{th} grade on test score growth in the pre-earthquake cohort, and θ_1 are cohort effects on grade effects, i.e. $\theta_1 = \Delta\lambda^{post} - \Delta\lambda^{pre}$. The model allows also for cohort effects on the coefficients of the test score production function: $\theta_3 = \alpha_1^{post} - \alpha_1^{pre}$ and $\theta_5 = \alpha_2^{post} - \alpha_2^{pre}$. Parameter θ_6 captures the correlation due to non-random location choices: $\theta_6 = \delta^{pre}$. The coefficient of interest is θ_7, equal to $\delta^{post} - \delta^{pre}$,

[57] These are from regressions with the full set of controls listed in the caption of Table 13. The p-value is 0.000. Full regression Tables available upon request. Notice that the coefficients on E_i in Table 13, columns (2) and (4), are very similar. This is expected, because, as shown below, they estimate the same object: the "effect" of the earthquake on test scores in the pre-earthquake cohort, δ^{pre}, due to non-random location choices.

which identifies the causal impact of the earthquake on test scores.[58] The identifying assumption is that residence in an earthquake region can be correlated with unobserved individual test score trends, however, the difference between regions in unobserved shock trends must be constant across cohorts:

$$\begin{aligned} & E[\Delta\epsilon_{ir}^{pre}|y_{ir,g-1}, X_{ig}, Z_{rg}, E_i \cdot d_g = 1] - E[\Delta\epsilon_{ir}^{pre}|y_{ir,g-1}, X_{ig}, Z_{rg}, E_i \cdot d_g = 0] \\ = \ & E[\Delta\epsilon_{ir}^{post}|y_{ir,g-1}, X_{ig}, Z_{rg}, E_i \cdot d_g = 1] - E[\Delta\epsilon_{ir}^{post}|y_{ir,g-1}, X_{ig}, Z_{rg}, E_i \cdot d_g = 0]. \end{aligned}$$

A.1.3 Effect of earthquake intensity: an application of Card (1992) and a placebo test of the identifying assumption

Card (1992) proposes a difference-in-differences estimator with continuous treatment, based on a fixed effect panel data model. The key identifying assumption is that treatment intensity can be correlated with individual permanent fixed effects, but not with individual unobserved trends.

In this context, consider data from the post-earthquake cohort. Moreover, restrict the sample to earthquake regions because the measure of treatment intensity (MSK-Intensity) is available only for those regions. Treatment occurs only in the second period ($d_g = 1$, i.e. 8^{th} grade), where all students are affected, but there is variation in the intensity of treatment, I_i.[59] The test score value added model becomes:

$$y_{irg} = \gamma y_{ir,g-1} + \alpha_1 X_{ig} + \alpha_2 Z_{rg} + \delta(I_i \cdot d_g) + \Delta\lambda + \Delta\epsilon_{ir}. \qquad (16)$$

where Δ is the operator performing weighted first differences defined in A.1.1. Notice that $I_i \cdot d_g$ is an interaction term, like $E_i \cdot d_g$, but here the interaction terms takes on a distinct value for each individual. Like in Card (1992), the identifying assumption is that $E[\Delta\epsilon_{ir}|y_{ir,g-1}, X_{ig}, Z_{rg}, I_i \cdot d_g] = 0$. That is, the differenced error term $\Delta\epsilon_{ir}$, capturing an individual trend in transitory shocks, must be uncorrelated with the intensity of treat-

[58] θ_7 can be expressed as

$$\begin{aligned} & \big[E[\Delta y|E = 1, T = 1, \cdot] - E[\Delta y|E = 0, T = 1, \cdot]\big] \\ - \ & \big[E[\Delta y|E = 1, T = 0, \cdot] - E[\Delta y|E = 0, T = 0, \cdot]\big]. \end{aligned}$$

[59] This variable plays the role of the fraction of teenagers in a state likely to be affected by the minimum wage increase in Card (1992), where the unit of observation is the state rather than the individual.

ment in 8^{th} grade, $I_i \cdot d_g$, conditional on lagged test score and on individual and classroom characteristics X and Z. Intensity $I_i \cdot d_g$ can be correlated with permanent unobserved heterogeneity, ϕ_i, but it cannot be correlated with individual specific unobserved trends in test scores, $\Delta\epsilon_{ir}$. Results are reported in the main text, in columns 1 and 2 in Table 3.

I perform a placebo test of the identifying assumption. I build a measure of future earthquake intensity in the pre-earthquake cohort of students, using the MSK formula based on location. I then estimate 16 in the pre-earthquake cohort of students. Because these students are never subject to the earthquake, the δ coefficient captures a correlation between household locations and unobserved test score trends, conditional on the other student and classroom characteristics. Simple descriptive statistics show that there is a correlation between earthquake intensity and student characteristics, with the intensity of shaking stronger in poorer areas.[60] However, the placebo test yields an insignificant $\hat{\delta}$ coefficient. Therefore, after conditioning on student observable characteristics such as parental income, a household's location has no residual predictive power on test score growth in a DD model like 16. This can be seen in columns 3 and 4 in Table 3, where I report results for the restricted sample of public school students to control for house type, as explained in section 3.1. This placebo test gives me confidence that the identifying assumption of the regression DD model with continuous treatment in 16 identifies the causal impact of earthquake intensity.

A.2 Effect of Peer Earthquake Exposure: Extending Card (1992) to Estimate Spill-overs in a Linear Model

First, I perform a placebo test that shows that identification would not be satisfied in a DD model like the one in the previous section, but where the treatment variables are now peer rather than individual intensity variables. Formally, I estimate the following regression on the pre-earthquake cohort of students:

$$y_{irg} = \gamma y_{ir,g-1} + \theta_0 + \theta_1 X_{ig} + \theta_2 Z_{rg} + \theta_3(\bar{I}_r \cdot d_g) + \theta_4(DI_r \cdot d_g) + \zeta_{irg} \tag{17}$$

where $\bar{I}_r$ is average damage in classroom r, DI_r is dispersion in damage in classroom r, and where the classroom characteristics vector Z_{rg} contains also MSK-Intensity in the

[60]For example, the correlation between MSK-Intensity and household income is -0.0201 and it is significant at the 0.001 level.

town of the school and, in some specifications, classroom composition variables such as mean and variance of income. There are two possible types of violations to the identifying assumption of the DD model: a statistically significant $\hat{\theta}_3$ would indicate correlation between average intensity and test score growth, while a statistically significant $\hat{\theta}_4$ would indicate correlation between dispersion in intensity and test score growth. Both types of violations would generate inconsistent estimates of the effects of interest if 17 was estimated on the post-earthquake cohort. Using standard deviation in MSK-Intensity as a measure of dispersion, Table 7 shows that $\hat{\theta}_3$ is statistically insignificant, while $\hat{\theta}_4$ is positive and statistically significant under all specifications.[61] Therefore, the ideal model should account for correlation between dispersion of MSK-Intensity in the classroom and trends in unobserved test score shocks.

Table 7: Placebo test: estimating the impact of mean and variance of future earthquake intensity in the pre-earthquake cohort. (Municipal schools sample)

	(1)	(2)	(3)	(4)
	Math TS	Math TS	Spanish TS	Spanish TS
$(\bar{I}_r \cdot d_g)$	0.002	-0.103	-0.012**	-0.109
	(0.004)	(0.108)	(0.004)	(0.178)
$(DI_r \cdot d_g)$	0.232***	0.137***	0.271***	0.178***
	(0.030)	(0.041)	(0.032)	(0.044)
Lagged TS	0.694***	0.672***	0.713***	0.699***
	(0.003)	(0.004)	(0.003)	(0.005)
Controls	No	Yes	No	Yes
Observations	45,814	26,145	46,127	25,628

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Controls: whether the student lives in the same town where the school is, gender, mother's education, father's education, household income, intensity of earthquake in home town and in school town, class size, whether the Math or Spanish teacher is female, has a postgraduate degree, has a permanent contract, her tenure at the school, her teaching experience. A constant is always included.

