6-2020

The effect of endogenously-determined bullying on test scores

Hisham bin Yacob Patel
Singapore Management University, hishamyp.2018@me.smu.edu.sg

THE EFFECT OF ENDOGENOUSLY-DETERMINED BULLYING ON TEST SCORES

Hisham Bin Yacob Patel

SINGAPORE MANAGEMENT UNIVERSITY

2020

The Effect of Endogenously-Determined Bullying on Test Scores

Hisham Bin Yacob Patel

A DISSERTATION

In

ECONOMICS

Presented to the Singapore Management University in Partial Fulfilment

of the Requirements for the Degree of Master of Philosophy in Economics

2020

____

Supervisor of Dissertation

____

MPhil in Economics, Programme Director

The Effect of Endogenously-Determined Bullying on Test Scores

By

Hisham Bin Yacob Patel

Presented to the Singapore Management University in Partial Fulfilment

of the Requirements for the Degree of Master of Philosophy in Economics

Dissertation Committee:

Huang Fali (Supervisor/Chair)
Associate Professor of Economics
Singapore Management University

Luca Facchinello
Assistant Professor of Economics
Singapore Management University

Steven Durlauf
Steans Professor in Educational Policy
University of Chicago Harris School of Public Policy

Singapore Management University
2020

Abstract

The problem of bullying is one of the main sources of adverse school environment. In this dissertation, I explore the effect of bullying on test scores, under the assumption that bullying is endogenously determined. Comparing my estimates with that from traditional OLS, I find that estimates from OLS grossly underestimates the effect of bullying on short-term student outcomes. I also find significant evidence of heterogeneity across various sub-groups; for example, male students are found to be less affected by bullying as compared to their female peers. Furthermore, male students are also found to be less affected by physical bullying while female students are found to be more affected by relational bullying. In addition, I find some evidence suggesting the possibility of an age-trend with respect to the effect of bullying, with older students being less affected than their younger peers, both within and across cohorts. Exploring possible mechanisms, I observe that bullying victims have a poorer perception of self and often negatively evaluate their teachers. Furthermore, these students are also more likely to perceive unfair treatment by their teachers, as well as report lower levels of interest in their studies.

JEL Classification: I20, I21, I28, J13

Keywords: Endogenous Treatment Effects, Bullying, Test Scores, School Environment

Table of Contents

I. Introduction and Literature Review

The general consensus is that a good education paves the way to a brighter future. This notion is supported by an abundance of evidence which show a strong relationship between education and future success in the labour market (Heckman 2000), better health (Blanchflower & Oswald 2004; Chevalier & Feinstein 2006; Grossman 2005), and other desirable outcomes in adult-life. In addition, there is also evidence of a persistence in the benefits reaped from a happy childhood (Heckman, et al 2006a). Drawing on evidence from various other fields also agree with the notion that child well-being – and behavioural traits – are important determinants of adult-life outcomes (Clark & Oswald 2002; Carneiro, Crawford & Goodman 2007; Feinstein 1999; Roberts, et al 2007). Similarly, unhappiness during childhood also produce adverse effects that propagate into adulthood; namely, depression and unemployment (Becker 2000; Reivich, et al 2005).

These broad considerations of general well-being in school have recently made its way into education policy; namely the *Every Child Matters* (ECM) initiative – which formed the basis of UK's Children Act 2004, USA's *No Child Left Behind* Act – which was repealed in 2015 and replaced, in spirit, by the *Every Student Succeeds* Act (ESSA) –, and the *Learn For Life* (LFL) initiative in Singapore. The ECM initiative has brought about important changes in educational services and aims to improve the overall well-being of children (DCSF 2004). An existing feature in the US education landscape, but only recently included in Singapore via the LFL initiative, is the removal of mandatory standardized testing in the first two years of primary school. This adjustment was made in the spirit of freeing up more time and space in schools to strengthen holistic development and provide an enjoyable educational experience (MOE 2018); in essence, these policies aim to promote greater general well-being of school-going children.

An influential aspect of a child's general well-being is the school environment (Samdal, et al 1998; Pittman & Richmond 2007). One of the main sources of adverse school environment is the problem of bullying (Beaty & Alexeyev 2008). Bullying is found to have an adverse effect on human capital accumulation both at, and beyond school. The adverse influence of bullying on educational attainment propagates into adulthood (Brown & Taylor 2008).

Looking at the broad literature on the spill-over effects of bad behaviour, most papers evaluate the impact over a considerably long horizon, in the form of labour market outcomes (Carrell, Hoekstra & Kuka 2018; Papageorge, Ronda & Zheng 2017; Segal 2013) and more recently, workplace performance (Murphy 2019). However, short-run outcomes warrant equal attention, especially since bullying has been tied to teenage suicides – bullying victims are 2 – 9 times more likely to consider suicide than non-victims – as well as homicides; homicide perpetrators were found to be twice as likely as homicide victims to have been bullied previously by peers in the US[1].

Drawing on findings in psychology and anthropology (Busch, et al 2014; Juvonen, Wang & Espinoza 2011; Nakamoto & Schwartz, 2010; Strøm, et al 2013; Wigderson and Lynch, 2013), I posit that bullying adversely affects scores of students in the short-run, either directly, through avenues such as increased incidences of physical and mental health issues (Due, et al 2007; Herge, La Greca & Chan 2016), or indirectly through higher absenteeism rates (Grinshteyn & Yang, 2017, Hutzell & Payne, 2012, Randa & Reyns, 2014).

Economic research on bullying is still very scarce, with most of the available literature focusing on the long-term effects of bullying. Brown & Taylor (2008), for example, estimated linear regression models and ordered probits to examine the associations between bullying and educational attainment, as well as wages in the UK. Their findings suggest that the effects of bullying propagate into adulthood. Individuals that were being bullied were found with lower educational attainment, and, as a result, lower wages in adulthood. Eriksen et al (2012) used administrative information from Denmark and reduced form regressions to document strong and empirically robust correlations between bullying and Grade 9 test scores,

[1] See stopbullying.gov

teenage pregnancy, the use of psychopharmacological medication, height and weight at age 18. While these papers represent novel efforts in empirically documenting the effect of bullying, they do little in dealing with the non-randomness of the incidence of bullying.

To that end, Eriksen et al (2014) attempted to remedy this by using an instrumental variable approach to account for the endogeneity of bullying. They use the proportion of classroom peers whose parents have criminal convictions as an instrument for the effect of teacher-parent reported bullying victimisation on Grade 9 GPA. Their estimates, however, suffer from the low variation of the instrument and also by the possibility of locational clustering of parents who have criminal convictions. This locational clustering might endogenize the sorting of students into schools and classrooms. Furthermore, since selection into the treatment – i.e. being a victim of bullying – is driven by unobserved heterogeneity (latent skills and characteristics), the IV approach does not necessarily identify causal effects (Heckman et al 2006b; Imbens & Angrist 1994). The approach adopted in this paper models the endogenous nature of bullying and attempts to account for as much of the unobserved heterogeneity as possible.

A recent study on the short-term effects of bullying showed that bullying affects a student's performance in school (Ponzo 2013). When using math scores as the outcome variable, the effect of bullying increases as students grow older – from Grade 4 to Grade 8. The opposite was observed when using science scores as the outcome variable, as the author found a decrease in the effect of bullying as students grow older. The author employed two specifications to test for the effect of bullying on test scores; a school fixed effects model, and propensity score matching. Both specifications produced similar results; i.e. the effect of bullying increases when moving from Grade 4 to Grade 8, when math scores were the chosen outcome variable, while the effect mostly decreases when using science scores as the outcome variable.

In this paper, I adopt the OECD definition of bullying. In the OECD PISA 2015 Report on Student Well-being, bullying is defined as a systematic abuse of power which can be inflicted directly, through physical (hitting, punching or kicking) and verbal (name-calling or mocking) abuse, or indirectly through relational bullying. Relational bullying refers to the phenomenon of social exclusion, where some children are ignored, excluded from games or parties, rejected by peers, or are the victims of gossip and other forms of public humiliation and shaming. Therefore, in this paper, a student is defined to be a victim of bullying if said student experiences any of the bullying-related incidents on a weekly basis – the highest frequency rate available in the dataset used in this paper. Bullying is modelled to be endogenously determined by the student's own characteristics, household features, as well as teacher and school characteristics. Comparing the estimates obtained from the endogenous bullying model with that of a traditional OLS model, I find that traditional OLS estimation procedure grossly underestimates the short-term effect of bullying on student test scores.

This paper makes several contributions to existing literature. First, I add to the growing literature on social interactions – specifically, the negative spill-over effects of bad behaviour (Carrell, et al 2018; Papageorge, et al 2017; Ponzo 2013; Murphy 2019; Segal 2013). Second, to the best of my knowledge, this is the first attempt to assess the short-term effect of bullying while dealing with its endogeneity. Therefore, my findings may provide valuable insights that can motivate policy interventions to reduce the incidence rate of bullying in schools. Third, owing to the richness of available data in the TIMSS dataset, I explore the separate effects of physical and relational bullying; a feature not yet present in the current economic literature, to the best of my knowledge. Lastly, I explore possible mechanisms through which bullying might affect a student's test scores.

The rest of the paper is constructed as follows. Section II describes the two data sets used in the analyses. Section III introduces the model of endogenous bullying, as well as the identification strategy. In Section IV, I present and interpret the main results, perform robustness checks, as well as check for heterogeneity across various sub-groups. Section V explores possible mechanisms by which bullying affects a student's test score, and Section VI concludes the paper.

II. Data

Trends in International Mathematics and Science Study (TIMSS)

TIMSS, conducted every 4 years since 1995 by the International Association for the Evaluation of Educational Achievement (IEA), is an international comparative study of student achievement. TIMSS is directed by the TIMSS International Study Centre at Boston College, Lynch School of Education, in collaboration with a worldwide network of organisations and representatives from the participating countries. It surveys students, teachers and school administrators to gain in-depth information about the various factors that contribute to educational outcomes. Currently, 70 countries participate in the assessments. TIMSS also collect extensive data about the contextual factors that affect learning, including school resources, student attitudes, instructional practices, and support at home. The TIMSS dataset allows for a multitude of controls, as well as the ability to extend the analysis beyond any single country. Another benefit of using TIMSS is that information on both Grade 4 and Grade 8 students are available in the dataset, although schools, and students, are randomly selected during each round, to make up a nationally representative sample for each country – i.e. a Grade 4 student who partakes in the TIMSS assessment when he is in Grade 4 is no more, or less likely, to be chosen for the next TIMSS assessment, 4 years later.

The TIMSS 2015 dataset was selected as it contained the richest of all the existing TIMSS datasets. The sub-sample of interest in this paper is the Grade 4 and Grade 8 sample of Singapore students.

There are 4 datafiles of interest for Grade 4 students in the TIMSS 2015 dataset for each country; namely, data relating to school background, student background, student's home background, teacher background data. Aside from features like age, test scores and some closed questions, most of the questions have 3 – 5 possible answers, measured on a Likert scale. The school background dataset collects information, provided by school administrators or principals, on a multitude of things, ranging from questions on the location of the school – i.e. urban or rural –, the general income level of their students, questions regarding shortages of supplies in school, the prevalence of problems – with students and teachers – encountered in schools, as well as information on the principal. The student background dataset, answered by the students themselves, on the other hand, collects a multitude of information, ranging from gender, general income level of the household, citizenship status of the student and parents, the school environment, feelings about mathematics and science, as well as the scores obtained from the TIMSS assessment. The student home background dataset, answered by their parents – available only for Grade 4 students – contains detailed information of the household, ranging from habits and activities at home, whether or not a child attended kindergarten, the general aptitude level of the child when entering Grade 1, the education level for both parents, as well as employment status of both parents. Lastly, the teacher background data, answered by teachers, collects information of the perceptions of the school and students, the years of experience, gender, age, major of study of teachers.

The data available for Grade 8 students in the TIMSS 2015 dataset is identical to what was reported above for Grade 4 students, with the exception of the student's home background data, which is only exclusive to the Grade 4 dataset.

TIMSS assess students in two areas; mathematics and science. I report the descriptive statistics of the two test scores – science and math – from the TIMSS dataset below in Table 1. For the empirical model introduced later, in Section III, I will utilise the standardised versions of the math and science scores, so as to ease comparability and interpretability.

	Mean	St.Dev	min	max	Mean	St.Dev	min	max
Grade 4 (N = 3779)					**Grade 8** (N = 3526)			
Math Score	628.75	76.39	332.98	837.74	626.30	76.66	349.33	790.67
Science Score	601.72	75.06	275.34	787.12	602.46	80.48	282.46	786.23

Table 1. Descriptive Statistics for Mathematics and Science Test Scores

As mentioned previously, in the OECD PISA 2015 Report on Student Well-being, bullying is defined as a systematic abuse of power which can be inflicted directly, through physical (hitting, punching or kicking) and verbal (name-calling or mocking) abuse, or indirectly through relational bullying. Relational bullying refers to the phenomenon of social exclusion, where some children are ignored, excluded from games or parties, rejected by peers, or are the victims of gossip and other forms of public humiliation and shaming.

The bullying variable defined in this paper is based the student's response to 8 key questions pertaining to the prevalence of bullying in the student's school life, such as how often a student is being *made fun of*, *left out of games*, *forced to do something*, *threatened*, etc. All questions have 4 possible answers ranging from *At least once a week* to *Never*. The bullying variable is binary in nature, taking the value 1 if a student reports experiencing any of the bullying-related incidents on a weekly basis, and 0 otherwise. The exact questions that contribute to the bullying variable can be found in the Appendix, in Table A0.

	Mean	St.Dev	min	max	Mean	St.Dev	min	max
Grade 4 (N = 3779)					**Grade 8** (N = 3526)			
Bullying	.375	.484	0	1	.26	.439	0	1

Table 2. Descriptive Statistics for Bullying

I report the descriptive statistics for the bullying variable, in Table 2. The incidence of bullying falls when comparing the Grade 4 samples with their Grade 8 counterparts – evidenced by both a lower mean and standard deviation. Referencing Figure A1 in the Appendix, I observe that Singapore ranked above the 50th percentile for both the Grade 4 and Grade 8 samples across all the countries present in the TIMSS 2015 sample. Furthermore, while Singapore students report a lower prevalence of bullying as compared to Hong Kong, it is significantly higher than what is reported by students in Taiwan in both the Grade 4 and Grade 8 samples. Interestingly, students in the Scandinavian countries report the lowest incidence rate of bullying.