Like in the model presented in A.1.2 to estimate the effect of the earthquake dummy, in this DDD model the last difference refers to the test score value added, a weighted first difference, and the first difference is across cohorts. Unlike that model, the second

[61] $\hat{\theta}_3$ is statistically significant but small only for Spanish test scores and only when no controls are used.

difference here is a derivative, because of the continuous nature of treatment. Specifically, I estimate the following regression model:

$$\begin{aligned} y_{irg} \;&=\; \gamma y_{ir,g-1} + \theta_0 + \theta_1 T_i + \theta_2 X_{ig} + \theta_3 X_{ig} T_i + \theta_4 Z_{rg} + \theta_5 Z_{rg} T_i + \\ &\quad \theta_6(\bar{I}_r \cdot d_g) + \theta_7(\bar{I}_r \cdot d_g)T_i + \theta_8(DI_r \cdot d_g) + \theta_9(DI_r \cdot d_g)T_i + \zeta_{irg} \end{aligned} \tag{18}$$

Parameter θ_1 captures direct cohort effects on test score growth, and θ_3 and θ_5 capture cohort effects on the coefficients of the test score production function. Together, they capture any change in the test score production function that occurred across cohorts, for example, induced by nation-wide policies such as the 2008 reform to the voucher system. The parameters of interest are the coefficients on the triple interactions: θ_7, the effect on own test score of a marginal increase in average earthquake intensity in the classroom, and θ_9, the effect on own test score of a marginal increase in intensity dispersion in the classroom.

The identifying assumption is that $E[\zeta_{irg}|T_i, X_{ig}, Z_{rg}, \bar{I}_r \cdot d_g, DI_r \cdot d_g] = 0$, which is equivalent to $E[\Delta\epsilon_{ir}^{post}|\bar{I}_r \cdot d_g, DI_r \cdot d_g, X_{ig}, Z_{rg}] = E[\Delta\epsilon_{ir}^{pre}|\bar{I}_r \cdot d_g, DI_r \cdot d_g, X_{ig}, Z_{rg}]$, where Δ is an operator performing weighted first differences as explained above. In particular, average and dispersion in damages in the classroom are allowed to be correlated with trends in temporary shocks, but this correlation must be constant across cohorts. Under this identifying assumption, θ_7 and θ_9 identify the causal impacts of the peer variables. They can be expressed as follows:

$$\begin{aligned} \theta_7 \;&=\; \frac{\partial E[y_{ir}|T_i = 1, DI_r \cdot d_g, \bar{I}_r \cdot d_g, X_{ig}, Z_{rg}]}{\partial \bar{I}_r \cdot d_g} - \frac{\partial E[y_{ir}|T_i = 0, DI_r \cdot d_g, \bar{I}_r \cdot d_g, X_{ig}, Z_{rg}]}{\partial \bar{I}_r \cdot d_g} \\ \theta_9 \;&=\; \frac{\partial E[y_{ir}|T_i = 1, \bar{I}_r \cdot d_g, DI_r \cdot d_g, X_{ig}, Z_{rg}]}{\partial DI_r \cdot d_g} - \frac{\partial E[y_{ir}|T_i = 0, \bar{I}_r \cdot d_g, DI_r \cdot d_g, X_{ig}, Z_{rg}]}{\partial DI_r \cdot d_g} \end{aligned}$$

where it is clear that potential biases are netted out.

A.2.1 Robustness of the findings from the linear model of spill-overs

Controlling for distribution of student characteristics in the classroom. Table 8 shows results from the linear model of spill-overs when student composition controls are included. Dispersion is measured by the standard deviation in MSK-Intensity. The first and third row in each panel demonstrate that the main results hold: dispersion in intensity has a large and negative impact on test scores even after controlling for variables that could be correlated with geographic dispersion in the classroom, such as standard deviation of parental incomes, parental education, and student academic ability. Moreover, the estimated effect of mean intensity is always insignificant in the preferred sample of Municipal schools. It is significant only in the sample of all schools, as seen in the fourth column, where, however, the point estimate is very small. Finally, the second row of each panel shows the pre-existing correlation between geographic dispersion and trends in temporary shocks, which is always positive, as discussed in the main text.

Controlling for school level dispersion in MSK-Intensity. To verify that the classroom is the relevant level of analysis rather than the school, I add to the model in 18 two regressors: variance of MSK-Intensity at the school level, and this variance interacted with the treatment indicator ($T_r \cdot d_g$). If the impact of the damage dispersion in the classroom is in fact due to damage dispersion at the school level, introducing these additional regressors should drive the coefficient θ_9 (effect of classroom damage dispersion) to zero. As shown in Table 9, while variance at the school level has a separate, imprecisely estimated and often insignificant effect, the damage dispersion effect at the classroom level remains highly significant and similar in magnitude to the model without controls for school-wide variance in damages, for all specifications.[62] Therefore, the classroom damage dispersion effect is not explained away by school level damage dispersion, indicating that the classroom is a relevant level of analysis.

A.2.2 Testing the identifying assumption of the linear model of spill-overs using data from regions not affected by the earthquake

The linear model finds an intensity dispersion effect on test scores. Identification hinges on the key assumption that the correlation between students' geographic dispersion and unobserved test score trends would be constant across cohorts in the absence of treatment, that is, if the earthquake did not happen. To test this assumption, I must be able to estimate in both cohorts the effect of geographic dispersion in the absence of damage dispersion. This is possible using data from the regions which were not affected by the

[62] First, compare the first row of the second panel in table 9, to the third row of the first panel of Table 4 in the main text. These estimates are the effect of damage standard deviation at the classroom level in the preferred sample of Municipal school, with and without controlling for the effect of damage standard deviation at the school level. They are very similar. Second, when controlling for school level damage dispersion, the effect of classroom level dispersion remains significant always except in the specifications with controls in the (least preferred) sample of all schools. Given that the main nonlinear analysis is performed on the sample of Municipal schools, this is not a concern.

Table 8: Linear model of spill-overs, controlling for student composition variables.

	(1) Math TS	(2) Math TS	(3) Spanish TS	(4) Spanish TS
Sample of all schools				
$(\bar{I}_r \cdot d_g) \times T_i$	0.004	-0.008	0.002	-0.016**
	(0.005)	(0.005)	(0.005)	(0.005)
$(DI_r \cdot d_g)$ (st. dev.)	0.086***	0.0803**	0.141***	0.127***
	(0.024)	(0.025)	(0.025)	(0.027)
$(DI_r \cdot d_g) \times T_i$ (st. dev)	-0.113***	-0.072*	-0.140***	-0.089**
	(0.030)	(0.031)	(0.032)	(0.034)
Observations	263,723	156,851	263,655	153,390
Sample of Municipal schools				
$(\bar{I}_r \cdot d_g) \times T_i$	0.008	-0.004	0.011	-0.008
	(0.006)	(0.007)	(0.007)	(0.007)
$(DI_r \cdot d_g)$ (st. dev.)	0.103**	0.125**	0.223***	0.190***
	(0.035)	(0.041)	(0.034)	(0.045)
$(DI_r \cdot d_g) \times T_i$ (st. dev)	-0.101*	-0.110*	-0.209***	-0.171**
	(0.044)	(0.049)	(0.047)	(0.053)
Observations	105,646	58,615	105,819	56,643
Composition controls	Yes	Yes	Yes	Yes
Additional controls	No	Yes	No	Yes

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Composition controls: classroom average and standard deviation of: father's education, mother's education, household income, lagged Math test scores and lagged Spanish test scores.

Additional controls: whether the student lives in the same town where the school is, mother's education, father's education, household income, intensity of earthquake in hometown and in school town, gender, lagged test score in Math or Spanish, class size, the teaching experience of the Math or Spanish teacher, whether he/she is female, has a postgraduate degree, has a permanent contract, her tenure at the school, average MSK-intensities among classmates, cohort dummy, and, for the sample of all schools, whether the school is public. A constant is always included.

earthquake. Ideally, one would estimate the same linear spill-over model from the main analysis using non-earthquake regions only. Evidence in favour of the identifying assumption would be a zero coefficient on the triple interaction $(DI_r \cdot d_g) \times T_i$. However, the measure of dispersion DI_r used in the main analysis is based on MSK-Intensity, which is not available for non-earthquake regions (Astroza, Ruiz, and Astroza 2012). Therefore, I need to find a measure of dispersion that is valid across both earthquake and non-

Table 9: Linear model of spill-overs, controlling for damage dispersion effects at the school level.

	(1) Math TS	(2) Math TS	(3) Spanish TS	(4) Spanish TS
Sample of all schools				
$(DI_r \cdot d_g) \times T_i$ (st. dev)	-0.212***	-0.104**	-0.228***	-0.037
	(0.042)	(0.037)	(0.042)	(0.039)
$(DI_R \cdot d_g) \times T_i$ (st. dev. school)	0.015	0.104*	0.001	-0.121*
	(0.057)	(0.049)	(0.057)	(0.053)
Observations	275,319	156,937	275,590	153,477
Sample of Municipal schools				
$(DI_r \cdot d_g) \times T_i$ (st. dev)	-0.219***	-0.184***	-0.273***	-0.182**
	(0.054)	(0.055)	(0.057)	(0.059)
$(DI_R \cdot d_g) \times T_i$ (st. dev. school)	-0.133 +	0.115	-0.181*	-0.006
	(0.075)	(0.074)	(0.079)	(0.079)
Observations	110,075	58,661	110,500	56,687
Controls for school-wide damage variance	Yes	Yes	Yes	Yes
Additional controls	No	Yes	No	Yes

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

School damage variance controls: school variance of MSK-Intensity, alone and interacted with treatment indicator $d_g \cdot T_i$.

Additional controls: same as in Table 8.

earthquake regions.

I define a classroom as geographically homogeneous if all students reside in the same town, and geographically dispersed if at least one student comes from a different town. In the entire sample, 55 percent of schools are geographically dispersed according to this measure. This figure is 58 percent for schools in earthquake regions, and 42 percent for schools in non-earthquake regions. To ensure that the choice of measure does not drive the results, I first repeat the main analysis on earthquake regions with a linear spill-over model that uses this new dispersion measure. The main difference with the main DDD model is that now dispersion is a discrete variable. Letting G_r indicate the discrete geographic dispersion measure, I estimate the following regression DDD model:

$$\begin{aligned} y_{ir} &= \gamma y_{ir,g-1} + \theta_0 + \theta_1 T_i + \theta_2 X_{ig} + \theta_3 X_{ig} T_i + \theta_4 Z_{ig} + \theta_5 Z_{rg} T_i + \\ &\quad \theta_6 (G_r \cdot d_g) + \theta_7 (G_r \cdot d_g) \times T_i + \zeta_{irg}. \end{aligned} \tag{19}$$

The results are qualitatively similar to the continuous treatment DDD model used in the main text, as can be seen in the bottom panel of Table 10. Unsurprisingly, the magnitudes of are different because of the different definition of geographic dispersion. In earthquake regions, a larger geographic dispersion is estimated to have a negative impact on test scores (the estimate is significant in all but one specifications). Because these regions are affected by the earthquake, this could reasonably be due to dispersion in damages within the classroom.

This is reassuring, because it indicates that the discrete measure of geographic dispersion is capable of detecting changes across cohorts in dispersion effects, if they are present. Therefore, if no effects are found in non-earthquake regions, there were no changes across cohorts in dispersion effects. Results from estimation on non-earthquake regions are in the top panel of Table 10. Reassuringly, the third row shows that the estimate of θ_7 is always small and always statistically insignificant in models with covariates (it is only significant in the model without covariates in columns 3). This indicates that in the richer models the identifying assumption of common correlation between geographic dispersion and test score trends across cohorts is justified.

B Appendix: Nonlinear Model. FOR ONLINE PUBLICATION ONLY

B.1 Estimation of the Semi-parametric Model

B.1.1 Algorithm for the Estimation of the Semi-parametric Single-Index Model (Notice: notation must be changed for consistency with main text)

1. Normalize to a constant one of the elements of θ, because only the ratios among the components of θ are identified. I normalize to -1 the coefficient on lagged test score (θ_1).

2. Make an initial guess for all the other elements of θ.

3. Form c_i $\forall i$ according to $c_i = -y_{i,t-1} + \theta_2 parental\_educ_i + \theta_3 income_i + \theta_4 gender_i$ if i belongs to the pre-earthquake cohort, and $c_i = -y_{i,t-1} + \theta_2 parental\_duc_i + \theta_3 income_i + \theta_4 gender_i + \theta_5 I_i + \theta_6 I_i X_i^-$ if i belongs to the post-earthquake cohort. X_i^- includes household income and student gender.

Table 10: Testing the linear model of spill-overs using non-earthquake regions. Sample of Municipal schools.

	(1) Math TS	(2) Math TS	(3) Spanish TS	(4) Spanish TS
Non-earthquake regions				
$(DI_r \cdot d_g)$	0.015	-0.038**	0.031*	-0.010
	(0.012)	(0.015)	(0.013)	(0.015)
T_i	0.025*	0.002	0.053***	0.006
	(0.011)	(0.011)	(0.011)	(0.012)
$(DI_r \cdot d_g) \times T_i$	-0.014	0.013	-0.030+	-0.010
	(0.017)	(0.018)	(0.018)	(0.019)
Lagged TS		0.664***		0.679***
		(0.005)		(0.005)
Female		-0.084***		0.143***
		(0.008)		(0.009)
Mother's Education		0.008***		0.010***
		(0.002)		(0.002)
Observations	47,396	23,473	47,253	23,298
Earthquake regions				
$(DI_r \cdot d_g)$	0.145***	0.027**	0.125***	0.010
	(0.008)	(0.008)	(0.008)	(0.009)
T_i	0.027***	0.008	0.031***	-0.009
	(0.008)	(0.008)	(0.009)	(0.008)
$(DI_r \cdot d_g) \times T_i$	-0.088***	-0.034**	-0.094***	-0.017
	(0.011)	(0.011)	(0.012)	(0.012)
Lagged TS		0.666***		0.684***
		(0.003)		(0.003)
Female		-0.096***		0.142***
		(0.005)		(0.006)
Mother's Education		0.010***		0.012***
		(0.001)		(0.001)
Observations	110,320	58,783	110,748	56,805
Controls	No	Yes	No	Yes

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Dispersion DI_r is a dummy variable equal to 1 if at least one student in the class comes from a different town.

Controls: whether the student lives in the same town of the school, father's education, household income, class size, the teaching experience and tenure at the school of of (resp.) the Math and Spanish teacher, whether he/she is female, has a postgraduate degree, has a permanent contract. A constant is included.

4. Estimate $E(y_i|c,l;\theta)\ \forall l$ by Nadaraya-Watson kernel regression with weights w_i:

$$\hat{h}_r(c;\theta) = \frac{\sum_{i\in r} w_i K\left(\frac{c_i-c}{b}\right) y_i}{\sum_{i\in r} w_i K\left(\frac{c_i-c}{b}\right)}$$

with a standard normal Kernel: $K(\psi) = (2\pi)^{-\frac{1}{2}} exp(-0.5\psi^2)$ and optimal bandwidth $b = 1.06\hat{\sigma}_c n^{-1/5}$, minimizing the Approximated Mean Integrated Squared Error (AMISE).[63] The weights w_i are such that only observations i where the p.d.f. of c at c_i exceeds a small positive number are used (see Ichimura (1993) and Horowitz (2010)). Observation i is excluded from the calculation of $\hat{h}$ at c_i.

5. Compute the sum of squared residuals in each r at the current guess for θ: $SSR_r(\theta) = \sum_{i\in r} w_i(y_i - \hat{h}_r(c_i;\theta))^2$. The weights are the same as those used in the kernel estimator of h.

6. Update guess for θ using Generating Set Search algorithm (HOPSPACK).

7. Repeat steps 1-6 until convergence to the minimizer of $\sum_r SSR_r(\theta)$.

Notice that unlike the standard semiparametric single-index model, here the $SSR(\theta)$ is computed in each classroom r, and its sum over classrooms is minimized. The dataset is clustered at the classroom level. While the functions h are allowed to differ by classrooms, the parameter θ is restricted to be identical in all classrooms. This assumption improves efficiency of the estimator of θ. To account for the clustered sample design in the estimation of the standard errors of the θ parameters, I bootstrap 100 samples stratified at the classroom level, and I estimate θ in each bootstrapped sample to obtain the standard errors.

The standard errors of $\widehat{TE(c)}$, which are needed to test the comparative statics result, cannot be easily bootstrapped for computational reasons.[64] Instead, I use the result in Ichimura (1993), who proves that the asymptotic variance of $\hat{h}_r(c)$ in the appropriately weighted semiparametric single-index model above is identical to the asymptotic variance of a non-parametric conditional mean estimator. The variance of such estimator is $V(\hat{h}_r(c)) = \frac{\sigma_r^2}{n_r h_r f_r(c)} \int K^2(\psi)d\psi + o(n^{-1}b_r^{-1})$, where σ^2 is the variance of ϵ_{ir}, b_r is the bandwidth, n_r is the size of classroom r (on average this is around 30), and $f_r(c)$ is the density

[63] The MISE is equal to $E\{\int[\hat{h}(c)-h(c)]^2dx\} = \int\left[(Bias\hat{h})^2 + V(\hat{h})\right]dc$, and AMISE substitutes the expressions for the bias and variance of $\hat{h}$ with approximations. See Pagan and Ullah (1999), p. 24.