Controls employed in this paper can be categorised in two manners; individual-specific controls, and class-specific controls. Individual-specific controls relate to characteristics such as citizenship status, education level of both parents, employment status of both parents, and a proxy for general income level. Class-specific controls refer to teacher characteristics that are commonly shared by students belonging to the same class. Since there are incidences of multiple teachers assigned to each class – as some schools have separate math and science teachers for Grade 4 students –, mean teacher characteristics are utilised. These characteristics include the teacher's education level, years of experience and class size.

The Grade 4 dataset also contains further information on the employment status of both parents. With that, I generate 3 dummy variables that individually take the value 1 if the father is working full-time, the mother is working full-time and if both parents are working full-time, and 0 otherwise. In addition, since the math and science teacher information are separately reported in the Grade 8 datasets, I separately classify them. All teacher characteristics are presented as averages since each class might have more than 1 teacher assigned to it; i.e. if they have separate teachers for mathematics and science. The summary statistics of all the controls employed are below in Table 3.

	Mean	St.Dev	min	max	Mean	St.Dev	min	max
Grade 4 (N = 3779)					Grade 8 (N= 3526)			
Gender	.497	.5	0	1	.468	.499	0	1
Citizen	.834	.372	0	1	.87	.336	0	1
Mother Citizen	.618	.486	0	1	.638	.481	0	1
Father Citizen	.701	.458	0	1	.754	.431	0	1
Father Education Level	5.783	1.723	1	8	4.242	1.831	1	7
Mother Education Level	5.629	1.641	1	8	4	1.764	1	7
Low Income	.03	.171	0	1	.015	.122	0	1
Middle Income	.489	.5	0	1	.461	.499	0	1
High Income	.481	.5	0	1	.524	.499	0	1
Father Full-time Working	.926	.262	0	1				
Mother Full-time Working	.631	.483	0	1				
Both Parents Full-time Working	.598	.49	0	1				
Average Teacher's Education Level	4.821	.674	2	6				
Average Teacher's Years of Experience	10.848	8.51	1	52				
Class Size	35.293	5.874	18	44				
Math Teacher's Average Education Level					5.071	.432	3	6
Math Teacher's Average Years of Experience					8.778	8.608	1	46
Math Class Size					35.57	6.552	5	44
Science Teacher's Average Education Level					5.16	.476	3	6
Science Teacher's Average Years of Experience					8.773	8.356	0	45
Science Class Size					35.799	6.251	10	44

Table 3. Descriptive Statistics for Individual-Specific and Class-Specific Control Variables

The gender binary variable takes the value 1 if student is male, 0 if female. The 3 proxies of income used in this paper – Low Income, Middle Income & High Income – are binary variables that relate to the student's response to a question querying whether the student has his/her own room, and/or internet access at home. Low income students are classified as students who responded no to both, while middle income students responded yes to either, and high income students responded yes to both. In all specifications presented in the paper, the low income students are the omitted category.

III. Model and Empirical Strategy

The effect of unobserved correlated factors can confound the underlying effect of bullying. Such correlation could arise if self-selection and sorting of students across schools are affected by school resources, or if there is a correlation between the quality of the school environment and how selective the school's admission process is. As such, I account for both these sources of confounding factors by utilising a school fixed effects model. Based on this approach, I present my model estimates of the effect of bullying on individual test scores, using the following specification,

$$Score_{ijk} = \alpha \cdot Bullying_{ijk} + \boldsymbol{\delta X}_{ij} + \boldsymbol{\theta C}_{ij} + \mu_j + \varepsilon_{ijk} \tag{1}$$

Where $Score_{ijk}$ is the test score of the ith student, in the jth school, in the kth subject – k = math or science –, the coefficient α capture the effect of bullying for the ith student, in the jth school on the kth subject test score respectively. X is a vector of individual characteristics of the ith student, in the jth school, while C is a vector of class-specific characteristics of the ith student, in the jth school. μ_j represents the school fixed effects, while ε_{ijk} represent the remaining unobserved error term.

There are several remarks about this identification strategy that merit additional comment. First, my identifying assumption is that the effect of bullying can vary for both subject. One possible scenario is that the effect of bullying might affect the social dynamics of the classroom, which in turn, will affect the interactions of students when doing group work. Since group work is much more prevalent in science than in mathematics, the effect of bullying might be stronger in science test scores, than in mathematics. The identification strategy employed in this paper allows for such a possibility. Second, the school fixed effects model supports the notion that schools, as a unit, vary in quality.

The endogenous bullying variable in (1) is modelled as follows:

$$Bullying_{ij} = \boldsymbol{\lambda_1 X_{ij}} + \boldsymbol{\lambda_2 Z}_{ij} + \boldsymbol{\lambda_3 C}_{ij} + \boldsymbol{\lambda_4 S}_j + \eta_{ij} \qquad (2)$$

The vectors X & C correspond to the same vectors of individual-specific and class-specific characteristics described above. The vector Z correspond to additional individual-specific characteristics that might inform of the probability of a student being a victim of bullying, while vector S corresponds to school characteristics. η_{ij} represents the remaining unobserved error term. The inclusion of school characteristics is pertinent to the model of endogenous bullying in two aspects. First, owing to the school fixed effects specification employed in (1), valuable information on the school quality and environment is unusable, owing to the estimation procedure. However, such measures of school quality and environment can potentially affect the likelihood of a student being a victim of bullying. These variables include the average income level of the student population – measured via the proportion of the student population belonging to either disadvantaged or affluent households –, the average level of bullying experienced in the school as well as the level of resources available in the school. In addition, I also include the incidence rate of other measures of adverse school environment – namely, late-coming, absenteeism, classroom disturbances, cheating, profanity, vandalism, theft, intimidation among students and physical injury – as such factors could potentially be correlated with the probability of a student being a victim of bullying, and thus, could provide information on the quality of the school environment.

There is one exogenous instrument included in Z that is assumed to be correlated with the probability a student becomes a victim of bullying but is uncorrelated with test scores. This variable is based on the response to a question asking to what degree the student like to see their classmates at school. A student who is being bullied is assumed to be less likely to enjoy seeing their classmates at school. Furthermore, it is plausible to assume that whether a student likes/dislikes their classmates has little bearing on his ability to do well academically. This variable is measured on a 4-point Likert scale; ranging from Disagree a lot, to Agree a lot, with no Neutral/Indifferent option available. The variable enters the estimation procedure through 4 indicator functions that correspond to the possible responses. The results do not materially differ if the variable is entered as a continuous variable or through 4 indicator functions – with one omitted category.

As a first check, including these variables as additional regressors in the OLS specification (1) yielded statistically insignificant coefficient estimates, even at 10% level for these parameters in question. In addition, all the indicator variables are found to be statistically significant at the 5% level, in the Probit regression of Equation (2). The instrument of choice can be viewed as a proxy of a non-cognitive characteristic of the student. As found in Sarzosa & Urzúa (2016), non-cognitive characteristics are found to be important determinants of bullying victimisation.

The methodology employed in this paper is based on the control function approach introduced by Heckman & Robb (1985), and further developed by Altonji, Elder & Taber (2005). The control function (CF) approach proceeds by noting that the correlation between the error terms ε_{ijk} & η_{ij} can be captured using a linear relationship, specified as follows:

$$\varepsilon_{ij} = \rho_1 \eta_{ij} + e_{ij} \tag{3}$$

The assumption is that ε_{ijk} & η_{ij} are both uncorrelated with Z. Therefore, it naturally follows that e_{ij} is also uncorrelated with Z, and thus e_{ij} must be uncorrelated with the outcome variable in Equation (1), $Score_{ijk}$. Therefore, a valid estimating equation can be obtained by plugging in Equation (3) into (1) to get the following equation:

$$Score_{ijk} = \alpha \cdot Bullying_{ij} + \boldsymbol{\delta X}_{ij} + \boldsymbol{\theta C}_{ij} + \mu_j + \rho_1 \eta_{ij} + e_{ijk} \tag{4}$$

Therefore, by including the generalised residuals $\hat{\eta}_{ij}$ obtained from the first-stage Probit regression of Equation (2), one obtains a new error term, e_{ijk}, that is uncorrelated with all other righthand-side variables, including the *bullying* 'treatment' variable. In effect, the inclusion of the fitted values η_{ij} 'controls for' the endogeneity of *Bullying*. In short, one can think of η_{ij} as proxying for the unobservable factors in ε_{ij} that are correlated with *Bullying*[2]. The exogenous variation induced by the excluded instrumental variables, Z, provides separate variation in the generalised residuals, $\hat{\eta}_{ij}$, obtained from Equation (2), and these residuals serve as the control functions.

In addition, the inclusion of η_{ij} produces a heteroscedasticity-robust Hausman (1978) test of the null hypothesis H_0: $\rho_1 = 0$, which implies that *Bullying* is exogenously determined. By comparison, the traditional form of the Hausman test that directly compares OLS and 2SLS is substantially harder to make robust to heteroscedasticity.

To check if the CF approach is applicable, one needs to check if the rank condition holds. The rank condition holds if $\boldsymbol{\lambda_2}$ & $\boldsymbol{\lambda_4}$ are jointly non-zero. To check, I run the first-stage Probit regression using Equation (2) and conduct a Wald Test with H_0: $\boldsymbol{\lambda_2} = \boldsymbol{\lambda_4} = \mathbf{0}$. I am able to reject H_0 at the 5% level, for both the Grade 4 and 8 samples. Thus, I can safely reject the

[2] For a more detailed exposition on control function methods, see Wooldridge (2015) for a synthesis of the various types of CF approaches.

null hypothesis at 5% significance level and conclude that the rank condition holds.

IV. Results

In this section, I present the estimated coefficients of bullying for Grade 4 and Grade 8 Singapore students, using the fixed effects specification with endogenous bullying described above. For each outcome variable – math and science scores –, I accompany estimates from the simple least-squares regression model, without accounting for the endogeneity of the bullying variable; i.e. results from executing Equation (1) alone, without regard of the dynamics introduced via Equation (2). The results are presented in Table 4.

In all specifications, the coefficient estimate for bullying is statistically significant and consistently negative in the Grade 4 sample, and in most specifications in the Grade 8 sample. A Grade 4 student who is classified as being a victim of bullying is observed to have lower standardised test scores in the magnitude of approximately 0.5 SD, regardless of the chosen outcome variable. In addition, whenever found to be statistically significant, in both samples, the coefficient estimates for bullying are over twice as large under the endogenous bullying framework, when compared to the traditional school fixed effects framework which assumes that the bullying treatment is exogenous. This implies that estimates obtained from OLS – under the assumption that bullying is exogenous – grossly under-estimates the effect of bullying on test scores.

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample				
Bullying	-0.2260***	-0.4908***	-0.1868***	-0.5556***
	(0.028)	(0.15)	(0.028)	(0.14)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3779	3779	3779	3779
R^2	0.462		0.491	
Panel II. Grade 8 Sample				
Bullying	-0.0968***	-0.2876*	-0.0743**	-0.0087
	(0.030)	(0.16)	(0.030)	(0.16)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3524	3524	3524	3524
R^2	0.601		0.604	

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression. Individual-specific controls include gender, citizenship of the student & parents, education level of both parents, the working status of both parents and the general income level of the household. Class-specific controls relate to the teacher's average education level, average years of experience and the class size.
In Panel I, in the Endogenous Bullying CF Approach, school fixed effects control for 163 out of a total of 179 schools. Controlling for the full set of schools, omitting 1, induces some form of multicollinearity in some indeterminate dimension which makes estimation impossible under the CF framework
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 4. Regression Estimates for Grade 4 Singapore Students (OLS vs Endogenous Bullying Model)

Under the endogenous bullying framework, the coefficient estimate for bullying is statistically indistinguishable from zero in the Grade 8 sample when science scores were chosen as the outcome variable.

In addition, if I were to divide the samples into 4 quartiles based on age, referencing Table A1 in the Appendix, I find that Grade 4 students who are older are less affected by bullying than their younger peers within the same cohort, with the oldest quartile of students being the least effected by bullying. This age differential is unlikely driven by a difference in the propensity of being bullied since students from all age groups have statistically indistinguishable rates of bullying across all 4 quartiles in the Grade 4 sample, as seen below in Table 5.

Variable: Bullying	N	Mean	St.Dev	min	max
1st Quartile	999	0.393	0.489	0	1
2nd Quartile	972	0.379	0.485	0	1
3rd Quartile	955	0.369	0.483	0	1
4th Quartile	853	0.358	0.480	0	1

Table 5. Descriptive Statistics for Age Quartiles for Grade 4 students

Furthermore, the estimated effect of bullying is consistently larger when comparing Grade 4 students with their Grade 8 peers, regardless of the chosen outcome variable and specification. All pair-wise differences between the two samples are statistically significant at 5% level, with the exception of the Grade 4 – Grade 8 pair from the CF approach with math scores as the outcome variable. This means that at the 5% level, I can additionally conclude that there is evidence suggesting that the effect of bullying is larger for Grade 4 students than their Grade 8 peers. I plot the coefficients of both bullying variables in Figure 1, to illustrate the differences.

Math Score as Outcome Variable | Science Score as Outcome Variable

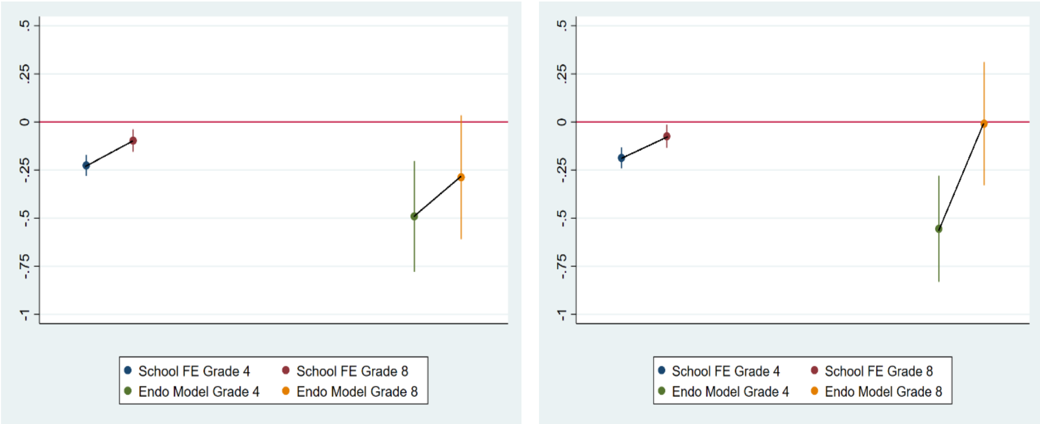


Figure 1. Plots for the coefficient of bullying variable

The CF approach allows me a heteroscedasticity-robust way to test the null hypothesis, H_0: $\rho_1 = 0$, which implies that *Bullying* is exogenously determined. I note that the Wald test statistic when math and science scores were the outcome variables in the Grade 4 sample are 3.27 and 6.84 respectively. That means, at the 10% level, I am able to reject the null hypothesis and conclude that there is sufficient evidence that suggest the two equations are not independent. This can be viewed as a falsification test of sorts, for the hypothesis of endogenous bullying. Similarly, the Wald test statistic in the Grade 8 sample was 1.35 & 0.16 when math and science scores were chosen respectively, with an accompanying p-value of 0.2445 & 0.6873. This means that in the Grade 8 sample, I fail to reject the null hypothesis of independent equations. Notwithstanding, this finding does not materially change the conclusion as the observed age trend persists, even under the school fixed effects model where bullying was assumed to be exogenously determined.