[64] This would require submitting around 4,000 jobs of duration 72 hours each.

at c in classroom r. The kernel $K(\cdot)$ is the normal kernel, resulting in $\int K^2(\psi)d\psi = 0.2821$. I estimate the asymptotic variance of $\hat{h}_r(c)\ \forall r$ on a fine grid for c. I substitute $f(c)$ with its kernel estimator, and σ_r^2 with its estimator obtained by averaging the squared residuals in each classroom: $\hat{\sigma}_r^2 = \frac{\sum_{i \in r}(y_i - \hat{y}_i)^2}{n_r - 1}$. I assume that the covariances between the $\hat{h}_r(c)$ belonging to different classrooms r are zero $\forall c$, and I obtain the following expression for the variance of $\widehat{TE(c)}$:

$$V\left(\widehat{TE(c)}\right) = \sum_{r=1}^{N^{pre}-1} \sum_{r'=l+1}^{N^{pre}} \sum_{s=1}^{N^{post}-1} \sum_{s'=s+1}^{N^{post}} \kappa_{rr'ss'}^2 \left(V\left(\hat{h}_s^{post}(c)\right) + V\left(\hat{h}_{s'}^{post}(c)\right) + V\left(\hat{h}_r^{pre}(c)\right) + V\left(\hat{h}_{r'}^{pre}(c)\right)\right).$$

The weights $\kappa_{rr'ss'}$ are given by:

$$\kappa_{rr'ss'} = \frac{\omega_{rr'ss'}}{\sum_{r=1}^{N^{pre}-1} \sum_{r'=r+1}^{N^{pre}} \sum_{s=1}^{N^{post}-1} \sum_{s'=s+1}^{N^{post}} \omega_{rr'ss'}}$$

where $\omega_{rr'ss'}$ is defined in equation 20 below.

B.1.2 Kernel Weighting

To ensure that the classrooms are similar, I assign increasing weights to quadruplets that are more similar in terms of W and δ. I construct weights using multivariate standard normal kernel functions. As in the main text, let ss' index a pre-earthquake classroom pair, and rr' a post-earthquake classroom pair. Letting $t = r, r', s, s'$ I assign the weight $\frac{1}{b}k\left(\frac{W_t - W_{t'}}{b}\right)$ to each of the pairs $tt' \in \{rr', ss', rs\}$. This ensures similarity between pairs within ($tt' = rr', ss'$) and across ($tt' = rs$) cohorts.[65] Finally, I build a weight that is declining in $|\delta_{ss'} - \delta_{rr'}|$, to guarantee that the pre- and post-earthquake pairs differ in terms of geographic dispersion δ in a similar way: $\frac{1}{b_\delta}k\left(\frac{\delta_{ss'} - \delta_{rr'}}{b_\delta}\right)$. The weight for the quadruple, $\omega_{rr'ss'}$, is the product of these four kernel weights:

$$\omega_{rr'ss'} = d_{rr'ss'} \frac{1}{b_\delta} k\left(\frac{\delta_{ss'} - \delta_{rr'}}{b_\delta}\right) \prod_{tt' \in \{rr', ss', sr\}} \frac{1}{b} k\left(\frac{W_t - W_{t'}}{b}\right) \tag{20}$$

where $d_{rr'ss'}$ is a dummy variable equal to one if $\delta_{rr'} > 0$ and $\delta_{ss'} > 0$, zero otherwise.

[65] I use a unique bandwidth b. Following Pagan and Ullah (1999), I normalize the elements in W_t so that they all have the same standard deviation and using a unique bandwidth is admissible.

B.2 Robustness to Fixed Effects

B.2.1 Robustness to Fixed Effects: Proofs

Proof of Proposition 6.1.
Suppose that the true Data Generating Process is the nonlinear model in 4, augmented with unobserved school×cohort effects α_{RT}:

$$\begin{aligned} Y_{irR} &= H(c^{T_r}(X_i), W_r, T_r, \sigma_r^2, R) + \epsilon_{irR} \\ &= h(c^{T_r}(X_i), W_r, T_r, \sigma_r^2) + \alpha_{RT} + \epsilon_{irR} \\ &= \phi(c_i, W_r) + \lambda^{GD}(c_i, W_r, \sigma_L^2) + T_r \cdot \left[\lambda^E(c_i, W_r) + \lambda^{DD}(c_i, W_r, \sigma_L^2)\right] + \\ &\quad + G_{r,r'} \cdot \lambda^{GD}(c_i, W_r, \delta) + T_r \cdot G_{rr'} \cdot \lambda^{DD}(c_i, W_r, \delta) + \alpha_{RT} + \epsilon_{irR} \end{aligned} \tag{21}$$

To keep track of the school each classroom is in, I use an upper case letter for the school, so, for example, classroom r is in school R, classroom r' is in school R', et cetera. The fixed effect α_{RT} is identified from the distribution of achievement Y conditional on student characteristics X in schools with multiple classrooms that are identical in terms of $W_r, T_r, \sigma_r^2, \forall r \in R$. Consider one quadruplet of classrooms, r, r' from the pre-earthquake cohort, l, l' from the post-earthquake cohort, sharing the same W and with $G_{rr'} = 1$, $G_{ss'} = 1$, and $\delta_{rr'} = \delta_{ss'} = \delta$. I drop the T subscript from the fixed effect because the school index now also uniquely identifies the cohort. Conditional on a point c, on W and on δ, the double difference now yields:

$$\begin{aligned} \Big(H_r(c_i) - H_{r'}(c_i)\Big) - \Big(H_s(c_i) - H_{s'}(c_i)\Big) &= \Big(\lambda^{GD}(c_i; W, \delta) + \lambda^{DD}(c_i; W, \delta)\Big) - \lambda^{GD}(c_i; W, \delta) + \\ &\quad + (\alpha_R - \alpha_{R'}) - (\alpha_S - \alpha_{S'}) \\ &= \lambda^{DD}(c_i; W, \delta) + (\alpha_R - \alpha_{R'}) - (\alpha_S - \alpha_{S'}) + \zeta_{iRR'SS'} \end{aligned} \tag{22}$$

There are two cases two consider. First, when $R = R'$ and $S = S'$, that is, when within cohorts the pairs of classrooms are in the same school, school×cohort fixed effects would not affect the unbiasedness of the nonlinear estimator, nor its efficiency, because the fixed effects cancel out. No restrictions on the fixed effects would have to be imposed. However, only 37 percent of schools in the sample have more than one classroom, hence, it is reasonable to expect that there is only a small number of within cohort matched classroom pairs that belong to the same school. Therefore, I consider the properties of the nonlinear model in the more empirically relevant case that $R \neq R'$ or $S \neq S'$ or both.

In this case, it is easy to see that the conditional expectation of the double difference in 22, $E\Big[\Big(H_r(c_i) - H_{r'}(c_i)\Big) - \Big(H_s(c_i) - H_{s'}(c_i)\Big)|c, W, \delta\Big]$, is equal to the conditional damage dispersion effect, $\lambda^{DD}(c_i; W, \delta)$, if and only if

$$E\Big[(\alpha_R - \alpha_{R'}) - (\alpha_S - \alpha_{S'})|c, W, \delta\Big] = 0 \quad \forall c, W, \delta, \tag{23}$$

where the expectation is taken with respect to the distribution of school×cohort fixed effects. In turn, when the condition in 23 is true, equation 7 identifies the treatment effect of interest, $TE(c)$, and the expectation of the sample mean in 8 is equal to $TE(c)$ for all values of c. That is, the nonlinear estimator is unbiased. To see why, notice that when the true DGP includes fixed effects, equation 8 is equivalent to:

$$\begin{aligned} T\hat{E}(c) \; = \; & \frac{\sum_{s=1}^{N^{pre}-1}\sum_{s'=s+1}^{N^{pre}}\sum_{r=1}^{N^{post}-1}\sum_{r'=r+1}^{N^{post}}\omega_{rr'ss'}(\Delta^{post}\hat{h}_{rr'}(c;W,\delta) - \Delta^{pre}\hat{h}_{ss'}(c;W,\delta))}{\sum_{s=1}^{N^{pre}-1}\sum_{s'=s+1}^{N^{pre}}\sum_{r=1}^{N^{post}-1}\sum_{r'=r+1}^{N^{post}}\omega_{rr'ss'}} + \\ & + \frac{\sum_{s=1}^{N^{pre}-1}\sum_{s'=s+1}^{N^{pre}}\sum_{r=1}^{N^{post}-1}\sum_{r'=r+1}^{N^{post}}\omega_{rr'ss'}(\hat{\alpha}_R - \hat{\alpha}_{R'}) - (\hat{\alpha}_S - \hat{\alpha}_{S'})}{\sum_{s=1}^{N^{pre}-1}\sum_{s'=s+1}^{N^{pre}}\sum_{r=1}^{N^{post}-1}\sum_{r'=r+1}^{N^{post}}\omega_{rr'ss'}}. \end{aligned}$$