IV.I – 2SLS vs CF

As noted in Vella & Verbeek (1999) & Wooldridge (2015), the 2SLS and CF approach both generates similar estimates. I confirm this by reporting both sets of estimates below in Table 6. The differences shown in Table 6 are measured in absolute terms. As expected, all differences are statistically insignificant, even at the 10% level. This implies that the 2SLS and CF produces statistically indistinguishable estimates for the bullying *treatment*.

	Mathematics Score			Science Score		
	CF	2SLS	|Difference|	CF	2SLS	|Difference|
Panel I. Grade 4 Sample						
Bullying	-0.4908^{***}	-0.5444^{**}	0.0536	-0.5556^{***}	-0.6065^{***}	0.0509
	(0.15)	(0.25)		(0.14)	(0.23)	
Panel II. Grade 8 Sample						
Bullying	-0.2876^{*}	-0.2053	0.0823	-0.0087	0.0867	0.0954
	(0.16)	(0.17)		(0.16)	(0.19)	

Note: Standard errors in parentheses are clustered by schools. Differences are measured in absolute terms.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 6. Regression Estimates (2SLS vs CF)

Notwithstanding their algebraic equivalence (Wooldridge 2015) and the empirical equivalence confirmed above, there are still qualitative differences between the two estimation procedures that make the CF approach attractive. First, and foremost, as previously mentioned, the CF approach produces a simple Hausman test (Hausman 1978) that is robust to heteroscedasticity and cluster correction of unknown form. This allows me to empirically test the assumption of endogenous bullying. Secondly, under the CF approach, there is less of a need to fully characterise the relationship that the *treatment* has on unobservables. Lastly, since the CF approach uses generalised residuals, the CF approach is likely more efficient than 2SLS because it exploits the binary nature of the treatment variable. However, in terms of consistency, the CF approach is usually less robust than IV approaches (Wooldridge 2015).

IV.II – Heterogeneous Effects of Bullying

To gain further insights into the effect of bullying, in this sub-section, I check for heterogeneous effects of bullying for different subgroups. I do this via the inclusion of interaction terms.

Gender Differences

In this section, I explore the possibility of heterogeneous effects of bullying, owing to gender differences. I do this via the inclusion of a gender term interacted with the bullying variable in Equation (1). Referencing Table A2 in the Appendix, I find that the interaction terms are consistently positive and statistically significant, at the 10% level, in the Grade 4 sample, regardless of the chosen outcome variable and specification. While the net-effect of being bullied is still negative for male students, this evidence suggests that, conditional on being bullied, male students are observed to be less affected by bullying than their female peers. This gender-differing effect is not observed in the Grade 8 sample, as all interaction terms are statistically insignificant even at the 10% level. It is worth noting that this gender differential was previously documented in both the medical (Sourander, et al 2009) and psychology literature (Wolke, Woods & Samara 2009).

Furthermore, pair-wise comparisons of the interaction terms in Panel I indicate that the coefficient estimates for the interaction terms are not statistically different from each other, at 5% significance level, under both approaches, regardless of the chosen outcome variable. This implies that the gender-differing effect estimated by the two different estimation procedures produced statistically indistinguishable estimates, at 5% significance level.

Citizenship Status

Naturally, another sub-group of interest are immigrants. As above, I explore the possibility of heterogenous effects of bullying, owing to citizenship status, via the inclusion of the relevant interaction terms. Results are presented in the Appendix, in Table A3. In the Grade 4 dataset, immigrants are found to be less affected by bullying, as evidenced by the consistently positive and statistically significant coefficient estimates, at 10% significance level, regardless of chosen outcome variable; no effect was found in the Grade 8 sample. One possible explanation for this observation is the idea of familialism – a cultural value that emphasizes close family relationships. This feature was previously reported to be more prevalent in

Hispanic and Asian immigrant families in the US (Campos, et al 2008; Toyokawa & Toyokawa 2013) as well as in Canada & Australia (Boucher 2013). The notion of familialism might induce a better social support structure in immigrant families, as immigrants are likely to have fewer social networks outside their immediate and extended families. This could be a potential explanation as to why immigrants are observed to be less affected by bullying.

Next, I explore the possibility of differing effects owing to the type of immigrants in Singapore; namely, Asian immigrants vs non-Asian immigrants. To do this, I interact the existing interaction term with a binary variable that corresponds to whether or not the student speaks English at home. As Asian immigrants are more likely to be learning English as a second language, it seems plausible to assume that Asian immigrants are more likely to speak their Mother Tongue at home, as opposed to non-Asian immigrants who are more likely to have English as their first language.

When math scores are the chosen outcome variable, this new interaction term yields statistically significant and negative point estimates in both the Grade 4 and Grade 8 sample, at the 10% level. This implies that immigrants who are native English speakers are observed to be more affected by bullying, conditional on being bullied. In addition, across both samples, the reported incidence rate of bullying among native English-speaking immigrants is higher than the sample average, while non-native English-speaking immigrants report a lower incidence rate of bullying, compared to the sample average, as presented below in Table 7.

Variable: Bullying	Mean	St.Dev	Mean	St.Dev
Grade 4			Grade 8	
Full Sample	0.3752	0.4842	0.2600	0.4387
Singapore Citizens	0.3725	0.4835	0.2587	0.4380
Immigrants that do not speak English at home	0.3748	0.4841	0.2574	0.4373
Immigrants that speak English at home	0.3793	0.4859	0.3035	0.4609

Table 7. Descriptive Statistics for the bullying variable across different citizenship sub-groups

Together, it suggests that discrimination might have a role to play in explaining why native English-speaking immigrants are observed to be more affected by bullying than their non-native English-speaking counterparts. In all, there exists considerable evidence that suggests heterogeneity in the effect of bullying, owing to citizenship status. Furthermore, these findings persist even after controlling for parental education level, as the majority of immigrants are observed to come from high parental education households; 52% of immigrants come from households with high parental education (defined in the next section below). Therefore, given that the effect persists after controlling for parental education level, it is less likely that these effects are driven by income, and more so by their citizenship status.

Income Level

Next, I explore the possibility that students belonging to households with different income levels might experience heterogeneous effects relating to bullying. The 3 proxies of income used in this paper – Low Income, Middle Income & High Income – are binary variables that relate to the student's response to a question querying whether the student has his/her own room, and/or internet access at home. Low income students are classified as students who responded no to both, while middle income students responded yes to either, and high income students responded yes to both. Regression estimates are presented in the Appendix, in Table A4a. I find no evidence of any heterogenous effects owing to this proxy of income, as all interaction terms are statistically indistinguishable from zero, at even the 10% level, in both samples.

However, I acknowledge that this proxy for income might, in fact, be too coarse a measure since Singapore has relatively a high internet penetration rate of 84.5%[3] as of 2019. Furthermore, given the fact that the housing situation in Singapore is such that approximately 81%[4] of the population live in apartments, if a household has 2 or more children – which applies to approximately 42%[5] of households in Singapore, as of 2018 –, then the child will likely be sharing rooms with a sibling.

Thus, I define a new variable that is arguably a better proxy for income: parental education. These variables are binary in nature. I classify households as ones with high parental education if the highest education level attained between both parents is at least a bachelor's degree. If at most

[3] Source: https://www.internetworldstats.com/
[4] Source: Housing Development Board (HDB) Key Statistics http://www10.hdb.gov.sg/ebook/AR2018/key-statistics.html
[5] Source: singstat.gov.sg

one parent has a bachelor's degree or higher, I classify them as middle parental education households. The remainder of the sample are then classified as low parental education households. I interact the bullying variable with these new parental education variables and report my findings in the Appendix, in Table A4b. Under this finer proxy for income level, I find some evidence for heterogeneous effects.

Firstly, in the Grade 4 sample of Singapore students, I find some evidence of heterogeneous effects of bullying, as the interaction terms for high parental education is consistently positive and statistically significant at the 5% level in the Grade 4 sample, regardless of the outcome variable of choice. No effect was observed in the Grade 8 sample. This suggests that Grade 4 students belonging to households with high parental education are observed to be less affected by bullying, conditional on being bullied.

As presented in Table 8 below, this effect is unlikely to be driven by an increased propensity to be bullied, as the incidence rate for students from high parental education households are, in fact, the lowest among the 3 groups.

Variable: Bullying	Mean	St.Dev		Mean	St.Dev
Grade 4			Grade 8		
Full Sample	0.3752	0.4842		0.2600	0.4287
Low Parental Education	0.4224	0.4941		0.2618	0.4397
Middle Parental Education	0.3628	0.4811		0.2872	0.4585
High Parental Education	0.3155	0.4649		0.2357	0.4247

Table 8. Descriptive Statistics for the bullying variable across different income sub-groups

The potential differences in the availability of home resources and support for bullied students could account for the positive interaction terms observed in the students belonging to high parental education households.

Adverse Environment

Next, I explore the possibility that the effect of bullying might be different for students in adverse school environments, as bullying might be perceived as a more commonplace phenomenon. I define an indicator variable that takes the value 1 if a student belongs to a school that has an average incidence rate of bullying that is higher than the average across all schools in Singapore. I then interact this new *Adverse Environment* variable with the bullying variable to check for any differing effects. I report my findings in the Appendix, in Table A5.

I find little evidence of such effects. Only 3 out of the 8 interaction terms are statistically significant at the 10% level, across all specifications in both the Grade 4 and Grade 8 sample. Therefore, there is insufficient evidence to suggest that being in an adverse school environment will dampen the effect of bullying on test scores.

Physical vs Relational Bullying

All the analyses presented in this paper thus far, has been based on a broad definition of the bullying variable; which takes the value 1 if a student experiences any of the 8 bullying-related incidents on a weekly basis.

Based on the data available, it is possible to separately analyse the effect of physical and relational bullying. I now generate a new *Physical Bullying* indicator variable that takes the value 1 if a student experiences any of the physical bullying incidents – Incidents 4, 5, 6 & 8, in Table A0 – on a weekly basis, 0 otherwise. Likewise, another new indicator variable, *Relational Bullying*, is created and takes the value 1 if a student experiences any of the relational bullying incidents – Incidents 1, 2, 3 & 7, in Table A0 – on a weekly basis, 0 otherwise. In addition, I also include gender interaction terms to test for the presence of gender differences. This is because it is documented in the psychology literature that male students are more likely to engage in physical bullying, whilst female students are more likely to engage in non-physical bullying (Olweus 1991).

	Mean	St.Dev	Mean	St.Dev
Panel I (Variable: Physical Bullying)				
Grade 4			Grade 8	
Full Sample	0.1961	0.3971	0.0797	0.2709
Males	0.2371	0.4254	0.1127	0.3163
Females	0.1556	0.3626	0.0507	0.2194
Panel II (Variable: Relational Bullying)				
Grade 4			Grade 8	
Full Sample	0.3271	0.4692	0.2436	0.4293
Males	0.3772	0.4848	0.3186	0.4660
Females	0.2776	0.4949	0.1776	0.3823

Table 9. Descriptive Statistics for the physical bullying variable for male and female students

Focusing on physical bullying, there are two points to note. First, referencing Panel I in Table 9 above, I note that male students are more likely to be report being a victim of physical bullying, than their female counterparts. This holds true for both the Grade 4 and Grade 8 sample.

Furthermore, male students account for over 60% of students who report being physically bullied; 60% in the Grade 4 sample and 66% in the Grade 8 sample.

This gender disparity persists even when turning our attention to relational bullying, in Panel II of Table 9. Male students report a higher incidence rate of relational bullying. Similarly, male students account for nearly 60% of victims of relational bullying; 57% in the Grade 4 sample, and 61% in the Grade 8 sample.

Therefore, the stark gender difference underpins why gender interaction terms are also included in this sub-section. Regression estimates for *Physical Bullying* & *Relational Bullying* can be found in the Appendix, in Tables A6 & A7 respectively.

Focusing first on physical bullying in Table A6, I find that the regression estimates for physical bullying are consistently negative and statistically significant at 5% level, in the Grade 4 sample, regardless of the outcome variable chosen, or the specification employed. Furthermore, in most instances, the interaction terms are positive and statistically significant at the 5% level. This suggests that, conditional on being a victim of physical bullying, males are found to be less affected by it than their female counterparts. Turning our attention to the Grade 8 sample, I find similar evidence that males are less affected by physical bullying, as the interaction terms are consistently positive, and statistically significant at the 10% level. Thus, I can safely conclude that there is sufficient evidence present to suggest heterogeneity in the effect of physical bullying owing to gender.

With regards to relational bullying, presented in the Appendix, in Table A7, I note that the regression estimates for relational bullying are consistently negative and statistically significant across all specifications, at the 5% level, invariant to the choice of outcome variable, in the Grade 4 sample. In addition, the interaction terms are also found to be statistically significant and consistently negative at the 5% level. This finding suggests that female students are more affected by relational bullying than their male peers, conditional on being a victim of relational bullying. Turning our attention to the Grade 8 sample, I find no evidence of this as all interaction terms are statistically indistinguishable from zero, at even 10% significance level. Furthermore, the regression estimates for relational bullying are found to only be statistically significant when math scores were the chosen outcome variable. Thus, I can conclude that there is some evidence suggesting heterogeneity in the effect of relational bullying, owing to gender.

However, I am cognizant of the possibility of measurement error; specifically, the under-reporting of bullying by female students. This is likely as my findings run counter with the findings in the psychology literature. Differences in the reporting habits might contribute to potential measurement errors, which could potentially understate the effect for female students.

Positively Selected Students

Lastly, in this sub-section, I explore if positively selected students are differentially affected by bullying. As standardised test scores are used in the analyses, positively selected students are defined as students who scored at least 1 SD higher than the sample mean, for both math and science. In the Grade 4 sample, 11.9% of students are characterised as positively selected, while 8.6% of students in the Grade 8 sample are characterised as positively selected. The results are presented in the Appendix, in Table A8a.

In the Grade 4 sample, the interaction terms are consistently positive and are statistically significant at the 5% level, regardless of the choice of outcome variable, or specification. This effect was not observed in the Grade 8 sample. This implies that positively selected Grade 4 students are observed to be less affected by bullying than their peers. Referencing Table 10 below, I note that these students are also found to be marginally less likely to experience bullying, evidenced by both a lower mean as well as standard deviation.