The second line is the empirical counterpart of $E\Big[E\Big[(\alpha_R - \alpha_{R'}) - (\alpha_S - \alpha_{S'})|c, W, \delta\Big]\Big]$, and, under condition 23, its expectation is equal to 0 by the central limit theorem. Additionally, as the number of schools goes to infinity, the second line of the expression above converges to zero (its population counterpart under 23) by the weak law of large numbers for independent and not identically distributed random variables.[66] Therefore, when there are school×cohort fixed effects in the DGP, condition 23 is a necessary and sufficient condition for the identification of the conditional treatment effect $\lambda^{DD}(c_i; W, \delta)$, and it is sufficient for the identification of the unconditional treatment effect of interest, $TE(c)$. Finally, using the definitions of R, R', S and S', condition 23 can be rewritten as it appears in the main text:

$$\begin{aligned} & \Big[E[\alpha_{MT}|T_m = 1, G_{MM'} = 1, c, W, \delta] - E[\alpha_{MT}|T_M = 1, G_{MM'} = 0, c, W, \delta]\Big] = \\ = \; & \Big[E[\alpha_{MT}|T_m = 0, G_{MM'} = 1, c, W, \delta] - E[\alpha_{MT}|T_M = 0, G_{MM'} = 0, c, W, \delta]\Big] \quad \forall c, W, \delta. \end{aligned}$$

[66] The elements of the average in the second line of the expression above are not identically distributed because the sampling variance of each element depends on the school size, that is, on the sample size on which the fixed effects that enter the double difference are calculated. It is the number of schools that must go to infinity for convergence because the expectation in 23 is taken with respect to the school fixed effect distribution.

In addition, notice that the school effects may have heterogeneous impacts on students. To see why, replace α_{RT} with a function $\alpha_{RT}(c)$ and all derivations above hold true.

Proof of Proposition 6.2.

The sufficient condition for identification with fixed effects in 9 must be true for every δ. In particular, it must necessarily be true for $\delta \to 0$, in which case, if the conditional expectation of the fixed effect is differentiable in σ_m^2, it must be that

$$\frac{\partial E[\alpha_M|T_M = 1, c, W, \sigma_m^2]}{\partial \sigma_m^2} = \frac{\partial E[\alpha_M|T_M = 0, c, W, \sigma_m^2]}{\partial \sigma_m^2} \quad \forall m \in M, \forall \sigma_m^2, c, W. \tag{24}$$

A specification for the conditional expectation of the fixed effects that is useful for testing is the special case in which this expectation is linear in σ^2 and T. In this case, condition 24 is also sufficient for identification of $TE(c)$ when there are school×cohort fixed effects in the DGP.[67] A further simplifying (but not necessary) assumption is that W and c enter in an additively separable way. Condition 24 under this model of fixed effects is equivalent to $\beta_3 = 0$ in:

$$\alpha_{MT} = \beta_0 + \beta_1\sigma_m^2 + \beta_2 T_M + \beta_3\sigma_m^2 T_M + g(W, c; \beta_4) + \epsilon_{mM}. \tag{25}$$

Therefore, $\beta_3 = 0$ is a sufficient condition for identification in the presence of fixed effects, when the fixed effects follow the specification in 25.

B.2.2 Robustness to Fixed Effects: Empirical Test

See Table 11.

B.3 Testing Monotonicity of $h(c)$

The procedure that I use is an application of Chetverikov (2013). It would be computationally unfeasible to perform the test in all classrooms. Therefore, I create 72 categories of classrooms that have similar distributions of c_i, and test monotonicity within each category. I consider classroom categories containing approximately 60 classrooms each. Classrooms in the same category share similar mean and variance of c_i. Therefore, the theoretical model predicts that the equilibrium achievement functions $y(c)$ should be very similar across classrooms within each category. The monotonicity of $h(c)$, the empirical

[67] To see why, notice that if $E[\alpha_M|T_M, c, W, \sigma_m^2]$ is linear in σ_m^2 and it has the same slope under $T_M = 1$ and $T_M = 0$, then condition 9 is satisfied for all values of δ and not only for $\delta \to 0$.

Table 11: Empirical test of identification in the presence of school×cohort fixed effects

	(1) Maths Predicted Fixed Effects	(2) Spanish Predicted Fixed Effects
β_1	-0.310	0.348
	(0.276)	(0.277)
β_2	1.428***	-0.352***
	(0.0475)	(0.0485)
β_3	0.471	-0.143
	(0.344)	(0.348)
Constant β_0	-2.284***	-1.269***
	(0.173)	(0.179)
School and student controls	Yes	Yes
Observations	1,810	1,778

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Controls: school level averages of father's education and income, teaching experience of (resp.) Math and Spanish teacher, and MSK-Intensity at the school.
β_3 in equation 11 is not statistically different from zero, therefore, the identifying condition under school×cohort fixed effects is not rejected by the data.
β_3 is insignificant even in the absence of controls (tables available upon request).

counterpart of $y(c)$, is tested within each one of these categories. Separating the sample in categories makes this procedure feasible from a computational point of view. In all categories, the null hypothesis that the h function is decreasing is not rejected at the $\alpha = 0.10$ significance level.

Technical details on the test's implementation and formulae can be found in the Online Supplementary Material on the author's webpage. I follow the choice of bandwidth recommended in Ghosal, Sen, and Van Der Vaart (2000), and I adopt the plug-in approach to simulate the critical values.

An important distinction with Chetverikov (2013) is that the c_i values in my sample are estimated (and not observed); $\hat{c}_i = \hat{\theta}^{T_r} X_i$. However, this additional noise is asymptotically negligible because the bandwidth used in the kernel weighting functions goes to zero as the sample size increases, and because $\hat{\theta}$ is root-n consistent (as shown in Ichimura (1993)), therefore, it is faster than the nonparametric rates appearing in the derivations in Chetverikov (2013).

B.4 Correlation between Earthquake Damage and Academic Skills

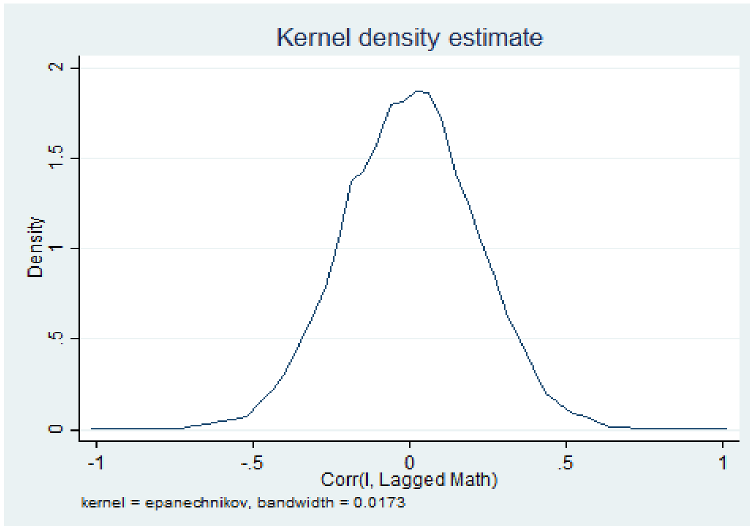

Figure 9: Distribution of the within classroom correlation between lagged Mathematics test scores and MSK-Intensity at the student's home. There is considerable sampling variation due to the finite size of classrooms. The distribution for the case of lagged Spanish test scores is very similar.

In the population, there is only a small correlation between a student's academic skills (as measured by lagged test scores) and the intensity of the earthquake in her home town. However, because of sampling variation, within many classrooms this correlation takes on larger values, as can be seen in Figure (9). I use this sampling variation to estimate the (causal) impact on achievement of the classroom correlation between earthquake damage and student academic skills. I estimate regressions of this form:

$$y_{ir} = \beta_0 + \beta_1(C_r \cdot d_g) + \beta_2 T_i + \beta_3(C_r \cdot d_g)T_i + \alpha_1 X_i + \alpha_2 Z_{ir} + \alpha_3 I_i + \epsilon_{ir} \tag{26}$$

where C_r is the correlation in classroom r between MSK-Intensity and student academic skills, as measured by lagged test scores. MSK-Intensity is available only in earthquake regions so this estimation is performed on those regions only. A positive correlation means that it is the more academically skilled students who sustain larger damages, while a negative correlation means that it is the least academically skilled ones who do so. The parameter of interest is β_3, the causal impact of larger correlation on achievement, which can be expressed as: $\beta_3 = \frac{\partial E[y|C,T,X,Z]}{\partial C \cdot d}|_{T=1} - \frac{\partial E[y|C,T,X,Z]}{\partial C \cdot d}|_{T=0}$. Like in the linear model in A.2, the common location effects assumption is invoked for identification. Here, location effects are defined as the (spurious) effect of C_r on test scores due solely to the location of students.

Table (12) reports various estimates for β_3 under different specifications of (26), and using different sub-samples of the data. While for Spanish test scores no patterns are detectable, for Mathematics an interesting pattern emerges. When it is the more highly skilled students who are affected more harshly by the earthquake, students do on average worse, *conditional* on the level of damage at a student's own home, on a student's academic skill and on the skill composition in the classroom.

Table 12: Effect on Math (Spanish) test scores of the correlation in the classroom between ability in Math (Spanish) and MSK-intensity.

	(1) All sample	(2) Quartile 1	(3) Quartile 2	(4) Quartile 3	(5) Quartile 4
Mathematics Municipal Schools					
Fewer Controls	-0.0687*	0.0488	-0.0515	-0.0555	-0.2664***
	(0.0230)	(0.0491)	(0.0546)	(0.0646)	(0.0793)
More Controls	-0.0743*	0.0185	-0.0510	-0.0615	-0.2297**
	(0.0336)	(0.0568)	(0.0613)	(0.0718)	(0.0851)
Spanish Municipal Schools					
Fewer Controls	0.0032	0.0474	0.0219	0.0275	-0.1184
	(0.0311)	(0.0524)	(0.0582)	(0.0655)	(0.0710)
More Controls	-0.0189	0.0389	-0.0482	0.0062	-0.0789
	(0.0354)	(0.0612)	(0.0666)	(0.0735)	(0.0877)

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Only earthquake regions included in the samples as MSK-Intensity is available only in these regions. The first line of each panel shows coefficient estimates from regressions where the only individual controls are lagged Mathematics (top) or Spanish (bottom) test score and MSK-intensity at the student's home. The second line refers to regressions with additional individual controls. The quartile columns refer to regressions on sub-samples defined by the quartiles of the lagged Math (top panels) or Spanish (bottom panels) test scores distribution in the sample regions. Adding student composition controls to these tables (mean and variance of lagged test scores, etc.) does not change the results. Full regression Tables available upon request.

B.5 Relationship with Nonlinear Difference in Differences Models

Nonseparability in X is due to the fact that student type c_i is determined differently in the pre- and post-earthquake cohort. As a result, the outcome function cannot be expressed as a component that only depends on damage dispersion status (high or low) and one that only depends on cohort, conditional on X. This is shown in subsection B.5.1 below. As described in footnote 44 in the main text, this feature of the model is important to make the correct comparisons between treated, post-earthquake students and control, pre-earthquake students. Separability is assumed in the QDID model, as well as in the models that, like the model in this paper, combine matching with differences in differences, that is, Blundell and Costa Dias (2000), Bell, Blundell, and Van Reenen (1999), Blundell, Costa Dias, Meghir, and Van Reenen (2004), Smith and Todd (2005), Heckman, Ichimura, Smith, and Todd (1998), and Abadie (2005). These models, therefore, are not well-suited in this context.

As described in section 4.2.1, the model in this paper accommodates the continuity of the treatment variable. This is not done within the frameworks of the Changes-in-Changes and quantile difference-in-differences models (Athey and Imbens 2006) nor within the framework of the propensity score matching models in Blundell and Costa Dias (2000), Bell, Blundell, and Van Reenen (1999), Blundell, Costa Dias, Meghir, and Van Reenen (2004), Smith and Todd (2005), and Heckman, Ichimura, Smith, and Todd (1998). On the other hand, the multi-level treatment case in Abadie (2005) and the model in D'Haultfoeuille, Hoderlein, and Sasaki (2015) accommodate continuous treatment. However, these last two models, like all other models, do not allow the researcher to collapse the student covariates in X_i into a single scalar, c_i. As explained in 6.3, this is an important feature of the model for testing the theoretical predictions.

B.5.1 Nonseparability

To exemplify non-separability, I describe it in the context of a QDID model applied to the context of this paper. Such a model would assume additive separability of the outcome function. Specifically, the outcome function in the absence of treatment (which is used to build the distribution of counterfactual outcomes for treated individuals) would be assumed to be: $Y^N = h(U, G, T, X) = h^G(U, G, X) + h^T(U, T, X)$ where G indicates dispersion (high or low) and T indicates cohort, U is an individual's unobservable, and

X is a vector of individual characteristics.[68] That is, function h would be composed of an outcome function that only depends on dispersion status h^G and one that only depends on cohort h^T, conditional on X. In the context of this paper this assumption is not satisfied if the vector X enters the outcome function as an index c_i and if the same vector X contributes to generate a student's type c_i differently in the pre- and post-earthquake cohort. This is the case when damage to a student's home has no effect on students' cost of effort before the earthquake (because damage has not occurred yet), but it does after the earthquake. Formally, the identifying assumption for QDID fails if what enters the outcome function is an index c which is a cohort-specific function of X. The data generating process in the absence of treatment would then be $Y^N = h(U, G, T, c^T(X)) = h^G(U, G, c^T(X)) + h^T(U, T, c^T(X))$, where it is clear that the first function depends on both G and T and, therefore, QDID would be misspecified, because additive separability would not be satisfied. The model presented in this paper relaxes the assumption of additive separability conditional on X. Additive separability holds only conditional on a value for the index c.

B.6 Additional Alternative Mechanisms

First, I consider the theory of social cognitive learning, which posits that students learn from similar classmates (Bandura 1986, Schunk 1996). Rank concerns, too, imply that students benefit from having similarly able classmates, because this triggers a healthy competition. However, the two theories can be distinguished. Specifically, social cognitive learning would require that low-cost students in both Mathematics and Spanish classes increase their test scores when damage dispersion increases, because of the larger proportion of low-cost classmates. However, I find that low-cost students in Spanish classes obtain lower test scores, as predicted by the rank concerns channel.

Second, I consider cooperative behavior between students affected by the earthquake. If students who were less affected by the earthquake helped the more affected ones, for example by hosting them at their less affected homes, then it could be possible that to classrooms with a larger variance in damages do not correspond classrooms with a larger variance in cost of effort. In spite of this potential attenuation mechanism, estimation

[68] I am switching the notation with respect to, for example, the notation in Athey and Imbens (2006). In particular, the control group here is $T = 0$, the pre-earthquake cohort, rather than $G = 0$. All results in Athey and Imbens (2006) are unchanged, one must only keep in mind that the time-trend there corresponds to the geographic dispersion effect here.

results do detect a worsening of student type and of achievement at higher earthquake intensity levels, and the variance in predicted cost of effort is increasing in the variance in MSK-Intensity at the estimated parameter values. Therefore, if such an insurance mechanism took place, the empirical findings can be thought of as lower-bounds in absolute value.

C Theoretical Appendix. FOR ONLINE PUBLICATION ONLY

C.1 Differential Equation

The first-order differential equation characterizing equilibrium strategies is obtained by rearranging the first order condition in 1, and substituting $c^{'}(e) = \frac{1}{e^{'}(c)}$:

$$\begin{aligned} e^{'}(c_i) &= \left(\frac{g(c_i)}{1 - G(c_i) + \phi}\right)\left(\frac{V(y(e), q(e,c))}{a(\mu)V_1 + V_2 \frac{\partial q}{\partial e}}\right). \\ &= \frac{g(c_i)}{1 - G(c_i) + \phi}\psi(e_i, c_i). \end{aligned} \tag{27}$$

C.2 Proof of Proposition 2.1

The proof is an adaptation of the proof in Hopkins and Kornienko (2004), where equilibrium strategies are strictly increasing and where the consumption and positional goods are two separate goods.

First, it is easy to show that the boundary conditions in the statement of the Proposition are optimal for the student with the highest cost, $\bar{c}$. The student with the highest type, $\bar{c}$, chooses the effort function that maxims utility V in the absence of rank concerns, as specified by the boundary condition in the statement of the Proposition. To see why, notice that in equilibrium her utility from rank is zero, therefore, she maximizes V because $V \times F + \phi \times V = V \times 0 + \phi \times V = \phi \times V$.