Variable: Bullying	Mean	St.Dev	Mean	St.Dev
Grade 4			Grade 8	
Full Sample	0.3752	0.4842	0.2600	0.4287
Positively Selected Students	0.2378	0.4262	0.2368	0.4258

Table 10. Descriptive Statistics for the bullying variable for positively-selected students

Next, I evaluate the validity of benchmarking against the sample mean. This is because students are unlikely to have information on the distribution of their national cohort across all schools in Singapore; in fact, they are more likely to be privy to the quality of students in their cohort in each school. Therefore, I repeated the above analyses with one minor change: positively selected students are now defined as students who score 1 SD higher than their respective sample school means, in both math and science. Under this new definition, 9.1% of students in the Grade 4 sample are classified as positively-selected and 7.4% of students in the Grade 8 sample are classified as positively-selected. I report the regression estimates in Table A8b.

In short, I obtain very similar findings under this new definition as before in Table A8a. The interaction terms are consistently positive, and statistically significant at the 5% level in the Grade 4 sample, regardless of the choice of outcome variable or specification.

Given the results obtained in this sub-section thus far, naturally, I explored to see if there are any differing effects observed by negatively-selected students. I mirror the earlier definition for positively-selected students and similarly define negatively-selected students as those who score 1 SD *lower* than the sample mean, in both math and science. In the Grade 4 sample, 9% of students are classified as negatively-selected while 12% of students in the Grade 8 sample are classified as negatively-selected. The interaction terms are all statistically insignificant in all specifications – results are not presented in the paper, but available upon request –, in both the Grade 4 and Grade 8 sample, regardless of the outcome variable of choice. Furthermore, I obtain statistically insignificant coefficient estimates even when adopting the alternative measure where I evaluate their scores against their respective school means. In short, I find no evidence that suggests the existence of heterogenous effects of bullying owing to negatively-selected students.

Thus, I can safely conclude that there is evidence pointing towards heterogeneity in the effect of bullying, owing to student ability.

IV.III Robustness Checks

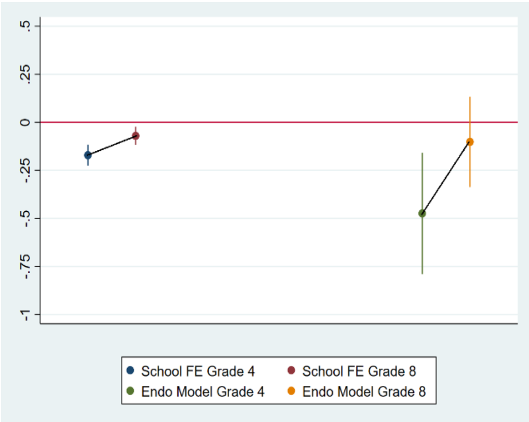
In this section, I present a set of robustness checks and alternative specifications to support my findings presented in this paper. Broadly speaking, I explore alternative definitions of bullying, as well as consider another alternative specification and an expanded set of instruments.

Broader Definition of Bullying

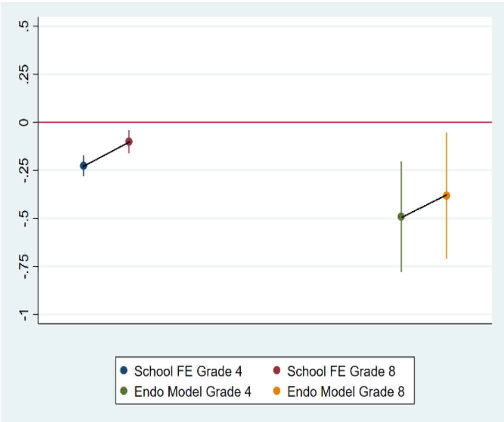
First, I explore to see if my findings are robust to how the bullying variable is defined in this paper. To do this, I broaden the definition of bullying variable to also include students who experience bullying related incidents on at least a monthly basis as well. Under this broader definition, I note that approximately 56% of students are characterised as being victims of bullying in the Grade 4 sample – of which 55% are male and 45% are female –, while approximately 49% of students are defined as being victims in the Grade 8 sample, with 55% of them being male, and 45% female as well.

Regression estimates are presented in the Appendix, in Table A9. I find that the regression coefficients of this broadly defined bullying variable are consistently negative and statistically significant at the 5% level, in the Grade 4 sample of students. However, in the Grade 8 sample of students, only the School FE model produces statistically significant coefficient estimates, while the CF approach does not result in statistically significant coefficient estimates, regardless of the outcome variable of choice. Notwithstanding, I plot the coefficient estimates below in Figure 2, and accompany them with the plots in Figure 1, for easy comparison. The vertical scale is maintained across all plots for ease of comparison.

Math Score as Outcome Variable (Broader Definition of Bullying)



Math Score as Outcome Variable (**Original** Definition of Bullying)



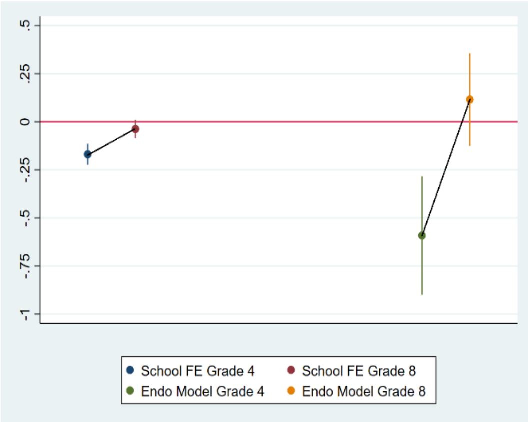
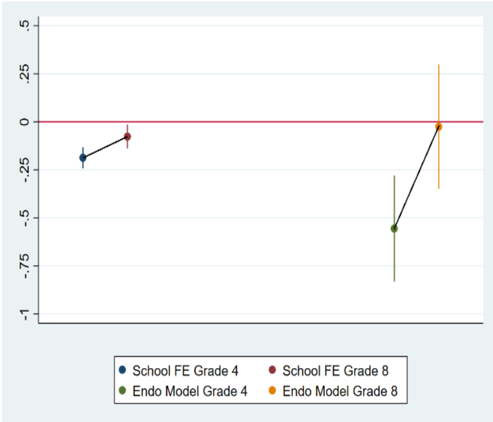
Science Score as Outcome Variable (Broader Definition of Bullying)	Science Score as Outcome Variable (**Original** Definition of Bullying)
	

Figure 2. Plots for the coefficient of broader vs original bullying variables

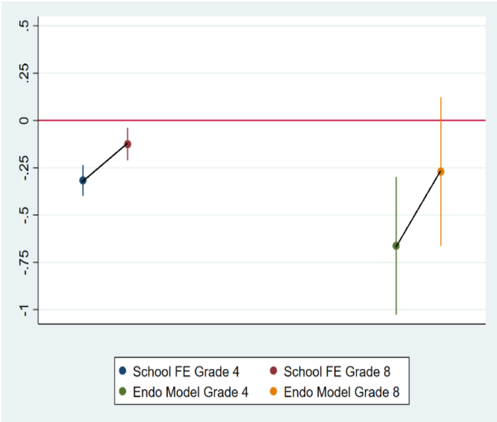
Referencing Figure 2, I note that under this broader definition of bullying, at 10% significance level, all pair-wise comparisons between Grade 4 and their Grade 8 counterparts are statistically significant. This implies that there is sufficient evidence to conclude that the effect of bullying is lower for Grade 8 students than their Grade 4 counterparts. This was also the conclusion drawn from the initial analysis, as a similar trend was observed. Thus, under this broader definition of bullying, I still find evidence suggesting the notion of age trend relating to bullying.

Stricter Definition of Bullying

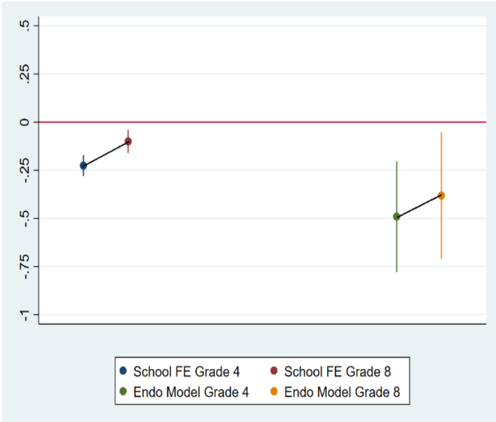
Next, I employ a stricter a stricter definition for bullying, such that only approximately 10% of the students from each sample are classified as being victims of bullying. This can be seen as a measure of the effect of severe bullying, in addition to being a robustness exercise. The results are reported in the Appendix, in Table A10.

I find that the coefficient estimates for bullying, under this new stricter definition, are consistently negative, and statistically significant at the 5% level, in all specifications in the Grade 4 sample, while they are found to be statistically significant only in the OLS specification, in the Grade 8 sample. I plot the coefficient estimates below in Figure 3, and parallel them with the plots using the original threshold for bullying.

Math Score as Outcome Variable (Stricter Definition of Bullying)



Math Score as Outcome Variable (**Original** Definition of Bullying)



Science Score as Outcome Variable (Stricter Definition of Bullying)

Science Score as Outcome Variable (**Original** Definition of Bullying)

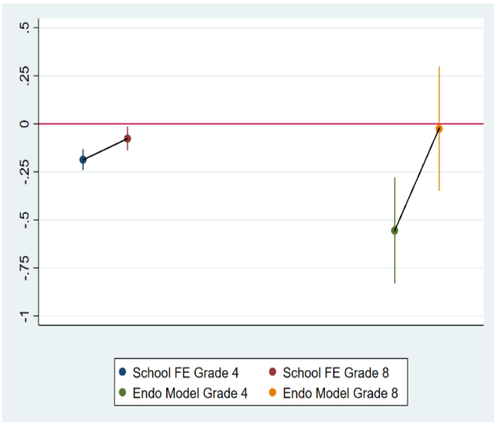


Figure 3. Plots for the coefficient of stricter vs original bullying variables

All pair-wise differences between Grade 4 and Grade 8 estimates are found to be statistically significant and different from zero, at the 5% level. Thus, even under this stricter definition for bullying, I am still able to obtain consistent results, as well as evidence of an age trend for bullying, with younger students observed to be more affected by bullying than their older peers.

Simultaneous Equation Specification

In this sub-section, I verify if the findings materially differ if I were to adopt a simultaneous equation model. I augment Equation (2) by also including standardised test scores as a regressor and employ both Equation (1) and (2) in a Simultaneous Equation (SE) setting. The regression estimates are presented in the Appendix, in Table A11, along with the estimates from the Control Function specification for comparison,

First, I note that in the Grade 4 sample, the estimates under both the SE & CF specifications are statistically significant. Furthermore, while the coefficient estimates obtained under SE are larger in magnitude in the Grade 4 sample, the differences are not statistically significant.

Pair-wise comparisons between the Grade 4 and Grade 8 estimates show that the differences are statistically significant only when science scores are the chosen outcome variable. Therefore, under the SE specification, there is some evidence of an age trend. I plot the coefficients below in Figure 4 to better illustrate the different specifications.

Math Score as Outcome Variable   Science Score as Outcome Variable

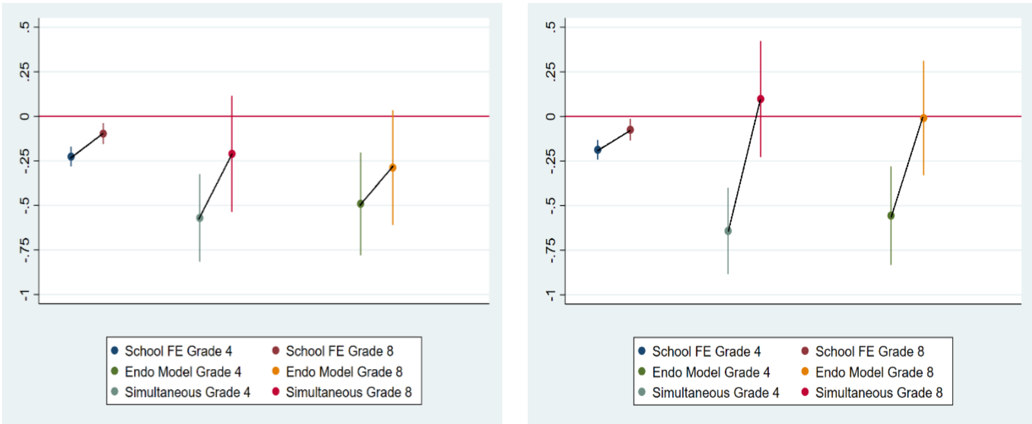


Figure 4. Plots for the coefficient of bullying variable for all specifications

Broader Class of Instruments

Throughout the paper, the instrument employed for the bullying variable is one that corresponds to the degree in which a student likes seeing their classmates at school. This variable was chosen as it was found to be correlated with the likelihood that a student is a victim of bullying, and uncorrelated with test scores.

In the TIMSS dataset, there are other similar qualitative measures of the student's environment which are informative and can therefore also act as instruments. Of them, 3 are of particular interest; in that, they are likely to be correlated with whether or not a student is a victim of bullying. These questions are: whether the student likes being in school, whether the student feels safe at school, and whether the student feels like they belong at school.

All questions are answered on a 4-point Likert scale, ranging from Disagree a lot, to Agree a lot. Descriptive statistics are provided below, in Table 11.

Variable (1 = Agree a lot, 4 = Disagree a lot)	Mean	St.Dev	Mean	St.Dev
Grade 4			Grade 8	
I like being in school	1.65	.756	1.785	.722
I feel safe when I am in school	1.616	.777	1.68	.726
I feel like I belong at this school	1.689	.845	1.858	.808
I like to see my classmates at school	1.318	.606	1.489	.687

Table 11. Descriptive Statistics for variables related to student's environment at school

In addition to the descriptive statistics presented above, I additionally present the pair-wise correlations between the individual variables with the bullying variable. Table 12 affirms their ability to act as instruments, with all pair-wise correlations being statistically significant at the 5% level.

Correlation Coefficients	Bullying (Grade 4)	Bullying (Grade 8)
I like being in school	0.0743***	0.0741***
I feel safe when I am in school	0.1342***	0.1386***
I feel like I belong at this school	0.1271***	0.1425***
I like to see my classmates at school	0.1737***	0.1519***

Table 12. Correlation coefficients for variables related to student's environment at school

The variables are added into the Z matrix in Equation (2) in the form of 3 indicator variables each – with the first category, corresponding to Agree a lot – being the omitted category for each variable. The regression estimates using this augmented Z matrix are presented in the Appendix, in Table A12. In the Grade 4 sample, the coefficient estimates for bullying increases in magnitude, under this larger set of instruments. Furthermore, under this broader set of instruments, the age trend for bullying becomes more pronounced, as can be seen below in Figure 5, with both pair-wise differences found to be statistically significant at the 5% level, under this broader set of instruments; previously, the pair-wise comparison under the CF approach, when math scores were the chosen outcome variable yielded statistically insignificant differences.

Math Score as Outcome Variable   Science Score as Outcome Variable

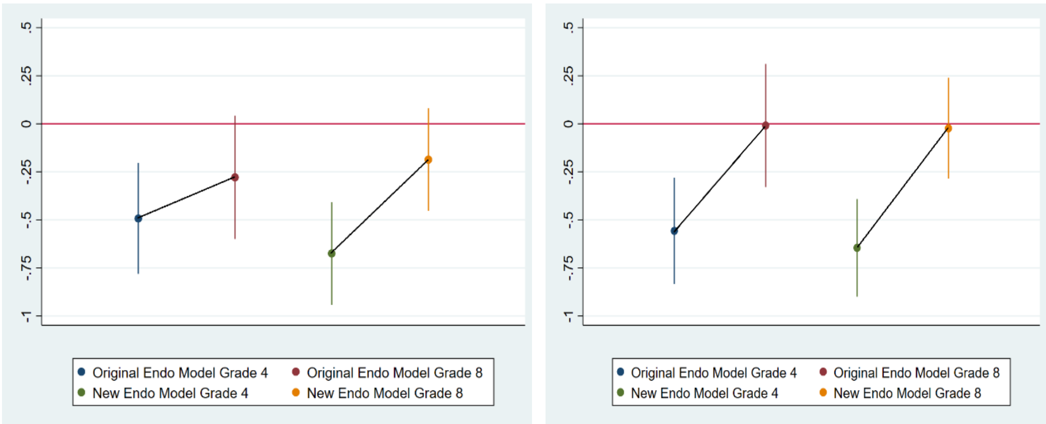


Figure 5. Plots for the coefficient of bullying variable under a larger set of instruments

Given the findings presented in this sub-section, I can therefore conclude that my main findings are not necessarily sensitive to the set of instruments used.

Survivorship Bias

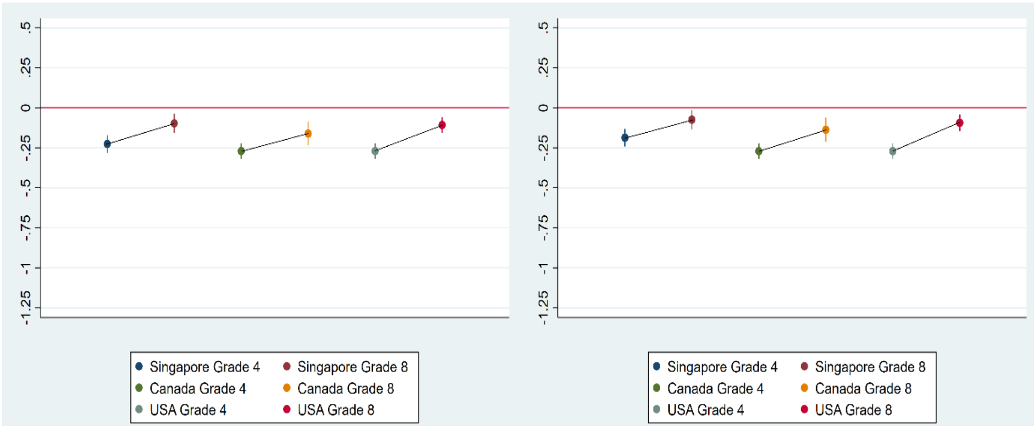
While over 99% of students in Singapore[6] continue on to secondary school, compulsory schooling only extends to Grade 6. Therefore, there is a possibility that the Grade 8 findings might be inaccurate due to survivorship bias, as some students might drop out of school because of being bullied (Cornell, et al 2013).

[6] Source: https://data.gov.sg/dataset/net-enrolment-ratio-for-primary-and-secondary-education

To that end, I replicate my main analysis using the TIMSS Canada sample of Grades 4 & 8, as well as the USA sample of Grades 4 & 8. The reason for this is because schooling up to at least Grade 8 is compulsory in both Canada, as well as in most states in the US. I plot the regression coefficients of bullying for the 3 countries – Singapore, Canada, and the USA – for comparison in Figure 6 below.

Math Score as Outcome Variable Science Score as Outcome Variable

School Fixed Effects



Math Score as Outcome Variable Science Score as Outcome Variable

Control Function

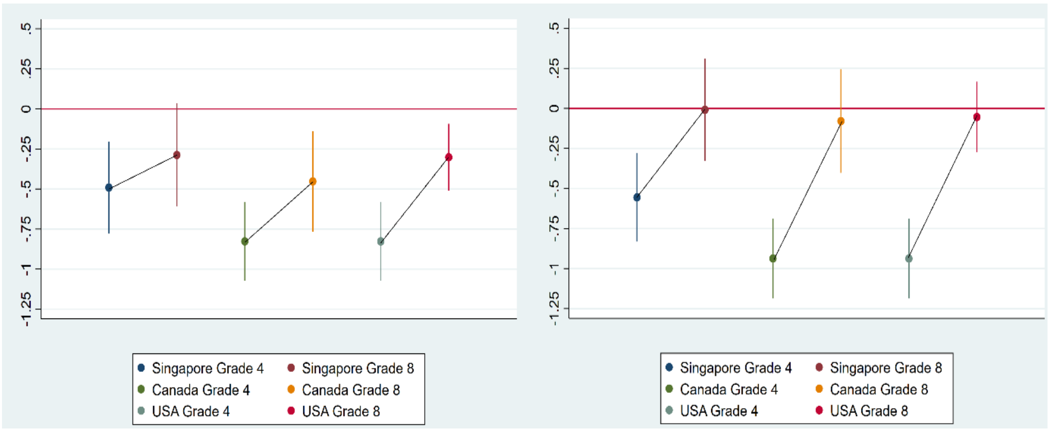


Figure 6. Plots for the coefficient of bullying variable under a larger set of instruments

In short, I observe very similar trends in both the Canada and the USA samples. However, while the similar findings might suggest that that survivorship bias might not be a severe problem, it would be premature to directly conclude as such, given that cultural differences and attitudes towards the importance of higher education are heterogeneous even across different states, let alone across countries. Therefore, this comparison can be seen as a falsification test of sorts, to verify the severity of the survivorship bias in the Singapore sample.

V. Mechanisms

The findings presented thus far show that bullying lowers the scholastic achievements of students. In this section, I explore possible mechanisms through which test scores can be affected by the act of being bullied. Broadly speaking, I explore 2 kinds of possible mechanisms; namely direct and indirect mechanisms. Understanding these mechanisms can help school administrators more efficiently tackle the problem of bullying via more efficient resource allocation, as educational institutions tend to adopt a multi-pronged approach when attempting to deal with the problem of bullying.

To do this, I examine possible channels by estimating Equation (1) using data from the student questionnaire. I explore the associations between bullying and absenteeism as a potential form of direct mechanism, as reported in the education literature (Grinshteyn & Yang, 2017; Hutzell & Payne, 2012; Randa & Reyns, 2014). As for the indirect mechanisms, I broadly classify the questions into three distinct categories: questions relating to perception of self, perception of teachers and perceived fairness. Bullying has been previously linked to lower perception of self, as well as low levels of psychological well-being (Rigby, 2003).

It is imperative to note that I am unable to provide causal interpretation to the correlations presented in this section as there might be other factors affecting both the observed mechanism, as well as test scores, or that there could potentially be reverse causation, in the sense that test scores might have an effect on the mechanism. Nevertheless, the students' subjective assessments of the school environment contain valuable information and were previously found to be associated with actual achievement in school (Fraser 1998; Freiberg 1999).

V.I – Direct Mechanism

Absenteeism

In this section, I explore the associations between bullying and absenteeism. There is 1 question related to absenteeism in the questionnaire, and it is measured on a 4-point scale, ranging from Once a week or more, to Never or almost never. I report the descriptive statistics below in Table 13.

1 = Once a week or more; 4 = Never or almost never	Mean	St.Dev	min	max
Grade 4				
How often are you absent from school?	3.654	0.784	1	4
Grade 8				
How often are you absent from school?	3.744	0.636	1	4

Table 13. Descriptive Statistics for Absenteeism Variable

Noting that the variable takes 4 discrete values, I standardise the variable against its sample mean, so as to obtain an easily interpretable coefficient estimate in the succeeding regression analysis. Using this measure of absenteeism as an outcome variable, I regress it on the bullying variable, as well as a set of individual characteristics X, class-specific characteristics C, as well as school fixed effects, to account for any school-level sentiments towards absenteeism. The coefficient estimates of bullying can be found in the Appendix, in Table A13.

First, focusing on the Grade 4 sample, I note that the coefficient of bullying is observed to be statistically significant and negatively associated with standardised absenteeism, at the 5% level. This suggests that students who are experience bullying are observed to have a 0.12 SD higher rate of absenteeism. However, the effect disappears when controlling for test scores. Interestingly, the same cannot be said when turning our attention to the Grade 8 sample, as all coefficients of bullying are observed to be statistically insignificant and indistinguishable from zero, at even the 10% level.

Therefore, the empirical evidence reported in this section supports previous findings in the education literature that bullying is tied to absenteeism. Thus, one potential direct mechanism through which bullying influences student test scores is through a higher absenteeism rate.

V.II – Indirect Mechanism

Perception of Self

There are 8 questions in the questionnaire that correspond to an individual's self-perception. The questions are all answered on a 4-point Likert scale, ranging from Disagree a lot, to Agree a lot. I present descriptive statistics for the individual questions below, in Table 14.

1 = Agree a lot, 4 = Disagree a lot	Mean	St.Dev	Mean	St.Dev
Panel I. Mathematics				
Grade 4			Grade 8	
I usually do well in Mathematics	1.958	0.881	2.216	0.971
Mathematics is harder for me than for many of my classmates	2.754	1.037	2.666	0.926
I am just not good at Mathematics	2.876	1.075	2.538	1.086
I learn things quickly in Mathematics	2.017	0.934	2.241	0.896
Mathematics makes me nervous	2.659	1.096	2.455	0.982
I am good at working out difficult Mathematics problems	2.415	0.994	2.563	0.903
Mathematics is harder for me than any other subject	2.867	1.113	2.801	1.054
Mathematics makes me confused	2.731	1.106	2.558	1.005
Panel II. Science				
I usually do well in Science	1.959	0.839	2.116	0.856
Science is harder for me than for many of my classmates	2.895	1.02	2.771	0.897
I am just not good at Science	2.969	1.023	2.648	0.995
I learn things quickly in Science	1.945	0.879	2.177	0.844
Science is harder for me than any other subject	2.947	1.067	2.84	0.924
Science makes me confused	2.968	1.053	2.715	0.951
I am good at working out difficult Science problems			2.41	0.878

Table 14. Descriptive Statistics for the Questions on Self-Perception

Noting that each variable only takes 4 discrete values, I standardise each variable against their individual sample means, so as to obtain a more straightforward interpretation for the regression coefficients. Next, I iteratively use each variable as an outcome variable, and regress each variable on the set of individual characteristics X, class-specific characteristics C, and school fixed effects, to account for any school-specific biases and sentiments which might affect in how students might evaluate teachers. The coefficient estimates can be found in Tables A14a & A14b for Math and Science respectively in the Appendix.

Included in the tables are also columns which present the coefficient estimates after controlling for the student's actual test score and test score squared. The reason for this is that sentiments towards self-worth might be influenced by test scores, as test scores provide partial information on a student's latent ability. For example, a student who has scored very well in Mathematics will be more likely to *Agree a lot* that they *usually do well in mathematics*, while a student that doesn't score well will be less likely to do so. That is indeed the case, given the statistically significant – at the 5% level – pair-wise correlation in the magnitude of –0.36. In addition, the squared test score variable is added to capture the existence of any non-linear effects in this regard.

Referencing Table A14a in the Appendix, students – in both the Grade 4 and Grade 8 sample – who experience bullying are observed to be more likely to rate themselves more negatively in multiple questions; specifically, they are more likely to think that math is tougher for them as compared to their peers as well as report that math makes them nervous and confused. More importantly, these associations persist even after controlling for actual test scores. This relationship also appears, to some degree, when Science were the chosen basis for the questions, with both Grade 4 and Grade 8 students observed to more likely to think Science is harder for themselves, as compared to their classmates, if they experience bullying. As before, the associations persist even after controlling for science test scores. The associations presented here therefore suggest that students who experience bullying have a poorer perception of self.

Perception of Teachers

Next, I explore how experiencing bullying might affect the student's perception of teachers. There are 9 questions available in the questionnaire that relate to perception of teachers. The questions are all scored on a 4-point Likert scale, with 1 corresponding to *Agree a lot*, and 4 corresponding to *Disagree a lot*. The descriptive statistics are presented below, in Table 15.

As before, since the variables take discrete values from 1 to 4, I standardise them against their sample means, so as to obtain a more straightforward interpretation of the results. Referencing Table A15a in the Appendix, focusing on Grade 4 students, I note that a student who experiences bullying are found to be more likely to negatively evaluate their math teachers in numerous aspects; namely, they do not think the teacher is easy to understand, are less interested in what the teacher has to say, do not think the teachers have clear answers to their questions, do not feel that teachers are good in explaining mathematics and believe that teachers do not listen to what they have to say.

1 = Agree a lot, 4 = Disagree a lot	Mean	St.Dev	Mean	St.Dev
Panel I. Mathematics				
Grade 4			Grade 8	
My teacher is easy to understand	1.577	0.739	1.791	0.792
I am interested in what my teacher says	1.709	0.787	2.003	0.823
My teacher gives me interesting things to do	1.753	0.842	2.211	0.857
My teacher has clear answers to my questions	1.518	0.726	1.76	0.78
My teacher is good at explaining Mathematics	1.41	0.666	1.69	0.775
My teacher lets me show what I have learned	1.87	0.878	1.936	0.788
My teacher does a variety of things to help us learn	1.492	0.722	1.908	0.808
My teacher tells me how to do better when I make a mistake	1.491	0.732	1.734	0.753
My teacher listens to what I have to say	1.637	0.835	1.799	0.778
Panel II. Science				
My teacher is easy to understand	1.56	0.75	1.825	0.788
I am interested in what my teacher says	1.597	0.779	1.817	0.797
My teacher gives me interesting things to do	1.573	0.779	1.883	0.799
My teacher has clear answers to my questions	1.53	0.751	1.762	0.761
My teacher is good at explaining Science	1.416	0.688	1.696	0.743
My teacher lets me show what I have learned	1.865	0.91	1.9	0.764
My teacher does a variety of things to help us learn	1.499	0.73	1.804	0.761
My teacher tells me how to do better when I make a mistake	1.558	0.784	1.796	0.75
My teacher listens to what I have to say	1.668	0.848	1.817	0.754

Table 15. Descriptive Statistics for the Questions on Perception of Teachers.