Next, I show that if the strategy $e^*(c)$ is a best response to other students' effort choices, then it is decreasing.If a student i of type c_i exerts effort $e_i = e^*(c_i)$ and this is a best response to the efforts of the other students as summarized by the effort distribution $F_E(\cdot)$, then it must be that $e_i \geq e_{nr}(c_i)$, where $e_{nr}(c_i)$ solves the first-order condition in the absence of rank concerns, i.e., $V_1 a(\mu)|_{e=e_{nr}} = -V_2 \frac{\partial q}{\partial e}|_{e_{nr}}$. This is because if $e <$

$e_{nr}(c_i)$, then $F_E(e)+\phi < F_E(e_n)+\phi$ and $V\left(y(e(c)), q(e(c), c)\right) < V\left(y(e_{nr}(c)), q(e_{nr}(c), c)\right)$. Therefore, $V\left(y(e), q(e, c)\right)\left(F_E(e) + \phi\right) < V\left(y(e_{nr}), q(e_{nr}(c), c)\right)\left(F_E(e_{nr}) + \phi\right)$, i.e., any level of effort below the no-rank-concerns level is strictly dominated by the no-rank-concerns level. Suppose that equality holds, so $e_i = e_{nr}(c_i)$. Then $e^*(\cdot)$ is decreasing because $e_{nr}(c_i)$ is decreasing. This follows from the assumptions on $V(\cdot)$ that $V_{11} = 0$, $V_{22} = 0$, $V_{ij} \leq 0$ for $i \neq j$, and from the assumptions on the cost of effort function that $\frac{\partial q}{\partial c} > 0$, $\frac{\partial q}{\partial e} > 0$, $\frac{\partial^2 q}{\partial^2 e} > 0$ and $\frac{\partial^2 q}{\partial e \partial c} \geq 0$. To see why, let $FOC(e, c) = V_1 a(\mu) + V_2 q_1$ and notice that by the Implicit Function Theorem:

$$\frac{de_{nr}}{dc} = -\frac{\partial FOC/\partial c}{\partial FOC/\partial e}.$$

The numerator is:

$$\frac{\partial FOC}{\partial c} = a(\mu)V_{12}\frac{\partial q}{\partial c} + V_{22}\frac{\partial q}{\partial e}\frac{\partial q}{\partial c} + V_2\frac{\partial^2 q}{\partial e \partial c} \leq 0.$$

The denominator is:

$$\frac{\partial FOC}{\partial e} = a(\mu)^2 V_{11} + a(\mu)V_{12}\frac{\partial q}{\partial e} + \left(a(\mu)V_{21} + V_{22}\frac{\partial q}{\partial e}\right)\frac{\partial q}{\partial e} + V_2\frac{\partial^2 q}{\partial^2 e} \leq 0.$$

As a result, $e^*(\cdot)$ is decreasing in c when it is equal to optimally chosen effort in the absence of rank concerns, because $\frac{de_{nr}}{dc} \leq 0$.

If equality does not hold, we want to show that if e_i is a best-response and $e_i > e_{nr}(c_i)$, then it is still the case that e_i is decreasing in c_i. First, I show that for any other choice $\tilde{e}_i \in (e_{nr}(c_i), e_i)$,

$$\frac{\partial V}{\partial c_i}\left(y(e_i), q(e_i, c_i)\right)\left(F_E(e_i) + \phi\right) < \frac{\partial V}{\partial c_i}\left(y(\tilde{e}_i), q(\tilde{e}_i, c_i)\right)\left(F_E(\tilde{e}_i) + \phi\right). \tag{28}$$

Rewrite the left-hand side as:

$$\frac{\partial V}{\partial c_i}\left(y(e_i), q(e_i, c_i)\right)\left(F_E(\tilde{e}_i) + \phi\right) + \frac{\partial V}{\partial c_i}\left(y(e_i), q(e_i, c_i)\right)\left(F_E(e_i) - F_E(\tilde{e}_i)\right).$$

The first term is smaller or equal to the right-hand side of equation 28, because $\frac{\partial V}{\partial c}$ is decreasing in e by the assumptions that $V_{21} \leq 0, V_{22} = 0, \frac{\partial q}{\partial c} > 0, V_2 < 0$, and $\frac{\partial^2 q}{\partial c \partial e} \geq 0$. To see why, notice that $\frac{\partial^2 V}{\partial c \partial e} = \left(V_{21}a(\mu) + V_{22}\frac{\partial q}{\partial e}\right)\frac{\partial q}{\partial c} + V_2\frac{\partial q}{\partial c \partial e} \leq 0$. The second term is strictly negative, because first, $\frac{\partial V}{\partial c_i}$ is strictly negative by virtue of the assumptions that $V_2 < 0$ and $\frac{\partial q}{\partial c} > 0$, and second, $(F_E(e_i) - F_E(\tilde{e}_i)) > 0$. To see why the latter is true, notice that for $e > e_{nr}$, $V\left(y(e), q(e,c)\right)$ is decreasing in e. Therefore, if e is a best-response, it must be the case that $F_E(e_i) > F_E(\tilde{e}_i)$, otherwise a student could lower effort and obtain a higher utility, while not lowering her status. This establishes the inequality in 28, so that at e_i, the overall marginal utility with respect to c ($\frac{\partial}{\partial c}\left(V(y,q)(F_E(e)+\phi)\right)$) is strictly decreasing in e. This implies that an increase in type c leads to a decrease in the marginal return to e, therefore, the optimal choice of effort e must decrease.

To show that if an effort function is an equilibrium strategy, then it must be continuous, suppose not. That is, suppose that that there was a jump downwards in the equilibrium effort function $e^*(c)$ at $\tilde{c}$, so that $\lim_{c \to \tilde{c}^+} e^*(c) = \tilde{e} < e^*(\tilde{c})$. Then, there would exist an $\epsilon > 0$ small enough, such that the student of type $\tilde{c} - \epsilon$ can reduce her effort to $\tilde{e}$, which is below $e^*(\tilde{c} - \epsilon)$, and obtain a discrete increase in utility because of the lower effort, while her rank would decrease by less, by continuity of the rank function $S(\cdot)$ at $\tilde{c}$. Therefore, there exists a student with an incentive to deviate, and such discontinuous $e^*(c)$ function cannot be an equilibrium strategy.[69]

Uniqueness of the solution to the differential equation in 27, and therefore uniqueness of the equilibrium, follows from the fundamental theorem of differential equations. The boundary condition pins down the unique solution.

Intuition for equilibrium uniqueness. Intuitively, uniqueness of the equilibrium follows from two key assumptions: achievement gives utility per se, i.e., irrespectively of the status (rank utility) it provides, and individuals have different costs of producing achievement. A common type of multiplicity of equilibria in this class of games is when all individuals exert the same amount of effort. If this were an equilibrium, there would be an infinite number of equilibria. However, all students playing the same level of effort e^* is not an equilibrium, because students with a high enough cost (i.e., with a cost above a certain cutoff that depends on e^*, i.e., $c > cutoff(e^*)$), have an incentive to reduce effort, obtain zero rank, and enjoy their private utility from achievement. Therefore, the classical problem of multiplicity of equilibria in coordination games does not arise.

C.3 Comparative Statics

Definition Two distributions G_A, G_B with support on $[\underline{c}, \bar{c}]$ satisfy the Unimodal Likelihood Ratio (ULR) order, $G_A \succ_{ULR} G_B$, if the ratio of their densities $L(c) = g_A(c)/g_B(c)$ is strictly increasing for $c < \tilde{c}$ and strictly decreasing for $c > \tilde{c}$ for some $\tilde{c} \in [\underline{c}, \bar{c})$ and if

[69] The remaining part of the proof, showing that the equilibrium strategy is strictly decreasing and differentiable, is a straightforward adaptation of the lengthy proofs in Hopkins and Kornienko (2004), and it is available from the author upon request.

$\mu_A \geq \mu_B$.

In particular, if B has the same mean but higher variance than A, then $G_A \succ_{ULR} G_B$. Define the cutoffs $\hat{c}^-$ and $\hat{c}^+$ as the extremal points of the ratio $(1-G_A(c)+\phi)/(1-G_B(c)+\phi)$ when $G_A \succ_{ULR} G_B$. It can be shown that these cutoffs are such that $\underline{c} < \hat{c}^- < \hat{c}^+ \leq \bar{c}$, and they can be conveniently interpreted as cutoffs that separate type categories.[70] Low c students are those with $c \in [\underline{c}, \hat{c}^-)$, middle c students as those with $c \in (\hat{c}^-, \hat{c}^+)$, and high c students as those with $c \in (\hat{c}^+, \bar{c}]$. The model has the following prediction:

Proposition C.1 (Adapted from Proposition 4 in Hopkins and Kornienko (2004)). *Suppose $e_A(c)$ and $e_B(c)$ are the equilibrium choices of effort for distributions G_A and G_B. If $G_A \succ_{ULR} G_B$ and $\mu_A = \mu_B$, then:*

- *$y(e_A(c))$ crosses $y(e_B(c))$ at most twice. Moreover, $y(e_A(c)) < y(e_B(c))$ for all $c \in [\hat{c}^+, \bar{c})$ with a crossing in $(\tilde{c}, \hat{c}^+)$ so that $y(e_A(c)) > y(e_B(c))$ for all $c \in [\hat{c}^-, \tilde{c}]$, with a possible crossing on $[\underline{c}, \hat{c}^-)$.*

Proof The proof is a straightforward adaptation of the lengthy proof in Hopkins and Kornienko (2004). It is available from the author upon request.