These negative associations are found to be statistically significant even after controlling for math test scores. With science as the basis of the questionnaire, a similar trend is observed; students who are subjected to bullying do not think teachers have clear answers to their questions and that they do not think teachers listen to what they have to say.

Interestingly, almost all associations are statistically insignificant in the Grade 8 sample, regardless of the subject of choice. However, that can be partly due to the fact that bullying is observed at a lower frequency in Grade 8 – 26% –, than in Grade 4 – 38%. Therefore, the current associations suggest that while bullied students have a negatively skewed evaluation of their teachers in Grade 4, this pessimism fades away as they get older.

Perceived Fairness

The decision to explore the notion of perceived fairness stems from recent findings in labour economics literature. In the labour market, it was found that labour supply decisions (Bracha, Gneezy & Loewenstein 2015), and more specifically, quit behaviour can be linked to the perception of relative wage unfairness amongst peers (Dube, Giuliano & Leonard 2019). Therefore, perceived fairness, or unfairness, in this case, could potentially influence the effort that students put into their education, which in turn, will affect their test scores.

In the questionnaire, there is one question which explicitly asks the student to evaluate the degree to which teachers at their school are fair to them, on a 4-point Likert scale, with responses ranging from *Agree a lot* to *Disagree a lot*. The descriptive statistics are reported below, in Table 16. This variable will hitherto be referred to as fairness.

1 = Agree a lot, 4 = Disagree a lot	Mean	St.Dev	Mean	St.Dev
Grade 4			Grade 8	
Teachers at my school are fair to me	1.611	0.773	1.807	0.76

Table 16. Descriptive Statistics for variables on perceived fairness

Referencing Table 17 below, I note that the bullying variable yielded consistently positive, and statistically significant coefficients when the fairness variable was the outcome variable of choice.

It is entirely plausible that students who are not doing well in school might be more inclined to dislike studying, and by extension, their teachers. However, the results indicate that conditional on individual and class characteristics, as well as obtained test scores and their squared counterpart – as a rough proxy for student aptitude in school –, students who are bullied are found to be more likely to perceive unfairness from their teachers. This association is found in both the Grade 4 and Grade 8 sample.

	Variable of Interest: Bullying			
	Grade 4		Grade 8	
	(1)	(2)	(3)	(4)
Outcome Variable				
Teachers at my school are fair to me	0.3178***	0.2956***	0.1669***	0.1555***
	(0.0.040)	(0.040)	(0.049)	(0.050)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Test Scores & Test Score Squared		✓		✓

Note: Standard errors in parentheses are clustered by schools. Standardised outcome variables are used in all regressions. * $p < .1$, ** $p < .05$, *** $p < .01$

Table 17. Regression estimates with various outcome variables measuring perceived fairness

Coupled with the previous findings in this section, this suggests that students who are bullied, on top of having poorer perception of self, as well as that of their teachers, they also are found to be more likely to think their teachers are treating them unfairly. Peterson, et al (2016) found that teachers' explicit and implicit attitudes are associated with educational achievement. Therefore, it is troubling if the act of being bullied somehow affects their attitudes towards their teachers and also how they perceive themselves. Like previously mentioned, perceived unfairness has been linked to quit behaviour in the labour market (Dube, Giuliano & Leonard 2019). Therefore, it might be the case that perceived unfairness affects the amount of effort a student puts into his schoolwork; since Grades 1 to 6 are compulsory in Singapore. I attempt to verify this hypothesis in the subsequent section.

Effort as a possible mechanism

To recap, I have found associations suggesting that bullied students are more likely to perceive themselves and their teachers more negatively. Furthermore, they are also observed to perceive unfairness by teachers. One possible mechanism through which these factors might influence test scores is through the level of effort that a student puts into their studies.

To do this, I evaluate associations between being bullied and responses on their interest in mathematics and science. There are 5 questions that corresponding to how a student feels about learning mathematics and science. The questions, answered on a 4-point Likert scale, measure their attitudes towards mathematics and science, and can therefore serve as a proxy for effort, as the positive associations between interest and effort has long been documented in the psychology literature (Dewey 1913). As always, the standardised versions of the variables are employed in the OLS regression, for ease of interpretability. The questions and their accompanying descriptive statistics are presented below in Table 18.

1 = Agree a lot, 4 = Disagree a lot	Mean	St.Dev	Mean	St.Dev
Panel I. Mathematics				
Grade 4			Grade 8	
I wish I did not have to study mathematics	3.14	1.039	2.802	1.058
Mathematics is boring	3.122	0.985	2.795	.949
I like mathematics	1.789	0.936	1.977	.931
I like any schoolwork that involves numbers	2.125	0.93	2.417	.897
I look forward to mathematics lessons	2.001	0.96	2.346	.941
Panel II. Science				
I wish I did not have to study science	3.288	0.954	3.046	.954
Science is boring	3.365	0.88	3.083	.903
I like science	1.633	0.844	1.786	.841
I look forward to learning science in school	1.695	0.861	1.905	.875
I like to do science experiments	1.257	0.603	1.601	.798

Table 18. Descriptive Statistics for the Questions on Effort

Referencing Table A16 in the Appendix, I find virtually no statistically significant correlations in the Grade 8 sample. Focusing on the Grade 4 sample in Panel I, however, there are a few interesting statistically significant associations. Firstly, bullied students are more likely to display lower levels of interest in mathematics, even after controlling for test scores; bullied students are more likely to respond that they feel mathematics is boring and that they wish they do not have to study mathematics. Second, bullied students are also observed to answer more negatively when queried if they like mathematics. However, I find no such associations when science was the chosen basis for these questions.

Thus, I can conclude that there is some evidence suggesting that bullying is associated with lower interests in studying, and that this could potentially be the mechanism through which bullying affects test scores in school. As mentioned earlier, I am unable to provide causal interpretations as these variables deal with highly intangible characteristics of students, and therefore clean identification proves challenging. Notwithstanding, the consistent associations presented in this and the preceding sections are informative in helping school administrators target their efforts in addressing the bullying problem in schools.

VI. Conclusion

In this paper, I estimate the effects of being bullied on student test scores. The paper contributes to the existing body of literature in three aspects: first, this paper looks at the short-term effect of bullying whilst dealing with its endogeneity through the use of a novel instrument; second, to the best of my knowledge, I am the first to explore the separate effects of physical and relational bullying; lastly, through means of an extensive survey on the school environment as well as the individual subjects, I explore possible direct and indirect mechanisms through which test scores are affected by bullying.

I find that the effect of bullying is over twice as large under the endogenous framework. I also find significant evidence suggesting an age-trend for bullying, with older students being less affected than their younger counterparts. I also find significant evidence of heterogeneity across various sub-groups; for example, male students are found to be less affected by bullying as compared to their female peers. Furthermore, male students are also found to be less affected by physical bullying while female students are found to be more affected by relational bullying.

With regards to potential mechanisms, I observe that bullying victims have a poorer perception of self and often negatively evaluate their teachers. Furthermore, these students are also more likely to perceive unfair treatment by their teachers, as well as report lower levels of interest in their studies. Worryingly, these associations persist even after controlling for test scores, thereby suggesting that these differing evaluations are not driven predominantly by their academic performance in school. Lastly, there is also some evidence suggesting that victims of bullying report lower levels of interest in studies.

Overall, the findings presented in this paper provide more insight into understanding the short-term effects of bullying and may have important implications for the design of educational policies. As evidenced in Figure 6, the age trend in bullying – with older students being less academically affected by bullying – is not necessarily a phenomenon specific to the Singapore education system.

There are several avenues worth exploring in the advancement of the economic literature on bullying. Firstly, in this paper, I am unable to disentangle the effect of being a victim of bullying from the effect of also concurrently being a bully towards others. Furthermore, it will also be informative to explore the possibility of a bullying cycle, where bullied students have a higher propensity to bully others in the future.

<u>References</u>

Altonji, J., Elder, T., & Taber, C. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy, 113*(1), 151-184.

Beaty, L. A., & Alexeyev, E. B. (2008). The problem of school Bullies: What the research tells us. *Adolescence*, 43 1-11.

Becker, D. (2000). A depression–prevention and life-skills program for children and adolescent, *Penn resiliency project*. University of Pennsylvania.

Blanchflower, D., & Oswald, A. (2004). Well-being over time in Britain and the USA. *Journal of Public Economics*, 88 1359–1386.

Boucher, A (2013) "Familialism and migrant welfare policy: Restrictions on social security provision for newly-arrived immigrants," *Policy and Politics*

Bracha, A., Gneezy, U. & Loewenstein, G. (2015). Relative Pay and Labor Supply, *Journal of Labor Economics,* vol 33, no. 2:297-315.

Brown, S. & Taylor, T. (2008). Bullying, education and earnings: Evidence from the National Child Development Study, *Economics of Education Review*, vol. 27(4).

Busch, V., Loyen, A., Lodder, M., Schrijvers, A. J. P., van Yperen, T. A., & de Leeuw, J. R. J. (2014). The Effects of Adolescent Health-Related

Behavior on Academic Performance: A Systematic Review of the Longitudinal Evidence. *Review of Educational Research*, *84*(2), 245–274.

Campos, B., Schetter, C.D., Abdou, C.M., Hobel, C.J., Glynn, L.M., and Sandman, C.A. (2008). Familialism, social support, and stress: Positive implications for pregnant Latinas. *Cultural Diversity and Ethnic Minority Psychology*, *14*(2), 155.

Carneiro, P., Crawford, C., & Goodman, A. (2007). The impact of early cognitive and non-cognitive skills on later outcomes, *CEE Discussion Paper 00092.*

Carrell, S. E., Hoekstra, M., & Kuka, E. (2018). The Long-Run Effects of Disruptive Peers. *American Economic Review* 108 (11): 3377-3415.

Chevalier, A., & Feinstein, L. (2006). Sheepskin or Prozac: The causal effect of education on mental health, London School of Economics, *Centre for the Economics of Education Discussion Paper 0071.*

Clark, A. & Oswald, A. (2002). A simple statistical method for measuring how life events affect happiness. *International Journal of Epidemiology*, 31 1139–1144.

Cornell, D., Gregory, A., Huang, F., & Fan, X. (2013). Perceived prevalence of teasing and bullying predicts high school dropout rates. *Journal of Educational Psychology, 105*(1), 138-149.

DCSF, United Kingdom. (2004). Every child matters: Change for children in schools, Department for Children, Schools and Families.

Dewey, J. (1913). *Riverside educational monographs. Interest and effort in education.* Boston, MA, US: Houghton Mifflin Company.

Dube, A., Giuliano, L. & Leonard, J. (2019). Fairness and Frictions: The Impact of Unequal Raises on Quit Behavior. *American Economic Review*, 109 (2): 620-63.

Due, P., Hansen, E. H., Merlo, J., Andersen, A., Holstein, B. E. (2007) Is victimization from bullying associated with medicine use among adolescents? A nationally representative cross-sectional survey in Denmark, *Pediatrics*, 120(1), 110-117

Eriksen, T., Nielsen, H., and Simonsen, M. (2012). The effects of bullying in elementary school. *Economics Working Papers,* 16.

Eriksen, T. L. M., Nielsen, H. S., and Simonsen, M. (2014). Bullying in elementary school. *Journal of Human Resources*, 49(4):839–71.

Fraser, J.B. (1998). Classroom environment instruments: development, validity, and applications, *Learning Environment Research*, 1(1):7–34.

Freiberg, H.J. (1999). School Climate: Measuring, Improving and Sustaining Healthy Learning Environments, Philadelphia: Falmer Press.

Grinshteyn, E., Yang, Y. T. (2017) The association between electronic bullying and school absenteeism among high school students in the United States, *Journal of School Health*, 87(2) 142-149

Grossman, M. (2005). In E. Hanushek, & F. Welch (Eds.), Education and non-market outcomes chapter 10 in the '*Handbook of The Economics of Education'*.

Hausman, J. A. (1978). "Specification Tests in Econometrics." *Econometrica,* 46, 1251-1271.

Heckman, J. (2000). Policies to foster human capital. *Research in Economics*, 54, 3–56.

Heckman, J. J., Robb, R. (1985). "Alternative Methods for Evaluating the Impact of Interventions". In Heckman, J.; Singer, B. (eds.). *Longitudinal Analysis of Labor Market Data*

Heckman, J., Stixrud, J. & Urzua, S. (2006a). The effects of cognitive and noncognitive abilities on labour market outcomes and social behavior. *Journal of Labor Economics* 24(3), 411–480.

Heckman, J. J., Urzua, S., and Vytlacil, E. (2006b). Understanding Instrumental Variables in Models with Essential Heterogeneity. Review of Economics and Statistics, 88(3):389–432.

Herge, W. M., La Greca, A. M., Chan, S. F. (2016) Adolescent peer victimization and physical health problems, *Journal of Pediatric Psychology*, 41(1), 15-27

Hutzell, K. L., Payne, A. A. (2012) The impact of bullying victimization on school avoidance, *Youth Violence and Juvenile Justice*, 10(4), 370-385

Imbens, G., & Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica, 62*(2), 467-475.

Juvonen, J., Wang, Y., & Espinoza, G. (2011). Bullying Experiences and Compromised Academic Performance Across Middle School Grades. *The Journal of Early Adolescence*, *31*(1), 152–173.

MOE, Singapore. (2018) 'Learn For Life' – Preparing Our Students To Excel Beyond Exam Results, Press Release 28th September 2018

Murphy, F. X. (2019). Does Increased Exposure to Peers with Adverse Characteristics Reduce Workplace Performance? Evidence from a Natural Experiment in the US Army, *Journal of Labor Economics,* 37(2), 435-466.

Nakamoto, J., Schwartz, D. (2010) Is peer victimization associated with academic achievement? A meta-analytic review, *Social Development*, 19(2) 221-242

Olweus, D. (1991). Bully/victim problems among schoolchildren: Basic facts and effects of a school based intervention program, *The development and treatment of childhood aggression*, 411-448

OECD (2017), PISA 2015 Results (Volume III): Students' Well-Being, PISA, OECD Publishing, Paris.

Papageorge, N. W., Ronda, V. & Zheng Y. (2017). The Economic Value of Breaking Bad: Misbehavior, Schooling and the Labor Market, *IZA Discussion Paper No. 10822*

Peterson, E., Rubie-Davies, C., Osborne, D. & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction*. 42. 123-140.