D Appendix: Additional Tables and Figures. FOR ONLINE PUBLICATION ONLY

[70]The proof is available upon request from the author. It is a modification of the lengthy proof in Hopkins and Kornienko (2004), there the c.d.f. functions, rather than their complement, appear in the ratio.

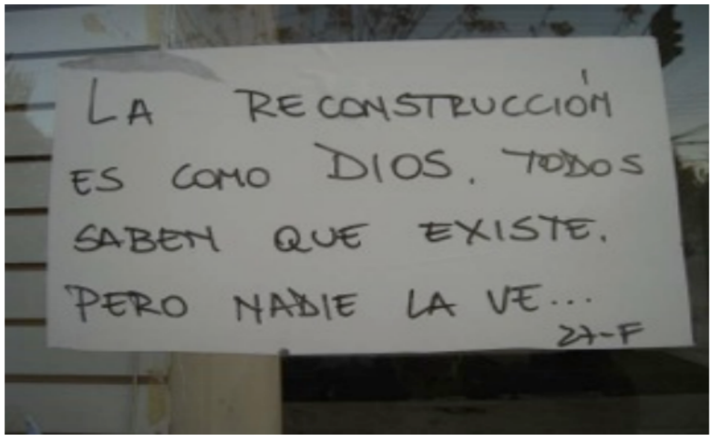

Figure 10: *Source*: Comerio (2013). Handmade sign found in Cauquenes, Chile, on February 2, 2012. Translation: "Reconstruction is like God. Everyone knows it exists, but nobody has seen it."

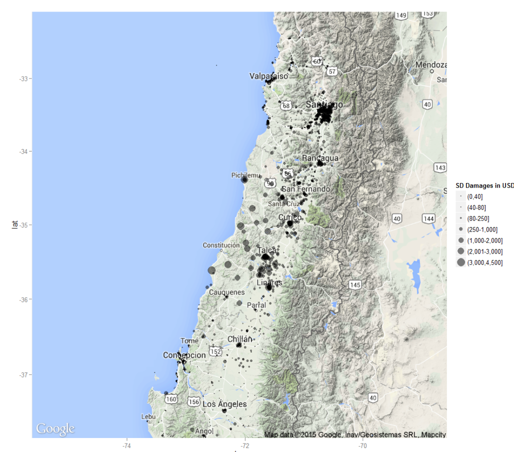

Figure 11: Schools where not all students are affected equally by the earthquake, and within school standard deviation in damages. There are three important features: 1. there are many dots on the map, i.e. many schools have a geographically dispersed student body; 2. dots have different sizes, i.e. there is across school variation in within school damage dispersion; 3. even schools close to each other suffered different damage dispersions.

Table 13: Effect of being exposed to the earthquake (binary variable) on Math and Spanish test scores (Municipal schools). DDD model.

	(1)	(2)	(3)	(4)
	Math TS	Math TS	Spanish TS	Spanish TS
$E_i \cdot d_g$	0.078***	0.084***	0.031***	0.028**
	(0.007)	(0.008)	(0.007)	(0.009)
T_i	0.025**	0.007	0.042***	-0.002
	(0.008)	(0.008)	(0.009)	(0.009)
$T_i \times E_i \cdot d_g$	-0.029**	-0.021*	-0.048***	-0.018+
	(0.010)	(0.010)	(0.011)	(0.011)
Lagged TS		0.667***		0.684***
		(0.003)		(0.003)
Controls	No	Yes	No	Yes
Observations	156506	81635	156801	79458

Standard errors in parentheses. $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Controls: whether the student lives in the same town where the school is, gender, mother's education, father's education, household income, class size, whether the Math or Spanish teacher is female, has a postgraduate degree, has a permanent contract, her tenure at the school, and her teaching experience. A constant is always included.

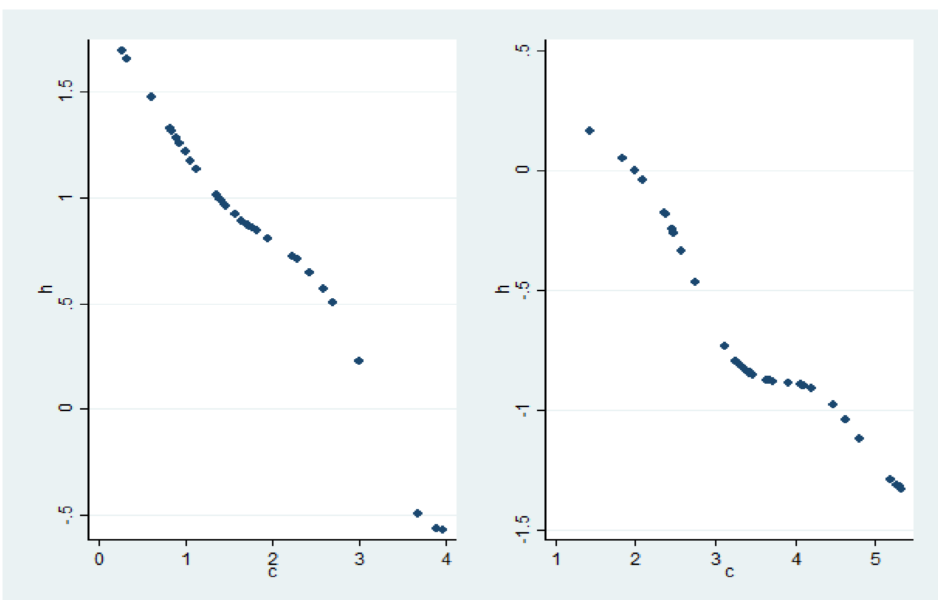

Figure 12: Examples of estimated $h(c)$ functions in two classrooms.

Table 14: Robustness to teacher productivity in the classroom.

Dependent Variable:	Curriculum Covered		Test Score	
	Mathematics	Spanish	Mathematics	Spanish
	(1)	(2)	(3)	(4)
β_1 (st. dev.)	-0.00246	0.00525	0.151***	0.217***
	(0.0412)	(0.0355)	(0.0454)	(0.0483)
β_2 (cohort)	-0.0105	0.0191	0.0454	0.0643
	(0.0439)	(0.0357)	(0.0460)	(0.0479)
β_3 (st. dev.$\times$ treatm. dummy)	0.0364	0.0264	-0.164**	-0.154**
	(0.0513)	(0.0435)	(0.0553)	(0.0586)
Curriculum Covered			0.276***	0.135***
			(0.0181)	(0.0223)
Constant β_0	0.555***	0.638***	-0.486***	-0.423***
	(0.0430)	(0.0337)	(0.0412)	(0.0435)
Controls	Yes	Yes	Yes	Yes
Observations	2,902	3,187	41,737	45,319

Standard errors in parentheses. $^{+}$ $p < 0.105$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. The sample is the same as the sample in the main analyses: Municipal schools.
Models (1) and (2) are regression 11. The unit of observation is the classroom. Included regressors are the classroom means of: MSK-intensity, also interacted with the cohort dummy, lagged Math and Spanish test scores, mother's and father's education, income; and the rich set of teacher and classroom characteristics as in the main linear model of spillovers.
In models (3) and (4) the unit of observation is the student. Included regressors are mean of MSK-intensity in the classroom, also interacted with the cohort dummy, damage at the student's home and the same rich set of student, teacher and classroom controls as in the main linear model of spillovers (see caption of Table 4).

Table 15: Probit regression, marginal probability estimates reported. Dependent variables: being at the top (1) or bottom (2) third of the distribution of elicited cost of effort.

	top 33 percent (1)	bottom 33 percent (2)
Lagged Math TS	-0.055^{***}	0.100^{***}
	(0.002)	(0.002)
Seismic intensity	0.013^{***}	-0.011^{***}
at student's home	(0.002)	(0.003)
SES Controls	Yes	Yes
Observations	46,059	46,059

Standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

SES controls: father's and mother's education, household income.