Pittman, L. D. & Richmond, A. (2007). Academic and Psychological Functioning in Late Adolescence: The Importance of School Belonging, *The Journal of Experimental Education*, 75:4 270-290

Ponzo, M., (2013). Does Bullying Reduce Educational Achievement? An Evaluation Using Matching Estimators". *Journal of Policy Modeling*, 35 (6), pp. 1057-1078.

Randa, R., Reyns B. W. (2014) Cyberbullying victimization and adaptive avoidance behaviors at school, *Victims & Offenders*, 9(3), 255-275, Reivich, K., Gillham, J., Shatte, A. & Seligam, M. (2005). A resilience initiative and depression prevention program for youth and their parents, *Penn resiliency program*. University of Pennsylvania.

Rigby, K. (2003). Consequences of Bullying in Schools. *The Canadian Journal of Psychiatry*, 48(9), 583–590.

Roberts, B. W., Harms, P. D., Caspi, A, & Moffitt, T. E. (2007). “Predicting the Counterproductive Employee in a Child-to-Adult Prospective Study.” *Journal of Applied Psychology*, 92 1427–1436.

Samdal, O., Nutbeam, D., Wold, B., Kannas, L. (1998). Achieving health and educational goals through schools—a study of the importance of the school climate and the students' satisfaction with school, *Health Education Research* 13-3, 383–397.

Sarzosa, M., & Urzúa, S. (2016) Bullying among Adolescents: The Role of Cognitive and Non-Cognitive Skills, Working Paper, Version: October 2016

Segal, C. (2013), Misbehavior, Education, and Labor Market Outcomes. *Journal of the European Economic Association* 11: 743-779.

Sourander A, Ronning J, Brunstein-Klomek A, et al. (2009) Childhood Bullying Behavior and Later Psychiatric Hospital and Psychopharmacologic Treatment: Findings From the Finnish 1981 Birth Cohort Study, *Arch Gen Psychiatry,* 66(9), 1005–1012.

Strøm, I. F., Thoresen, S., Wentzel-Larsen, T., Dyb, G. (2013) Violence, bullying and academic achievement: A study of 15-year-old adolescents and their school environment, *Child Abuse & Neglect*, 37(4), 243-251.

Toyokawa, N., & Toyokawa, T. (2013). The construct invariance of family values in Asian and Hispanic immigrant adolescents: An exploratory study. *Asian American Journal of Psychology,* 4(2), 116-125.

Vella, F., & Verbeek, M. (1999). Estimating and Interpreting Models with Endogenous Treatment Effects. *Journal of Business & Economic Statistics, 17*(4), 473-478.

Wigderson, S., Lynch, M. (2013) Cyber- and traditional peer victimization: Unique relationships with adolescent well-being, *Psychology of Violence*, 3(4), 297-309.

Wolke, D., Woods, S. & Samara, M. (2009), Who escapes or remains a victim of bullying in primary school? *British Journal of Developmental Psychology*, 27: 835-851.

Wooldridge, J. M. (2015) Control Function Methods in Applied Econometrics, *Journal of Human Resources*, 50(2), 420-445.

Appendix

Components of Bullying Variable

1) How often are you made fun of?
2) How often are you left out of games?
3) How often do others spread lies about you?
4) How often do people steal something from you?
5) How often are you hurt by others?
6) How often are you forced to do something?
7) How often do people spread embarrassing information about you?
8) How often are you being threatened?

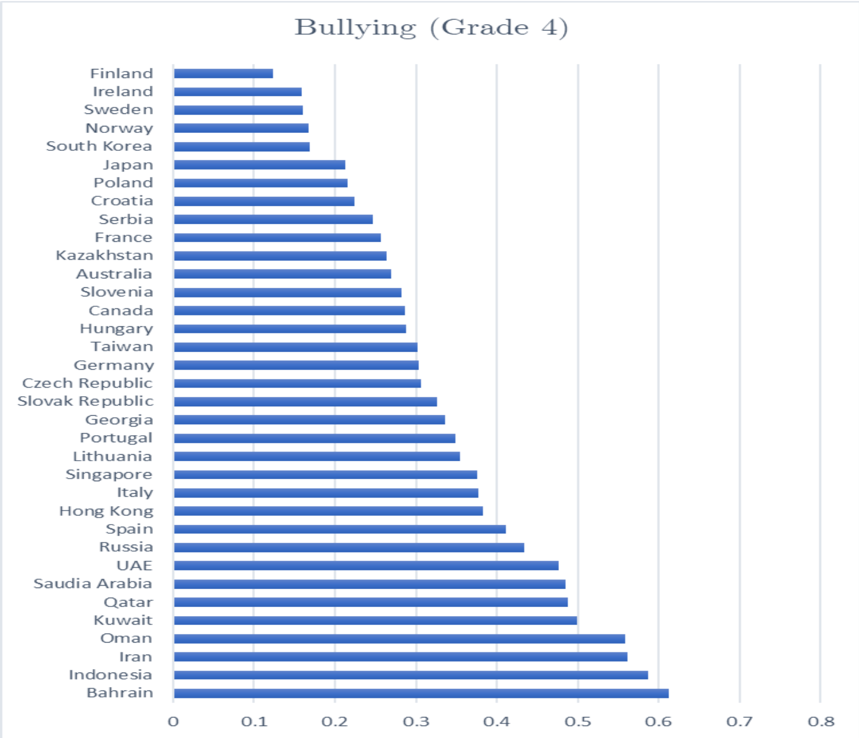
Questions about the school environment

To what degree does the school have a problem with:

1. Student late-coming?
2. Absenteeism?
3. Classroom disturbances?
4. Cheating?
5. Profanity?
6. Vandalism?
7. Theft?
8. Intimidation among students?

Physical fighting?

Table A0. Components of Bullying Variable & School Environment Measures



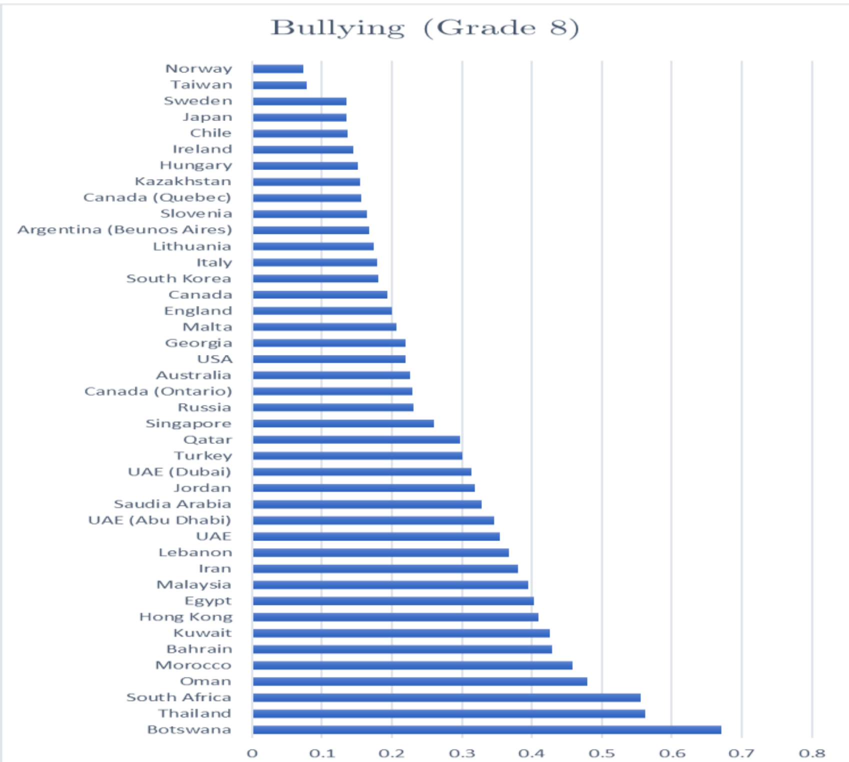


Figure A1. Prevalence of Bullying across various countries in the TIMSS 2015 sample

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample				
Bullying	-0.3088***	-0.5440***	-0.2255***	-0.6054***
	(0.044)	(0.15)	(0.043)	(014)
Age Interaction Terms				
2nd Quartile x Bullying	0.0469	0.0539	0.0160	0.0258
	(0.056)	(0.055)	(0.052)	(0.052)
3rd Quartile x Bullying	0.1219**	0.1184**	0.0723	0.0699
	(0.056)	(0.056)	(0.053)	(0.055)
4th Quartile x Bullying	0.1896***	0.1887***	0.0783	0.0767
	(0.062)	(0.059)	(0.060)	(0.057)
Endogenous Bullying		✓		✓
N	3779	3779	3779	3779
R^2	0.464		0.492	
Panel II. Grade 8 Sample				
Bullying	-0.0935**	-0.2848*	-0.0649	-0.6969**
	(0.044)	(0.17)	(0.043)	(0.18)
Age Interaction Terms				
2nd Quartile x Bullying	0.0102	0.0116	-0.0007	-0.0282
	(0.058)	(0.060)	(0.059)	(0.065)
3rd Quartile x Bullying	-0.0041	-0.0040	-0.0087	0.0076
	(0.058)	(0.064)	(0.065)	(0.065)
4th Quartile x Bullying	-0.0239	-0.0226	-0.0350	-0.0392
	(0.062)	(0.064)	(0.061)	(0.067)
Endogenous Bullying		✓		✓
N	3524	3524	3524	3524
R^2	0.601		0.604	

Note: Standard errors in parentheses are clustered by schools.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A1. Regression Estimates with Within-Cohort Age Interaction Terms

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample				
Bullying	-0.3021***	-0.5729***	-0.2382***	-0.6379***
	(0.038)	(0.15)	(0.038)	(0.14)
Male x Bullying	0.1435***	0.1436***	0.0969*	0.0967**
	(0.056)	(0.051)	(0.053)	(0.049)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3779	3779	3779	3779
R2	0.464		0.492	
Panel II. Grade 8 Sample				
Bullying	-0.1163***	-0.3473**	-0.1144***	-0.0939
	(0.043)	(0.18)	(0.042)	(0.18)
Male x Bullying	0.0348	0.0499	0.0674	0.0659
	(0.057)	(0.054)	(0.054)	(0.054)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3524	3522	3524	3522
R2	0.601		0.596	

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A2. Regression Estimates with Gender Interaction Terms

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample				
Bullying	-0.2525***	-0.4917***	-0.2089***	-0.5870***
	(0.028)	(0.14)	(0.029)	(0.14)
Immigrant x Bullying	0.3089**	0.3006**	0.1615*	0.1548*
	(0.086)	(0.082)	(0.091)	(0.084)
Immigrant x Bullying x Speak English at Home	-0.2822***	-0.3074***	-0.0601	-0.0849
	(0.100)	(0.095)	(0.11)	(0.098)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3779	3779	3779	3779
R2	0.464		0.492	
Panel II. Grade 8 Sample				
Bullying	-0.0943***	-0.2839*	-0.0783**	-0.0165
	(0.031)	(0.16)	(0.032)	(0.16)
Immigrant x Bullying	0.0964	0.1016	0.0401	0.0405
	(0.094)	(0.087)	(0.096)	(0.089)
Immigrant x Bullying x Speak English at Home	-0.2315**	-0.2317**	-0.0178	-0.0156
	(0.11)	(0.12)	(0.10)	(0.12)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3524	3522	3524	3522
R2	0.601		0.604	

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A3. Regression Estimates with Citizenship Interaction Terms

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample				
Bullying	-0.3120*	-0.6737***	-0.3032*	-0.8129***
	(0.17)	(0.20)	(0.16)	(0.19)
High Income x Bullying	0.0785	0.0745	0.1065	0.1165
	(0.18)	(0.16)	(0.17)	(0.15)
Middle Income x Bullying	0.0995	0.1117	0.1346	0.1527
	(0.18)	(0.16)	(0.17)	(0.15)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3779	3779	3779	3779
R2	0.462		0.491	
Panel II. Grade 8 Sample				
Bullying	-0.2917	-0.4646	-0.1479	-0.0686
	(0.25)	(0.29)	(0.28)	(0.32)
High Income x Bullying	0.1955	0.1913	0.0327	0.0302
	(0.25)	(0.24)	(0.28)	(0.28)
Middle Income x Bullying	0.2006	0.1935	0.1270	0.1277
	(0.25)	(0.25)	(0.29)	(0.28)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3524	3522	3524	3522
R2	0.601		0.604	

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A4a. Regression Estimates with Income Interaction Terms

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample				
Bullying	-0.2662***	-0.6205***	-0.2189***	-0.7111***
	(0.038)	(0.14)	(0.037)	(0.14)
High Parental Education x Bullying	0.1320**	0.1483***	0.1207**	0.1345***
	(0.051)	(0.54)	(0.049)	(0.052)
Middle Parental Education x Bullying	-0.0020	-0.0005	-0.0260	-0.0244
	(0.054)	(0.056)	(0.056)	(0.054)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3779	3779	3779	3779
R2	0.463		0.492	
Panel II. Grade 8 Sample				
Bullying	-0.0710*	-0.2528	-0.0599	0.0108
	(0.038)	(0.17)	(0.037)	(0.17)
High Parental Education x Bullying	-0.0400	-0.0367	0.0085	0.0082
	(0.059)	(0.057)	(0.054)	(0.058)
Middle Parental Education x Bullying	-0.0969	-0.0945	-0.0870	-0.0907
	(0.065)	(0.061)	(0.060)	(0.060)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3524	3524	3524	3524
R2	0.601		0.604	

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A4b. Regression Estimates with Income Interaction Terms (based on parental education level)

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample				
Bullying	-0.2903***	-0.6681***	-0.2234***	-0.7130***
	(0.042)	(0.15)	(0.039)	(0.15)
Adverse Environment x Bullying	0.1152**	0.0780	0.0657	0.0293
	(0.054)	(0.052)	(0.054)	(0.051)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3791	3791	3791	3791
R2	0.463		0.491	
Panel II. Grade 8 Sample				
Bullying	-0.0312	-0.1581	-0.0142	0.1485
	(0.044)	(0.19)	(0.042)	(0.19)
Adverse Environment x Bullying	-0.0982*	-0.0872	-0.0899	-0.1060*
	(0.058)	(0.055)	(0.058)	(0.055)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3524	3522	3524	3522
R2	0.601		0.604	

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A5. Regression Estimates with Adverse Environment Interaction Terms

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample				
Physical Bullying	-0.4166***	-0.9234***	-0.3065***	-0.9357***
	(0.052)	(0.20)	(0.051)	(0.19)
Male x Physical Bullying	0.1843***	0.2111***	0.0995	0.1333**
	(0.070)	(0.065)	(0.072)	(0.065)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3779	3779	3779	3779
R2	0.467		0.493	
Panel II. Grade 8 Sample				
Physical Bullying	-0.2179***	-0.4505	-0.1899**	-0.0457
	(0.083)	(0.30)	(0.085)	(0.30)
Male x Physical Bullying	0.1931*	0.2183*	0.2043*	0.1901**
	(0.11)	(0.096)	(0.11)	(0.096)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3524	3522	3524	3522
R2	0.600		0.604	

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A6. Regression Estimates of Physical Bullying

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample				
Relational Bullying	-0.1205***	-0.5886***	-0.1009***	-0.7079***
	(0.040)	(0.15)	(0.038)	(0.15)
Female x Relational Bullying	-0.1499***	-0.1663***	-0.1129**	-0.1306**
	(0.058)	(0.054)	(0.055)	(0.052)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3779	3779	3779	3779
R2	0.460		0.489	
Panel II. Grade 8 Sample				
Relational Bullying	-0.0819**	-0.3363**	-0.0598	-0.0363
	(0.041)	(0.17)	(0.042)	(0.17)
Female x Relational Bullying	-0.0393	-0.0579	-0.0526	-0.0516
	(0.059)	(0.056)	(0.055)	(0.055)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3524	3522	3524	3522
R2	0.601		0.604	

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A7. Regression Estimates of Relational Bullying

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample				
Bullying	-0.1937***	-0.7737***	-0.1539***	-0.8437***
	(0.029)	(0.13)	(0.028)	(0.12)
Positively Selected x Bullying	0.2164***	0.2330***	0.1979***	0.2162***
	(0.069)	(0.057)	(0.069)	(0.057)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3779	3779	3779	3779
R2	0.564		0.588	
Panel II. Grade 8 Sample				
Bullying	-0.0934***	-0.2662*	-0.0672**	0.0247
	(0.031)	(0.16)	(0.031)	(0.16)
Positively Selected x Bullying	0.0653	0.0632	0.0378	0.0398
	(0.059)	(0.053)	(0.059)	(0.055)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3524	3522	3524	3522
R2	0.627		0.634	

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A8a. Regression Estimates with Interaction Terms for Positively Selected Students (Sample-Level)

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample				
Bullying	-0.2140***	-0.4869***	-0.1719***	-0.5462***
	(0.028)	(0.13)	(0.026)	(0.13)
Positively Selected x Bullying	0.2280***	0.2171***	0.1954***	0.1804***
	(0.065)	(0.060)	(0.061)	(0.059)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3779	3779	3779	3779
R2	0.549		0.579	
Panel II. Grade 8 Sample				
Bullying	-0.1048***	-0.3818***	-0.0813***	-0.0952
	(0.030)	(0.15)	(0.030)	(0.15)
Positively Selected x Bullying	0.1013	0.1055*	0.0915	0.0909
	(0.062)	(0.062)	(0.059)	(0.057)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3524	3522	3524	3522
R2	0.660		0.664	

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A8b. Regression Estimates with Interaction Terms for Positively Selected Students (School-Level)

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample				
Bullying	-0.1709***	-0.4742***	-0.1689***	-0.5913***
	(0.028)	(0.16)	(0.028)	(0.16)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3779	3779	3779	3779
*R*2	0.458		0.490	
Panel II. Grade 8 Sample				
Bullying	-0.0706***	-0.1016	-0.0376	0.1156
	(0.024)	(0.12)	(0.024)	(0.12)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3524	3522	3524	3522
*R*2	0.600		0.603	

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A9. Regression Estimates for Broader Definition of Bullying

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample				
Bullying	-0.3169***	-0.6625***	-0.2194***	-0.7019***
	(0.041)	(0.19)	(0.041)	(0.18)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3779	3779	3779	3779
R2	0.462		0.488	
Panel II. Grade 8 Sample				
Bullying	-0.1249***	-0.2700	-0.1021**	0.0460
	(0.044)	(0.20)	(0.044)	(0.20)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying		✓		✓
N	3524	3522	3524	3522
R2	0.601		0.604	

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A10. Regression Estimates for Stricter Definition of Bullying

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample	CF	Simultaneous	CF	Simultaneous
Bullying	-0.4908***	-0.5698***	-0.5556***	-0.6417***
	(0.15)	(0.13)	(0.14)	(0.12)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying	✓		✓	
N	3779	3779	3779	3779
R2		0.411		0.423
Panel II. Grade 8 Sample	CF	Simultaneous	CF	Simultaneous
Bullying	-0.2876*	-0.2101	-0.0087	0.0488
	(0.16)	(0.17)	(0.16)	(0.16)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying	✓		✓	
N	3524	3522	3524	3522
R2		0.599		0.601

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A11. Regression Estimates for Simultaneous Equation Specification

	Outcome Variable			
	Mathematics Score		Science Score	
Panel I. Grade 4 Sample	Original CF	New CF	Original CF	New CF
Bullying	-0.4928***	-0.6748***	-0.5573***	-0.6456***
	(0.15)	(0.14)	(0.14)	(0.13)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying	✓	✓	✓	✓
N	3778	3778	3778	3778
R2				
Panel II. Grade 8 Sample	Original CF	New CF	Original CF	New CF
Bullying	-0.2786*	-0.1864	-0.0087	-0.0232
	(0.16)	(0.14)	(0.16)	(0.13)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Endogenous Bullying	✓	✓	✓	✓
N	3522	3507	3522	3507
R2				

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A12. Regression Estimates using larger set of instruments

	Variable of Interest: Bullying			
	Grade 4		Grade 8	
Panel I. Mathematics	(1)	(2)	(3)	(4)
Outcome Variables				
How often are you absent from school?	-0.1228***	-0.0426	-0.0171	0.0136
	(0.043)	(0.041)	(0.046)	(0.043)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Test Scores & Test Score Squared		✓		✓
Panel II. Science				
Outcome Variables				
How often are you absent from school?	-0.1228***	-0.0431	-0.0174	0.0065
	(0.043)	(0.040)	(0.045)	(0.041)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Test Scores & Test Score Squared		✓		✓

Note: Standard errors in parentheses are clustered by schools. Standardised outcome variables are used in all regressions.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A13. Regression Estimates with Absenteeism as the outcome variable

	Variable of Interest: Bullying			
	Grade 4		Grade 8	
Panel I. Mathematics	(1)	(2)	(3)	(4)
Outcome Variables				
I usually do well in Mathematics	0.1333***	0.0077	0.0499	-0.0057
	(0.038)	(0.035)	(0.045)	(0.043
Mathematics is harder for me than for many of my classmates	-0.1965***	-0.0897***	-0.1460***	-0.1005***
	(0.035)	(0.034)	(0.046)	(0.044)
I am just not good at Mathematics	-0.1440***	-0.0345	-0.0586	-0.0062
	(0.037)	(0.035)	(0.043)	(0.042)
I learn things quickly in Mathematics	0.1156***	0.0216	-0.0214	-0.0720*
	(0.037)	(0.036)	(0.043)	(0.041)
Mathematics makes me nervous	-0.2455***	-0.1748***	(0.1395***	-0.1021**
	(0.037)	(0.037)	(0.043)	(0.042)
I am good at working out difficult Mathematics problems	0.1314***	0.0450	0.0138	-0.0335
	(0.033)	(0.033)	(0.045)	(0.042)
Mathematics is harder for me than any other subject	-0.2139***	-0.1049***	-0.0581	-0.0042
	(0.035)	(0.032)	(0.039)	(0.038)
Mathematics makes me confused	-0.2399***	-0.1458***	-0.1319***	-0.0890**
	(0.037)	(0.036)	(0.039)	(0.038)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Test Scores & Test Score Squared		✓		✓

Note: Standard errors in parentheses are clustered by schools. Standardised outcome variables are used in all regressions.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A14a. Regression Estimates with various outcome variables measuring self-perception (Math)

	Variable of Interest: Bullying			
	Grade 4		Grade 8	
Panel II. Science	(1)	(2)	(3)	(4)
Outcome Variables				
I usually do well in Science	0.0077	-0.0357	0.0317	0.0021
	(0.040)	(0.039)	(0.041)	(0.039)
Science is harder for me than for many of my classmates	-0.1479***	-0.0804**	-0.1010**	-0.0705*
	(0.042)	(0.040)	(0.040)	(0.038)
I am just not good at Science	-0.1267***	-0.0689*	-0.0745*	-0.0415
	(0.042)	(0.040)	(0.040)	(0.038)
I learn things quickly in Science	-0.0238	-0.0593	-0.0026	-0.0284
	(0.042)	(0.042)	(0.040)	(0.040)
Science is harder for me than any other subject	-0.0838**	-0.0242	-0.0811**	-0.0479
	(0.038)	(0.036)	(0.040)	(0.038)
Science makes me confused	-0.1623***	-0.1086***	-0.0909**	-0.0657
	(0.038)	(0.035)	(0.042)	(0.040)
I am good at working out difficult Science problems			0.0191	-0.0013
			(0.044)	(0.043)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Test Scores & Test Score Squared		✓		✓

Note: Standard errors in parentheses are clustered by schools. Standardised test scores are used in this regression.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A14b. Regression Estimates with various outcome variables measuring self-perception (Science)

	Variable of Interest: Bullying			
	Grade 4		Grade 8	
Panel I. Mathematics	(1)	(2)	(3)	(4)
Outcome Variables				
My teacher is easy to understand	0.1388***	0.0887**	0.0661	0.0469
	(0.041)	(0.039)	(0.044)	(0.044)
I am interested in what my teachers says	0.1114***	0.0901**	0.0168	0.0016
	(0.041)	(0.041)	(0.043)	(0.043)
My teacher gives me interesting things to do	0.0589	0.0552	0.0839*	0.0770*
	(0.042)	(0.042)	(0.044)	(0.043)
My teacher has clear answers to my questions	0.1851***	0.1367***	0.0855*	0.0697
	(0.039)	(0.039)	(0.044)	(0.044)
My teacher is good at explaining Mathematics	0.1392***	0.0997***	0.0582	0.0436
	(0.039)	(0.037)	(0.044)	(0.044)
My teacher lets me show what I have learned	0.0343	0.0384	0.0597	0.0519
	(0.040)	(0.040)	(0.046)	(0.046)
My teacher does a variety of things to help us learn	0.0808**	0.0608	0.0304	0.0280
	(0.040)	(0.040)	(0.042)	(0.042)
My teacher tells me how to do better when I make a mistake	0.0479	0.0387	0.0344	0.0287
	(0.039)	(0.038)	(0.043)	(0.043)
My teacher listens to what I have to say	0.3023***	0.2719***	0.0823*	0.0703
	(0.040)	(0.040)	(0.042)	(0.043)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Test Scores & Test Score Squared		✓		✓

Note: Standard errors in parentheses are clustered by schools. Standardised outcome variables are used in all regressions.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A15a. Regression Estimates with various outcome variables measuring perception of math teachers

	Variable of Interest: Bullying			
	Grade 4		Grade 8	
Panel II. Science	(1)	(2)	(3)	(4)
Outcome Variables				
My teacher is easy to understand	0.0753**	0.0513	-0.0127	-0.0224
	(0.036)	(0.036)	(0.044)	(0.043)
I am interested in what my teachers says	-0.0015	-0.0169	-0.0073	-0.0212
	(0.038)	(0.038)	(0.045)	(0.044)
My teacher gives me interesting things to do	0.0182	0.0058	-0.0386	-0.0441
	(0.040)	(0.040)	(0.048)	(0.048)
My teacher has clear answers to my questions	0.0930**	0.0687*	0.0181	0.0091
	(0.038)	(0.038)	(0.045)	(0.045)
My teacher is good at explaining Science	0.0483	0.0251	-0.0143	-0.0247
	(0.037)	(0.038)	(0.044)	(0.043)
My teacher lets me show what I have learned	0.0011	0.0039	0.0298	0.0313
	(0.034)	(0.038)	(0.044)	(0.045)
My teacher does a variety of things to help us learn	0.0251	0..0123	-0.0039	-0.0067
	(0.035)	(0.035)	(0.045)	(0.045)
My teacher tells me how to do better when I make a mistake	0.0322	0.0226	-0.0072	-0.0080
	(0.036)	(0.038)	(0.043)	(0.043)
My teacher listens to what I have to say	0.1933***	0.1799***	0.0445	0..0378
	(0.037)	(0.038)	(0.041)	(0.041)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Test Scores & Test Score Squared		✓		✓

Note: Standard errors in parentheses are clustered by schools. Standardised outcome variables are used in all regressions.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A15b. Regression Estimates with various outcome variables measuring perception of math teachers

	Variable of Interest: Bullying			
	Grade 4		Grade 8	
Panel I. Mathematics	(1)	(2)	(3)	(4)
Outcome Variables				
I wish I did not have to study mathematics	-0.2239***	-0.1662***	-0.0635	-0.0217
	(0.037)	(0.037)	(0.044)	(0.041)
Mathematics is boring	-0.1701***	-0.1130***	-0.0256	0.0107
	(0.042)	(0.042)	(0.046)	(0.044)
I like mathematics	0.1346***	0.0658*	0.0126	-0.0300
	(0.038)	(0.038)	(0.040)	(0.038)
I like any schoolwork that involves numbers	0.0901**	0.0576	0.0345	0.0023
	(0.043)	(0.042)	(0.044)	(0.042)
I look forward to mathematics lessons	0.0825**	0.0388	0.0509	0.0242
	(0.041)	(0.041)	(0.046)	(0.045)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Test Scores & Test Score Squared		✓		✓
Panel II. Science				
Outcome Variables				
I wish I did not have to study science	-0.0905**	-0.0598	-0.0346	-0.0093
	(0.042)	(0.041)	(0.045)	(0.043)
Science is boring	-0.0602	-0.0275	-0.0675	-0.0409
	(0.039)	(0.038)	(0.043)	(0.040)
I like science	-0.0088	-0.0359	-0.0130	-0.0367
	(0.042)	(0.042)	(0.045)	(0.044)
I look forward to learning science in school	-0.0212	-0.0421	0.0189	-0.0016
	(0.042)	(0.044)	(0.046)	(0.045)
I like to do science experiments	0.0153	-0.0119	-0.0626	-0.0767*
	(0.037)	(0.037)	(0.046)	(0.045)
Controls:				
Individual & Class-specific Characteristics	✓	✓	✓	✓
School Fixed Effects	✓	✓	✓	✓
Test Scores & Test Score Squared		✓		✓

Note: Standard errors in parentheses are clustered by schools. Standardised outcome variables are used in all regressions.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table A16. Regression Estimates with various outcome variables measuring subject interest